Biochem. J. (1991) 277, 11–15 (Printed in Great Britain)

NADH:ubiquinone oxidoreductase from bovine mitochondria

cDNA sequence of a 19 kDa cysteine-rich subunit

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

The sequence of a 19 kDa subunit of NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria has been determined by a new strategy based on the polymerase chain reaction. The subunits of the enzyme were resolved in a polyacrylamide gel by two-dimensional isoelectric focusing and electrophoresis under denaturing conditions, transferred to a poly(vinylidene difluoride) membrane, and the *N*-terminal sequence was determined on the stained 19 kDa protein up to residue 27. This information was used to design two mixed oligonucleotide primers and a mixed oligonucleotide probe. With total bovine heart cDNA as template, overlapping cDNAs extending to sequences corresponding to both the 5′ and 3′ extremities of the mRNA coding for the 19 kDa subunit were synthesized in three polymerase chain reactions. These cDNAs were cloned and sequenced and encode a 171-amino-acid mature protein preceded by a methionine residue. The mature protein contains eight cysteine residues spaced at regular intervals through the protein, but the cysteine-rich motifs that are often associated with tetranuclear or binuclear centres in other proteins are not present. However, all eight cysteine residues are strictly conserved in a related protein from *Neurospora crassa*, suggesting that they have structural and/or functional significance in complex I.

INTRODUCTION

Mitochondrial NADH:ubiquinone oxidoreductase (also known as complex I) is a very complicated membrane-bound structure (reviewed by Ragan, 1987). On the basis of two-dimensional gel electrophoresis and isoelectric focusing combined with protein-sequencing experiments (J. M. Skehel & J. E. Walker, unpublished work), the bovine enzyme appears to contain more than 30 different subunits, including the seven that are encoded in mitochondrial DNA (Chomyn *et al.*, 1985, 1986). Within this assembly are bound a number of prosthetic groups, including a molecule of FMN, which is thought to be associated non-covalently with the 51 kDa subunit (Ingledew & Ohnishi, 1980; Krishnamoorthy & Hinkle, 1988); the substrate NADH also binds to the same protein (Chen & Guillory, 1981; Deng *et al.*, 1990). In addition, five or six iron–sulphur centres have been defined by e.p.r. experiments (summarized by Ragan, 1987), and a major task is to determine the polypeptide chains that provide the ligands for these centres. Some clues can be gained from the sequences of the constituent subunits. Often three of the ligands of binuclear centres are provided by cysteine residues in the sequence CXXXCXXC (where 'C' is cysteine) (Yasunobu & Tanaka, 1980; Cammack, 1983), and the sequence CXXCXXC has been found to be associated with tetranuclear centres (Stout, 1982). The fourth ligand is frequently provided by a cysteine residue located four amino acids away in a *C*-terminal direction from the cluster; in other cases it is provided by the cysteine residue in the sequence Cys-Pro located distantly in the primary structure on either side of the cysteine-rich cluster. In other iron–sulphur proteins, e.g. the Rieske iron–sulphur protein from the mitochondrial cytochrome bc_1 complex, the cysteine ligands are not clustered (Harnisch *et al.*, 1985). The sequence of a subunit of bovine complex I described below contains eight cysteine residues. The cysteine-rich motifs described above are not present, but the strict conservation of these cysteine residues in an otherwise weakly conserved protein identified in the *Neurospora crassa* complex I (Videira *et al.*, 1990) argues for a significant role for these residues in the function and/or the structure of the enzyme.

MATERIALS AND METHODS

Protein sequence analysis of subunits of complex I

The enzyme was purified from bovine heart mitochondria (Hatefi, 1978; Ragan *et al.*, 1987). Its subunits were separated by two-dimensional isoelectric focusing and electrophoresis in polyacrylamide gels (O'Farrell, 1975), transferred to a poly(vinylidene difluoride) (PVDF) membrane, stained with Coomassie Blue dye and their *N*-terminal sequences were determined as described elsewhere (Fearnley *et al.*, 1989; Dupuis *et al.*, 1991; Pilkington *et al.*, 1991*a*,*b*).

Oligonucleotide synthesis

The synthesis and purification of oligonucleotides, for use both in PCRs and DNA-sequencing experiments, and conditions for radiolabelling oligonucleotides and for their use as hybridization probes, were described previously (Powell *et al.*, 1989; Walker *et al.*, 1989; Pilkington *et al.*, 1991*a*,*b*). In order to facilitate cloning of products of PCRs into M13 vectors, some oligonucleotides were made with a linker sequence containing various restriction-enzyme sites (*Eco*RI, *Hin*dIII or *Bam*HI) on the 5′ ends.

