Thomann, H., Dalton, L. R., & Dalton, L. A. (1984) in *Biological Magnetic Resonance* (Berliner, L. J., & Reuben, J., Eds.) Vol. 6, pp 143–186, Plenum Press, New York.
Trebst, A., Harth, E., & Draber, W. (1970) *Z. Naturforsch. 25b*, 1157–1159.
Tsang, H.-T., Batie, C. J., Ballou, D. P., & Penner-Hahn, J. E. (1989) *Biochemistry 28*, 7233–7240.
Tsukihara, T., Fukuyama, K., Nakamura, M., Katsube, Y., Tanaka, N., Kakudo, M., Wada, K., Hase, T., & Matsubara, H. (1981) *J. Biochem. 90*, 1763–1773.
Wikström, M., & Krab, K. (1980) *Curr. Top. Bioenerg. 10*, 51–101.
Wikström, M., Krab, K., & Saraste, M. (1981) *Annu. Rev. Biochem. 50*, 623–655.
Yu, C.-A., & Yu, L. (1980) *Biochim. Biophys. Acta 591*, 409–420.
Yu, L., Mei, Q.-C., & Yu, C.-A. (1984) *J. Biol. Chem. 259*, 5752–5760.

The 30-Kilodalton Subunit of Bovine Mitochondrial Complex I Is Homologous to a Protein Coded in Chloroplast DNA†

Stephanie J. Pilkington,‡ J. Mark Skehel, and John E. Walker*
Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.
Received June 21, 1990; Revised Manuscript Received October 29, 1990

ABSTRACT: In cattle, 7 of the 30 or more subunits of the respiratory enzyme NADH:ubiquinone reductase (complex I) are encoded in mitochondrial DNA, and potential genes (open reading frames, orfs) for related proteins are found in the chloroplast genomes of *Marchantia polymorpha* and *Nicotiana tabacum*. Homologues of the nuclear-coded 49- and 23-kDa subunits are also coded in chloroplast DNA, and these orfs are clustered with four of the homologues of the mammalian mitochondrial genes. These findings have been taken to indicate that chloroplasts contain a relative of complex I. The present work provides further support. The 30-kDa subunit of the bovine enzyme is a component of the iron–sulfur protein fraction. Partial protein sequences have been determined, and synthetic oligonucleotide mixtures based on them have been employed as hybridization probes to identify cognate cDNA clones from a bovine library. Their sequences encode the mitochondrial import precursor of the 30-kDa subunit. The mature protein of 228 amino acids contains a segment of 57 amino acids which is closely related to parts of proteins encoded in orfs 169 and 158 in the chloroplast genomes of *M. polymorpha* and *N. tabacum*. Moreover, the chloroplast orfs are found near homologues of the mammalian mitochondrial genes for subunit ND3. Therefore, the plant chloroplast genomes have at least two separate clusters of potential genes encoding homologues of subunits of mitochondrial complex I. The bovine 30-kDa subunit has no extensive sequences of hydrophobic amino acids that could be folded into membrane-spanning α-helices, and although it contains two cysteine residues, there is no clear evidence in the sequence that it is an iron–sulfur protein.

The chloroplast genomes of the liverwort, *Marchantia polymorpha*, and the tobacco plant, *Nicotiana tabacum*, are 121 024 and 155 844 bp long, and they are thought to contain 119 and 122 different genes, respectively (Ohyama et al., 1986, 1988; Shinozaki et al., 1986). Both genomes have a duplicated region, and the large difference between the sizes of the two genomes is accounted for primarily by the lengths of these inverted DNA repeats. About 90 genes have been identified in each genome, often by homology of predicted protein sequences encoded in the chloroplast DNAs with those of known proteins (Ohyama et al., 1988). They fall broadly into two categories: those involved in transcription and translation in the chloroplast and those required for bioenergetic functions, such as photosynthetic electron transport and ATP synthesis. About 60 genes are in the former group and about 20 in the latter. In addition, seven open reading frames (orfs) in the liverwort code for protein sequences that are homologous to seven components of mitochondrial NADH:ubiquinone reductase (complex I) in mammals and other species [see Fearnley et al. (1989) for a summary]. These proteins are encoded in mitochondrial DNA and are known as subunits ND1–ND6 and ND4L (Chomyn et al., 1985, 1986). The liverwort chloroplast orfs are named *ndh*1–*ndh*6 and *ndh*4L; six homologues (*ndh*A, *ndh*B, *ndh*C, *ndh*D, *ndh*E, and *ndh*F) are also present in tobacco chloroplasts, and two tobacco open reading frames (orfs 138 and 99B) correspond to liverwort *ndh*6 if a frame shift is introduced in the tobacco DNA sequence (Ohyama et al., 1988). These unexpected findings have led to suggestions that the putative chloroplast proteins encoded in the orfs are subunits of an NADH or NADPH:plastoquinone reductase component of a chloroplast respiratory electron transport activity and that this enzyme complex will be rather closely related to the mitochondrial NADH:ubiquinone reductase complex. This view is strengthened by the recent finding that the 49- and 23-kDa subunits, both nuclear-encoded components of the mitochondrial enzyme, also have homologues encoded in orfs in both chloroplast genomes (Fearnley et al., 1989; Dupuis et al., 1991), and the orfs are both in a gene cluster that also contains four of the homologues of the mitochondrial gene products of complex I.

