The EMBO Journal vol.8 no.3 pp.665-672, 1989

A homologue of the nuclear coded 49 kd subunit of bovine mitochondrial NADH-ubiquinone reductase is coded in chloroplast DNA

Ian M.Fearnley, Michael J.Runswick and John E.Walker

The Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Communicated by J.E.Walker

The mitochondrial NADH-ubiquinone reductase (complex I) is an assembly of ~26 different polypeptides. In vertebrates and invertebrates, seven of its subunits are the products of genes in the mitochondrial DNA, and homologues of these genes have been found previously in the chloroplast genomes of *Marchantia polymorpha* and *Nicotiana tabacum*, although their function in the chloroplast is unknown. The remainder of the subunits of the mitochondrial complex are nuclear gene products that are imported into the organelle, amongst them the 49 kd subunit, a component of the iron-sulphur subcomplex of the enzyme. In the present work, the N-terminal sequence of this protein has been determined, and this has been used to design two mixtures of synthetic oligonucleotides, each containing 32 different sequences 17 bases long. These mixtures have been used as hybridization probes to isolate cDNA clones from a bovine library. The DNA sequences of these clones have been determined and they encode the mature 49 kd protein, with the exception of amino acids 1 and 2. The protein sequence of 430 amino acids is closely related to those of proteins that are encoded in open reading frames (ORFs) present in the chloroplast genomes of *M.polymorpha* and *N.tabacum*. Only one cysteine is conserved and the sequences provide no indication that the 49 kd protein contains iron-sulphur centres. These ORFs are found in the single copy regions of chloroplast DNA in close proximity to four of the homologues of the mammalian mitochondrial genes that encode subunits of complex I. Thus, it appears that both of the chloroplast DNAs contain a cluster of genes encoding proteins that probably assemble together in the same enzyme complex. This chloroplast enzyme complex is likely to be closely related in structure to the mitochondrial NADH-ubiquinone reductase. A plausible candidate is NADH-plastoquinone reductase. Its presence has been detected in chloroplasts of *Chlamydomonas reinhardtii*, and it participates in a respiratory electron transport pathway, which would therefore appear to be present also in plant chloroplasts.

Key words: NADH-ubiquinone reductase/iron-sulphur complex/49 kd subunit

Introduction

Mitochondrial NADH-ubiquinone reductase (complex I) is found in the inner membrane and is the first enzyme in the respiratory chain. It catalyses the oxidation of NADH and transfers electrons to ubiquinone (reviewed by Ragan, 1987), and for each electron transferred between NADH and ubiquinone two protons are pumped from the mitochondrial matrix (Wikström, 1984). Complex I is aptly named, and preparations of the enzyme isolated from both mammalian and fungal sources appear to contain ~26 different polypeptides, as well as FMN and six to eight iron-sulphur centres as redox groups (Heron *et al.*, 1979; Ise *et al.*, 1985; Ragan, 1987). Seven hydrophobic subunits of the enzyme are encoded in mitochondrial DNA (Chomyn *et al.*, 1985, 1986), and the remainder are presumed to be nuclear gene products that are imported into the organelle. The enzyme can be broken down in the presence of perchlorate to yield two distinct soluble subcomplexes, the flavoprotein (FP) and iron-sulphur (IP) fragments (Galante and Hatefi, 1979). The former contains all of the FMN, six iron atoms and three polypeptides (51, 24 and 10 kd) in equal molar amounts. The IP fragment is not homogeneous, but it contains 9-10 of the total of >20 iron atoms per FMN in complex I, and has a low flavin content that probably arises because of contamination with the FP fragment. This contamination by the FP fragment is confirmed by the presence of minor amounts of the 51, 24 and 10 kd polypeptides; the major protein components of the IP fragment are the 75, 49, 30, 18, 15 and 13 kd polypeptides, and variable amounts of other minor components are also seen (Ragan, 1987). The 49, 30 and 13 kd subunits together form a simpler subcomplex that contains a tetranuclear and a binuclear iron-sulphur cluster, but attempts to resolve this subcomplex resulted in denaturation, and prevented assignment of the clusters to specific polypeptides. As described in this paper we have determined the primary structure of the 49 kd polypeptide by a combination of direct protein sequence analysis and cDNA cloning and sequencing. This sequence is homologous to proteins that are encoded in unidentified reading frames in chloroplast DNA in *Nicotiana tabacum* (Shinozaki *et al.*, 1986) and *Marchantia polymorpha* (Ohyama *et al.*, 1986). The presence in the chloroplast genome of potential genes for homologues of the components of complex I that are encoded in mitochondrial DNA has been noted previously (Ohyama *et al.*, 1986; Shinozaki *et al.*, 1986; Kohchi *et al.*, 1988), but the present work now provides an example of a gene in the plastid DNA encoding a homologue of a nuclear-coded component of the mitochondrial enzyme. This finding lends support to the view that chloroplasts contain an enzyme complex that is closely related to the mitochondrial complex I and that participates in a chlororespiratory pathway (Bennoun, 1982; Umesono and Ozeki, 1987).

Results and discussion

Amino-terminal sequence of the 49 kd subunit

The subunits of complex I were separated by polyacrylamide gel electrophoresis in the presence of SDS and were

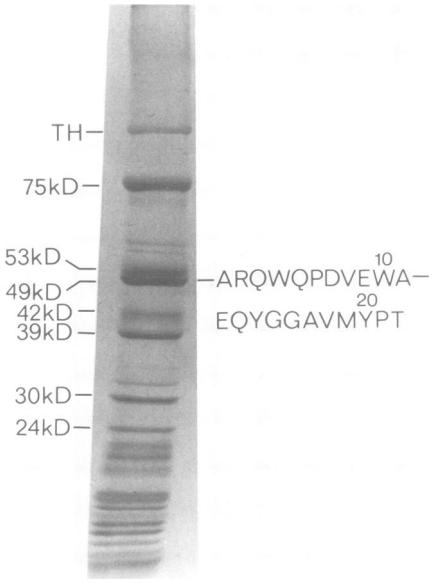

Fig. 1. The separation of the subunits of complex I from bovine heart mitochondria by polyacrylamide electrophoresis under denaturing conditions, and the N-terminal sequence of the 49 kd subunit. The proteins shown in the photograph were stained with Coomassie blue dye. The sequence was determined on material obtained by electrotransfer of the separated subunits of the complex to a polyvinylidene difluoride membrane (see Materials and methods for details). Some of the subunits of complex I are indicated according to their estimated sizes, and the position of the contaminant transhydrogenase (TH) is also shown.

