NADH:ubiquinone oxidoreductase from bovine heart mitochondria

cDNA sequences of the import precursors of the nuclear-encoded 39 kDa and 42 kDa subunits

Ian M. FEARNLEY, Moshe FINEL, J. Mark SKEHEL and John E. WALKER*

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

The 39 kDa and 42 kDa subunits of NADH:ubiquinone oxidoreductase from bovine heart mitochondria are nuclear-coded components of the hydrophobic protein fraction of the enzyme. Their amino acid sequences have been deduced from the sequences of overlapping cDNA clones. These clones were amplified from total bovine heart cDNA by means of the polymerase chain reaction, with the use of complex mixtures of oligonucleotide primers based upon fragments of protein sequence determined at the *N*-terminals of the proteins and at internal sites. The protein sequences of the 39 kDa and 42 kDa subunits are 345 and 320 amino acid residues long respectively, and their calculated molecular masses are 39115 Da and 36693 Da. Both proteins are predominantly hydrophilic, but each contains one or two hydrophobic segments that could possibly be folded into transmembrane α -helices. The bovine 39 kDa protein sequence is related to that of a 40 kDa subunit from complex I from *Neurospora crassa* mitochondria; otherwise, it is not related significantly to any known sequence, including redox proteins and two polypeptides involved in import of proteins into mitochondria, known as the mitochondrial processing peptidase and the processing-enhancing protein. Therefore the functions of the 39 kDa and 42 kDa subunits of complex I are unknown. The mitochondrial gene product, ND4, a hydrophobic component of complex I with an apparent molecular mass of about 39 kDa, has been identified in preparations of the enzyme. This subunit stains faintly with Coomassie Blue dye, and in many gel systems it is not resolved from the nuclear-coded 39 kDa subunit.

INTRODUCTION

NADH:ubiquinone oxidoreductase (or complex I) is a multi-subunit assembly found in the inner membranes of mitochondria (for reviews see Ragan, 1987; Weiss *et al.*, 1991). It transfers electrons from NADH to ubiquinone, and for each electron transferred two protons are translocated from the matrix across the mitochondrial inner membrane (Wikström, 1984). The bovine enzyme contains at least 30 different proteins (J. M. Skehel, M. Finel, I. M. Fearnley, M. J. Runswick & J. E. Walker, unpublished work), including the seven that are encoded in mitochondrial DNA (Chomyn *et al.*, 1985, 1986), and also one molecule of FMN and several Fe-S clusters. Two water-soluble subcomplexes, known as the flavoprotein and iron-sulphur protein fractions, can be split out of complex I with chaotropic agents. The water-insoluble residue is referred to as the hydrophobic protein fraction (Galante & Hatefi, 1979). The flavoprotein fraction, a subcomplex of three polypeptides (apparent molecular masses 51, 24 and 10 kDa), can transfer electrons from NADH to ferricyanide, but not to ubiquinone. The 51 kDa subunit appears to bind both FMN and NADH, and its sequence (Pilkington *et al.*, 1991b) and spectral evidence (Ohnishi *et al.*, 1985) suggest that it contains a [4Fe-4S] centre. Six polypeptides (apparent molecular masses 75, 49, 30, 18, 15 and 13 kDa) are present in the iron-sulphur protein fraction, and their sequences provide evidence for only one [4Fe-4S] cluster and possibly one [2Fe-2S] centre in the 75 kDa subunit (Fearnley *et al.*, 1989; Runswick *et al.*, 1989; Pilkington *et al.*, 1991a; S. J. Pilkington, M. J. Runswick & J. E. Walker, unpublished work). The hydrophobic protein fraction is the most complex of the three fractions. It has no known enzymic activity and contains at least 20 proteins. The primary structures of the 39 kDa and 42 kDa subunits, its two most prominent components in

Coomassie Blue-stained PAGE gels, are described below. They were obtained by partial protein sequence analysis of the subunits, and sequence analysis of cDNAs obtained with the aid of the polymerase chain reaction (PCR). The sequence of the 39 kDa subunit is related to that of a 40 kDa subunit found in complex I in *Neurospora crassa* (Röhlen *et al.*, 1991). Otherwise, neither sequence is related significantly to any known protein sequence, including those of iron-sulphur proteins, and of two polypeptides involved in import of proteins in mitochondria. Therefore the functions of the 39 kDa and 42 kDa subunits of complex I are obscure. Only one hydrophobic region with potential to form a membrane-spanning α -helix is present in the 39 kDa subunit, and possibly two are present in the 42 kDa subunit. The latter subunit is comparatively loosely bound and can be removed from the enzyme complex under mild conditions.