The work presented provides further evidence for a complex I like assembly in chloroplasts. We have determined from among the 30 or more subunits of complex I isolated from bovine heart mitochondria the primary structure of the 30-kDa subunit, a component of the simpler iron–sulfur (IP) fraction

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05314.
* To whom correspondence should be addressed.
‡ Supported by a Medical Research Council research studentship.

0006-2960/91/0430-1901$02.50/0

that can be prepared from the complex with chaotropic agents [reviewed by Ragan (1987)]. Part of the sequence of the 30-kDa protein is homologous to proteins encoded in orfs in the plant chloroplast DNAs. Moreover, these orfs are found in close proximity to reading frames that encode a homologue of the complex I mitochondrial gene product, ND3. Also, the sequence shows that the 30-kDa subunit is probably not an iron–sulfur protein.

Materials and Methods

Purification of Complex I. Complex I was purified from bovine heart mitochondria (Hatefi et al., 1962) and assayed as described previously (Fearnley et al., 1989). Its activity was about 50 units/mg of protein where 1 unit reduces 1 μmol of ferricyanide in 1 min at 30 °C.

Protein Chemical Methods. Complex I (100 μg) was fractionated on gradient gels (10–25% acrylamide; Laemmli, 1970). The 30-kDa subunit was recovered by electroelution (Runswick et al., 1989) and concentrated by precipitation with 90% ethanol at −20 °C overnight. A sample of it (1–10 μg) was cleaved for 18 h with cyanogen bromide, and another (5–10 μg) was digested with *Staphylococcus aureus* V8 protease (0.05 mg/mL) for 45 min at 18 °C. The products of both reactions and a sample of complex I were fractionated on a minigel and a 10–25% gradient gel, respectively, transferred to poly(vinylidene difluoride) membranes (Matsudaira, 1987), and sequenced as described elsewhere (Fearnley et al., 1989).

Screening the cDNA Library. Approximately 5×10^5 recombinants in a bovine plasmid library in pUC8 (Vieira & Messing, 1982; Gay & Walker, 1985a) were screened with 2 mixtures, each of 64 oligonucleotides 17 bases long, based on the overlapping protein sequences MFGVFF and FFANHP, determined by N-terminal sequence analysis of a cyanogen bromide fragment of the 30-kDa subunit. Recombinants in the library were grown, transferred to filters, and prehybridized as described before (Fearnley et al., 1989). Hybridizations with probes were carried out on replica sets of filters at 37 and 39 °C (5 °C below minimum dissociation temperatures; Gay & Walker, 1985b; Fearnley et al., 1989). About 200 recombinants gave a strong signal with both probes. Eight were grown up, and three classes were distinguishable from restriction fragment sizes. One from each group, isolates pBovCI-30.1, pBov CI-30.2, and pBovCI-30.3, was selected for further study.

Subcloning and DNA Sequencing. Inserts were excised from plasmids with *Eco*RI and *Pst*I and gave fragments of about 850, 500, and 1000 bp, respectively, from pBovCI-30.1, -.2, and -.3. Their DNA sequences were determined by the modified dideoxy method (Sanger et al., 1977; Biggin et al., 1983) after cloning them into M13mp18 and -mp19 vectors. The complete sequences in both orientations of pBovCI-30.1 and pBovCI-30.2 were determined with the aid of three internal primers on the entire inserts and on subfragments obtained by digestion at internal and polylinker *Bam*HI and *Hin*dIII sites. The sequence of pBovCI-30.3 was determined from the *Eco*RI–*Pst*I fragment cloned in M13mp19 only, from fragments arising from digestion at the internal and polylinker *Bam*HI and *Hin*dIII sites, and from fragments obtained by digestion with *Bam*HI and *Sac*II together and *Bam*HI and *Eag*I together. Overlapping sequence was obtained across all restriction enzyme sites, and the sequence was determined in both sensees throughout.

Polymerase Chain Reactions. Amplification of cDNA using the polymerase chain reaction was performed with unique 34-base synthetic oligonucleotide primers consisting of 9 bases of linker sequence and 25 bases complementary to the cDNA. Forward and reverse primers were made with *Eco*RI or *Pst*I linkers, respectively. Template single- or double-stranded cDNA was made from poly(A+) mRNA (see below) using a cDNA synthesis kit (Amersham International, Amersham, U.K.). Single-stranded cDNA derived from mRNA (1.5 μg) was tailed at its 3′ end with a run of dA residues using terminal deoxynucleotide transferase (36 units; Pharmacia, Milton Keynes, U.K.) in a buffer (15 μL) containing 100 mM potassium cacodylate, pH 6.9, 0.1 mM dithiothreitol, 1 mM cobalt chloride, 1 mM dATP, and 0.5 μg of bovine serum albumin. The reaction was performed at 37 °C for 30 min. End-repaired double-stranded cDNA (100 ng) was circularized by incubation at 15 °C for 12 h in a buffer (30 μL) containing 33 mM Tris–acetate, pH 7.5, 66 mM potassium acetate, 10 mM magnesium acetate, 3 mM spermidine, 0.1 mg/mL bovine serum albumin, 0.75 mM ATP, 240 units of T4 DNA ligase, and 160 units of T4 RNA ligase. Part of this solution (5 μL) was used in each reaction with the following schedule: denaturation (94 °C) for 1 min, annealing (55 °C) for 2 min, synthesis (72 °C) for 2 min, repeated for 35 cycles. This was followed by a single incubation at 72 °C for 7 min. For other details, see Runswick et al. (1990). The *Thermus aquaticus* DNA polymerase was added after preincubation at 94 °C for 2.5 min followed by 5 min at 60 °C. Sequences from circularized cDNA templates were amplified for 30 cycles at a low primer concentration (20 nM). A fresh aliquot of enzyme and more of the same primers (final concentration 1 μM) were added, and a further 30 cycles of synthesis were performed. The products were analyzed on 1% or 1.4% high melting point agarose gels. Fragments were recovered and digested with *Eco*RI and *Hin*dIII or *Pst*I and cloned into M13 mp18 or M13 mp19 vectors.