transferred by electrophoresis to a polyvinylidene difluoride membrane. The stained band corresponding to the 49 kd subunit was excised and then the protein was subjected to the automated Edman degradation. This generated the N-terminal sequence of residues 1–22 (see Figure 1). In order to confirm that this sequence did indeed originate from the 49 kd component, the experiment was repeated with the 49 kd protein present in the IP fragment of complex I. The same protein sequence was obtained, thereby excluding the possibility that the sequence originated from a contaminating protein present in complex I that co-migrated in the polyacrylamide gel with the 49 kd subunit. Nor does it originate from the 51 kd subunit, which has a distinctly different N-terminal sequence (I.M.Fearnley and J.E. Walker, unpublished results), from the α or β subunit of ATP synthase (Walker *et al.*, 1985) which are contaminants present in preparations of complex I, or from the core proteins of complex III (González-Halphen *et al.*, 1988) which also migrate in this region of denaturing polyacrylamide gels.

Cloning and DNA sequence analysis

Residues 3–8 and 9–14 of the protein sequence were used to design two mixtures of oligonucleotides, each containing 32 different sequences. These are referred to as probes 1 and 2, and they were used to screen a bovine cDNA library in the plasmid pUC8. Of the five clones that hybridized strongly to both mixtures of oligonucleotides, four rescreened with the same two probes. Restriction analysis showed that two of them, named pBovCI-49.1 and pBovCI-49.2, appeared to be identical. The third isolate pBovCI-49.3 was shown subsequently to encode malate dehydrogenase (I.M.

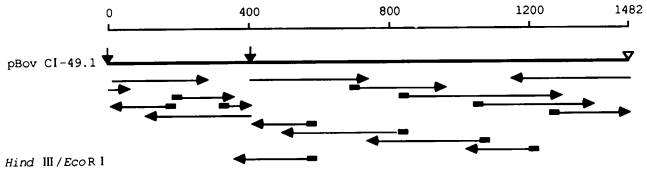

Fig. 2. Sequence analysis of cDNA clones encoding the 49 kd subunit of NADH-ubiquinone reductase from bovine mitochondria. The thick line represents the inserted DNA in the isolate pBovCI-49.1. It is flanked in the polylinker of the vector by a *Bam*HI site, ↓, (and an adjacent *Hin*dIII site) and an *Eco*RI site, ∇; there is a second *Bam*HI site within the insert. The horizontal arrows indicate the extent and direction of the sequences that were determined, and the positions of hybridization of the synthetic oligonucleotides that were employed as primers in the sequencing reactions are denoted by black rectangles. The scale is in bases.

Fearnley, unpublished results), and the fourth has not been characterized further. Digestion of plasmids pBovCI-49.1 and pBovCI-49.2 with *Eco*RI and *Bam*HI together released a *Bam*HI fragment and an *Eco*RI–*Bam*HI fragment in each case. The latter fragments were all ~1 kb in length, and the former were ~400 bp. These restriction fragments were cloned into appropriate vectors and sequences at their 5′ and 3′ ends were determined. The fragments derived from isolate pBovCI-49.1 were completely sequenced in both senses of the DNA by the use of synthetic oligonucleotide primers in the sequencing reactions. These experiments showed that the *Bam*HI fragment contained a sequence encoding the known N-terminal sequence of the 49 kd subunit of complex I, except for amino acids 1 and 2, and that the *Bam*HI–*Eco*RI fragment has a poly(A) sequence at the end distal from the *Bam*HI site, confirming that the *Bam*HI fragment represents the 5′ end of the mRNA. In order to establish the overlap across the internal *Bam*HI site the inserted DNA from pBovCI-49.1 was excised as one fragment with *Hin*dIII and *Eco*RI. The fragment was cloned into M13mp19 that had been digested with the same two restriction enzymes, and the sequence across the *Bam*HI site was determined by the use of an appropriate primer in the sequencing reactions. The sequences at the ends of the *Eco*RI–*Bam*HI fragment from isolate pBovCI-49.2 were identical to those of the fragment from the isolate pBovCI-49.1, and no evidence was obtained of different cDNA clones encoding different, but closely related, isoforms of the 49 kd protein, such as has been found for a number of other mitochondrial proteins (Gay and Walker, 1985a; Walker *et al.*, 1987a; Powell *et al.*, 1989; Walker *et al.*, 1989), although this remains a possibility.

The final sequence of the bovine cDNA is shown in Figure 3. It is 1482 bases long and lacks the first five bases at the 5′ end of the coding region of the mature 49 kd subunit of mitochondrial complex I. At its 3′ end is found the sequence A_{20} separated by 18 bases from the sequence AATAAA, a typical signal for polyadenylation of mRNA (Proudfoot and Brownlee, 1976).

