ciated with lysine residues. By comparing the ^{13}C NMR data from the reductively methylated CH2 fragment, the reductively methylated pFc′ and tFc′ fragments, and the reductively methylated Fc fragment, we were able to assign dimethyllysyl resonances in the Fc fragment to domains. For example, resonances 2 and 8 each represent single lysine residues in the CH2 domain, and we have suggested candidate lysine residues for each of these resonances.

The assignments will be extended and verified in additional studies of this system. For example, we plan to use the salt dependence of the dimethyllysyl resonances to confirm the identity of lysyl resonances in ion pair interactions. An ultimate goal of these studies is to use the dimethylated Fc fragment to study its interactions with effector molecules such as C1q and the Fc cell surface receptor.

ACKNOWLEDGMENTS

We thank Josephine Secnik for aid in isolating the IgG-Tu and for providing her data on the behavior of dimethylated N-terminal glycine residues and John Cameron for assistance in data analysis.

REFERENCES

Deisenhofer, J. (1981) *Biochemistry 20*, 2361–2378.
Dorrington, K. J., & Painter, R. H. (1974) *Prog. Immunol. 2*, 75–84.
Edelman, G. M., Cunningham, B. A., Gall, W. E., Gallieb, P. D., Rutishauser, U., & Wardal, M. J. (1969) *Proc. Natl. Acad. Sci. U.S.A. 63*, 78–85.
Ellerson, J. R., Yasmeen, D., Painter, R. H., & Dorrington, K. J. (1972) *FEBS Lett. 24*, 318–322.
Ellerson, J. R., Yasmeen, D., Painter, R. H., & Dorrington, K. J. (1976) *J. Immunol. 116*, 510–517.
Gerken, T. A. (1984) *Biochemistry 23*, 4688–4697.
Gerken, T. A., Jentoft, J. E., Jentoft, N., & Dearborn, D. G. (1982) *J. Biol. Chem. 257*, 2894–2900.
Isenman, D. E., Ellerson, J. R., Painter, R. H., & Dorrington, K. J. (1977) *Biochemistry 16*, 233–240.
Jentoft, J. E., Jentoft, N., Gerken, T. A., & Dearborn, D. G. (1979) *J. Biol. Chem. 254*, 4366–4370.
Jentoft, J. E., Gerken, T. A., Jentoft, N., & Dearborn, D. G. (1981) *J. Biol. Chem. 256*, 231–236.
Jentoft, J. E., Dearborn, D. G., & Bernier, G. M. (1982) *Biochemistry 21*, 289–294.
Jentoft, N., & Dearborn, D. G. (1979) *J. Biol. Chem. 254*, 4359–4365.
Lee, B. K., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Minta, J. O., & Painter, R. H. (1972) *Immunochemistry 9*, 1041–1048.
Natvig, J. B., & Turner, M. W. (1971) *Clin. Exp. Immunol. 8*, 685–700.
Okafor, G. O., Turner, M. W., & Hay, F. C. (1974) *Nature 248*, 228–230.
Porter, R. R. (1959) *Biochem. J. 73*, 119–126.
Turner, M. W., & Bennich, H. (1968) *Biochem. J. 107*, 171–178.
Yasmeen, D., Ellerson, J. R., Dorrington, K. J., & Painter, R. H. (1973) *J. Immunol. 110*, 1706–1709.
Yasmeen, D., Ellerson, J. R., Dorrington, K. J., & Painter, R. H. (1976) *J. Immunol. 116*, 518–526.

Mitochondrial NADH–Ubiquinone Reductase: Complementary DNA Sequences of Import Precursors of the Bovine and Human 24-kDa Subunit

Stephanie J. Pilkington‡ and John E. Walker*
Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.
Received November 1, 1988; Revised Manuscript Received December 22, 1988

ABSTRACT: The 24-kDa subunit of mitochondrial NADH–ubiquinone reductase (complex I) is an iron–sulfur protein that is present in the flavoprotein or NADH dehydrogenase II subcomplex. It is a nuclear gene product and is imported into the organelle. A group of human patients with mitochondrial myopathy have been shown to have reduced levels of subunits of complex I in skeletal muscle mitochondria, and in one patient the 24-kDa subunit appears to be absent (Schapira et al., 1988). To investigate the genetic basis of this type of myopathy, cDNA clones have been isolated from a bovine library derived from heart and liver mRNA by hybridization with two mixtures of 48 synthetic oligonucleotides 17 bases in length that were designed on the basis of known protein sequences. The recombinant DNA sequence has been determined, and it encodes a precursor of the mature 24-kDa protein. The N terminus of the mature protein is preceded by a presequence of 32 amino acids that has properties that are characteristic of mitochondrial import sequences. The sequence of the mature protein deduced from the cDNA contains a segment of nine amino acids that was not determined in an earlier partial protein sequence analysis. The bovine clone has been employed as a hybridization probe to identify cDNA clones of the human homologue of the 24-kDa protein. Its DNA sequence has also been determined, and it codes for a protein that is closely related to the bovine protein. Three conservative substitutions are found in the mature protein, and the human and bovine presequences differ at a further five positions. The bovine cDNA clone has been used as a hybridization probe with digests of bovine and human DNA to investigate the genetic complexity of the protein.

Mitochondrial NADH–ubiquinone reductase (complex I) is embedded in the inner membrane of the organelle. It is the first enzyme of the respiratory chain and catalyzes the reoxidation of NADH and transfer of electrons to ubiquinone [for a review see Ragan (1987)]. For each electron transferred between NADH and ubiquinone two protons are pumped from the mitochondrial matrix (Wikström, 1984). Complex I is by

* To whom correspondence should be addressed.
‡ S.J.P. is supported by an MRC research studentship.

