Volume 313, number 1, 80–84 FEBS 11762 November 1992

Complementary DNA sequences of two 14.5 kDa subunits of NADH:ubiquinone oxidoreductase from bovine heart mitochondria

Completion of the primary structure of the complex?

Jesús M. Arizmendi*, J. Mark Skehel, Michael J. Runswick, Ian M. Fearnley and John E. Walker

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

Received 2 October 1992

The amino acid sequences of two nuclear-encoded subunits of complex I from bovine heart mitochondria have been determined. Both proteins have an apparent molecular weight of 14.5 kDa and their *N*-α-amino groups are acetylated. They are known as subunits B14.5a and B14.5b. Neither protein is evidently related to any known protein and their functions are obscure. A total of 34 nuclear-encoded subunits of bovine complex I have now been sequenced and it is thought that the primary structure of the complex is now complete, although with such a complicated structure it is difficult to be certain that there are no other subunits remaining to be sequenced. Seven additional hydrophobic subunits of the enzyme are encoded in mitochondrial DNA, and therefore bovine heart complex I is an assembly of about 41 different proteins. If it is assumed that there is one copy of each protein in the assembly, these polypeptides contain 7,955 amino acids in their sequences, more than are found in the *Escherichia coli* ribosome, which contains 7,336 amino acids in its 32 polypeptides.

Complex I; Bovine heart mitochondria; cDNA sequence

1. INTRODUCTION

NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria is a highly complex membrane-bound multisubunit assembly which is now thought to be made of 41 different polypeptides [1,2]. Seven of them are encoded in mitochondrial DNA [3–6], and the remaining subunits are nuclear gene products that are synthesised in the cytoplasm and are imported into the organelle. The sequences of 32 of the nuclear-encoded subunits have been described previously [2,7–17], and they have given many clues about the topography of the enzyme, and about the functions of some of the subunits [1]. Complex I appears to have two major structural domains, namely a hydrophobic membrane domain and attached to it an extrinsic membrane domain made of globular subunits. The enzyme can be split with the detergent lauryldimethylamine oxide into two sub-complexes, known as Iα and Iβ, which correspond approximately to the extrinsic and intrinsic membrane domains, respectively [18]. The sequences of two further nuclear-encoded subunits that form part of subcomplex Iα are presented below. Both proteins have modified *N*-α-amino groups and the same apparent molecular weight, 14.5 kDa, and they are known as subunits B14.5a and B14.5b. The sequences show that they are not related to any known protein and their hydrophobicity profiles are consistent with their presence in subcomplex Iα. It is thought that the two protein sequences complete the primary structure of complex I from bovine heart mitochondria, although it is difficult to be absolutely certain at this stage that of all the subunits of this complicated assembly have been characterized.

Correspondence address: J.E. Walker, Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. Fax: (44) (223) 412178.

**Present address:* Department of Biochemistry and Molecular Biology, University of the Basque Country, P.O. Box 644, 48080 Bilbao, Spain.

The EMBL Data Library accession numbers for the sequences of proteins B14.5a and B14.5b presented in this paper are X68585 and X68647, respectively.

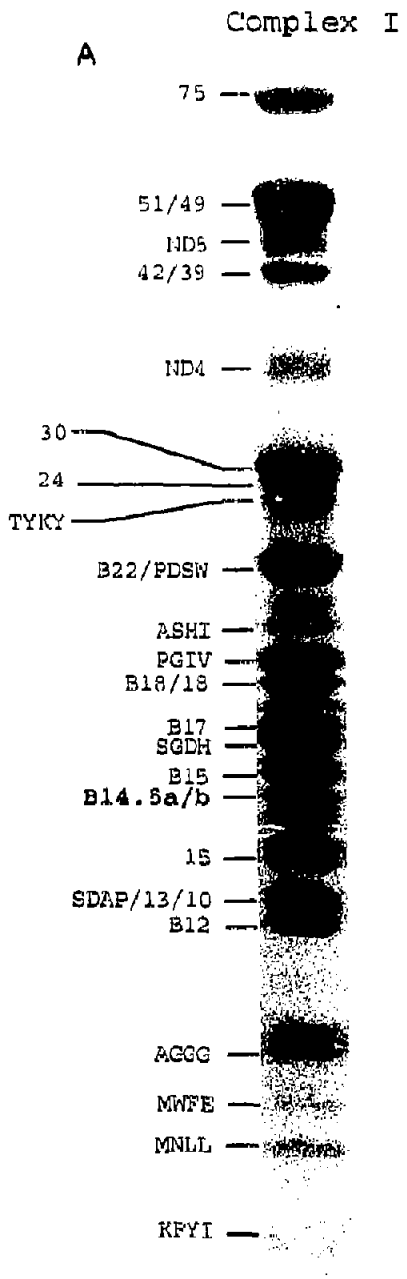
2. MATERIALS AND METHODS

2.1. *Isolation and protein sequencing of subunits 14.5a and 14.5b from bovine complex I*