Sequence of the 49 kd protein

Taken together, the cDNA sequence and the N-terminal sequence of amino acids 1–22 of the 49 kd subunit of complex I provide the primary structure of the protein, and over residues 3–22 the data from protein and cDNA sequencing experiments are entirely in agreement with each other. The mol. wt of the protein calculated from the

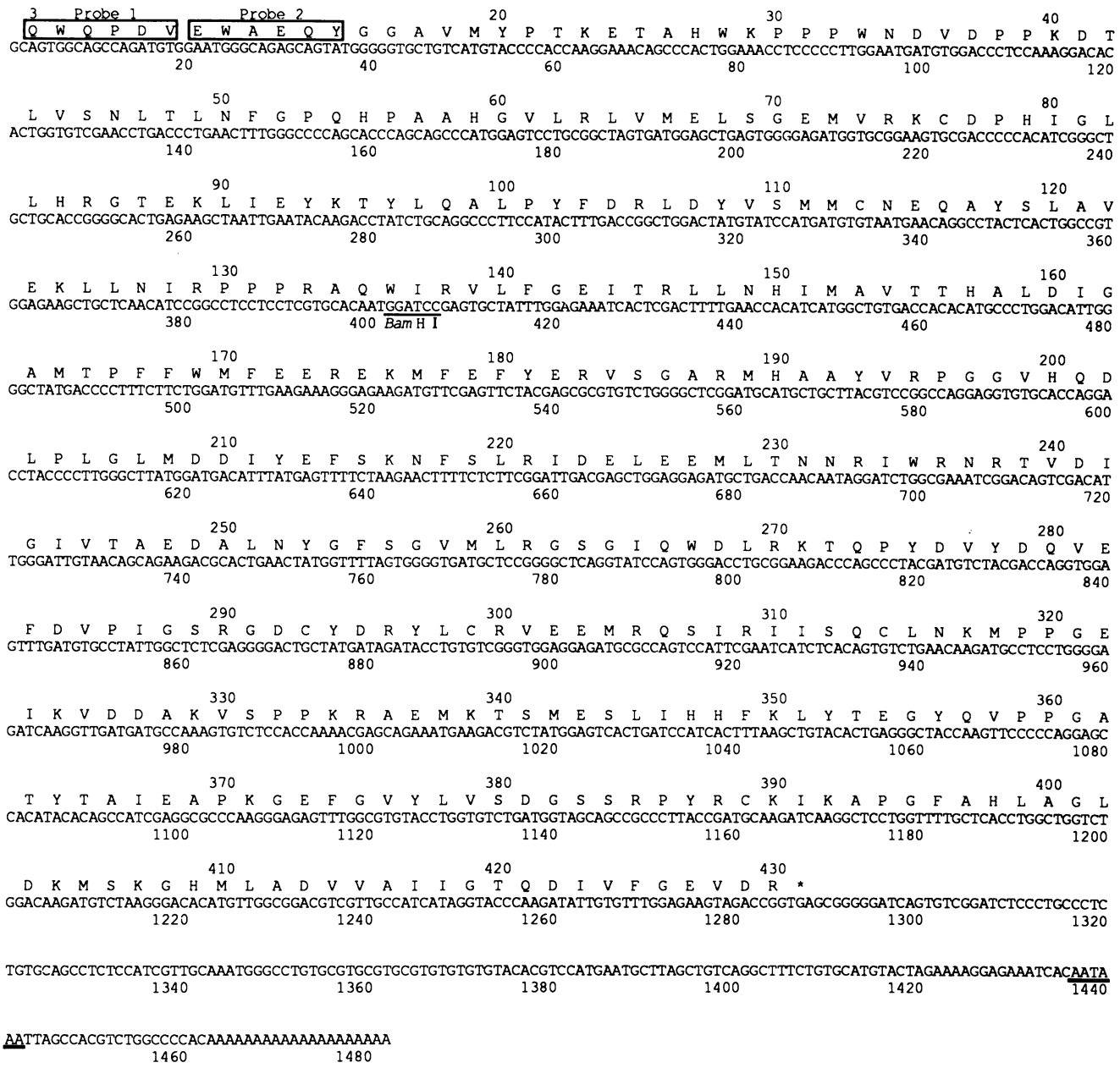

Fig. 3. The cDNA sequence encoding the 49 kd subunit of NADH-ubiquinone reductase from bovine mitochondria. The cDNA lacks 5 bases at its 5′ end and so the codons for amino acids 1 and 2 of the mature protein are unknown. The N-terminal sequence of residues 1–22 was obtained by direct sequence analysis of the protein (see Figure 1). The oligonucleotide probes that were used to isolate the cDNA were designed on the basis of the protein sequences of amino acids 3–8 and 9–14 and these sequences are boxed. A sequence near to the 3′ end of the cDNA that could serve as a signal for polyadenylation of the corresponding mRNA (Proudfoot and Brownlee, 1976) has been underlined. An internal *Bam*HI site is indicated.

sequence is 49 175, which is consistent with the value of 49 kd estimated by gel electrophoresis in the presence of SDS (Heron *et al.*, 1979).

The bovine 49 kd subunit of complex I, together with the 30 kd and 13 kd proteins, can be isolated in a subcomplex of the enzyme that also contains two iron–sulphur clusters, but further resolution of this subcomplex results in changes in the shape of associated EPR spectra, and so it is not known with which of these three subunits the clusters are associated (Ragan, 1987). The sequence of the bovine 49 kd protein contains five cysteine residues, but the motifs CysXXCys-XXCysXXXCysPro and CysXXXXCysXXCys that are associated with 4Fe–4S and 2Fe–2S centres, respectively, from ferredoxins of bacteria and chloroplasts (Yasunobu and Tanaka, 1980), are not present, and so there is no evidence from the sequence that the protein contains an iron–sulphur centre.