0006-2960/89/0428-3257$01.50/0

far the most complicated of the respiratory complexes, and preparations of the enzyme isolated from both bovine hearts and *Neurospora crassa* contain about 26 different polypeptides, as well as an FMN and 6–8 iron–sulfur centers as redox groups (Heron et al., 1979; Ise et al., 1985; Ragan, 1987). From this assembly a simpler fragment can be prepared that catalyzes NADH oxidation by a variety of electron acceptors. This is type II NADH dehydrogenase or the flavoprotein (FP) fragment. It contains three of the subunits of complex I, namely, the 51-, 24-, and 10-kDa polypeptides (Galante & Hatefi, 1979; Heron et al., 1979). FMN and a tetranuclear iron–sulfur cluster are associated with the 51-kDa protein [evidence summarized in Ragan (1987)], and a binuclear iron–sulfur cluster is in the 24-kDa component (Ohnishi et al., 1985). The amino acids that form part of this center have not been identified, but the almost complete sequence of the 24-kDa protein contains five cysteine residues grouped in two pairs and a single residue (von Bahr-Lindström et al., 1983) that could be involved. Interest in mammalian complex I has been heightened by the finding that the enzyme is defective in skeletal muscle mitochondria of a number of patients with myopathies (Morgan-Hughes et al., 1985). Two of these patients have a generalized deficiency of subunits of the enzyme in skeletal muscle mitochondria and, in addition, in one of these, the 24-kDa component is completely absent (Schapira et al., 1988). To investigate the genetic basis of tissue-specific mitochondrial myopathies in this class, we have isolated and sequenced cDNA clones encoding both the bovine and human 24-kDa proteins as described in this paper. The bovine clone was recognized by hybridization with two mixed oligonucleotide probes designed on the basis of the protein sequence, and then the bovine clone was employed to isolate a human cDNA. The protein sequences encoded in the cDNAs are very similar; they differ in five positions in their N-terminal import sequences, which are lost during entry to the organelle, and in only three positions of the mature protein. Hybridization experiments conducted on digests of bovine and human genomic DNA suggest that both genomes contains more than one sequence related to the cDNAs. This may indicate the presence of multiple genes (possibly including pseudo-genes) for the 24-kDa subunit of complex I.

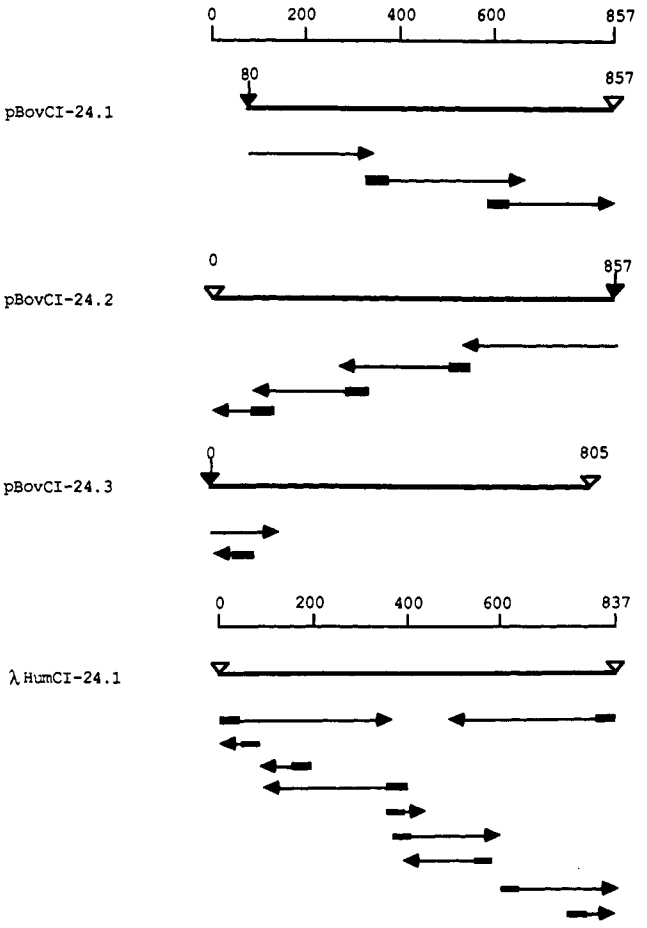

FIGURE 1: Sequence analysis of cDNA clones encoding the bovine and human precursors of the 24-kDa subunit of mitochondrial NADH–ubiquinone reductase. The thick lines indicate the inserted DNAs in the isolates pBovCI-24.1–3 and λHumCI-24.1. The bovine inserts are flanked by *Eco*RI and *Bam*HI sites in the polylinker of the vector, and the human insert is flanked by *Eco*RI sites. The horizontal arrows represent the extent and direction of the DNA sequences that were determined. Synthetic oligonucleotide primers used in the sequencing are indicated by black boxes. The scale is bases. (▽) *Eco*RI; (↓) *Bam*HI.

Materials and Methods

Reagents. The sources of chemicals, biochemicals, and enzymes used in the experiments described here have been given previously (Gay & Walker, 1985a,b; Walker et al., 1987b,c).

Oligonucleotide Synthesis. Mixtures of oligonucleotides and unique oligonucleotides were synthesized by automated phosphoramidite chemistry in an Applied Biosystems 380B synthesizer. A mixture of 48 oligonucleotides 17 bases in length with the sequence 5' AARTAYCAYATY/ACARGT 3' (probe 1) was designed from the protein sequence KYHIQV. These are residues 113–118 in the sequence determined by von Bahr-Lindström et al. (1983) or residues 97–102 of the protein sequence deduced from the cDNA sequence presented in Figure 3. Calculations based on contributions of 4 and 2 °C, respectively, by each G·C and A·T base pair (Suggs et al., 1981) give a minimum dissociation temperature of 40 °C for the components of this mixture. A second mixture of 48 oligonucleotides also 17 bases in length was synthesized in the same way. This is referred to as probe 2 and has the sequence 5' ATGGTNCARATY/AAAYGA 3'. It was based upon the protein sequence MVQIND, residues 169–174 of the protein sequence reported by von Bahr-Lindström et al. (1983) or residues 153–158 of the protein sequence deduced from the cDNA sequence (Figure 3). The minimum dissociation temperature estimated for this mixture is 42 °C. Unique 17-base synthetic oligonucleotides were used as sequencing primers as indicated in Figure 1.

cDNA Libraries. Two libraries were investigated. A plasmid library in the vector pUC8 (Vieira & Messing, 1982) was made earlier by using mRNA isolated from bovine heart and liver (Gay & Walker, 1985a). A human T cell cDNA library (HUT 78) cloned in λgt10 (Hyunh et al., 1985) was obtained from Dr. T. H. Rabbitts.