Preparation of cDNA

First-strand cDNA synthesis from polyadenylated [poly(A)$^+$] bovine heart mRNA (Walker *et al.*, 1987) was performed as described previously (Pilkington *et al.*, 1991*b*), except that it was primed with an oligonucleotide with the sequence TAGGAATTCGGATCCAAGC$(T)_{17}$. Thus the double-stranded cDNA product has a polylinker sequence at one end containing *Eco*RI, *Bam*HI and *Hin*dIII sites. This has

Abbreviations used: poly(A)$^+$, polyadenylated; PVDF, poly(vinylidene difluoride).

* Present address: INSERM, Laboratoire de Biochimie, BP 85X, 38041 Grenoble, France.
† To whom correspondence should be sent.

two uses; first, PCRs can be primed with an oligonucleotide with the same polylinker sequence. This is preferable to the alternative of priming directly on the poly(A) tail, since the PCR can be performed at a higher temperature, thereby increasing specificity. Secondly, the polylinker facilitates cloning of products of PCRs into appropriate M13 vectors. The addition of adenosine residues to the 5′ end of single-stranded cDNA with terminal transferase was described previously (Dupuis *et al.*, 1991).

PCRs

Conditions for PCRs, for the fractionation of products by gel electrophoresis, for the detection of products by hybridization with synthetic oligonucleotides, and for the recovery and cloning of cDNAs in M13 vectors, were described previously (Runswick *et al.*, 1990; Pilkington *et al.*, 1991*a*; Walker *et al.*, 1991). When degenerate primers or oligo(dT) primers were employed (see PCR 3 in Fig. 2), the specificity of the reaction was increased by carrying out the first 30 cycles at a low primer concentration (10–25 nM), followed by an additional 30 cycles at higher primer concentration (1 mM).

DNA sequence analysis

DNA sequences were determined by the modified chain-termination procedure (Sanger *et al.*, 1977; Biggin *et al.*, 1983). All sequences were determined completely in both senses of the DNA, and compressions were resolved by the use of deazaguanidine or dITP in sequencing reactions. Data were compiled and analysed with the computer programs DBUTIL (Staden, 1982*a*) and ANALYSEQ (Staden, 1985). The protein sequence of the 19 kDa subunit was compared with those in the PIR, PIRNEW and EMBL databases with the program FASTA (Lipman & Pearson, 1985). Pairwise comparisons of protein sequences were made with DIAGON (Staden, 1982*b*), and hydrophobicity profiles were calculated using HYDROPLOT, a version of SOAP (Kyte & Doolittle, 1982).

RESULTS AND DISCUSSION

N-Terminal sequence of the 19 kDa subunit

The subunits of complex I from bovine heart mitochondria were separated under denaturing conditions by two-dimensional isoelectric focusing and electrophoresis in a polyacrylamide gel. As shown in Fig. 1, after transfer of proteins to a PVDF membrane and sequence analysis of stained spots, several subunits of known sequence could be recognized from their *N*-terminal sequences (Fearnley *et al.*, 1989; Pilkington & Walker, 1989; Runswick *et al.*, 1989; Pilkington *et al.*, 1991*a*,*b*; Dupuis *et al.*, 1991). The sequence of the 19 kDa subunit was determined up to residue 27 (see Fig. 2).

Cloning and sequence analysis of cDNAs encoding the 19 kDa subunit

Clones for the 19 kDa subunit of complex I were isolated by following a strategy based upon the PCR. This strategy has been developed in the course of sequence studies on subunits of complex I and other mitochondrial proteins (Runswick *et al.*, 1990; Pilkington *et al.*, 1991*a*,*b*; Dupuis *et al.*, 1991; Walker *et al.*, 1991). It involves, first, the determination of a protein sequence of at least 18 consecutive amino acids either at the *N*-terminus or at an internal site in the protein. The hexapeptide sequences at the *N*- and *C*-terminal extremities of this sequence then serve as the basis for the synthesis of degenerate oligonucleotides 17 bases long, to be used respectively as the forward and reverse primers in a PCR using bovine heart cDNA as template. The protein sequence between the hexapeptide extremities is used to design a third degenerate oligonuleotide, which is employed as a hybridization probe in the identification of a short cDNA encoding the protein sequence from the products of the PCR. The sequence of this short cDNA is determined, but, because of the degeneracy of the primers in the PCR, only the segment of sequence between the primers is accurate. Unique synthetic primers based on this segment of DNA sequence are then used in further PCRs to produce cDNAs that extend to the 3′ and 5′ extremities of the mRNA. In the former experiment,