Homology of the protein sequence of the 49 kd subunit with chloroplast gene products

In order to try to delineate sequences in the 49 kd subunit that might illuminate its function, the protein sequence was compared with entries in the PIR database. Thereby a strong homology was detected with a protein encoded in the

Fig. 4. Comparison of the protein sequence of the bovine 49 kd subunit of complex I with protein sequences encoded in ORFs 392 and 393 present in the chloroplast genomes of *M.polymorpha* and *N.tabacum* respectively. The sequences were aligned by the computer program FASTP (Lipman and Pearson, 1985). Amino acids that are conserved in all three proteins are indicated by an asterisk, and where alignment has been improved by the introduction of insertions this is shown by a dash. A region of hydrophobicity that is present in all three sequences is underlined.

chloroplast genome of the liverwort, *M.polymorpha* (Ohyama *et al.*, 1986, 1988; Kohchi *et al.*, 1988) and with a closely related homologue encoded in the chloroplast genome of the tobacco plant, *N.tabacum* (Shinozaki *et al.*, 1986). These regions of the genomes of the two chloroplasts have been designated as open reading frames (ORFs), that encode proteins of hitherto unknown function with protein chains of 392 (*M.polymorpha*) or 393 amino acids (*N. tabacum*). The sequence relationship between the bovine mitochondrial 49 kd protein and chloroplast proteins extends throughout the sequence of the latter (see Figure 4). The three sequences are identical in 150 amino acids out of 392, or ~38% of the chloroplast sequences, and conservative changes are found in all three sequences in a further 62 positions, or ~17%. This alignment requires the introduction of only two insertions in the chloroplast sequences and one in the bovine protein. It also shows that of the five cysteines that are present in the 49 kd subunit, only Cys76 is conserved, and so it is unlikely that the non-conserved cysteines play any central role in the function of this protein family.

Studies of the arrangement of subunits within complex I show that the 49 kd protein, along with other components of the iron–sulphur fragment, is readily accessible to hydrophilic reagents. Other labelling studies on intact mitochondria and submitochondrial particles (inverted vesicles from the inner membrane) with hydrophilic reagents that do not penetrate the inner membrane, form the basis of the proposal that the 49 kd protein is a transmembrane protein, but that it is exposed more on the cytoplasmic surface than on the matrix side (Smith and Ragan, 1980; Patel *et al.*, 1988). However, it is not labelled by hydrophobic probes, and so transmembrane stretches would have to be protected from reaction by other proteins. The hydrophobic profiles of the 49 kd protein and its chloroplast homologues are consistent with the view that the protein is largely globular, and only one hydrophobic region of sufficient length to be folded into a membrane-spanning α-helix is present (see Figure 5). However, this region also contains five basic residues and two acidic amino acids that are conserved in all three sequences and these would have to be accommodated within the lipid bilayer. So this does not appear to be a convincing candidate for a membrane segment. The exposure of the 49 kd protein on both sides of the inner membrane, and the lack of a convincing hydrophobic region in the primary structure of the protein, can be reconciled by models of the complex in which an extensive hydrophilic transmembrane domain is protected from interaction with the lipid environment of the membrane by other more hydrophobic proteins (Ragan, 1987; Patel *et*

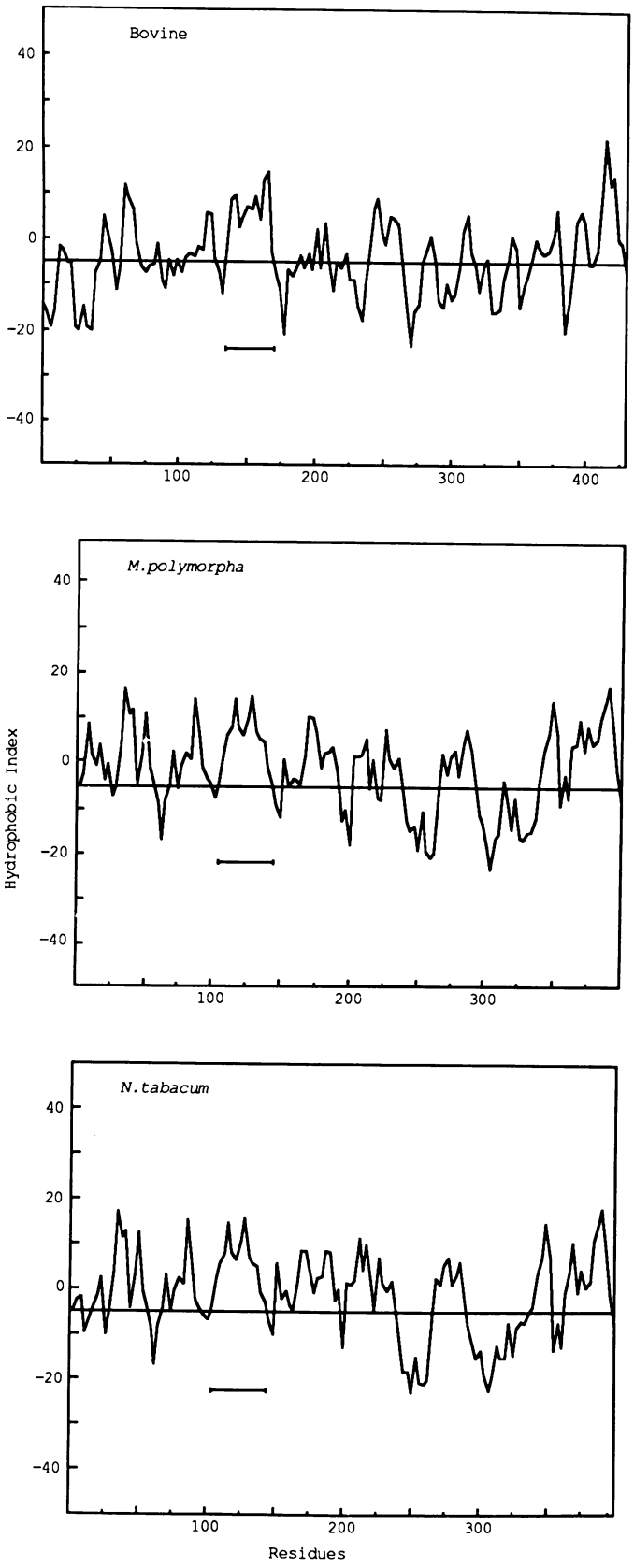

Fig. 5. Comparison of the hydrophobic profiles of the 49 kd subunit of bovine mitochondrial complex I and of the protein encoded in ORFs 392 and 393 respectively, present in the chloroplast genomes of *M.polymorpha* and *N.tabacum*. The calculations were made with HYDROPLOT using a window of 11 amino acids. The most hydrophobic span is indicated with a bar (see also Figure 4).

al., 1988), although the evidence for such a structure is weak and has no precedent.