Screening the cDNA Libraries. Approximately 0.5×10^6 recombinants in the plasmid library were grown on 20 Pall Biodyne A nylon membranes (13.2-cm diameter) placed on agar in Petri dishes (13.5-cm diameter). Two replicas of bacterial colonies on each plate were transferred onto membranes as follows. A dry membrane was placed for 1 min on the membrane bearing the colonies. Then it was removed and placed on an agar plate containing chloramphenicol (50 μg/mL), and the bacteria were grown for 24 h at 37 °C. The original membrane was moved onto a fresh agar plate and incubated for 12 h at 37 °C before the second replica membrane was made. Portions (50 pmol) of each oligonucleotide mixture were radiolabeled by using polynucleotide kinase and [γ-^{32}P]ATP (specific activity approximately 3000 Ci/mmol)

and purified by electrophoresis through a 20% polyacrylamide gel containing 6 M urea, 90 mM Tris (pH 8.3), 90 mM boric acid, and 25 mM EDTA. The oligonucleotides were eluted from the gel and purified further by centrifugation through a Sephadex G-50 column (1 mL). Hybridizations using probes 1 and 2 with replica sets of filters were carried out at 35 and 37 °C, respectively (5 °C below the estimated minimum dissociation temperatures). In each case, the purified probe was added to the hybridization solution (100 mL) in which 20 filters had been immersed. Otherwise, treatment of the membranes and hybridization conditions were as described previously (Gay & Walker, 1985b). Recombinants that gave a strong signal with both probes were rescreened with probe 1. Those isolates that still gave a positive response, pBovC1-24.1, pBovC1-24.2, and pBovC1-24.3, were grown and their recombinant plasmids isolated and characterized further.

Recombinants in the human T cell cDNA library were grown for 6–7 h at 37 °C on *Escherichia coli* C600HF1 in 13.5-cm-diameter Petri dishes. Two nitrocellulose filters were placed sequentially on each plate and left for 1 min. The filters were treated as described by Benton and Davis (1977) and then screened by using a "prime cut" probe (Farrell et al., 1983) derived from the full length of clone pBovC1-24.2 by using the terminal *Eco*RI and *Bam*HI sites. They were prehybridized for 1 h at 65 °C in 5× Denhardt's solution [1× Denhardt's solution contains 200 mg/L each of poly(vinylpyrrolidone), bovine serum albumin (fraction V), and Ficoll], 6× SSC [1× SSC contains 0.15 M sodium chloride and 0.015 M sodium citrate], 0.5% *N*-laurylsarcosine, and sonicated salmon sperm DNA (100 μg/mL). Hybridization was carried out at 65 °C for 18 h in 100 mL of the same solution supplemented with 10% dextran sulfate and the radioactive probe. Filters were washed at 65 °C in a solution that contained 2× SSC and 0.5% *N*-laurylsarcosine. Then they were autoradiographed in the same way as for the plasmid library screen. Hybridizing plaques were rescreened under the same conditions until the plaque was pure. DNA was prepared from these isolates by using the plate lysate method (Maniatis et al., 1982). Approximately 0.5×10^6 recombinants were screened on 10 plates. Over 100 hybridizing recombinants were identified, and 2 of them, λHumC1-24.1 and λHumC1-24.2, were investigated in detail.

Subcloning and DNA Sequencing. DNA sequences were determined by the dideoxy method (Sanger et al., 1977) as modified by Biggin et al. (1983). Inserts from the pUC8 recombinants were excised by using *Bam*HI and *Eco*RI together and then cloned into M13mp18 and M13mp19. Inserts from λgt10 recombinants were excised with *Eco*RI and cloned into M13mp18 and pUC18 (Yanisch-Perron et al., 1985). When the *Eco*RI fragment from recombinant λHumC1-24.2 was cloned into the latter vector and then recloned into M13mp18, it appeared to have rearranged and so it was not sequenced further. The flanking primer LMB2 (Duckworth et al., 1981) and six internal primers were used in the determination of the bovine sequence. Ten internal primers were used for the human sequence, including four of those used for sequencing the bovine cDNA. Sequences were compiled by using the computer programs DBAUTO and DBUTIL (Staden, 1982a).

Data Analysis. DNA sequences were analyzed with the computer program ANALYSEQ (Staden, 1985). The protein sequence of the bovine 24-kDa protein was compared with sequences in the PIR data base. The significance of homologies between proteins was assessed by pairwise comparisons with the aid of DIAGON (Staden, 1982b), and the hydrophobic profile of the 24-kDa protein was calculated with HYDROPLOT, a version of SOAP (Kyte & Doolittle, 1982).

Preparation of Genomic DNA. Bovine DNA was prepared from liver (Walker et al., 1987b). Human DNA was obtained from fresh samples (approximately 30 mL) of blood. A solution of 5% EDTA was added and the plasma removed by centrifugation. The red blood cells were lysed by the repeated addition of buffer (10 mM Tris-HCl, 5 mM magnesium chloride, 10 mM sodium chloride, pH 7.6) up to a final volume of 50 mL, followed by centrifugation, until a pellet of white blood cells only was obtained. These were washed again with the same solution and then resuspended in 2 mL of the buffer. The white blood cells were lysed by gentle shaking overnight at 42 °C in 15 mL of a solution containing 10 mM Tris-HCl, pH 7.6, 10 mM EDTA, 50 mM sodium chloride, 0.2% sodium dodecyl sulfate, and 200 μg/mL proteinase K (Gross-Bellard et al., 1973). The solution was extracted twice by gentle shaking for 10 min to give a complete emulsion, with a mixture of phenol (3 volumes) and chloroform:isoamyl alcohol (1 volume, 24:1 v/v). A final extraction was done with the chloroform:isoamyl alcohol, and the DNA was precipitated with propan-2-ol, rinsed with ethanol, and dried briefly in vacuo. It was redissolved to a concentration of approximately 500 μg/mL in a buffer that contained 10 mM Tris-HCl, pH 7.6, and 0.1 mM EDTA and was stored over chloroform at 4 °C.