MATERIALS AND METHODS

Purification of subunits of complex I and protein sequence analysis

Complex I was isolated from bovine heart mitochondria (Hatefi, 1978; Ragan *et al.*, 1987). Its subunits were separated by PAGE in the presence of SDS, then transferred to a poly(vinylidene difluoride) membrane (Matsudaira, 1987) with a semi-dry blotting apparatus (Pilkington *et al.*, 1991b), and *N*-terminal protein sequences of both subunits were determined as described previously (Fearnley *et al.*, 1989; Dupuis *et al.*, 1991a,b; Pilkington *et al.*, 1991a,b). Samples of the 42 kDa and 39 kDa polypeptides were recovered from gels by electro-elution in 0.1 M-ammonium bicarbonate buffer, pH 8.1, containing 0.1% SDS. The proteins were precipitated from the eluate with 20 vol. of ethanol cooled to -20°C or, alternatively, low-molecular-mass impurities were extracted from the protein

* To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X59418 and X59419.

Fig. 1. Separation of the subunits of complex I from bovine mitochondria by PAGE in the presence of SDS

The gels contain a 10–20% gradient of acrylamide and were run with the buffers described by Laemmli (1970). The positions of molecular-mass marker proteins and their estimated molecular masses are indicated at the left-hand side, and the N-terminal sequences of various subunits are shown on the right.

solution with a chloroform/methanol (1:4, v/v) mixture (Wessel & Flügge, 1984). The purified proteins were dissolved in 70% (v/v) formic acid and treated either with CNBr at room temperature for 4–24 h or with N-bromosuccinimide at 55 °C for 2 h (Wachter & Werhahn, 1979; Runswick *et al.*, 1989). The digests of the proteins were fractionated on polyacrylamide minigels according to either the procedure of Laemmli (1970) or that of Schägger & von Jagow (1987), and then transferred to poly(vinylidene difluoride) membranes and their N-terminal sequences determined as above. Other samples of proteins

recovered from PAGE gels were treated with methanolic 1.5 M-HCl for 2 h at 37 °C before sequence analysis.

Isolation of RNA and cDNA synthesis

Total RNA was extracted from bovine heart tissue by using the procedure of Chirgwin *et al.* (1979), and polyadenylated mRNA was prepared from it with the use of mAP paper (Amersham International, Amersham, Bucks., U.K.; Wreschner & Herzberg, 1984). Double-stranded cDNA was synthesized by using polyadenylated mRNA with a cDNA synthesis kit from Amersham International (see Pilkington *et al.*, 1991a). Single-stranded cDNA was synthesized by using mRNA with a 36-base primer comprised of 15 T residues and a linker sequence containing the restriction-enzyme sites for *Bam*HI, *Hind*III and *Eco*RI. Alternatively, synthesis was primed with random hexanucleotides, and A residues were added to the 5'-end of the product with terminal transferase (Dupuis *et al.*, 1991a; J. E. Walker, A. Dupuis, I. M. Fearnley, M. Finel, S. M. Medd, S. J. Pilkington, M. J. Runswick & J. M. Skehel, unpublished work). These single-stranded cDNA preparations were used as templates in PCRs as described below.

Oligonucleotide synthesis

Oligonucleotides were synthesized by automated phosphoramidite chemistry with the aid of an Applied Biosystems 380B DNA synthesizer. Primers for use in PCRs were synthesized with nine bases of linker sequence at their 5'-ends in order to facilitate the cloning of products into M13 vectors. Forward primers included a restriction-enzyme site for *Eco*RI in the linker and reverse primers a *Hind*III site. Because of the presence of *Eco*RI, *Hind*III and *Bam*HI sites in the cDNA for the 42 kDa subunit some of the primers were made with a *Bam*HI site in their linker sequences.

Polymerase chain reactions

Amplifications of segments of cDNA by means of the PCR were performed as described previously (Runswick *et al.*, 1990a,b; Dupuis *et al.*, 1991a,b; J. E. Walker, A. Dupuis, I. M. Fearnley, M. Finel, S. M. Medd, S. J. Pilkington, M. J. Runswick & J. M. Skehel, unpublished work) except that they were carried out in two stages. Primer concentrations of 25 nM were used in the first

Table 1. Amino acid sequences determined on the 39 kDa and 42 kDa subunits of bovine heart complex I and on their fragments

X denotes that no amino acid was identified.

Subunit	Sample	Determined sequence	Position in sequence
39 kDa	Intact protein	LHHAVIPHGKGGRSSVSGIV	1–20
	N-Bromosuccinimide	ETQNFDVFVKIPQAIQV	100–121
	N-Bromosuccinimide	IGRLFXISPFQXPXT	266–279
	N-Bromosuccinimide	TTRDKVXRIHTTDKIL†	279–294
	N-Bromosuccinimide	NGRDKDSHRAVXH	74–87
	N-Bromosuccinimide	FGGVPLI	188–194
	CNBr	GDLGQIIFMD	63–72
	CNBr	GSQVIV	42–47
	CNBr*	PDARGKTFA	221–229
42 kDa	Intact protein	LQYGPLAYILGEKATKKM	1–18
	CNBr	LVQDKLEVLNXTXIPVFLPEVTGVA	262–286
	CNBr	YRQGFIRKQXVDHYNQ	139–154
	CNBr	KITXAYLQDI†	193–202
	N-Bromosuccinimide	EAQDAEXVVE†	227–236

* The peptide was produced by acid cleavage of the Asp–Pro peptide bond.
† The N-terminal sequence was also present.