The bovine sequence is longer than those of the chloroplast proteins and extends in an N-terminal direction by an additional 36 amino acids. This raised the possibility that a homologue of this region could be encoded in a separate exon elsewhere within the chloroplast genome. Split genes have been found for a number of chloroplast-coded proteins, including the *ndh* 1 and *ndh* A genes in *M.polymorpha* and *N.tabacum* respectively (Ohyama *et al.*, 1986; Shinozaki *et al.*, 1986). These encode hydrophobic proteins that are related to ND1, a subunit of complex I encoded in mitochondrial DNA (Chomyn *et al.*, 1985). Also, *trans*-splicing of exons from different primary transcripts has been shown to occur in chloroplasts (Koller *et al.*, 1987; Zaita *et al.*, 1987). Therefore, the DNA sequence of the *M.polymorpha* chloroplast genome was translated in all three phases of the DNA and in both of its senses, and these protein sequences were compared with that of residues 1–35 of the bovine 49 kd protein. However, a homologue of this N-terminal sequence was not detected in protein sequences coded in either strand of the *M.polymorpha* chloroplast DNA, and so this extra N-terminal extension is peculiar to the mitochondrial protein.

A gene cluster in chloroplast DNA encoding homologues of complex I subunits

The sequence of chloroplast DNA from liverwort contains seven ORFs *ndh* 1, *ndh* 2, *ndh* 3, *ndh* 4, *ndh* 4L, *ndh* 5 and *ndh* 6, that code for homologues of subunits ND1, ND2, ND3, ND4, ND4L, ND5 and ND6 respectively of mitochondrial NADH dehydrogenase (Ohyama *et al.*, 1986). These proteins are encoded in the mitochondrial genome of vertebrates (Anderson *et al.*, 1981, 1982; Bibb *et al.*, 1981; Görtz and Feldmann, 1981; Grosskopf and Feldmann, 1981; Koike *et al.*, 1982; Roe *et al.*, 1985) and invertebrates (Clary and Wolstenholme, 1985; Cantatore *et al.*, 1987; Wolstenholme *et al.*, 1987; Himeno *et al.*, 1987; Jacobs *et al.*, 1988), and some of them at least are encoded in fungal (Brown *et al.*, 1985; Burger and Werner, 1986; Ise *et al.*, 1985; Peñalva and Garciá, 1986; Nelson and Macino, 1987), algal (Pratje *et al.*, 1984) and higher plant mitochondria (Stern *et al.*, 1986). Homologues of the first six of these have also been identified in the tobacco chloroplast genome, where they are named respectively *ndh* A, *ndh* B, *ndh* C, *ndh* D, *ndh* E and *ndh* F (Shinozaki *et al.*, 1986), and two tobacco ORFs, ORF138 and ORF99B, together make up a homologue of *ndh* 6 if a frame shift is introduced into the sequence (Ohyama *et al.*, 1988). In liverwort, *ndh* 1, *ndh* 6, *ndh* 4L and *ndh* 4, and their respective tobacco homologues, are found close to each other in the small single-copy region of the chloroplast genome, and if they are expressed they would be transcribed in the same sense of the DNA, although transcription of these regions has not been demonstrated. As shown in Figure 6, the ORFs encoding the homologues of the bovine 49 kd protein immediately precede exon 1 of the *ndh* 1 gene in both species, and it now seems quite possible that other ORFs in this vicinity will also encode homologues of other as yet uncharacterized subunits of the mitochondrial complex I. In the liverwort genome, this includes a gene *frx* B which encodes a protein that is related to the 4Fe–4S ferredoxins, but not *frx* A, which has been shown to be closely related