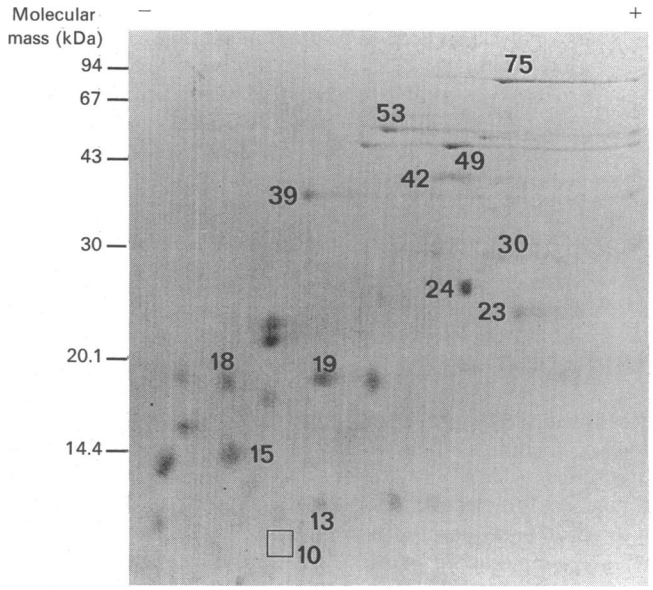

Fig. 1. Separation of the subunits of bovine complex I by two-dimensional isoelectric focusing and electrophoresis in a polyacrylamide gel after transfer to PVDF membrane

Isoelectric focusing was carried out in the horizontal direction, and the positions of the anode and cathode are denoted by the + and − respectively. In the second dimension, a 16% polyacrylamide gel was employed. The proteins were transfered to a PVDF membrane (see the Materials and methods section) and detected by staining with Coomassie Blue dye. The positions of various subunits determined by *N*-terminal sequence analysis are indicated by their estimated molecular masses. The positions of molecular-mass markers are given on the left-hand side.

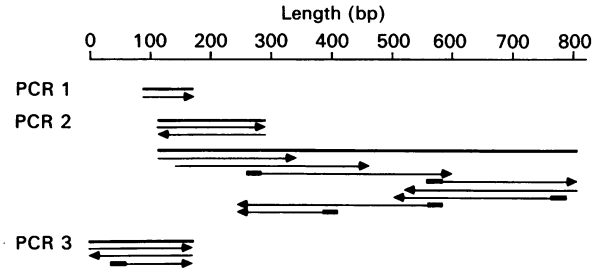

Fig. 2. Generation by the PCR and sequence analysis of cDNA clones encoding the 19 kDa subunit of complex I from bovine heart mitochondria

PCR 1–3 denote the cloned partial cDNAs generated by the polymerase reactions. The heavy lines represent these cDNAs, and the arrows indicate the directions and the extents of the sequences obtained from them. The boxes attached to the arrows correspond to the positions of synthetic oligonucleotide primers used in sequencing reactions.