35 cycles, and then 80 % of the products of this reaction mixture was amplified for a further 35 cycles with primer concentrations of 1 μ M. In one experiment concerning the 39 kDa subunit, two adjacent nested unique forward primers (4AF) and 4BF) were employed in the two rounds of amplification.

Bovine cDNA library screening

A bovine cDNA library (Gay & Walker, 1985) was screened with a unique oligonucleotide sequence from a partial cDNA for the 39 kDa subunit obtained by the PCR (see the Results and discussion section). Details of procedures employed in these experiments have been described previously (Runswick *et al.*, 1989).

DNA sequence analysis

DNA sequence analysis was performed by the modified dideoxy chain-termination method (Sanger *et al.*, 1977; Biggin *et al.*, 1983). The flanking primer LMB2 (Duckworth *et al.*, 1981) and 19 internal unique primers 17 bases in length were used in these experiments. The cDNA sequences of the 39 kDa and 42 kDa subunits of complex I were determined completely in both senses of the DNA, and, in addition, the sequences of both cDNAs were established in at least one direction by reactions containing 7-deaza-dGTP in place of dGTP (Mizusawa *et al.*, 1986) in order to remove compressions and artifactual bands from sequencing gels. One 'compression' (at position 157 in the sequence of the 39 kDa cDNA) was resolved by the use of dITP in place of dGTP (Mills & Kramer, 1979). Data were compiled and analysed with the aid of the computer programs Sequence Alignment Program (SAP) and Nucleotide Interpretation Program (NIP; Staden, 1990). At least three independent clones from each PCR were characterized and each nucleotide residue in the cDNAs for the 39 kDa and 42 kDa subunits was determined an average of 9.6 and 8.9 times respectively. Some point differences in sequence were observed between independent clones generated in PCRs. The sequences reported at these positions represent the majority consensus. In almost all cases this is greater than or equal to a 5:1 majority.

Data analysis

Protein sequences were compared with protein sequence databases by using FASTA (Lipman & Pearson, 1985). Hydrophobic profiles were calculated by using the SOAP program (Kyte & Doolittle, 1982) and pairwise comparison of proteins with DIAGON (Staden, 1982), these being contained within the programs Protein Interpretation Program (PIP) and Similarity Investigation Program (SIP) respectively (Staden, 1990).

RESULTS AND DISCUSSION

Protein sequence analyses

The 39 kDa and 42 kDa subunits were resolved from other subunits of complex I by PAGE (see Fig. 1). Unique *N*-terminal sequences were obtained for both polypeptides (see Table 1), and the same sequences were obtained after resolution of the complex on two-dimensional gels (results not shown). A number of other subunits of the enzyme were also identified from their *N*-terminal sequences (see Fig. 1). Internal sequences in both polypeptide chains were obtained from CNBr and *N*-bromosuccinimide digests (summarized in Table 1). Once the cDNA sequences had been established, it was apparent that the sequences observed in these digests had arisen from cleavages after three out of four methionine residues and five out of six tryptophan residues in the 39 kDa subunit, and after three out of five methionine residues

and one out of six tryptophan residues in the 42 kDa sequences (see Table 1). A fragment resulting from acid cleavage of an Asp-Pro bond in the 39 kDa subunit was present in the CNBr digest. In some experiments more than one sequence was observed (see Table 1), but eventually, once the sequences of the proteins had been completed, all of them could be accounted for.

Isolation of cDNAs encoding precursors of the 39 kDa and 42 kDa subunits of bovine complex I

Parts of the *N*-terminal sequences of both subunits were used to design degenerate oligonucleotides 17 bases long for use as hybridization probes in the isolation of cDNA clones from a bovine cDNA library (Gay & Walker, 1985). No clones were isolated for either subunit.

Fig. 2. Generation by the PCR and sequence analysis of cDNAs encoding the import precursors of the 39 kDa subunit (a) and the 42 kDa subunit (b) of complex I

The thick lines represent partial cDNAs obtained with the aid of the PCR, and are labelled according to the primers used in their amplification. The black boxes represent synthetic oligonucleotide primers that were used in the sequencing reactions, and horizontal arrows the directions and extents of the sequences that were determined. Except where indicated the inserts are flanked by *Eco*RI and *Hind*III restriction sites in the linker sequences of the primers used in PCRs. In (a) pCI-39.1 is a partial cDNA clone for the 39 kDa subunit (nucleotide residues 416–804) obtained from a bovine cDNA library. In (b) vertical bars with E, H and B are *Eco*RI, *Hind*III and *Bam*HI restriction sites respectively. H* denotes the sequence AAGTTT, a site also cleaved by *Hind*III.