Computer Methods. The protein sequence of the 30-kDa subunit of complex I was compared with sequences in the PIR database with FASTP (Lipman & Pearson, 1985). The sequences of chloroplast DNA from *M. polymorpha* (Ohyama et al., 1986, 1988; Kohchi et al., 1988) and *N. tabacum* (Shinozaki et al., 1986) were translated in all phases and in both senses of the DNA. Predicted protein sequences greater than 14 amino acids in length were compiled into a database in the PIR format, and screened with the protein sequence of the 30-kDa subunit of complex I, using FASTA (Lipman & Pearson, 1985). Pairwise comparisons between the 30-kDa subunit and others were made with DIAGON (Staden, 1985). Secondary structure predictions were made by the method of Garnier et al. (1978). Hydrophobic profiles were calculated with SOAP (Kyte & Doolittle, 1982).

Hybridization with Human Genomic DNA. Human DNA was prepared from blood (Pilkington & Walker, 1989), and samples (20 μg/slot) were digested at 37 °C in 400 μL of buffer with each of the restriction endonucleases *Eco*RI, *Pst*I, *Hin*dIII, and *Xba*I. Hybridization was carried out with the radiolabeled "prime cut" probes (Farrell et al., 1983) for 18 h at 65 °C. The filters were washed twice at room temperature in a mixture of 2 × SSC and 0.5% and *N*-laurylsarcosine, twice again in the same solution at 65 °C, and then twice more at 65 °C in a solution that contained 0.2 × SSC and 0.5% *N*-laurylsarcosine. For other details including prehybridization and hybridization conditions and radioautography, see Walker et al. (1987a).

Results and Discussion

Protein Sequence Analysis of the 30-kDa Subunit. The sequence of amino acids 1–26 (Table I) was determined on a sample obtained from complex I (Figure 1), and was iden-

Table I: Protein Sequences Determined on the 30-kDa Subunit of Complex I and on Digestion Products

sample[a]	apparent molecular mass (kDa)	determined sequence	position in protein
intact protein	30	ESSAADTRPTVRPRNDVAHKQLSAFG	1–26
CNBr (band d)	15	FGVFFANHPDLXRILTXYGFEXXPF	138–162
CNBr (band c)	14	ESSAADTR	1–8
V8 protease (band b)	16	LTPIEXSVPVXKAAN	114–128

[a] For bands b, d, and e, see Figure 1. Band c contained a mixture of sequences from the N terminus and from band b.

tical with that obtained by analysis of the 30-kDa band in the iron–sulfur fraction of the enzyme. The identity of residue 19 was tentative (although the histidine assignment proved to be correct). The sequence did not contain a segment of six consecutive amino acids suitable for design of an oligonucleotide mixture of low degeneracy for screening a cDNA library. Therefore, fragments of the protein produced by cleavage with cyanogen bromide and by enzymic digestion were purified by polyacrylamide gel electrophoresis and sequenced. Although the bands in the protease digest in particular were rather faint, N-terminal sequence was obtained without difficulty. The apparent molecular masses of the two fragments from the former were about 15 and 14 kDa (see Figure 1B). Therefore, the 30-kDa subunit probably contains one methionine residue midway in its polypeptide chain. The partial protease digest contained two major bands with apparent molecular weights of about 14 and 16 kDa (Figure 1B, Table I).

cDNA Sequence of the 30-kDa Subunit. Recombinants pBovCI-30.1 and pBovCI-30.2 overlap and gave the sequence of nucleotides 51–920 (see Figure 2). Isolate pBovCI-3 proved to be more problematical. The insert released by digestion with *Eco*RI and *Pst*I was about 1 kb in length, but no clones were obtained in one orientation. Also, several M13 isolates in the reverse orientation extended to different points in a region about 200 bp from the 5′ end of the insert in pBov CI-30.1, and there was no recognizable adjacent cloning site. Therefore, deletions had occurred during cloning, but from its estimated size, it appeared likely that the fragment extended beyond the 5′ end of the insert in pBovCI-30.1. It was shown to extend 174 bp beyond the 5′ end of the insert in pBovCI-30.1, and the 123 bp at its 5′ end were found to be an inverted repeat of residues 155–278. No poly(A) sequence was found at the 3′ end (see Figure 2 for further details).