Hybridization with Human and Bovine Genomic DNA. This was performed by the method of Southern (1975) as described in Walker et al. (1987b), except that Hybond N nylon membranes (Amersham International plc) were employed. Bovine or human DNA (20 μg/slot) was digested at 37 °C in 400 μL of buffer with the restriction endonuclease *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sac*I, or *Xba*I. About 20 units of enzyme was added initially and an additional 20 units after 2 h. Digestion was continued for a further 2 h, and then the DNA was precipitated at −20 °C with ethanol and 0.3 M sodium acetate. After the DNA had been dried briefly, it was redissolved in 30 μL of water and fractionated by agarose gel electrophoresis overnight, and the separated fragments were then transferred to membranes (Walker et al., 1987b). The membranes were prehybridized for 1 h at 65 °C in a solution containing 6× SSC, 5× Denhardt's solution, sonicated salmon sperm DNA (100 μg/mL), and 0.5% *N*-laurylsarcosine. Hybridization was carried out in the same solution with the addition of 10% dextran sulfate and the radiolabeled prime cut probe for 18 h at 65 °C. The probe corresponded to the entire region of pBovCI-24.2 between the flanking *Eco*RI and *Bam*HI sites. The filters were washed twice at room temperature in a mixture of 2× SSC and 0.5% *N*-laurylsarcosine, twice again in the same solution at 65 °C, and then twice more at 65 °C in a solution that contained 0.2× SSC and 0.5% *N*-laurylsarcosine. Filters were autoradiographed at −70 °C for 12–48 h with preflashed film (Fuji RX 150).

Results and Discussion

Cloning and DNA Sequence Analysis. The bovine plasmid cDNA library was screened with two mixtures of 48 oligonucleotides representing amino acids 97–102 and 169–174, respectively, of the protein sequence determined by von Bahr-Lindström et al. (1983) or amino acids 97–102 and 153–158 of the protein sequence deduced from the cDNA (see Figures 2 and 3). Three clones, pBovCI-24.1, pBovCI-24.2, and pBovCI-24.3, hybridized strongly to both probes, and restriction analysis showed that they all contained inserts of approximately 800 base pairs. The fragment from isolate pBovCI-24.1 was found to have the sequence A_{13} at its 3′ end

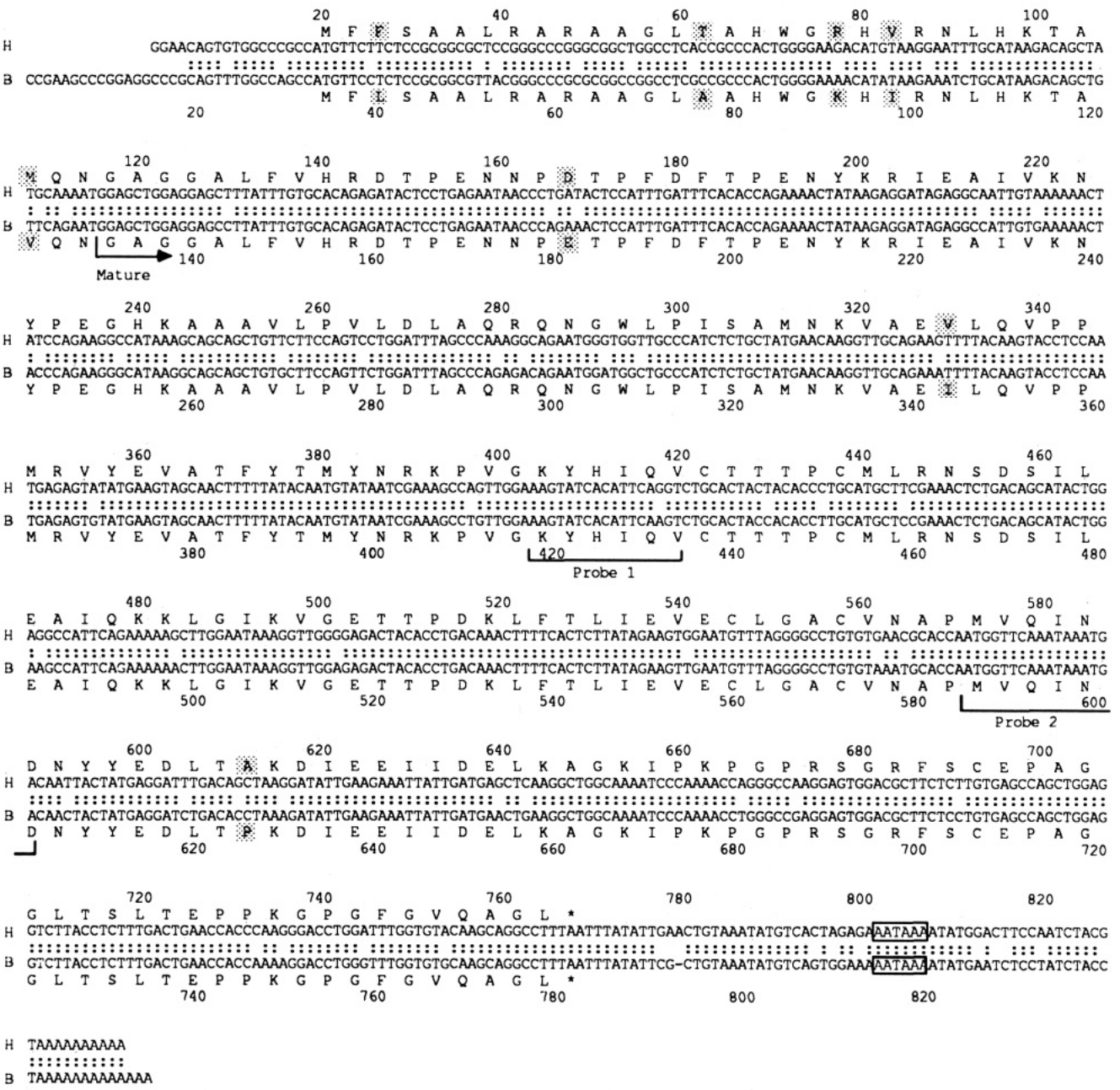

FIGURE 2: Comparison of the human bovine cDNA sequences encoding precursors of the 24-kDa subunit of mitochondrial NADH–ubiquinone reductase. The cDNA sequences, human (H, upper) and bovine (B, lower) are numbered. They have been aligned with the computer program NUCALN (Wilbur & Lipman, 1983), and identities between them are indicated by colons. Toward their 3′ ends, sequences that probably serve as signals for polyadenylation of transcripts are boxed. The human and bovine protein sequences encoded in the cDNAs are shown above and below the DNA sequences, respectively, and differences between them have been shaded. The N terminus of the mature bovine protein is indicated, as are the two segments of protein sequences that were used to design the synthetic oligonucleotide probes 1 and 2.

and also contains sequence coding for the entire mature 24-kDa subunit, but not for an intact mitochondrial import sequence. The inserted DNA in clone pBovCI-24.2 also had the sequence A_{13} at its 3′ terminus and contained a sequence coding for the mature protein. However, it extended in a 5′ direction 80 bases further than the insert in pBovCI-24.1 and encodes the complete mitochondrial import precursor of the protein. Isolate pBovCI-24.3 has the same 5′ terminus as pBovCI-24.2, but does not extend beyond base 805. A 30-base insertion of sequence from bacteriophage M13 was found after position 792, indicating that rearrangement had occurred at this end of the sequence, probably during subcloning, and so the 3′ end of this clone was not sequenced further. Sequences determined in the three isolates are summarized in Figure 1, and taken together they provide the DNA sequence in both senses. With the exception of the rearrangement in pBovCI-24.3 discussed above, the sequences determined in the three isolates are entirely concordant.