Complex I was purified from bovine heart mitochondria [19]. Subcomplex Iα consisting of about 22 subunits [18] and subcomplex Iλ, a related but somewhat simpler subcomplex consisting of 15 subunits (M. Finel et al., unpublished work), were prepared from complex I after treatment with lauryldimethylamine oxide. The preparation of subcomplex Iλ differs from that of Iα only in that Iα is prepared from the detergent-treated enzyme by ion exchange chromatography [18], whereas subcomplex Iλ is made by centrifugation on a sucrose density gradient in the presence of 0.5 M phosphate and detergent (M. Finel et al., unpublished work).

The subunits of complex I and its subcomplexes were separated by

 Published by Elsevier Science Publishers B.V.



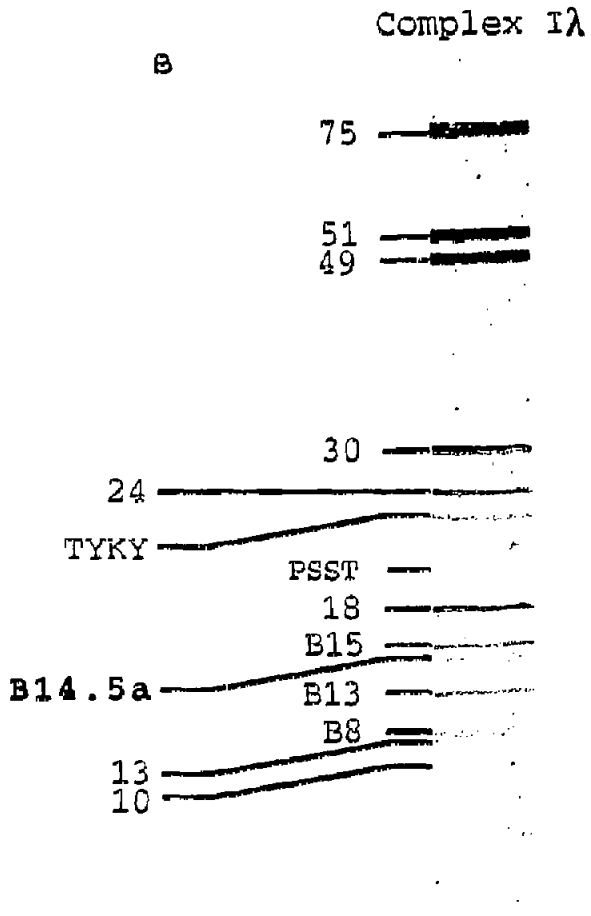


Fig. 1. Separation of the subunits of complex I and of subcomplex Iλ from bovine heart mitochondria. In (A), complex I subunits fractionated by gel electrophoresis in 16.5% acrylamide according to Schägger and von Jagow [21]; (B) subcomplex Iλ subunits separated by gel electrophoresis in a 10–20% acrylamide gradient gel according to Laemmli [20]. The positions of subunits are indicated at the sides.

gel electrophoresis in the presence of sodium dodecyl sulphate in the absence [20] and in the presence [21] of urea. They were transferred to a poly(vinylidene difluoride) membrane and sequences were determined at their N-terminals where this was possible. The band containing subunits B14.5a and B14.5b, which were not resolved from each other, was recovered by electro-elution, and treated with cyanogen bromide in 70% formic acid. The cleavage products were separated by 2-dimensional isoelectric focussing and gel electrophoresis in the presence of sodium dodecyl sulphate, as described before [15]. Then the cylindrical gel was placed on a 16.5% polyacrylamide gel prepared in urea [21], and embedded in a 4% stacking gel. The separated fragments were transferred to a poly(vinylidene difluoride) membrane and N-terminal sequences were determined. The mixture of B14.5a and B14.5b was also kept for 12 h at 37°C in 70% formic acid. The resulting fragments were fractionated by gel electrophoresis in the presence of sodium dodecyl sulphate and urea [21].