Fig. 3. cDNA sequence and deduced protein sequence of the 39 kDa subunit of bovine mitochondrial complex I

The site of cleavage of the mitochondrial import precursor is indicated and the mature sequence is labelled from 1 to 345. Protein sequences determined by sequence analysis of the intact protein and of fragments obtained by cleavage with either CNBr or N-bromosuccinimide are shaded. Boxed protein sequences were used to design mixtures of oligonucleotides employed as primers in PCRs. Boxed nucleotide sequences are those of unique primers. For clarity, the box denoting the position of primer 3R is beneath the nucleotide sequence. Horizontal arrows pointing either leftwards or rightwards show that the primers were synthesized either to the sequence shown or to its complement respectively. A potential polyadenylation signal is underlined.

Therefore it was decided to use a cloning strategy based on the PCR that has been developed and used to characterize cDNAs for several other mitochondrial membrane proteins (see Runswick *et al.*, 1990a,b; Pilkington *et al.*, 1991a,b; Dupuis *et al.*, 1991a,b). This strategy has been described in detail elsewhere (Walker *et al.*, 1991), and requires knowledge of the sequence of a minimum of 18 consecutive amino acid residues in the protein of interest. In a first round of PCR a short cDNA

fragment is amplified, with the use of a complex cDNA population as template and degenerate oligonucleotide primers based on the N- and C-terminal hexapeptide extremities of the known amino acid sequence. The reaction products are cloned into an M13 vector, and clones containing the desired sequence are recognized by hybridization with a third degenerate synthetic oligonucleotide based on a hexapeptide sequence in the central portion of the known segment of amino acid sequence. Thus a

Fig. 4. cDNA sequence and deduced protein sequence of the 42 kDa subunit of complex I

The mature protein sequence is numbered from 1 to 320. Otherwise, the symbols are the same as those described in the legend to Fig. 3. The positions of the internal *Eco*RI, *Hind*III and *Bam*HI restriction sites are underlined.

short unique cDNA sequence is produced that encodes the segment of known amino acid sequence, although only the central part of this sequence between the two priming sites of the degenerate primers is accurate. Then this unique central region of DNA sequence is employed as primer in further PCRs to extend the sequence in both 5' and 3' directions. The simplest way to complete the sequence is to generate partial cDNAs extending to the 3'- and 5'-extremities in two further PCRs. In the former experiment the unique sequence in the short cDNA is employed as forward primer and oligo(dT) is used as reverse primer. In the latter experiment the primers are the complement of the unique sequence in the short cDNA and oligo(dT), with as

template a single-stranded cDNA population with poly(A) added to its 5'-end with terminal transferase. However, as described below, a related but more complex route was followed in the generation of the complete cDNA sequences of both the 39 kDa and 42 kDa subunits of complex I (see Fig. 2).

In the case of the 39 kDa subunit, inadequate *N*-terminal sequence was obtained, and so an internal segment of amino acid sequence (residues 100–121) from an *N*-bromosuccinimide cleavage product was used to design two degenerate primers 17 bases in length. These primers (1F and 1R) contain 64 and 192 different sequences respectively, and were employed in the amplification of a short sequence from total bovine heart cDNA, and

Table 2. Amino acid compositions and calculated molecular masses of the 39 kDa and 42 kDa subunits of bovine mitochondrial complex I

Amino acid	Amino acid composition (residues/molecule)	
	39 kDa subunit	42 kDa subunit
Aspartic acid	17	20
Asparagine	8	11
Threonine	17	12
Serine	19	18
Glutamic acid	22	25
Glutamine	8	18
Proline	23	15
Glycine	26	20
Alanine	21	17
Cysteine	1	5
Valine	31	26
Methionine	4	5
Isoleucine	28	14
Leucine	25	32
Tyrosine	10	20
Phenylalanine	20	10
Histidine	14	8
Lysine	20	29
Arginine	25	10
Tryptophan	6	5
Total	345	320
Calculated molecular mass	39 115 Da	36 693 Da

Fig. 5. Comparison of the sequences of the 39 kDa subunit of bovine complex I and the 40 kDa subunit of *N. crassa* complex I

The comparison was made with the computer program DIAGON. The points on the graph represent scores within a window of 25 amino acid residues that exceed a threshold score of 280.

oligonucleotide 1P based on amino acid residues 108–113 was used to identify the desired cDNA product. This was sequenced and the unique central portion of the sequence was used as a primer 2R with mixed primer 2F to generate a 331 bp cDNA encoding the protein sequence from the *N*-terminal region up to residue 112. In a third PCR a cDNA was produced that extended from primer 3F to the 3'-terminal poly(A) tail. Then a cDNA of approx. 320 bp extending towards the 5'-extremity was amplified from a single-stranded cDNA population tailed at its 5'-end with

a run of adenine residues. However, sequence analysis showed that it extended in a 5' direction only 31 bases beyond the start of the coding region for the mature protein. Therefore, in order to amplify sequences extending further in a 5' direction another PCR was carried out with a reverse primer (6R), which annealed closer the 5'-end of the known sequence. Finally, other regions of the cDNA for the 39 kDa subunit were made in order to confirm the sequence (see Figs. 2a and 3).