The human cDNA library was screened with a prime-cut probe containing the entire sequence of isolate pBovCI-24.2. This was produced by excision at the flanking *Eco*RI and *Bam*HI restriction sites. Over 100 strongly hybridizing plaques were identified in the initial screening experiment, and two of them, λHumCI-24.1 and λHumCI-24.2, were selected for further investigation. Their inserts were excised with *Eco*RI and were about 800 base pairs long. The former was cloned directly into M13mp18 and the latter first into the *Eco*RI site of pUC18 and from there into M13mp18. However, sequence analysis showed that the insert had rearranged, and so it was not characterized further. Isolate λHumCI-24.1 had the 3′

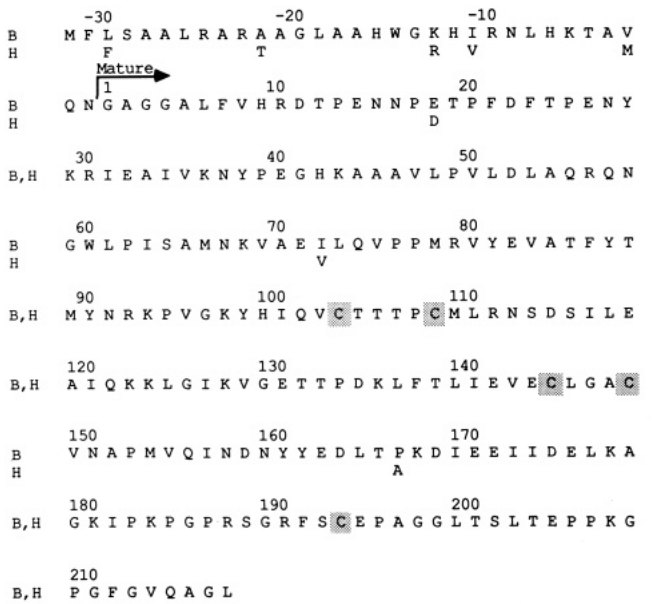

FIGURE 3: Deduced protein sequences of the precursors of the bovine and human 24-kDa proteins. Where the two proteins differ, the bovine sequence (B) is shown above the human sequence (H). The mature protein is numbered from 1 to 217 and the mitochondrial import sequence from −1 to −32. The N terminus of the mature bovine protein is indicated. Cysteine residues are shaded.

sequence A_{10} and encodes a mitochondrial import precursor of the human 24-kDa protein that is very similar in sequence to the bovine precursor protein. The coding sequence is preceded by a short 5′ noncoding region. The sequencing strategy used to obtain the human cDNA sequence in both of its senses is shown in Figure 1.

The final sequences of the cDNAs encoding the bovine and human 24-kDa proteins are compared in Figure 2. The bovine sequence is 14 bases longer at its 5′ end and 1 base shorter in its 3′ noncoding region. Each has a potential polyadenylation signal AATAAA (Proudfoot & Brownlee, 1975), 21 bases before the poly(A) sequence. The 3′ noncoding region is unusually short, and the presence of the sequence ATTTA immediately following the termination codon might indicate that the mRNA is unstable (Shaw & Kamen, 1986). In their 5′ regions both cDNAs have a preponderance of C and G residues, which might indicate that CpG islands are associated with the 5′ ends of the genes. The coding regions are identical in length. Over the region of alignment, the nucleotide sequences are very similar and differ in only 70 positions of 827. Of these, 40 are within the mature protein coding region, 13 in the region encoding the N-terminal extension, 6 in the 5′ untranslated region of overlap, and 11 within the 3′ untranslated region.

Deduced Protein Sequences. The human and bovine protein sequences of the 24-kDa subunit of complex I (Figure 3) differ in only three positions within the mature protein, all three being conservative amino acid substitutions resulting from single nucleotide changes. A further 5 amino acid changes are present within the 32-amino acid N-terminal presequences, again caused by single nucleotide substitutions.

Residues 1–56 and 66–217 of the bovine protein sequence encoded in the cDNA are identical with residues 1–56 and 82–233, respectively, of the sequence reported by von Bahr-Lindström et al. (1983). The previously reported protein sequence was incomplete; no overlaps were established between residues 56–57, 67–68, and 81–82 and, in addition, residues 68–81 were presented as a partial sequence deduced from amino acid composition. Also, it has been pointed out previously that residues 57–67 of this earlier sequence are identical with residues 130–140 of bovine trypsin (see entry DEECR in the PIR data base); presumably this peptide arose by autolysis during tryptic digestion of the 24-kDa protein. Our sequences show that there are nine residues only between amino acids 56 and 83 of the earlier version of the sequence and that the mature protein chain is 217 amino acids long. The amino acid composition of the final bovine sequence is in good agreement with the experimentally derived composition published earlier (von Bahr-Lindström et al., 1983; see Table I). The calculated molecular mass is 23 830, close to the value of 24 kDa estimated in the presence of sodium dodecyl sulfate by polyacrylamide gel electrophoresis. Assuming that cleavage of the mitochondrial import sequence is at the same position as in the bovine protein, the human 24-kDa protein is the same length as the bovine homologue, and its molecular mass is 23 761.