Verification of the sequence of the cDNA for the 30-kDa subunit was obtained by sequencing appropriate cDNAs produced by the polymerase chain reaction. First, unique oligonucleotide primers were used to amplify nucleotides 36–822 of the sequence determined from plasmid isolates. Secondly, the 5′ and 3′ ends of the cDNA sequence were confirmed by using unique oligonucleotide primers either on cDNA that had been homopolymer-tailed using terminal deoxynucleotide transferase or on circularized cDNA. By this means, nucleotides −2 to 127 (see legend to Figure 2) and 746–893 were amplified.

The final sequence presented in Figure 3 is 920 bp long. The sequences of the plasmid isolates differ from each other and from polymerase chain reaction products in three places, and these are boxed in Figure 3. At positions 53–54, 165, and 324–325, isolate pBovCI-30.1 has the sequences AA, A, and GT, whereas GT, C, and AC are found in these positions in pBovCI-30.3. At these sites of disagreement, the polymerase chain reaction products had the sequences GT, A, and AC, respectively, and these are the sequences that appear in Figure 3. At the 3′ end, the polyadenylation signal AATAAA (Proudfoot & Brownlee, 1976) is found 14 bp before the start of the poly(A) region. Near the 5′ end, a preponderance of G and C residues suggests that a CpG-rich island (Bird, 1986) may be associated with the 5′ end of the gene.

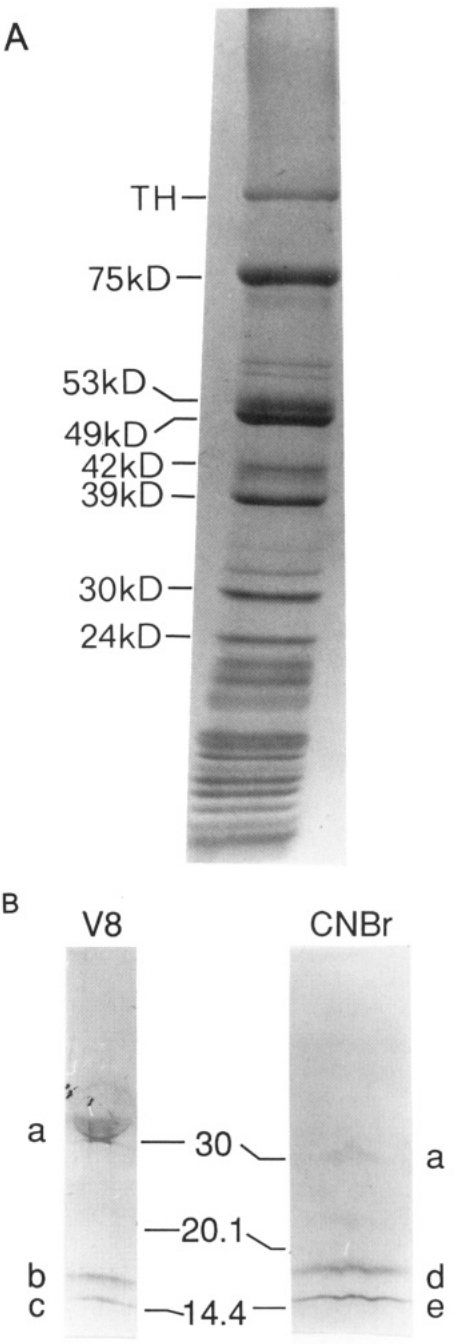

FIGURE 1: Separation by polyacrylamide gel electrophoresis of the subunits of complex I from bovine heart mitochondria and of fragments of its 30-kDa subunit. (A) Complex I; the positions of several subunits are indicated, and also contaminating transhydrogenase, TH. (B) *Staph. aureus* V8 protease partial cleavage products and the cyanogen bromide cleavage products of the 30-kDa subunit. Band a is the uncleaved protein, and bands b–d are cleavage products. The positions of molecular mass markers are shown.

Sequence of the 30-kDa Protein. The cDNA sequences provide the primary structure for the mitochondrial import precursor of the 30-kDa protein (Figure 3). At residue 16,