In other experiments the bovine cDNA library was re-examined for more extensive clones of the 39 kDa subunit, with as hybridization probe the short-segment cDNA from the first round of the PCR (nucleotide residues 420–478 in Fig. 3). At least 57 positively hybridizing clones were identified; ten of them were characterized partially, including one that contained the sequence of nucleotide residues 416–804 of the cDNA for the 39 kDa protein. This provided independent corroboration of this sequence. Further characterization of the pUC clones from the cDNA library was hindered by a deletion towards the 3'-end of the insert that included the flanking polylinker. This deletion had occurred in all ten clones examined. The complete cDNA sequence encoding the precursor of the 39 kDa subunit was compiled from the overlapping partial cDNAs obtained with the aid of the PCR together with the partial cDNA clone derived from a library (Fig. 3). The final sequence is 1263 nucleotide residues in length and contains a potential polyadenylation signal, ATTAAA (Proudfoot & Brownlee, 1976), 14 bases to the 5' side of the site of poly(A) addition.

The strategy used for cloning cDNAs encoding the 42 kDa subunit (Fig. 2b) was similar to that described above for the 39 kDa subunit. However, experiments were complicated by the presence in the cDNA sequence of restriction-enzyme sites also present in the linker sequences of oligonucleotide primer. The degenerate oligonucleotide primers (1F and 1R) for the first PCR were designed from the segment of 18 amino acid residues of *N*-terminal sequence (Table 1). As described above for the 39 kDa subunit, a short cDNA of 54 nucleotide residues corresponding to bases 100–153 in the completed sequence was amplified from total cDNA, cloned into M13 vectors and selected by plaque hybridization with an oligonucleotide probe (1P) corresponding to amino acid residues 7–12 of the 42 kDa subunit (Fig. 2b). Then part of the unique 20-base sequence between the priming sites was used to design a 17-base forward primer (2F). This was used to amplify a cDNA between primer 2F and a degenerate primer 2R containing a mixture of 128 sequences designed from an amino acid sequence obtained from CNBr cleavage of the intact protein. This generated a cDNA sequence of 431 bp, and provided the basis of further unique oligonucleotide primers 3F, 4F and primer 4R. Two adjacent primers 3F and 4F were used with primer 3R, in independent PCRs, to isolate and sequence the remaining cDNA sequence to the polyadenylation site. Primer (4R) was used on a single-stranded cDNA template to obtain the cDNA sequence 5' to the region encoding the mature protein. As shown in Fig. 2(b), primer 4R was synthesized with a *Bam*HI linker to enable the product to be cloned into M13 vector in one piece and the nucleotide sequence obtained across both the internal *Eco*RI and *Hind*III restriction sites in the sequence. The complete sequence (Fig. 4) is 1335 nucleotide residues in length and a typical polyadenylation signal, AATAAA, is found 14 bases to the 5' side of the poly(A) tail (Proudfoot & Brownlee, 1976).

Protein sequences of the 39 kDa and 42 kDa subunits

The mature 39 kDa and 42 kDa subunits of complex I contain 345 and 320 amino acid residues and have calculated molecular masses of 39 115 Da and 36 693 Da respectively (see Table 2). Therefore the former, but not the latter, agrees with the value

Fig. 6. Alignment of the protein sequence of the 39 kDa subunit of bovine complex I with that of the 40 kDa subunit of *N. crassa* complex I

The alignment is based upon the comparison of the proteins made with the computer program DIAGON (see Fig. 5). Dashes indicate insertions that have been introduced to improve the alignment, and asterisks denote identical amino acid residues in both sequences. The *N. crassa* complex I sequence is taken from Röhlen *et al.* (1991).

estimated by PAGE. Discrepancies between molecular masses calculated from sequences and apparent molecular masses estimated by gel electrophoresis have been noted with other polypeptides in complex I. For example, the 51 kDa subunit of the FP fraction has a calculated molecular mass of 48416 Da (Pilkington *et al.*, 1991b), and that calculated for the 30 kDa subunit is 26432 Da (Pilkington *et al.*, 1991a). One explanation for such discrepancies may be the persistence of secondary structure in the protein in the presence of SDS. It should be noted also that estimates of molecular masses by gel electrophoresis may depend upon the gel system used; for example, the 39 kDa and 42 kDa subunits co-migrate in the gel system described by Schägger & von Jagow (1987), as do the 49 kDa and 51 kDa polypeptides of complex I. The molecular mass of the 30 kDa subunit has been measured by electrospray m.s. and the value obtained is in excellent agreement with that calculated from the sequence (I. M. Fearnley & J. E. Walker, unpublished work). The calculated values for the molecular masses of the 39 kDa and 42 kDa subunits have yet to be confirmed by the same method.