The subunits of subcomplexes Iα and Iλ were fractionated by reverse phase chromatography on an Aquapore C_4 column (100 × 2.1 mm i.d.) in 0.1% trifluoroacetic acid with an acetonitrile gradient. Subunits B14.5a and B14.5b eluted at acetonitrile concentrations of 37% and 54%, respectively, and were identified by electrospray mass spectrometry (see below). B14.5a isolated from both subcomplex Iα and Iλ was contaminated with minor amounts of subunit 18 kDa (IP), a protein of known sequence [2]. It was digested with trypsin and the peptides were sequenced.

2.2. *Amplification and sequencing of cDNAs encoding the B14.5a and B14.5b subunits*

cDNAs for the two subunits were amplified from total bovine heart cDNA by the polymerase chain reaction using in the first instance mixtures of synthetic oligonucleotides as primers and hybridization probes. These oligonucleotides mixtures were based upon the partial amino acid sequences of peptides prepared as described above. Further details of the amplification and sequencing experiments are given in the legends to Figs. 2 and 3. The strategy employed in this part of the work has been used extensively in the analysis of cDNAs for the other nuclear-encoded subunits of complex I and is described in detail elsewhere [2]. DNA sequences were determined by the modified chain termination procedure [22,23], and were established in both senses of the DNA.

2.3. *Electrospray mass spectrometry*

Subunits B14.5a and B14.5b were subjected to electrospray mass spectrometry in a VG BIO-Q triple quadrupole instrument with electrospray ionisation (VG Biotech, Altrincham, Cheshire, UK) as described previously [2].

3. RESULTS AND DISCUSSION

3.1. *Identification of subunits B14.5a and B14.5b in bovine complex I and in its subcomplexes*

The subunits of complex I were separated by gel electrophoresis in the presence of sodium dodecyl sulphate, and those with free N-terminals were identified from their N-terminal sequences (see Fig. 1). The positions of the subunits with modified N-terminals have been ascertained by protein chemical and mass spectrometric analysis (J.M. Skehel, I.M. Fearnley and J.E. Walker, unpublished work), and subunits B14.5a and B14.5b both appear to belong to this category. In the gel system of Schägger and von Jagow [21], they were unresolved from each other and migrated to a position between subunits B15 and 15 kDa (IP), with apparent molecular weights of about 14 kDa (see Fig. 1A). In the gel system of Laemmli [20] the positions of the two subunits have not been ascertained, but given the large number of subunits in complex I with molecular weights in the range 13–16,000 [1], it is most unlikely that they occupy unique positions.

Subunit B14.5a (but not B14.5b) is present in subcomplex Iλ as demonstrated by chromatographic separation of the subunits followed by measurement of their molecular masses by electrospray mass spectrometry, and by sequence analysis of the tryptic peptides of purified B14.5a. Subunit B14.5a is readily resolved from the other subunits of subcomplex Iλ in the Laemmli gel system (see Fig. 1B). Both B14.5a and B14.5b have been detected in subcomplex Iα by mass spectrometry. The measured mass of B14.5a is 12,589.2 ± 1.0. Two values were obtained for B14.5b, namely 14,141.6 ± 1.1 and 14,099.4 ± 0.4. They differ by 42 mass units (see below).

3.2. *Sequences of bovine B14.5a and B14.5b subunits*

At first, it was not recognized that the band observed at 14.5 kDa in gels of bovine complex I (Fig. 1a) contains two different proteins. Therefore, the cyanogen bromide digest and the acid cleavage was performed on mixtures of the two proteins recovered from gels. However, once subunit B14.5a had been isolated from subcomplex Iλ and its tryptic peptides had been sequenced, it seemed likely that the original material contained more than one protein, and this was confirmed once the complete cDNA sequence had been determined from overlapping partial cDNAs (see Fig. 2). This showed that the mature B14.5a protein is 112 amino acids long and that the molecular mass estimated from the sequence (12,587.4), assuming that the initiator methionine is removed and the *N*-α-amino group of the adjacent alanine is acetylated, agrees with that obtained by electrospray mass spectrometry. This sequence accounted for all of the tryptic peptides that were characterized from B14.5a isolated from subcomplex Iλ, and for a sequence starting at residue 64 of the complete protein (Fig. 2), derived from a cyanogen bromide fragment obtained from the material isolated from complex I. However, neither a sequence of another cyanogen bromide fragment nor another sequence seen after the acidic cleavage was present.