Table I: Amino Acid Composition of the Mature 24-kDa Subunit of NADH–Ubiquinone Reductase from Bovine Mitochondria

amino acid	by amino acid analysis[a]	from the DNA sequence
aspartic acid[b]	21.0	20
threonine	12.9	14
serine	5.8	6
glutamic acid[c]	26.3	24
proline	22.6	21
glycine	19.3	18
alanine	16.5	16
cysteine	ND[d]	5
valine	14.7	15
methionine	4.9	5
isoleucine	13.3	14
leucine	17.4	17
tyrosine	7.7	8
phenylalanine	6.4	7
histidine	2.9	3
lysine	14.3	15
arginine	9.0	8
tryptophan	ND[d]	1
total		218

[a] Values quoted from von Bahr-Lindström et al. (1983). [b] Sum of aspartic acid and asparagine. [c] Sum of glutamic acid and glutamine. [d] ND, not determined.

The import presequences are both 32 amino acids in length. Neither is evidently extensively related in primary structure to other mitochondrial import sequences that have been described (Nicholson & Neupert, 1988). In general, these sequences are very diverse, but are characterized by having a net positive charge, and rarely contain acidic amino acids. Both of these features are present in the putative import sequences of the 24-kDa protein. Neither contains an acidic amino acid, and both presequences have a net positive charge of 7.5 (assuming values of +1 for each lysine and arginine and the N-terminal methionine and +0.5 for each histidine residue). It has also been proposed that import sequences contain a segment that can fold into an amphipathic α-helix, which can then interact with the mitochondrial membrane (Roise et al., 1986; von Heijne, 1986). The method of Garnier et al. (1978) of predicting secondary structures of proteins suggests that residues −32 to −18 have a strong potential for forming an α-helix, but helical wheel plots of this region do not reveal any convincing amphipathic features.

The arrangement of the various subunits within complex I has been studied by chemical labeling techniques. These show that the flavoprotein (FP) fraction, which contains the 24-kDa protein, is inaccessible to reaction with hydrophilic probes (Ragan, 1987), presumably therefore being shielded

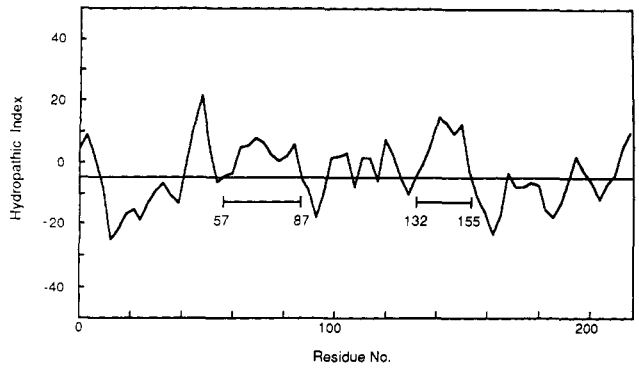

FIGURE 4: Hydrophobic profile of the mature bovine 24-kDa protein. This was calculated by the method of Kyte and Doolittle (1982) using a span length of 11 residues. Hydrophobic regions appear above the line (which denotes the average hydropathy of 84 fully sequenced soluble proteins). The positions of residues at the ends of hydrophobic regions are marked.

from reaction by other subunits or possibly by interaction with membrane phospholipids. On the other hand, the FP fraction can be released from the complex by chaotropes and is water soluble, and therefore it is unlikely that it contains extensive intrinsic membrane domains. The hydrophobic profile of the 24-kDa protein contains two regions of fairly high hydrophobicity (see Figure 4) that if folded into an α-helix would be long enough to span a membrane. However, the stretch between residues 57 and 87 contains three proline residues that would disrupt an α-helix, and the region from residue 132 to residue 155 also has prolines at positions 133 and 152 and a glutamic acid residue at position 141 that, in a transmembrane α-helix, would have to be accommodated within the lipid bilayer. So, these regions are unconvincing candidates for membrane spanning α-helices. It is perhaps more likely that they are found at hydrophobic interfaces with other proteins or that they form interior structures in the folded 24-kDa protein.

The presence of a binuclear iron–sulfur cluster in the 24-kDa protein (Ohnishi et al., 1985) has led to proposals about the possible role of its five cysteine residues in these centers and of the possible relationship of the 24-kDa protein to other iron–sulfur proteins, including the Rieske iron–sulfur protein of mitochondrial complex III (Harnisch et al., 1985). Ragan et al. (1986) have pointed out that the positioning of cysteine residues in the bovine 24-kDa protein shows some homology with the Rieske iron–sulfur protein of *Neurospora* (Harnisch et al., 1985). This presumably refers to the separation of four amino acids in each case of the cysteine pairs 103/108 in the former and 142/147 in the latter. Also, it has been suggested by von Bahr-Lindström et al. (1983) that regions of the 24-kDa protein show some relationship to the single subunit NADH dehydrogenase of *E. coli* (Young et al., 1981).

We have reexamined these proposals and, in addition, we have searched for relationships between the 24-kDa protein and other proteins by comparing its sequence with current entries in the PIR data base. The significance of relationships has been assessed by pairwise comparisons made with the computer program DIAGON. The most significant region of homology between the 24-kDa protein and the Rieske iron–sulfur protein of mitochondrial complex III detected by DIAGON encompasses the region of the 24-kDa protein that contains its second cysteine pair, 144/148, but the related sequence in the Rieske protein contains no cysteines (see Figure 5a). Homologies involving regions that contain cysteine pairs in both proteins were not found in this calculation. However, visual inspection of the immediate sequences around cysteine pairs 103/108 (24 kD) and 174/179 (Rieske Fe-S) indicates that the sequences are related weakly over about eight amino acids, but not beyond this short span of sequence (see Figure 5b). The proposed relationship between the 24-kDa protein and the NADH dehydrogenase from *E. coli* was not detected by DIAGON, and the most significant, albeit weak, homology involved other regions of the two proteins (see Figure 5c).