The 5' region of the cDNA sequence of the 39 kDa subunit contains two ATG codons in phase with the sequence coding for the mature protein. In the absence of further information it is assumed that the ATG closest to the 5'-end is the translational initiator, and if this is so the mature protein is preceded at its *N*-terminus by a pre-sequence 35 amino acid residues. Similarly, the 42 kDa subunit has an *N*-terminal extension of 23 amino acid residues. The proposed initiation codon in the latter case can be identified with more certainty, as an in-phase stop codon is found

at nucleotide residues 7–9. These *N*-terminal extensions have features that suggest that they act as mitochondrial import sequences; they have a net basic charge, a basic residue is found close to the junction with the mature protein, and neither contains any acidic residues (von Heijne *et al.*, 1989; Roise *et al.*, 1986).

Possible functions of the 39 kDa and 42 kDa subunits

The 39 kDa and 42 kDa subunits contain one and five cysteine residues respectively (see Table 2), and the cysteine-rich sequences associated with some binuclear and tetranuclear Fe-S centres (Yasunobu & Tanaka, 1980; Stout, 1982; Cammack, 1983) are not found in either protein. Therefore it is rather unlikely that either protein contains such redox centres. Also, their sequences were compared with sequences in the PIR, EMBL and GenBank libraries, but no significant relationships were found. Therefore the proposal that a 42 kDa subunit of bovine heart complex I is ferrochelatase (Taketani *et al.*, 1986) is not substantiated, as there is no relationship between the sequences of mouse or yeast ferrochelatase (Labbe-Bois, 1990; Taketani *et al.*, 1990) and either the 39 kDa subunit or the 42 kDa subunit of complex I. Homologues of other nuclear encoded subunits of complex I are encoded in chloroplast genomes of higher plants (Fearnley *et al.*, 1989; Dupuis *et al.*, 1991a; Pilkington *et al.*, 1991b), in mitochondrial genomes of lower eukaryotes (Pritchard *et al.*, 1990), in kinetoplast maxicircles of trypanosomes and *Leishmania tarantolae* and in prokaryote operons for hydrogenases (Böhm *et al.*, 1990; Tran-Betcke *et al.*, 1990; Pilkington *et al.*, 1991b). Therefore the sequences of the 39 kDa and 42 kDa subunits were

Fig. 7. Hydrophobic profiles of the 39 kDa subunit (a) and the 42 kDa subunit (b) of bovine complex I

The calculations were made with HYDROPLLOT, with a span of 11 amino acid residues.

also compared with all translational reading frames more than 15 amino acid residues long present in the six phases of the chloroplast DNAs of rice, tobacco and liverwort, and in the mitochondrial genome of *Paramecium*. No significant matches were detected.

The sequence of the bovine 39 kDa subunit is related to that of the 40 kDa subunit of complex I from *Neurospora crassa* mitochondria (see Figs. 5 and 6), and therefore they are likely to have similar functions. The proteins are identical in 117 positions (approx. 34%), and the relationship is strongest in the *N*-terminal halves of the proteins. It has been proposed that the sequence of the *Neurospora* 40 kDa protein is related to those of the mitochondrial processing peptidase and the processing-enhancing peptidase (Röhlen *et al.*, 1991). These sequence-related proteins co-operate in the proteolytic cleavage of mitochondrial precursor proteins (Pollock *et al.*, 1988; Hawlitschek *et al.*, 1988), and belong to a family of proteins that includes core proteins I and II of mitochondrial cytochrome *c* reductase (Schulte *et al.*, 1989). We have re-examined the proposed sequence similarity between the *Neurospora* 40 kDa protein and the mitochondrial processing peptidase and the processing-enhancing peptidase by using the computer program DIAGON, and we did not find any statistically significant extensive relationship of the 40 kDa subunit with either protein, nor can we find any evidence of a relationship between either the 39 kDa or the 42 kDa subunits of the bovine enzyme with the mitochondrial processing peptidase and processing-enhancing peptidase. Therefore the roles of these two proteins in bovine complex I remain obscure.

On the basis of photoaffinity labelling studies with an analogue of NADH, the 42 kDa subunit has been proposed to be responsible for a transhydrogenase activity in the complex that is distinct from that of the mitochondrial 120 kDa transhydrogenase protein (Chen & Guillory, 1984). However, the 42 kDa subunit was not identified unambiguously as the site of reaction; in the

analysis of the reaction products by gel electrophoresis, the 42 kDa and 39 kDa proteins were not well resolved, and, in addition, it is now known that the mitochondrial gene product, ND4, also runs in the same region of the gel as these two proteins (see below). The sequence of the 42 kDa subunit provides no indication of the presence of an NAD-binding site; the glycine-rich sequence motif often associated with the binding site of the ADP moiety of NAD is not present, and the 42 kDa subunit is not related to either the 51 kDa subunit of complex I (Pilkington *et al.*, 1991a) or *Escherichia coli* NADH dehydrogenase (Young *et al.*, 1981).