The residual cyanogen bromide fragment sequences

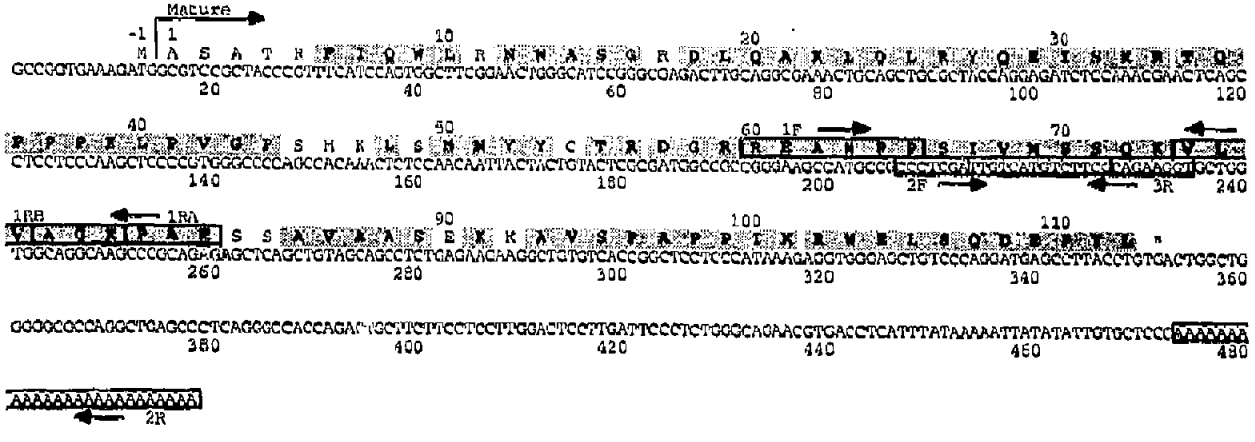


Fig. 2. Analysis by protein and DNA sequencing of subunit B14.5a of complex I from bovine heart mitochondria. The shaded regions were determined by direct protein sequencing of peptides (see section 2). Overlapping cDNAs were isolated from total bovine cDNA by a strategy based on the polymerase chain reaction. The boxed sequences 1F, 1RA and 1RB were used to design synthetic oligonucleotide mixtures which were employed in the initial polymerase chain reaction as forward primer and as overlapping nested reverse primers, respectively. The two boxed nucleotide sequences served as unique complementary primers in two further reactions in which the sequence was extended to the 3′ and 5′ extremities. Forward primer 2F was used with reverse primer 2R in the former reaction. An oligonucleotide complementary to a synthetic homopolymer 5′G tail, which had been added to the cDNA [15], was used with reverse primer 3R in the latter reaction. The hybridising product of this reaction was isolated and used as template in another polymerase chain reaction with the same primers. The senses of the various primers are indicated by arrows.