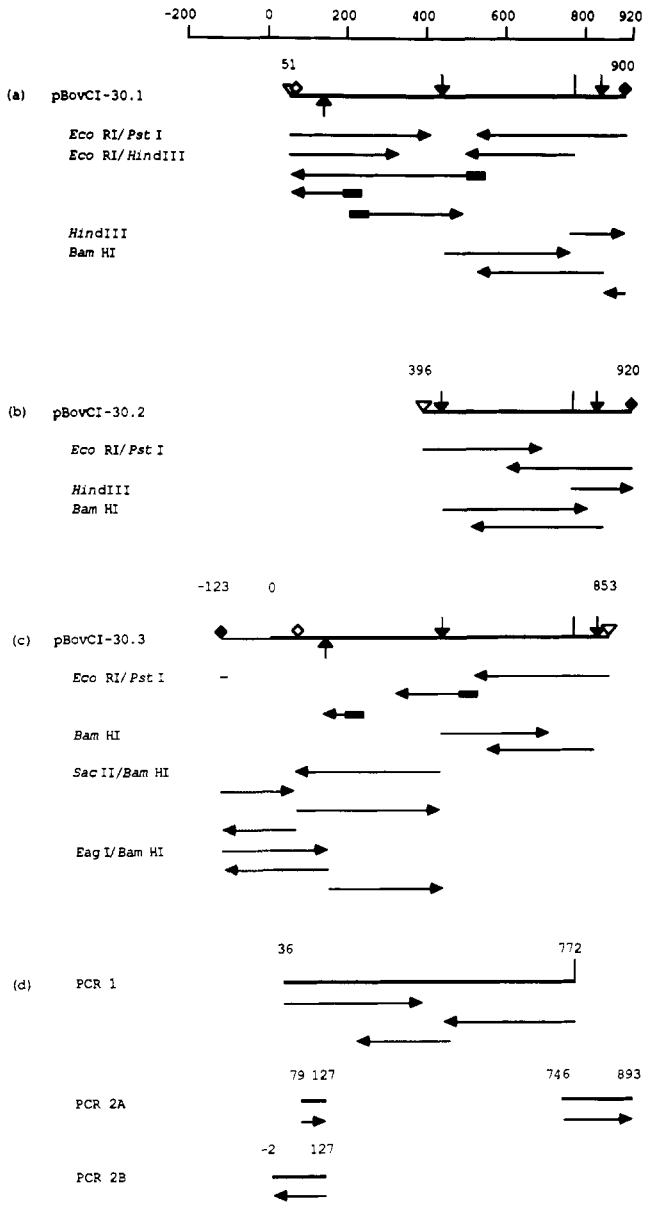

FIGURE 2: Sequence analysis of cDNA clones for the 30-kDa subunit of complex I from bovine heart mitochondria. Sections a, b, c, and d refer to the three plasmid isolates pBovCI-30.1, pBovCI-30.2, and pBovCI-30.3 and products amplified by using the polymerase chain reaction (PCR 1, 2A, and 2B), respectively. The thick lines represent the inserted DNAs in the plasmid isolates pBovCI-30.1, pBovCI-30.2, and pBovCI-30.3 or products amplified by using the polymerase chain reaction (PCRs 1, 2A, and 2B). The plasmid inserts are flanked by *Eco*RI at one end and by *Bam*HI, *Pst*I, and *Hin*dIII sites in the polylinker of the vector at the other. The horizontal arrows indicate the extent and direction of the determined DNA sequences. The positions of oligonucleotide primers used in sequencing reactions are indicated by black boxes. The grey shaded lines represent an inverted repeat of internal sequence (not shown in the final cDNA sequence in Figure 3). As shown in section d, PCR 2B extended two bases in the 5′ direction beyond the authentic cDNA sequence determined from pBovCI-30.3. These bases are not shown in the final sequence since they obtained from one product only, and hence they are numbered with relation to the cDNA sequence of pBovCI-30.3. The scale is in bases. Restriction enzyme sites are denoted as follows: (▽) *Eco*RI; (↓) *Bam*HI; (|) *Hin*dIII; (↑) *Eag*I; (◇) *Sac*II; (◆) polylinker.

both alanine and aspartic acid were detected by protein sequencing, while at this point (nucleotide 165), pBovCI-30.1 and the polymerase chain reaction product encode aspartic acid, and pBovCI-30.3 encodes alanine. Thus, both protein sequencing and cDNA sequencing provide evidence for isoforms, although the ambiguity in the protein sequencing experiments could also represent different alleles in the bovine population. At amino acid 69 (outside the regions of directly determined protein sequence), pBovCI-30.1 encodes a serine residue, whereas pBovCI-30.3 and the polymerase chain reaction product encode an asparagine. There are no other differences within the mature protein sequence, but a further ambiguity occurs at amino acid −22 of the presequence (nucleotides 53–54). A lysine is predicted from the DNA sequence of pBovCI-30.1 and a valine from pBovCI-30.3 and the pcr product. However, this sequence lies 2 bp from the end of the insert in pBovCI-30.1, where errors in cDNA synthesis are more likely, and so the valine is more plausible.

The molecular weight of the mature protein calculated from the sequence is 26 432 rather than the estimated value of 30K. This difference may be due to the persistence of strong secondary structure in the protein under these conditions, and secondary structure prediction (Garnier et al., 1978) revealed a region with a strong potential to form an α-helix between residues 173 and 197 which might be responsible. It is also noteworthy that estimates of molecular weight depend on the gel system, and in urea/SDS gels (Schägger & von Jagow, 1987), the apparent molecular weight of the 30-kDa subunit is about 27K.

The molecular mass of the mature protein has been determined by electrospray mass spectrometry to be 26 433 (I. M. Fearnley and J. E. Walker, unpublished results). This agrees with the value of 26 432 calculated from the sequence in Figure 3 and appears to eliminate the alternate assignments at residues 16 and 69 of alanine and serine, respectively.

Many mitochondrial import sequences share no obvious sequence homology but are characterized by a net positive charge, and acidic amino acids are usually, but not invariably, absent (Nicholson & Neupert, 1988). The presequence of the 30-kDa subunit is 38 amino acids in length. It contains no acidic residues and has a positive charge of 7 (if residue −22 is valine) or 8 (if residue −22 is lysine), assuming that each arginine or lysine residue and the N-terminal methionine each contribute a charge of +1. The ability to form an amphiphatic α-helix has been suggested as a feature of presequences (Roise et al., 1986; von Heijne, 1986), but prediction of secondary structure did not reveal any regions in the 30-kDa presequence with a high probability to form α-helical structure nor was any amphiphatic feature detected by plotting the sequence on a helical wheel.