Locations of the 39 kDa and 42 kDa subunits in complex I

The 39 kDa and 42 kDa proteins are present in the hydrophobic protein fraction of complex I, but their hydropathy profiles (Fig. 7) indicate that both of them are likely to be predominantly globular proteins. The 39 kDa subunit contains one hydrophobic segment of sequence (residues 15–37) of sufficient length to form a transmembrane α -helix, but in the 40 kDa subunit of *N. crassa* complex I the analogous sequence is rather less hydrophobic, and so it is doubtful that it represents a membrane-embedded region in either protein. A 39 kDa component of complex I was found to react with the membrane-penetrant photo-affinity label arylazidophosphatidylcholine (Earley & Ragan, 1981). However, it is possible that the labelled protein was the hydrophobic product of mitochondrial gene ND4 (see below) and not the nuclear coded 39 kDa protein described in the present paper. The hydropathy profile of the 42 kDa subunit suggests that it has two relatively hydrophobic regions (amino acid residues 157–178 and 263–284). Both of them contain at least two proline residues and also a number of charged residues, and neither segment is a convincing candidate for a transmembranous α -helix. Therefore it appears that both proteins lie mostly outside the lipid bilayer, as chemical labelling studies with hydrophilic probes also suggest (Ragan, 1987; Patel *et al.*, 1988). The 42 kDa subunit is rather loosely bound to complex I, and can be removed in the presence of mild detergents without evidently disrupting the rest of the complex (M. Finel, unpublished work).

Nuclear-coded 39 kDa subunit and the ND4 protein

The bovine mitochondrial ND4 gene product is a hydrophobic protein with a calculated molecular mass of 52 103 Da (Anderson *et al.*, 1982). It has been purified from chloroform/methanol extracts of bovine mitochondria, and shown to have an apparent molecular mass on PAGE of about 39 kDa (Fearnley & Walker, 1987). A product with a molecular mass of 39 kDa has also been observed in bovine mitochondria radiolabelled in the presence of cycloheximide (Gibb & Ragan, 1990). Therefore it has been proposed that the 39 kDa band observed in Coomassie Blue-stained gels of complex I and the hydrophobic protein fraction is a mitochondrial gene product (Ragan, 1987), whereas we have characterized a nuclear-coded subunit of complex I from the 39 kDa band. This apparent contradiction has been resolved by recovering the 39 kDa band from complex I and conducting *N*-terminal sequence analysis after a treatment of the recovered protein with methanolic HCl. As expected, this revealed a sequence originating from the nuclear coded 39 kDa subunit, and also a second sequence corresponding to that of the *N*-terminus of the ND4 gene product. Therefore it appears that in some gels the nuclear coded 39 kDa subunit and the ND4 protein are not resolved from each other. Under slightly different experimental conditions a faintly staining band is visible below the heavily staining band denoted as the 39 kDa subunit (see Fig. 1), and this is probably the ND4 protein. A sequence originating from another mitochondrial gene product was not found in a

methanolic-HCl-treated sample of the 42 kDa band, thereby making the proposal that the 42 kDa band contains a mitochondrial gene product (Ragan, 1987) rather implausible. The failure to distinguish between the ND4 protein and the nuclear coded 39 kDa protein, and in some cases the 42 kDa subunit also, raises doubts mentioned above about interpretations that have been made of chemical labelling studies of complex I in which proteins with molecular-mass values around 39 kDa have been found to react (Earley & Ragan, 1981; Chen & Guillory, 1984).

M.F. was supported by a long-term E.M.B.O. Fellowship. We are grateful to Professor H. Weiss for sending a preprint of a paper about the *N. crassa* 40 kDa subunit.