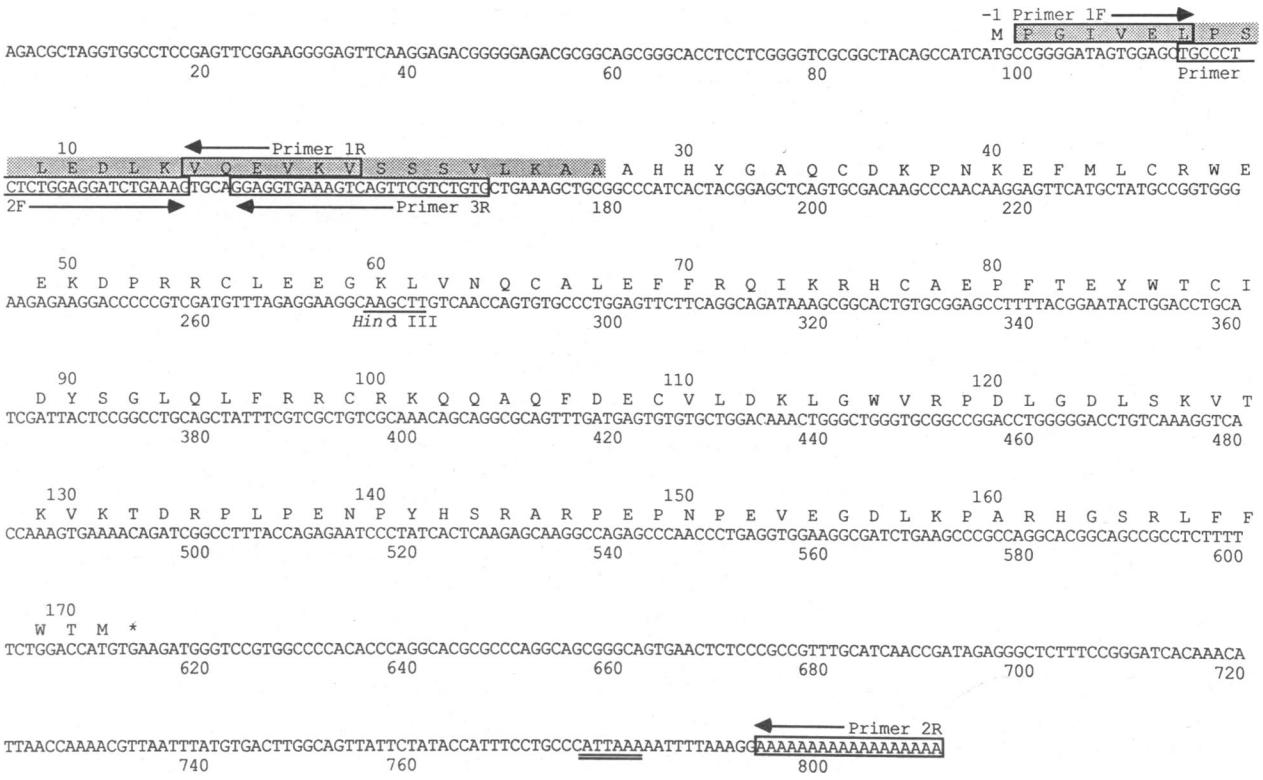

Fig. 3. Complete cDNA sequence compiled from partial cDNAs encoding the 19 kDa subunit of bovine heart mitochondrial complex I

The *N*-terminus of the mature protein is numbered as position 1 and is preceded by a methionine residue (numbered as −1), assumed to be the translational initiator. Sequence determined directly on the mature protein is shaded. Boxed nucleotide sequences were used in PCR experiments PCR 2 and PCR 3 and are numbered accordingly. F and R denote forward and reverse primers respectively. In experiment PCR 2, the reverse primer corresponds to the polylinker added during first-strand cDNA synthesis (see the Materials and methods section). In experiment PCR 3, the forward primer is oligo(dT) complementary to the poly(A) tract added to the 5′ end of the cDNA with terminal transferase. A potential polyadenylation signal is underlined twice, and an internal *Hin*dIII site is also indicated.

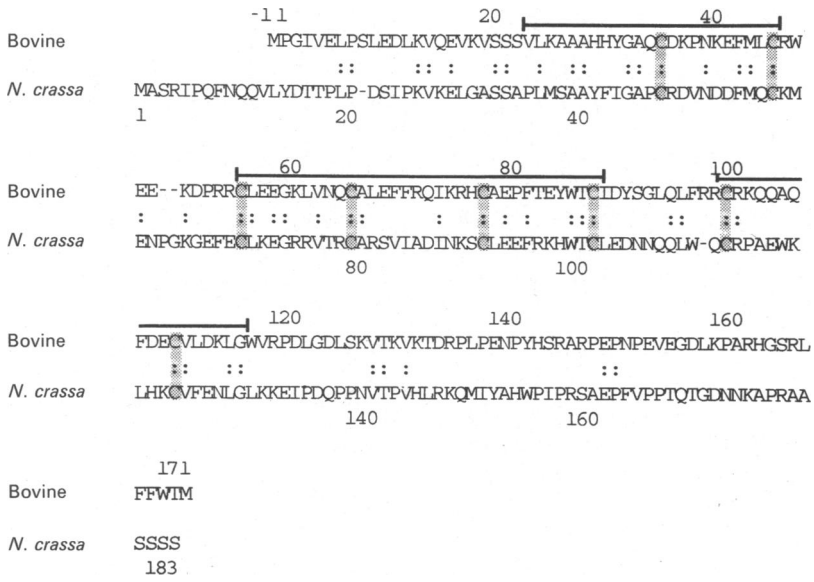

Fig. 4. Alignment of the sequences of the 19 kDa and 22 kDa subunits of mitochondrial complex I from bovine heart and *N. crassa* respectively

Colons denote identical residues in both sequences, and cysteine residues are shaded. Conserved regions detected by DIAGON (results not shown) are marked by a thick line.

the second primer was the 19-base oligonucleotide TAGGAATTCGGATCCAAGC corresponding to the polylinker sequence introduced during first-strand cDNA synthesis (see the Materials and methods section). In the latter experiment, oligo(dT) is used as the second primer, with template cDNA tailed at its 5′ end with adenine residues.