The 30-kDa subunit is a component of the iron–sulfur protein (IP) fraction of complex I, which has six major constituents, the 75-, 49-, 30-, 18-, 15-, and 13-kDa subunits, and contains at least three iron–sulfur clusters. A 4Fe–4S center and a 2Fe–2S center are present in a subfraction containing the 49-, 30-, and 13-kDa polypeptides, but further resolution leads to loss of characteristic EPR line spectra and hence more precise assignment of the clusters was not possible. The primary structure of the 49-kDa polypeptide contains five cysteine residues (Fearnley et al., 1989), but none of the structural motifs associated with 4Fe–4S or 2Fe–2S clusters in bacteria or chloroplasts were present, and so the sequence provided no convincing evidence for the association of an iron–sulfur center with this subunit. The primary structure of the 30-kDa subunit contains only two cysteines, separated by six residues. So it also appears unlikely that the 30-kDa subunit contains iron–sulfur clusters, although the possibility remains that it could contribute ligands to one or more clusters jointly with other subunits of complex I.

Chemical labeling experiments show that the 30-kDa subunit and other components of the IP fraction are accessible

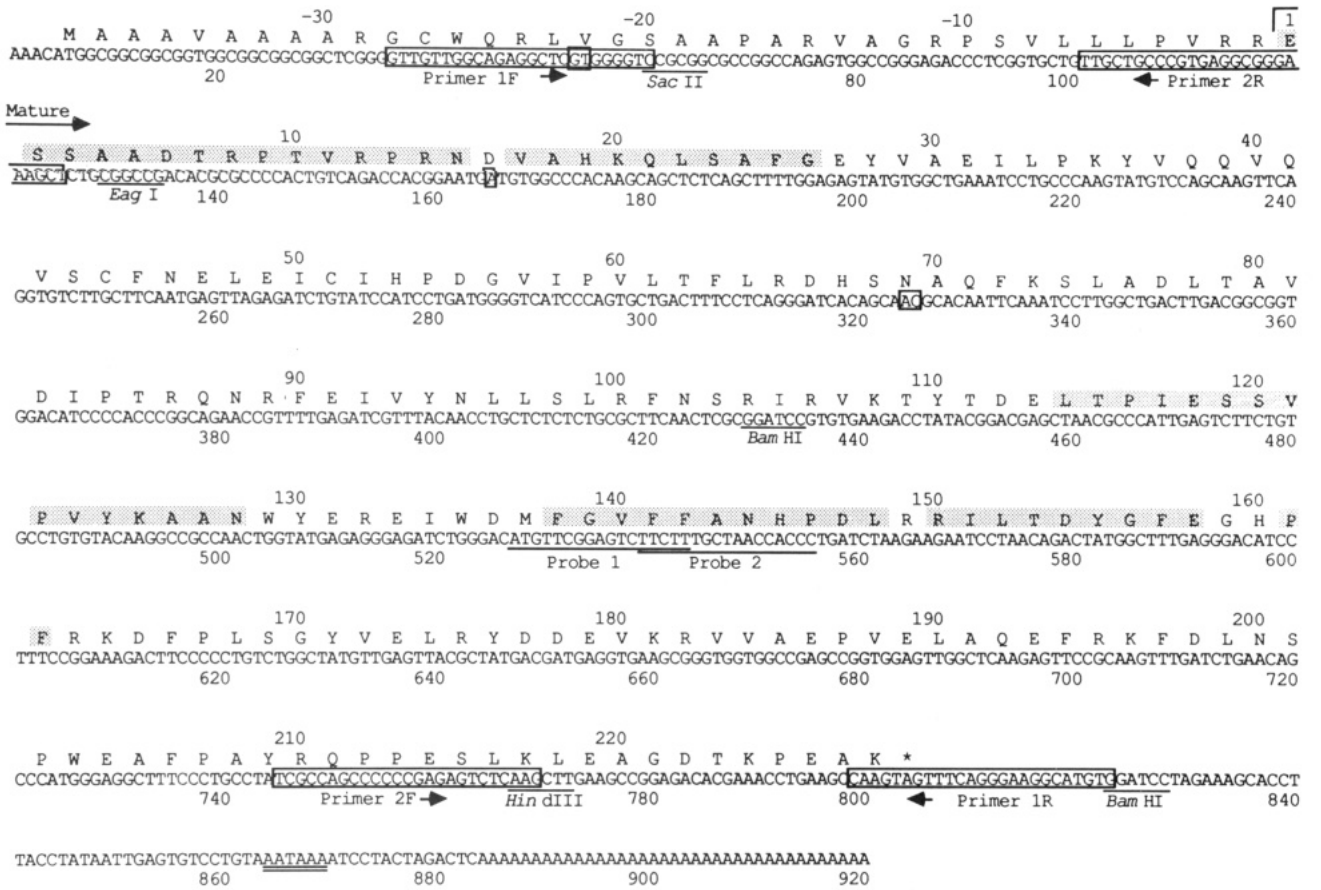

FIGURE 3: cDNA sequence encoding the 30-kDa subunit of complex I from bovine heart mitochondria. The sequence shown is that obtained from the products of polymerase chain reactions. Nucleotides where the sequence differs in isolates pBovCI-30.1 or pBovCI-30.3 are boxed. Restriction enzyme sites used in the process of sequencing are indicated. A possible polyadenylation signal near the 3′ end is underlined twice. The deduced protein sequence is shown above the DNA sequence, and directly determined protein sequences are shaded. The regions that were used to design oligonucleotide probes 1 and 2 are indicated. Boxed nucleotide sequences are those of unique primers used in polymerase chain reactions. Horizontal arrows pointing to the right or left indicate that the primer was synthesized with the sequence shown or its complement, respectively.

to hydrophilic probes in isolated complex I and that in intact mitochondria and inverted submitochondrial vesicles the 30-kDa subunit is exposed to similar extents on both sides of the membrane (Patel et al., 1988). This conclusion is supported by the different patterns of cross-linking to other subunits that are obtained, depending on which side of the membrane is exposed to the cross-linking agent (Patel et al., 1988; Patel & Ragan, 1988). In contrast, no labeling was observed with hydrophobic probes (Earley & Ragan, 1980, 1981), indicating that any transmembranous segments in the 30-kDa proteins must be shielded by other proteins. The hydrophobic profile of the protein (not shown) contains no convincing evidence of a hydrophobic region long enough to form a membrane-spanning helix, and the topography of the protein requires further experimental investigation.