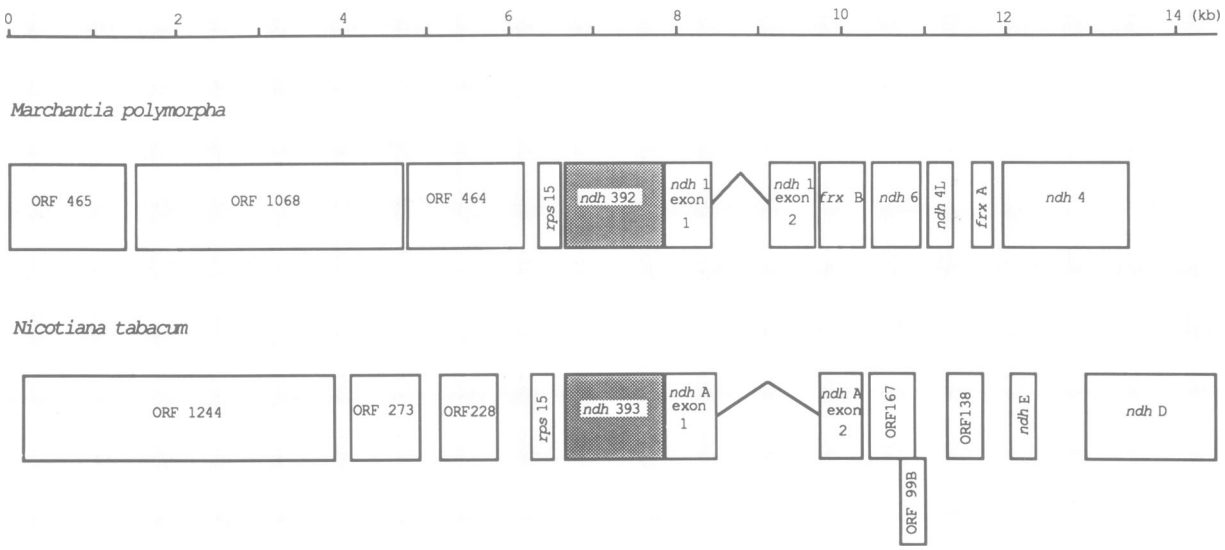

Fig. 6. Arrangement of genes and ORFs in the vicinities of ORFs 392 and 393 respectively in the chloroplast genomes of *M.polymorpha* and *N.tabacum*. The diagram is based upon published data (Kohchi *et al.*, 1988; Shinozaki *et al.*, 1986). ORFs are unidentified ORFs (potential genes), the associated number in each case indicating the length of the polypeptide chain that is encoded therein. The genes named *rps* 15 are homologues of the *Escherichia coli* gene for ribosomal protein S15, *frx* A and *frx* B encode iron–sulphur proteins, and *ndh* designates genes coding for homologues of subunits of mitochondrial NADH dehydrogenase. On the basis of evidence presented in this paper it is proposed that ORFs 392 and 393, which are shaded, be renamed *ndh* 392 and *ndh* 393, respectively. In *N.tabacum* the introduction of a frameshift in the region of ORF 167 and ORF 99B produces a reading frame that encodes a homologue of the mitochondrial protein ND6 (Ohyama *et al.*, 1988). The introns between exons 1 and 2 in *ndh* 1 and *ndh* A are indicated by inverted V shapes. The scale is in kb.

in sequence to the 8 kd subunit of spinach photosystem I (Oh-oka *et al.*, 1987; Høj *et al.*, 1987). Neither of these genes seems to be present in the tobacco chloroplast genome. The only other nuclear-coded subunit of the mitochondrial NADH dehydrogenase complex to have been sequenced to date is the 24 kd subunit (von Bahr-Lindström *et al.*, 1983; Pilkington and Walker, 1989), but we have not detected sequences coding for homologues of this protein in the chloroplast genomes. The remaining genes coding for homologues of complex I subunits (*ndh* 2, *ndh* 3 and *ndh* 5 and *ndh* B, *ndh* C and *ndh* F in liverwort and tobacco respectively) are at separate loci and are not evidently associated with other homologues of subunits of complex I.

Clusters of genes in chloroplast DNA coding for proteins with related function have been noted before (Cozens *et al.*, 1986; Umesono and Ozeki, 1987) and linkage relationships between genes for subunits of ATP synthase in chloroplasts and bacteria are comparable to some extent and support an endosymbiotic origin for chloroplasts (Cozens and Walker, 1987). Conversely, the presence in chloroplast genomes of the cluster coding for homologues of complex I subunits suggests that related clusters are likely to be present in bacterial genomes.

The regions of the chloroplast genomes that contain the homologues of subunits of mitochondrial complex I contain neither in-phase stop codons nor frame-shifts and so in these respects they do not resemble pseudo-genes, and it is likely that they are expressed. If so, the question arises of what the function might be of these gene products in chloroplasts, and in explanation it has been pointed out that respiration has been demonstrated in the chloroplasts of *Chlamydomonas reinhardtii* (Bennoun, 1982). This electron transport system appears to involve a NADH-plastoquinone oxidoreductase activity, which remains active when the photoelectron transport pathway between photosystems I and II has been prevented by an inhibitor of plastoquinone oxidation (Maione and Gibbs, 1986). So it appears to be likely that the chloroplast genes encode components of this enzyme, although direct confirmation, requiring characterization of the enzyme, has not been carried out.

Materials and methods

Reagents

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

Isolation of complex I

Mitochondria were prepared from bovine hearts as described by Smith (1967), and complex I was purified from them by the procedure of Hatefi *et al.* (1962) as described by Ragan *et al.* (1987). The enzyme was assayed at 30°C by following the oxidation of NADH in the presence of ferricyanide; its activity was ~50 U/mg of protein, where 1 U is the amount of enzyme required to reduce 1 μmol of ferricyanide in 1 min at 30°C (Hatefi, 1978). Protein concentrations were measured in the presence of 1% SDS by the method of Lowry *et al.* (1951).

Electrophoretic transfer of proteins to polyvinylidene difluoride membranes

Samples of complex I (100 μg) were dried *in vacuo* and then were dissolved in 5% SDS (5 μl). To this solution were added water (20 μl) and a solution of dyes (2 μl) containing 50% sucrose, 90 mM Tris, 90 mM boric acid, 2.5 mM EDTA, 0.3% xylene cyanol FF and 0.3% bromophenol blue. Eight of these samples were fractionated in separate lanes of a polyacrylamide gradient gel (10–25% acrylamide. 0.5 mm thick) in the presence of 0.1% SDS (Laemmli, 1970). Then the proteins were transferred by electrophoresis to a polyvinylidene difluoride membrane (Immobilon, Millipore) as described by Matsudaira (1987), except that the transfer buffer (10 mM CAPS, pH 11 or 12) containing 0.05% SDS. Proteins were stained on the membrane with 0.1% PAGE blue 83 dye (BDH, Poole, UK) in 50% methanol for a maximum period of 5 min, and after a brief destaining with 50% methanol the bands were excised with a scalpel and stored at −20°C before sequence analysis.

Protein sequence analysis

The protein that had been transferred to polyvinylidene difluoride membrane was sequenced in an Applied Biosystems 470A gas phase protein sequencer

with 'on-line' detection of phenylthiohydantoin amino acids. An optical sensor has been incorporated into the sample loop of the HPLC chromatograph. This permits 80% of the sample produced after each Edman cycle to be transferred reproducibly to the HPLC system and analysed, and increases the effective sensitivity of the sequencer by a factor of 2. Pieces of membrane to which proteins had been transferred by electrophoresis were placed in the reaction chamber of the sequencer in the presence of a Teflon seal only, and neither a glass fibre filter disc nor polybrene was employed in these experiments.

Oligonucleotide synthesis

Two mixtures of oligonucleotides and 11 unique oligonucleotides were synthesized by automated phosphoramidite chemistry by Mr T.V.Smith with the aid of an Applied Biosystems 380B synthesizer. The mixture contained oligonucleotides that were 17 bases long and were based upon the protein sequences QWQPDV and EWAEQY determined by N-terminal sequence analysis of the 49 kd component (summarized in Figure 1). Their sequences were 5′ CARTGGCARCCNGAYGT 3′ (probe 1) and 5′ GARTGGGC-NGARCARTA 3′ (probe 2), and both mixtures contain 32 different oligonucleotides. Calculations based on contributions of 4°C and 2°C respectively by each GC and AT base pair (Suggs *et al.*, 1981) give minimum dissociation temperatures of 50°C (probe 1) and 46°C (probe 2) for the components of these mixtures. The unique 17-base synthetic oligonucleotides were used as primers in DNA sequencing experiments as indicated in Figure 1. After synthesis each oligonucleotide or oligonucleotide mixture was desalted by reverse phase chromatography on a C_{18} Sep-Pak cartridge (Millipore).