REFERENCES

- 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
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
- Böhm, R., Sauter, M. & Böck, A. (1990) *Mol. Microbiol.* **4**, 231–243
- Cammack, R. (1983) *Chem. Scr.* **21**, 87–95
- Chen, S. & Guillory, R. J. (1984) *J. Biol. Chem.* **259**, 5124–5131
- Chirgwin, J. M., Przbyla, A. F., MacDonald, A. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- 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
- Chomyn, A., Cleeter, M. W. J., Ragan, C. I., Riley, M., Doolittle, R. F. & Attardi, G. (1986) *Science* **234**, 614–618
- Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G. F., Singh, M. & Titmas, R. (1981) *Nucleic Acids Res.* **9**, 1691–1706
- Dupuis, A., Skehel, J. M. & Walker, J. E. (1991a) *Biochemistry* **30**, 2954–2960
- Dupuis, A., Skehel, J. M. & Walker, J. E. (1991b) *Biochem. J.* **277**, 11–15
- Earley, F. G. P. & Ragan, C. I. (1981) *FEBS Lett.* **127**, 45–47
- Fearnley, I. M. & Walker, J. E. (1987) *Biochemistry* **26**, 8247–8251
- Fearnley, I. M., Runswick, M. J. & Walker, J. E. (1989) *EMBO J.* **8**, 665–672
- Galante, Y. M. & Hatefi, Y. (1979) *Arch. Biochem. Biophys.* **192**, 559–568
- Gay, N. J. & Walker, J. E. (1985) *Biochem. J.* **225**, 707–712
- Gibb, G. M. & Ragan, C. I. (1990) *Biochem. J.* **265**, 903–906
- Hatefi, Y. (1978) *Methods Enzymol.* **53**, 11–14
- Hawltischek, G., Schneider, H., Schmidt, B., Tropschug, M., Hartl, F. U. & Neupert, W. (1988) *Cell* **53**, 795–806
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Labbe-Bois, R. (1990) *J. Biol. Chem.* **265**, 7278–7283
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441
- Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038
- Mills, D. R. & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2232–2235
- Mizusawa, S., Nishimura, S. & Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319–1324
- Ohnishi, T., Ragan, C. I. & Hatefi, Y. (1985) *J. Biol. Chem.* **260**, 2782–2788
- Patel, S. D., Cleeter, M. W. J. & Ragan, C. I. (1988) *Biochem. J.* **256**, 529–535
- Pilkington, S. J., Skehel, J. M. & Walker, J. E. (1991a) *Biochemistry* **30**, 1901–1908
- Pilkington, S. J., Skehel, J. M., Gennis, R. B. & Walker, J. E. (1991b) *Biochemistry* **30**, 2166–2175
- Pollock, R. A., Hartl, F. U., Cheng, M. Y., Ostermann, J., Horwich, A. & Neupert, W. (1988) *EMBO J.* **7**, 3493–3500
- Pritchard, A. E., Seilhamer, J. J., Sable, C. L., Venuti, S. E. & Cummings, D. J. (1990) *Nucleic Acids Res.* **18**, 173–180
- 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., Wilson, M. T., Darley-Usmar, V. M. & Lowe, P. N. (1987) in *Mitochondria: A Practical Approach* (Darley-Usmar, V. M., Rickwood, D. & Wilson, M. T., eds.), pp. 79–112, IRL Press, Oxford and Washington
- Röhlen, D. A., Hoffmann, J., van der Pas, J. C., Nehls, U., Preis, D., Sackman, U. & Weiss, H. (1991) *FEBS Lett.* **278**, 75–78
- Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H. & Schatz, G. (1986) *EMBO J.* **5**, 1327–1334
- Runswick, M. J., Gennis, R. B., Fearnley, I. M. & Walker, J. E. (1989) *Biochemistry* **28**, 9452–9459
- Runswick, M. J., Medd, S. M. & Walker, J. E. (1990a) *Biochem. J.* **266**, 421–426
- Runswick, M. J., Walker, J. E., Bisaccia, F., Iacobazzi, V. & Palmieri, F. (1990b) *Biochemistry* **29**, 11033–11040
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Schägger, H. & von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
- Schulte, U., Arretz, M., Schneider, H., Tropschug, M., Wachter, E., Neupert, W. & Weiss, H. (1989) *Nature (London)* **339**, 147–149
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961
- Staden, R. (1990) *CABIOS* **4**, 387–393
- Stout, C. D. (1982) in *Iron Sulphur Proteins* (Spiro, T. G., ed.), pp. 97–146, John Wiley and Sons, New York
- Taketani, S., Tanaka-Yoshioka, A., Masali, R., Tashito, Y. & Tokunaga, R. (1986) *Biochim. Biophys. Acta* **833**, 277–283
- Taketani, S., Nakahashi, Y., Osumi, T. & Tokunaga, R. (1990) *J. Biol. Chem.* **265**, 19377–19380
- Tran-Betcke, A., Warnecke, U., Böcker, C., Zabarosch, C. & Friedrich, B. (1990) *J. Bacteriol.* **172**, 2920–2929
- von Heijne, G., Steppuhn, J. & Herrmann, R. G. (1989) *Eur. J. Biochem.* **180**, 535–545
- Wachter, E. & Werhahn, R. (1979) *Anal. Biochem.* **97**, 56–64
- Weiss, H., Friedrich, T., Hofhaus, G. & Preis, D. (1991) *Eur. J. Biochem.* **197**, 563–576
- Wessel, D. & Flügge, U. I. (1984) *Anal. Biochem.* **138**, 141–143
- Wikström, M. (1984) *FEBS Lett.* **169**, 300–304
- Wreschner, D. H. & Herzberg, M. (1984) *Nucleic Acids Res.* **12**, 1349–1359
- Yasunobu, K. & Tanaka, M. (1980) *Methods Enzymol.* **69**, 228–238
- Young, I. G., Rogers, B. L., Campbell, H. G., Jaworowski, A. & Shaw, D. C. (1981) *Eur. J. Biochem.* **116**, 165–170

Received 27 February 1991/27 March 1991; accepted 2 April 1991