Homology of the Protein Sequence of the 30-kDa Subunit with Chloroplast Gene Products. In order to try and obtain clues to the function of the 30-kDa subunit, its protein sequence was compared with entries in the PIR database. This revealed a highly significant relationship with a protein encoded in a potential gene (orf 169) in the chloroplast genome of the liverwort, *M. polymorpha* (Ohyama et al., 1986, 1988; Kohchi et al., 1988), and with its homologue (orf 158) encoded in the chloroplast genome of the tobacco plant, *N. tabacum* (Shinozaki et al., 1986). These predicted proteins are 169 and 158 amino acids, respectively, and hitherto no known function has been ascribed to them. The N-terminal amino acids (1–12) of the protein encoded in orf 169 are not represented in orf 158 (with the exception of the N-terminal methionine), but

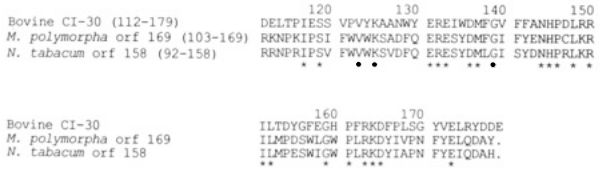

FIGURE 4: Comparison of the protein sequence of the 30-kDa subunit of complex I from bovine heart mitochondria with protein sequences encoded in orfs 169 and 158 found in the chloroplast genomes in *M. polymorpha* and *N. tabacum*. The regions of homology were detected by DIAGON. Identical amino acids in all three sequences are indicated by an asterisk.

the remainder of their sequences are identical at 122 positions (or 77%), and they have conservative replacements at a further 25 positions (16%), without any necessity to introduce insertions. Pairwise comparison of the sequences of the two chloroplast orfs with the 30-kDa subunit using the computer program DIAGON (not shown) confirmed the significance of the relationship between the mitochondrial and chloroplast proteins. The region of homology extends over 57 amino acids, 23 amino acids being identical in this region and a further 19 being conservatively substituted (Figure 4). Thus, the chloroplast-coded proteins and the mitochondrial 30-kDa subunit appear to contain a very similar domain of unknown function. The possibility was considered that a separate exon or exons elsewhere in the chloroplast genome might encode the remainder of the 30-kDa protein. Therefore, the entire chloroplast genomes of both plant species were translated in all three phases and in both senses of the DNA, and the translated