Comparison of the sequence of the bovine 24-kDa protein with protein sequences in the PIR data base revealed two further relationships that DIAGON confirms as being significant. These again concern the region around the cysteine pair

a)	Neurospora Rieske FeS	67-85	N M L F S Y F M V G T M G A I T A A G
	24 kD	134-152	D K L F T L I E V E C L G A C V N A P
b)	Neurospora Rieske FeS	173-180	V C T H L G C V
	24 kD	102-109	V C T T T P C M
c)	E.coli NADH dehydrogenase	340-353	C A M N N I L A Q M K G K P
	24 kD (bovine)	64-77	S A M N K V A E I L Q V P P
d)	FEMYFS	36-61	P D E C V D C G A C E P V C P V E A I Y Y E D D V P
	24 kD (bovine)	141-166	E V E C L G A C V N A P M V Q I N D N Y Y E D L T P
e)	D.yakuba ND 4L	88-96	D Y F Q S F S I M *
	Bovine ND 4L	88-98	D Y V Q N L N L L Q C *
	Human ND 4L	88-98	D Y V H N L N L L Q C *
	X.laevis ND 4L	88-98	D K L F S L N L L E C *
	24 kD	134-144	D K L F T L I E V E C

FIGURE 5: Regions of homology between the 24-kDa protein and other proteins. With the exception of (b) the relationships were detected by pairwise comparisons with DIAGON. A score of 280 and a window length of 25 were employed in these calculations. (a) Most significant region of homology with the *Neurospora* Rieske iron–sulfur protein. (b) Alignment of sequences around iron–sulfur pairs in Rieske iron–sulfur protein and bovine 24-kDa protein. This homology was detected by visual inspection. (c) Most significant region of homology with the *E. coli* NADH dehydrogenase (Young et al., 1981). This does not correspond with the region suggested by von Bahr-Lindström et al. (1983) to be related to the 24-kDa protein. In (d) and (e) regions of homology were found by comparison with entries in the PIR data base and assessed by pairwise comparison using DIAGON. (d) Alignment with FEMYFS, the 2Fe–2S ferredoxin from *M. smegmatis* (Hase et al., 1979). (e) Comparison with the ND4L proteins of *D. yakuba* (Clary & Wolstenholme, 1985), cow (Anderson et al., 1982), man (Anderson et al., 1981), and *X. laevis* (Roe et al., 1985). The mouse (Bibb et al., 1981) and rat (Grosskopf & Feldmann, 1981) sequences are identical with the bovine sequence in this region. A star (★) indicates the C terminus of the protein.

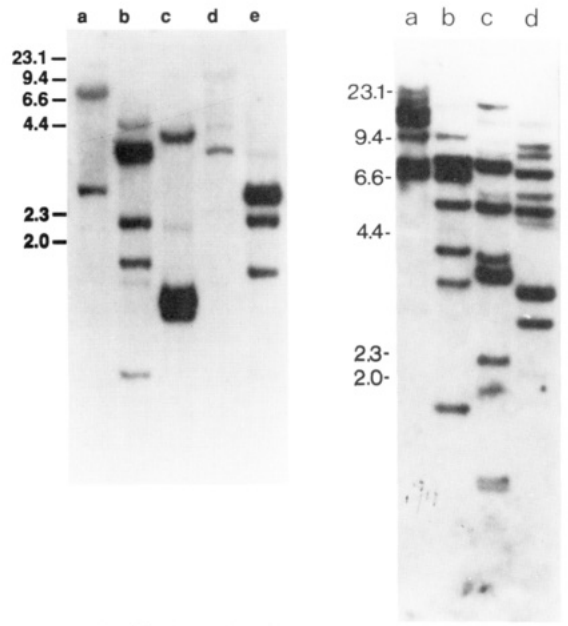

FIGURE 6: Hybridization of restriction digests of human and bovine genomic DNA with the cDNA for the bovine 24-kDa protein. A prime-cut probe extending between the terminal *Eco*RI and *Bam*HI sites of pBovCI-24.2 was employed. This sequence contains the entire coding region and the 3′ noncoding region. The human DNA was digested with the following restriction enzymes: (a) *Eco*RI; (b) *Hin*dIII; (c) *Pst*I; (d) *Sac*I; (e) *Xba*I. The bovine DNA was digested with (a) *Bam*HI, (b) *Eco*RI, (c) *Hin*dIII, and (d) *Pst*I. After hybridization, the filter was washed in 0.2× SSC at 65 °C. The positions of DNA size markers produced by digestion of DNA from bacteriophage λ with *Hin*dIII are shown at the left side of the filters.

144/148 in the 24-kDa protein. First, it is related to a region in the 7S ferredoxin from *Mycobacterium smegmatis* (Hase et al., 1979) that includes three cysteines that ligand the 4Fe-4S center (residues 39, 42, and 45) and one cysteine (residue 49) that is involved in the 3Fe center (Howard et al., 1983). Second, it is homologous to the C-terminal region of the ND4L protein in mammals, *Drosophila yakuba*, and *Xenopus laevis* (see Figure 5e). The ND4L protein is also a subunit of mitochondrial complex I. It is a small hydrophobic protein of unknown function that is encoded in mitochondrial DNA (Anderson et al., 1981, 1982; Chomyn et al., 1985). The sequence homology includes a C-terminal cysteine that is conserved in mammals and *Xenopus* but not in other species that have been investigated. The significance of these rather weak relationships between the 24-kDa protein and other iron–sulfur proteins is hard to assess at present, but the finding that both a ferredoxin and the Rieske protein are related to a region containing one of the cysteine pairs of the 24-kDa protein could be taken as indicating that both or one of the cysteines may be involved in the iron–sulfur cluster and that the structures of the three proteins may be related, at least in these regions.

Number of Bovine and Human Genes for the 24-kDa Protein of Complex I. To investigate the number of sequences in the bovine and human genomes with homology to the isolated cDNAs for the 24-kDa subunit, restriction digests of genomic DNA were hybridized with a probe that contained the entire sequence of the bovine cDNA. After washing at low and high stringencies, the same number of bands was observed. In the human case, two to five bands were present in the various digests, and seven to nine bands could be seen in the bovine digests. These experiments can be interpreted as indicating the presence of several sequences in each genome with a high degree of homology to the cDNA and as suggesting that there is more than one expressed gene or pseudogene for the 24-kDa component in both species. However, it is also likely that more than one band arises from the same gene, so estimates of the size of the gene family for the 24-kDa protein are imprecise. Investigations of other mammalian mitochondrial proteins including the proteolipid (Gay & Walker, 1985b; Dyer & Walker, 1988; Dyer et al., 1988) and α-subunits (Walker et al., 1988) of ATP synthase and ADP/ATP translocase (Battini et al., 1987; Walker et al., 1987a; Neckleman et al., 1987; Houldsworth et al., 1988; Powell et al., 1988) have shown that each has at least two, and in the last case at least three, expressed genes. The proteolipid has also been shown to have numerous spliced pseudo-genes in both man and cow (Dyer & Walker, 1988; Dyer et al., 1988). In all three examples, the related genes are expressed differently in various bovine tissues. In the cases of other mitochondrial proteins, Southern blotting experiments suggest that they also may have more than one human and bovine gene (Walker et al., 1987b). Should the 24-kDa component of complex I also prove to have two or more expressed human genes subject to different expression in various tissues, in principle this would provide an explanation of tissue-specific defects in this protein.