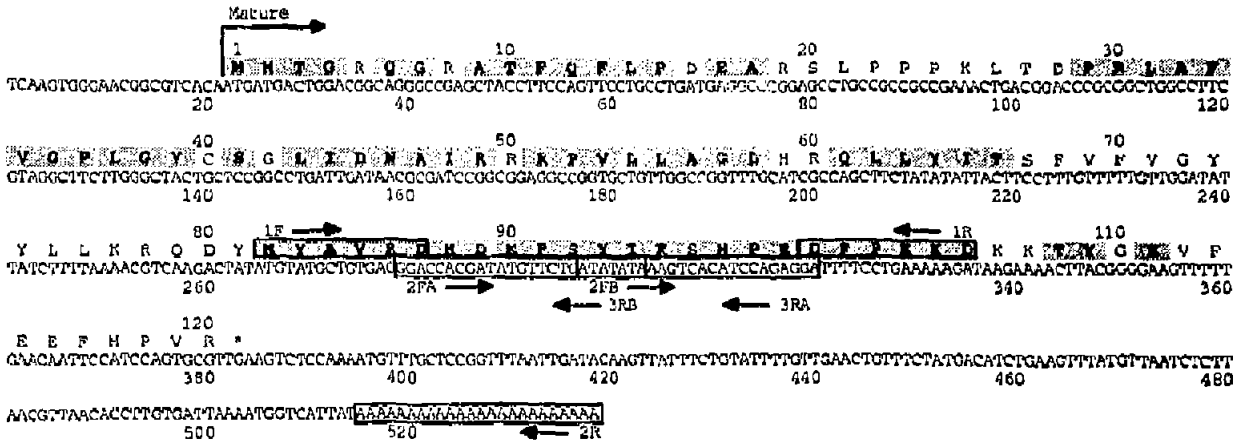


Fig. 3. Analysis by protein and DNA sequencing of subunit B14.5b of complex I from bovine heart mitochondria. For the meaning of the symbols and details of the strategy see legend to Fig. 2 and section 2.

were used to obtain the first partial clone of B14.5b, and its completed protein sequence (see Fig. 3) also contained the partial sequence from the acid cleavage reaction. If it is assumed that methionine-1 is the initiator, and that it becomes post-translationally modified, the mature protein is 120 amino acids long. Its molecular weight calculated from the sequence is 14,138.3 (assuming acetylation of the N-terminal methionine), close to one of the masses (14,141.6 ± 1.1) observed by mass spectrometry. The second value observed by mass spectrometry (14,099.4 ± 0.4) corresponds to the non-acetylated species (calculated mass 14,096.3), and suggests that a minor fraction of subunit B14.5b is unmodified. Re-examination of protein sequencing data showed that the N-terminal sequence was present at a low level, and that the non-acetylated protein is the minor form. It has been observed by in vitro processing of mutants of human β-globin, that a protein starting with two consecutive methionines did not undergo cleavage of the initiator methionine and was partially acetylated [24]. Subunit B12 of complex I is also partially acetylated, but here the initiator methionine is removed and the adjacent alanine is the partially modified amino acid [2]. In addition to B14.5a and B14.5b, nine other nuclear-encoded subunits of complex I have modified N-terminals. All of them appear to be *N*-α-acetylated, except for B18 which is myristylated on the glycine following its initiator methionine. With the exception of B14.5b, the modification requires removal of the initiator methionine, and is found on an immediately adjacent alanine or serine residue.

The cDNA sequences that were determined do not extend in a 5′ direction far beyond the coding sequence. Nonetheless, there is an in-frame termination codon upstream of the proposed initiator methionine of B14.5a, verifying the choice of initiation codon, and demonstrating that this subunit has no cleavable N-terminal extension to direct it into the mitochondrion. There is no in-frame termination codon upstream of the proposed initiator of B14.5b, but if the assignment is correct as seems probable, B14.5b also lacks a cleavable N-terminal extension. Nine further nuclear-encoded subunits of bovine complex with modified N-terminals also lack such presequences [2], as do an additional 6 subunits with free N-termini [2,14]. Therefore, a total of 17 out of the 34 nuclear-encoded subunits of complex I have no cleavable presequences to direct their import into the organelle. The signals are presumably present in the mature protein and may be close to their N-terminal regions as demonstrated in the case of the d subunit of mitochondrial ATP synthase (S.M. Medd, V.L.J. Tybulewicz, and J.E. Walker, unpublished work). The ADP/ATP translocase also has no processed presequence to direct it into the inner membrane of the mitochondria. Here also targetting information is found in the N-terminal region [25], and additional stop transfer sequences are present at internal sites [26]. The other 17 nuclear-encoded subunits have cleavable presequences and are presumably imported by the general mechanism of import of inner membrane proteins [27]. Neither of the mature B14.5a and B14.5b proteins is notably hydrophobic, which is consistent with their presence in the extramembrane domain of the enzyme. Subunits B14.5a and B14.5b are related to each other over residues 33–60 and 19–51, respectively. Neither protein is significantly related to any other subunit of bovine complex I, nor to any other known protein sequence. Their functions in complex I are obscure.