In the case of the 19 kDa subunit of bovine complex I, a cDNA representing amino acids 1–23 was produced using degenerate primers 17 nucleotides long based on amino acids 1–6 (forward primer 1F) and 14–19 (reverse primer 1R). Their complexities were 768 and 128 respectively. The cDNA was identified by hybridization with a degenerate oligonucleotide probe containing 4096 sequences based on amino acids 8–13. The sequence of the cloned fragment was 56 bases long and, because of primer degeneracy, only nucleotides 18–30 from the central part of this sequence were accurate. In a second PCR a cDNA was synthesized that extended from this known sequence to the 3′ poly(A) tail with primers 2F and 2R (see Fig. 3 and PCR 2 in Fig. 2). The reaction product was estimated to be about 650 bases long and hybridized strongly with a degenerate probe based on the established unique codons for amino acids 12 and 13 and amino acids 14–17. The products of this reaction were digested with *Hin*dIII and were cloned into an appropriate M13 vector. It was found that the cDNA contained an internal *Hin*dIII site (see Fig. 3), and so the reaction was repeated with primers with appropriate linker sequences. This time the products were cloned into the *Bam*HI site of M13, and a clone was obtained that extended through the *Hin*dIII site up to the poly(A) tail present in the mRNA. In a third reaction, a cDNA was made that extended to the 5′ end of the sequence using primers 3R and oligo(dT) with 5′-tailed cDNA.

The complete cDNA sequence presented in Fig. 3 is 811 nucleotides in length and is terminated by a poly(A) tail 12 nucleotides to the 3′ side of the sequence ATTAAA, which can serve as a polyadenylation signal (Proudfoot & Brownlee, 1976).

Sequence of the 19 kDa subunit of bovine complex I

The mature 19 kDa subunit encoded in the cDNA is 171 amino acids in length and its molecular mass, calculated from the sequence, is 19960 Da, which agrees reasonably well with the value of 18.5 kDa estimated by SDS/PAGE. The codon for the *N*-terminal proline residue is preceded by the sequence ATG. As this is the only in-phase potential ATG codon in the 5′ sequence, it is reasonable to assume that this is the translational initiator, although there are no in-phase termination codons to the 5′ side of it. Therefore it appears that the 19 kDa protein belongs to a relatively rare group of nuclear-encoded mitochondrial proteins that have no processed *N*-terminal import sequence, and mitochondrial import presumably is specified by sequences in the mature protein. Other examples of nuclear-encoded mitochondrial proteins lacking a processed import sequence are ADP/ATP translocase (Powell *et al.*, 1989), the d-subunit of ATP synthase (Walker *et al.*, 1987) and the oxoglutarate/malate-carrier protein (Runswick *et al.*, 1990).

Conservation of cysteine residues in the 19 kDa subunit

The most striking feature of the sequence of the 19 kDa subunit of bovine complex I is the presence of eight cysteine residues. These are spaced evenly through the sequence with gaps of nine amino acids between the first four (cysteine residues 35, 45, 55, and 65) and gaps of 11, 8, 12 and 9 amino acids between residues 65 and 77, 77 and 86, 86 and 99, and 99 and 109, respectively. In the related 22 kDa subunit from *N. crassa* complex I, the cysteine residues are all conserved (see Fig. 4). The sequence similarity between the two proteins extends over the central regions of the two proteins, and their *N*- and *C*-terminal extremities have no obvious relationship. Therefore, it is clear that the bovine counterpart of the *N. crassa* 23 kDa subunit is the 19 kDa subunit and not, as has been suggested (Videira *et al.*, 1990), the 24 kDa subunit. The sequence of the bovine 24 kDa subunit (Pilkington & Walker, 1989) is not significantly related to that of either the bovine 19 kDa or the *N. crassa* 22 kDa proteins. Moreover, the 24 kDa subunit is part of the flavoprotein subcomplex, whereas the 19 kDa subunit is found in neither the flavoprotein nor the iron–sulphur protein subcomplexes of complex I. It is probably in the residue left after removal of the flavoprotein and iron–sulphur subcomplexes. This is referred to as the 'hydrophobic protein' fraction of complex I (see Ragan, 1987), although the hydrophobic profile of the 22 kDa subunit (not shown) has no extensive hydrophobic segments of sequence with the potential to be folded into a membrane-spanning α-helixes.