Added in Proof

The cDNA sequence of the 24-kDa subunit of rat NADH–ubiquinone reductase has been published recently [see Nishikimi et al. (1988)].

Acknowledgments

We thank Dr. T. H. Rabbitts for a sample of the human T-cell cDNA library, HUT 78, and T. V. Smith for synthesizing oligonucleotides.

Registry No. NADH–ubiquinone reductase, 9028-04-0; cDNA bovine clone, 119593-03-2; cDNA human 24-kDa clone, 119593-02-1; human 24-kDa precursor protein, 119593-07-6; human 24-kDa protein, 119593-08-7; bovine precursor protein, 119593-09-8; bovine protein, 119593-10-1.

References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. D., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., & Young, I. G. (1981) *Nature (London) 290*, 457–465.

Anderson, S., de Bruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F., & Young, I. G. (1982) *J. Mol. Biol. 156*, 683–717.

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

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–182.

Bibb, M. J., van Etten, R. A., Wright, C. T., Walberg, M. W., & Clayton, D. A. (1981) *Cell (Cambridge, Mass.) 26*, 167–180.

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

Chomyn, A., Mariottini, P., Cleeter, M. W. J., Ragan, C. I., Matsuno-Yagi, A., Hatefi, Y., Doolittle, R. F., & Attardi, G. (1985) *Nature (London) 314*, 592–597.

Clary, D. O., & Wolstenholme, D. R. (1985) *J. Mol. Evol. 22*, 252–271.

Duckworth, M. L., Gait, M. J., Goelet, P., Fang Hong, G., Singh, M., & Titmas, R. (1981) *Nucleic Acids Res. 9*, 1691–1706.

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).
Farrell, P. J., Deininger, P. L., Bankier, A., & Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1565–1569.
Galante, Y. M., & Hatefi, Y. (1979) *Arch. Biochem. Biophys. 192*, 559–568.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
Gay, N. J., & Walker, J. E. (1985a) *Biochem. J. 225*, 707–712.
Gay, N. J., & Walker, J. E. (1985b) *EMBO J. 4*, 3519–3524.
Gross-Bellard, M., Oudet, P., & Chambon, P. (1973) *Eur. J. Biochem. 36*, 32–38.
Grosskopf, R., & Feldmann, H. (1981) *Curr. Genet. 4*, 151–158.
Harnisch, U., Weiss, H., & Sebald, W. (1985) *Eur. J. Biochem. 149*, 95–99.
Hase, T., Wakabayashi, S., Matsubara, H., Imai, T., Matsumoto, T., & Tobari, J. (1979) *FEBS Lett. 103*, 224–228.
Heron, C., Smith, S., & Ragan, C. I. (1979) *Biochem. J. 181*, 435–443.
Houldsworth, J., & Attardi, G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 377–381.
Howard, J. B., Lorsbach, T. W., Ghosh, D., Melis, K., & Stout, C. D. (1983) *J. Biol. Chem. 258*, 508–522.
Huynh, T. V., Young, R. A., & Davis, R. (1985) in *DNA cloning: a practical approach* (Glover, D. M., Ed.) Vol. 1, pp 49–76, IRL Press, Oxford, U.K.
Ise, W., Haiker, H., & Weiss, H. (1985) *EMBO J. 4*, 2075–2080.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Morgan-Hughes, J. A., Hayes, D. J., Cooper, M., & Clark, J. B. (1985) *Biochem. Soc. Trans. 13*, 648–650.
Neckelmann, N., Li, K., Wade, R. P., Shuster, R., & Wallace, D. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7580–7584.
Nicholson, D. W., & Neupert, W. (1988) in *Protein transfer and organelle biogenesis* (Das, R. C., & Robbins, P. W., Eds.) pp 677–746, Academic Press, New York.
Nishikimi et al. (1988) *Biochem. Biophys. Res. Commun. 157*, 914–920.
Ohnishi, T., Ragan, C. I., & Hatefi, Y. (1985) *J. Biol. Chem. 260*, 2782–2788.
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.
Ragan, C. I. (1987) *Curr. Top. Bioenerg. 15*, 1–36.
Ragan, C. I., Ohnishi, T., & Hatefi, Y. (1986) in *Frontiers of Iron–Sulfur Protein Research*, pp 220–231, Japan Scientific Societies Press, Tokyo.
Roe, B. A., Ma, D. P., Wilson, R. K., & Wong, J. F. H. (1985) *J. Biol. Chem. 260*, 9759–9774.
Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H., & Schatz, G. (1986) *EMBO J. 5*, 1327–1334.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schapira, A. H. V., Cooper, J. M., Morgan-Hughes, J. A., Patel, S. D., Cleeter, M. J. W., Ragan, C. I., & Clark, J. B. (1988) *Lancet 8584*, 500–503.
Shaw, G., & Kamen, R. (1986) *Cell (Cambridge, Mass.) 46*, 659–667.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Staden, R. (1982a) *Nucleic Acids Res. 10*, 4731–4751.
Staden, R. (1982b) *Nucleic Acids Res. 10*, 2951–2961.
Staden, R. (1985) in *Genetic Engineering: Principles and Methods* (Setlow, J. K., & Hollaender, A., Eds.) Vol. 7, pp 67–114, Plenum Publishing, New York.
Suggs, S. V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J., Itakura, K., & Wallace, R. B. (1981) in *Developmental Biology Using Purified Genes* (Brown, D. D., Ed.) pp 683–693, Academic Press, New York.
Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.
von Bahr-Lindström, H., Galante, Y. M., Persson, M., & Jörnvall, H. (1983) *Eur. J. Biochem. 134*, 145–150.
von Heijne, G. (1986) *EMBO J. 5*, 1335–1342.
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).
Wikström, M. (1984) *FEBS Lett. 169*, 300–304.
Wilbur, W. J., & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 726–730.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103–119.
Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A., & Shaw, D. C. (1981) *Eur. J. Biochem. 116*, 165–170.