Screening the cDNA libraries

A plasmid library in the vector pUC8 (Vieira and Messing, 1982) was made using bovine mRNA isolated from bovine heart and liver (Gay and Walker, 1985b). Recombinants in this library were grown up on Pall Biodyne A nylon membranes placed on agar plates, from which replica membranes were prepared. Portions (50 pmol) of each of the oligonucleotide mixtures were radiolabelled using polynucleotide kinase and [γ-^{32}P]ATP (sp. act. ~3000 Ci/mmol) and purified by electrophoresis through a 20% polyacrylamide gel containing 7.6 M urea, 90 mM Tris (pH 8.3), 90 mM boric acid and 2.5 mM EDTA. The oligonucleotides were extracted from macerated acrylamide gel by incubation at 37°C for 1 h in buffer containing 10 mM Tris–HCl, pH 8.0, and 0.1 mM EDTA. Filters were treated at 55°C for 2 h with a solution containing 90 mM Tris–HCl, pH 7.8, 0.9 M NaCl, 5 × Denhardt's solution, 0.5% sarkosyl, and boiled and sonicated salmon sperm DNA (0.1 mg/ml). Hybridizations were carried out on replica sets of filters with probes 1 and 2 at 45°C and 41°C respectively (5°C below the estimated minimum dissociation temperatures). In each case the purified probe was added to the hybridization solution (100 ml) described previously (Gay and Walker, 1985b) in which 10 filters had been immersed. Recombinants that gave a strong signal with both probes were rescreened with both probes. Those that still gave a positive response were grown up and their recombinant plasmids isolated and characterized further. These are isolates pBovCI-49.1, pBovCI-49.2 and pBovCI-49.3.

Subcloning and DNA sequencing

Inserts from the pUC8 recombinants were excised by digestion with *Bam*HI and *Eco*RI together. From recombinants pBovCI-49.1 and pBovCI-49.2 this produced a *Bam*HI fragment (estimated by agarose gel electrophoresis to be 450 bp) and an *Eco*RI–*Bam*HI fragment (estimated size ~1000 bp) which were then cloned into appropriate sites in the vectors M13mp9, M13mp8 and M13mp19. DNA sequences were determined by the dideoxy method (Sanger *et al.*, 1977) as modified by Biggin *et al.* (1983). The flanking primer LMB2 (Duckworth *et al.*, 1981) and 11 internal primers were used in the determination of the cDNA sequence in both directions. In order to establish the overlap between the *Bam*HI and *Bam*HI–*Eco*RI fragments, a *Hin*dIII–*Eco*RI fragment containing the entire cDNA was cloned into an appropriately prepared M13 vector, and then the overlapping sequence through the *Bam*HI site was determined with the use of a synthetic primer (see Figure 2).

Computer methods

DNA sequences were compiled using the programs DBAUTO and DBUTIL (Staden, 1982) and analysed using ANALYSEQ (Staden, 1985). The protein sequence of the 49 kd subunit of complex I was compared with sequences in the PIR database with the computer program FASTP (Lipman and Pearson, 1985). Hydrophobic profiles were calculated with the computer program HYDROPLOT, which is based upon SOAP (Kyte and Doolittle, 1982).

Acknowledgements

We thank Mr T.V.Smith for synthesizing oligonucleotides, Dr J.M.Cooper for the gift of a sample of the IP fragment of bovine complex I, and Dr R.Staden for his help in searching the chloroplast sequences.

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. and Young,I.G. (1981) *Nature*, **290**, 457–465.

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

Bennoun,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4352–4356.

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

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.

Brown,T.A., Waring,R.B., Scazzocchio,C. and Davies,R.W. (1985) *Curr. Genet.*, **9**, 113–117.

Burger,G. and Werner,S. (1986) *J. Mol. Biol.*, **186**, 231–242.

Cantatore,P., Roberti,M., Morisco,P., Rainaldi,G., Gadaleta,M.N. and Saccone,C. (1987) *Gene*, **53**, 41–54.

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

Chomyn,A., Cleeter,M.W.J., Ragan,C.I., Riley,M., Doolittle,R.F. and Attardi,G. (1986) *Science*, **234**, 614–618.

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

Cozens,A.L. and Walker,J.E. (1987) *J. Mol. Biol.*, **194**, 358–383.

Cozens,A.L., Walker,J.E., Phillips,A.L., Huttly,A.K. and Gray,J.C. (1986) *EMBO J.*, **5**, 217–222.

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

Galante,Y.M. and Hatefi,Y. (1979) *Arch. Biochem. Biophys.*, **192**, 559–568.

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

Gay,N.J. and Walker,J.E. (1985b) *Biochem. J.*, **225**, 707–712.

González-Halphen,D., Lindorfer,M.A. and Capaldi,R.A. (1988) *Biochemistry*, **27**, 7021–7031.

Görtz,G. and Feldmann,H.(1982) *Curr. Genet.*, **5**, 221–225.

Grosskopf,R. and Feldmann,H. (1981) *Curr. Genet.*, **4**, 151–158.

Hatefi,Y. (1978) *Methods Enzymol.*, **53**, 11–14.

Hatefi,Y., Haavik,A.G. and Griffiths,D.E. (1962) *J. Biol. Chem.*, **237**, 1676–1680.

Heron,C., Smith,S. and Ragan,C.I. (1979) *Biochem. J.*, **181**, 435–443.

Himeno,H., Masaki,H., Kawai,T., Ohta,T., Kumagai,I., Miura,K.-I. and Watanabe,K. (1987) *Gene*, **56**, 219–230.