3.3. *Molecular weight of complex I from bovine heart mitochondria*

The subunits of complex I isolated from bovine heart mitochondria have been subject to extensive analysis by protein chemical and mass spectrometric techniques [2], and it now appears likely that the sequences of all of its known nuclear-encoded subunits have been determined. Many of the subunits can be resolved by 2-dimensional isoelectric focussing and electrophoresis in poly-

acrylamide gels [14], but because many of the subunits have molecular masses in the range of 10–20,000 and have N-terminal modifications, not all of components detected by such analyses can yet be assigned with certainty. This should be possible once suitable antibodies recognizing each individual subunit have been prepared. Therefore, until this analysis is complete, and until further mass spectrometric analyses have been completed, it remains possible that other subunits of the enzyme remain undiscovered. The 34 completed nuclear-encoded subunits contain 5,837 amino acids, with a corresponding molecular mass of 668,364. If unit stoichiometries for all the subunits are assumed, the 34 nuclear-encoded subunits together with the 7 subunits of complex I that are encoded in mitochondrial DNA contain a total of 7,955 amino acids, and their combined molecular mass is 906,826, including all known post-translational modifications, except for Fe-S clusters. For comparison, the *E. coli* ribosome contains 7,336 amino acid residues and its proteins have a combined molecular mass of about 820,000 [28].

Acknowledgements: We thank Mrs. J. Fogg and Mr. T.V. Smith for synthesizing oligonucleotides. J.M.A. was supported by a long term EMBO Fellowship.

REFERENCES

[1] Walker, J.E. (1992) Q. Rev. Biophys. (in press).
[2] Walker, J.E., Arizmendi, J.M., Dupuis, A., Fearnley, I.M., Finel, M., Medd, S.M., Pilkington, S.J., Runswick, M.J. and Skehel, J.M. (1992) J. Mol. Biol. 226, 1051–1072.
[3] 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.
[4] Chomyn, A., Cleeter, M.W.J., Ragan, C.I., Riley, M., Doolittle, R.F. and Attardi, G. (1986) Science 234, 614–618.
[5] 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.
[6] Fearnley, I.M. and Walker, J.E. (1992) Biochim. Biophys. Acta (in press).
[7] Fearnley, I.M., Runswick, M.J. and Walker, J.E. (1989) EMBO J. 8, 665–672.
[8] Fearnley, I.M., Finel, M., Skehel, J.M. and Walker, J.E. (1991) Biochem. J. 278, 821–829.
[9] Pilkington, S.J. and Walker, J.E. (1989) Biochemistry 28, 3257–3264.
[10] Pilkington, S.J., Skehel, J.M. and Walker, J.E. (1991) Biochemistry 30, 1901–1908.
[11] Pilkington, S.J., Skehel, J.M., Gennis, R.B. and Walker, J.E. (1991) Biochemistry 30, 2166–2175.
[12] Runswick, M.J., Gennis, R.B., Fearnley, I.M. and Walker, J.E. (1989) Biochemistry 28, 9452–9459.
[13] Runswick, M.J., Fearnley, I.M., Skehel, J.M. and Walker, J.E. (1991) FEBS Lett. 286, 121–124.
[14] Dupuis, A., Skehel, J.M. and Walker, J.E. (1991) Biochem. J. 277, 11–15.
[15] Dupuis, A., Skehel, J.M. and Walker, J.E. (1991) Biochemistry 30, 2954–2960.
[16] Skehel, J.M., Pilkington, S.J., Runswick, M.J., Fearnley, I.M. and Walker, J.E. (1991) FEBS Lett. 282, 135–138.
[17] Arizmendi, J.M., Runswick, M.J., Skehel, J.M. and Walker, J.E. (1992) FEBS Lett. 301, 237–242.
[18] Finel, M., Skehel, J.M., Albracht, S.J.P., Fearnley, I.M. and Walker, J.E. (1992) Biochemistry (in press).
[19] Ragan, C.I., Wilson, M.T., Darley-Usmar, V.M. and Lowe, P.N. (1987) in: Mitochondria: A Practical Approach (Darley-Usmar, V.M., Rickwood, D. and Wilson, M.T., eds.) IRL Press, Oxford and Washington, pp. 79–112.
[20] Laemmli, U.K. (1970) Nature 227, 680–685.
[21] Schägger, H. and von Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[22] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[23] Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
[24] Boissel, J.-P., Kasper, T.J. and Bunn, H.F. (1988) J. Biol. Chem. 263, 8443–8449.
[25] Pfanner, N. and Neupert, W. (1987) J. Biol. Chem. 262, 7528–7536.
[26] Pfanner, N., Hoeben, P., Tropschug, M. and Neupert, W. (1987) J. Biol. Chem. 262, 14851–14854.
[27] Pfanner, N. and Neupert, W. (1990) Annu. Rev. Biochem. 59, 331–353.
[28] Wittmann, H.G. (1982) Annu. Rev. Biochem. 51, 155–183.