At present, the roles in complex I of the conserved cysteine residues in the 19 kDa subunit are unknown, but it is possible that they may be involved in providing ligands for iron–sulphur centres, although there is no supporting evidence. Further experimentation is needed to demonstrate directly the presence of the centres, and to determine the correspondence with the spectroscopically defined centres. In order to further these and other studies of complex I, several of the bovine subunits, including the 19 kDa subunit, have been expressed individually in *Escherichia coli* (A. Dupuis, S. M. Medd & J. E. Walker, unpublished work).

Note added in proof (received 16 April 1991)

The molecular mass of the 19 kDa subunit determined by electrospray m.s., namely 19958 Da, is close to the value of 19960 Da calculated from the sequence, and so provides independent corroboration of the sequence (I. M. Fearnley & J. E. Walker, unpublished work).

We are grateful to Mr. T. V. Smith and Mrs. J. Fogg for the synthesis of oligonucleotides. A. D. was supported by INSERM and by the Conseil Régional Rhône-Alpes, France.

REFERENCES

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965
Cammack, R. (1983) Chem. Scr. **21**, 87–95
Chen, S. & Guillory, R. J. (1981) J. Biol. Chem. **256**, 8318–8323
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
Deng, P. S. K., Hatefi, Y. & Chen, S. (1990) Biochemistry **29**, 1094–1098
Dupuis, A., Skehel, J. M. & Walker, J. E. (1991) Biochemistry **30**, 2954–2960
Fearnley, I. M., Runswick, M. J. & Walker, J. E. (1989) EMBO J. **8**, 665–672
Harnisch, U., Weiss, H. & Sebald, W. (1985) Eur. J. Biochem. **149**, 95–99
Hatefi, Y. (1978) Methods Enzymol. **53**, 11–14
Ingledew, W. J. & Ohnishi, T. (1980) Biochem. J. **186**, 111–117
Krishnamoorthy, G. & Hinkle, P. C. (1988) J. Biol. Chem. **263**, 17566–17575
Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
Lipman, D. J. & Pearson, W. R. (1985) Science **227**, 1435–1441
O'Farrell, P. H. (1975) J. Biol. Chem. **250**, 4007–4021
Pilkington, S. J. & Walker, J. E. (1989) Biochemistry **28**, 3257–3264
Pilkington, S. J., Skehel, J. M. & Walker, J. E. (1991*a*) Biochemistry **30**, 1901–1908
Pilkington, S. J., Skehel, J. M., Gennis, R. B. & Walker, J. E. (1991*b*) Biochemistry **30**, 2166–2175
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., 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

Runswick, M. J., Gennis, R. B., Fearnley, I. M. & Walker, J. E. (1989) Biochemistry **28**, 9452–9459

Runswick, M. J., Walker, J. E., Bisaccia, F., Iacobazzi, V. & Palmieri, F. (1990) Biochemistry **29**, 11033–11040

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467

Staden, R. (1982*a*) Nucleic Acids Res. **10**, 4731–4751

Staden, R. (1982*b*) Nucleic Acids Res. **10**, 2951–2961

Staden, R. (1985) in Genetic Engineering: Principles and Methods (Setlow, J. K. & Hollander, A., eds.), pp. 67–114, Plenum Publishing Corporation, New York and London

Stout, C. D. (1982) in Iron Sulfur Proteins (Spiro, T. G., ed.), pp. 97–146, John Wiley and Sons, New York

Videira, A., Tropschüg, M., Wachter, E., Schneider, H. & Werner, S. (1990) J. Biol. Chem. **265**, 13060–13065

Walker, J. E., Runswick, M. J. & Poulter, L. (1987) J. Mol. Biol. **197**, 89–100

Walker, J. E., Powell, S. J., Viñas, O. & Runswick, M. J. (1989) Biochemistry **28**, 4702–4708

Walker, J. E., Dupuis, A., Fearnley, I. M., Finel, M., Medd, S., Pilkington, S. J., Runswick, M. J. & Skehel, J. M. (1991) DNA Sequence, in the press

Yasunobu, K. & Tanaka, M. (1980) Methods Enzymol. **69**, 228–238

Received 19 November 1990/21 January 1991; accepted 4 February 1991