Høj,P.B., Svendsen,I., Scheller,H.V. and Møler,B.L. (1987) *J. Biol. Chem.*, **262**, 12676–12684.

Ise,W., Haiker,H. and Weiss,H. (1985) *EMBO J.*, **4**, 2075–2080.

Jacobs,H.T., Elliott,D.J., Math,V.B. and Farquharson,A. (1988) *J. Mol. Biol.*, **202**, 185–217.

Kohchi,T., Shirai,H., Fukuzawa,H., Sano,T., Komano,T., Umesono,K., Inokuchi,H., Ozeki,H. and Ohyama,K. (1988) *J. Mol. Biol.*, **203**, 353–372.

Koike,K., Kobayashi,M., Yaginuma,K., Taira,M., Yoshida,E. and Imai,M. (1982) *Gene*, **20**, 177–185.

Koller,B., Fromm,H., Galun,E. and Edelman,M. (1987) *Cell*, **48**, 111–119.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.

Lowry,O.H., Rosebrough,N.J., Farr,A.L. and Randall,R.J. (1951) *J. Biol. Chem.*, **193**, 265–275.

Maione,T.E. and Gibbs,M. (1986) *Plant Physiol.*, **80**, 364–368.

Matsudaira,P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.

Nelson,M.A. and Macino,G. (1987) *Mol. Gen. Genet.*, **206**, 307–317.

Oh-oka,H., Takahashi,Y., Wada,K., Matsubara,H., Ohyama,K. and Ozeki,H. (1987) *FEBS Lett.*, **218**, 52–54.

Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. (1986) *Nature*, **322**, 571–574.

Ohyama,K., Fukuzawa,H., Kohchi,T., Sano,T., Sano,S., Shirai,H., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. (1988) *J. Mol. Biol.*, **203**, 281–298.

Patel,S.D., Cleeter,M.W.J. and Ragan,C.I. (1988) *Biochem. J.*, **256**,

529–535.
Peñalva,M.A. and Garciá,J.L. (1986) *Curr. Genet.*, **10**, 797–801.
Pilkington,S.J. and Walker,J.E. (1989) *Biochemistry*, in press.
Powell,S.J., Medd,S.M., Runswick,M.J. and Walker,J.E. (1989) *Biochemistry*, in press.
Pratje,E., Schnierer,S. and Dujon,B. (1984) *Curr. Genet.*, **9**, 75–82.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Ragan,C.I. (1987) *Curr. Topics Bioenerget.*, **15**, 1–36.
Ragan,C.I., Wilson,M.T., Darley-Usmar,V.M. and Lowe,P.N. (1987) In Darley-Usmar,V.M., Rickwood,D. and Wilson,M.T. (eds), *Mitochondria: A Practical Approach*. IRL Press, Oxford, UK, pp. 79–112.
Roe,BA., Ma,D.P., Wilson,R.K. and Wong,J.F.H. (1985) *J. Biol. Chem.*, **260**, 9759–9774.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Shinozaki,K., Ohme,M., Tanaka,M., Wakasugi,T., Hayashida,N., Matsubayashi,T., Zaita,N., Chunwongse,J., Obokata,J., Yamaguchi-Shinozaki,K., Ohto,C., Torazawa,K., Meng,B.Y., Sugita,M., Deno,H., Kamogashira,T., Yamada,K., Kusuda,J., Takaiwa,F., Kato,A., Tohdoh, N., Shimada,H. and Sugiura,M. (1986) *EMBO J.*, **5**, 2043–2049.
Smith,A.L. (1967) *Methods Enzymol.*, **10**, 81–86.
Smith,S. and Ragan,C.I. (1980) *Biochem. J.*, **185**, 315–326.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 4731–4751.
Staden,R. (1985) In Setlow,J.K. and Hollaender,A. (eds), *Genetic Engineering: Principles and Methods*. Plenum, New York, Vol. 7, pp. 67–114.
Stern,D.B., Bang,A.G. and Thompson,W.F. (1986) *Curr. Genet.*, **10**, 857–869.
Suggs,S.V., Hirose,T., Miyake,T., Kawashima,E.H., Johnson,M.J., Itakura,K. and Wallace,R.B. (1981) In Brown,D.D. (ed.), *Developmental Biology using Purified Genes*. Academic Press, New York, pp. 683–693.
Umesono,K. and Ozeki,H. (1987) *Trends Genet.*, **3**, 281–287.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.
von Bahr-Lindström,H., Galante,Y.M., Persson,M. and Jörnvall,H. (1983) *Eur. J. Biochem.*, **134**, 145–150.
Walker,J.E., Fearnley,I.M., Gay,N.J., Gibson,B.W., Northrop,F.D., Powell,S.J., Runswick,M.J., Saraste,M. and Tybulewicz,V.L.J. (1985) *J. Mol. Biol.*, **184**, 677–701.
Walker,J.E., Cozens,A.L., Dyer,M.R., Fearnley,I.M., Powell,S.J. and Runswick,M.J. (1987a) *Chem. Scr.*, **27B**, 97–105.
Walker,J.E., Gay,N.J., Powell,S.J., Kostina,M. and Dyer,M.R. (1987b) *Biochemistry*, **26**, 8613–8619.
Walker,J.E., Runswick,M.J. and Poulter,L. (1987c) *J. Mol. Biol.*, **197**, 89–100.
Walker,J.E., Powell,S.J., Viñas,O. and Runswick,M.J. (1989) *Biochemistry*, in press.
Wikström,M. (1984) *FEBS Lett.*, **169**, 300–304.
Wolstenholme,D.R., McFarlane,J.R., Okimoto,R., Clary,D.O. and Wahleithner,J.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1324–1328.
Yasunobu,K. and Tanaka,M. (1980) *Methods Enzymol.*, **69**, 228–238.
Zaita,N., Torazawa,K., Shinozaki,K. and Sugiura,M. (1987) *FEBS Lett.*, **210**, 153–156.

Received on December 6, 1988