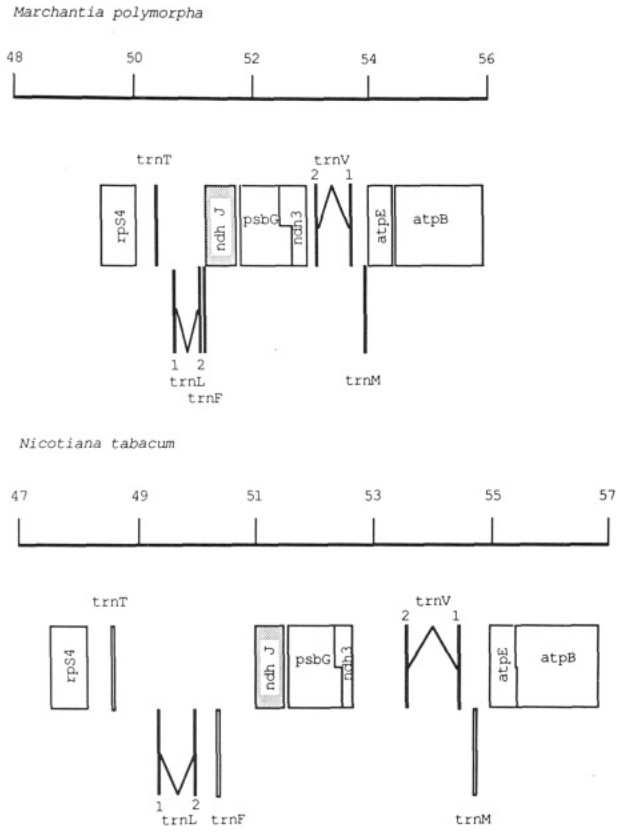

FIGURE 5: Arrangement of genes and orfs in the vicinities of orfs 169 and 158 in the chloroplast genomes of *M. polymorpha* and *N. tabacum*, respectively. The scale is in kilobases, and the diagram is derived from the published sequences (Kohchi et al., 1988; Shinozaki et al., 1986). V shapes denote introns. The unassigned open reading frames (orfs) have been renamed *ndh*J; *ndh*H and *ndh*I are the chloroplast equivalents of the 49- and 23-kDa subunits (Dupuis et al., 1991). The genes labeled *rps*4 encode homologues of the ribosomal protein S4 from *Escherichia coli*, *psa*A encodes a subunit of photosystem I; *psb*G was thought formerly to encode a component of photosystem II, but it has been suggested that it encodes a homologue of a complex I component (Mayes et al., 1990). *atp*B and *atp*E code for the β- and ϵ-subunits of the ATP synthase complex. *ndh*3 encodes the homologue of the mitochondrial ndh 3 component of complex I, and *trn* denotes tRNA genes. The orfs showing homology with the 30-kDa subunit of bovine mitochondrial complex I are shaded.

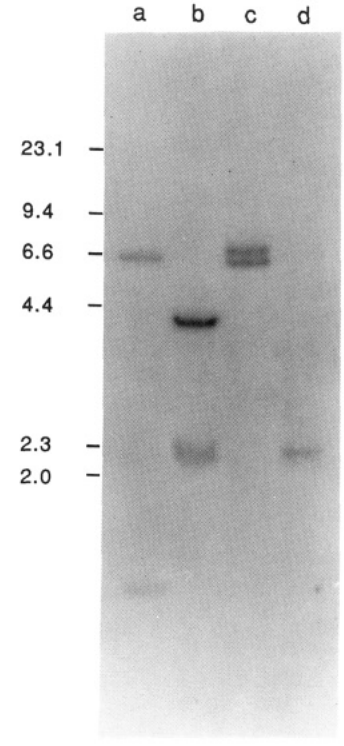

FIGURE 6: Hybridization of restriction digests of human genomic DNA with the cDNA for the bovine 30-kDa protein. The "prime cut" probe was derived from the full length of PCR 1 (nucleotides 36–772; see Figure 2). The DNA was digested with the following restriction enzymes: (a) *Eco*RI; (b) *Hin*dIII; (c) *Pst*I; (d) *Xba*I. After hybridization, the filter was washed in 0.2 × SSC at 65 °C. The positions of DNA size markers produced by digestion of DNA from bacteriophage λ with *Hin*dIII are shown at the left side of the filters.

sequences were compared with that of the 30-kDa protein. However, no homologous sequence was detected. The 49- and 23-kDa subunits of complex I also have homology with chloroplast proteins (Fearnley et al., 1989; Dupuis et al., 1991), but in both cases, the homology extends over the full length of the chloroplast proteins, and there is also homology of the 49-kDa protein with orf products encoded in the maxicircle (mitochondrial) DNA of protozoa (Simpson et al., 1987), which is not the case for the 30-kDa subunit.

Chloroplast genomes contain genes that encode homologues for the seven components of complex I encoded in vertebrate and invertebrate mitochondria (Anderson et al., 1981, 1982; Bibb et al., 1981; Görtz & Feldmann, 1982; Grosskopf & Feldmann, 1981; Koike et al., 1982; Roe et al., 1985; Clary & Wolstenholme, 1985; Canatatore et al., 1987; Wolstenholme et al., 1987; Himeno et al., 1987; Jacobs et al., 1988). In both liverwort and tobacco, four of these homologues are found in a gene cluster with the homologues of the 49- and 23-kDa subunits (Fearnley et al., 1989; Dupuis et al., 1991). The orfs related to the 30-kDa subunit are found in close proximity to, and are transcribed in the same direction as, the *ndh*3 genes in both liverwort and tobacco. No other homologues of complex I components have been identified in the same region (see Figure 5), but it has been suggested from indirect evidence that *psb* G is a component of a chloroplast assembly related to complex I (Mayes et al., 1990).

The identification of chloroplast-encoded homologues of a second nuclear-encoded component of mitochondrial complex I provides further evidence for the presence of a closely related respiratory complex. The most likely candidate is an NADH:plastoquinone oxidoreductase activity, which has been demonstrated in *Chlamydomonas reinhardtii* chloroplasts (Bennoun, 1982; Maione & Gibbs, 1986).

Number of Bovine Genes for the 30-kDa Subunit. In order to estimate the number of sequences in the human genome related to the cDNA for the 30-kDa subunit, restriction digests of genomic DNA were transferred to membranes and hybridized with a probe that includes nucleotides 36–772 of the cDNA sequence. One, two, or three bands were observed in each lane both at high and at low stringencies (see Figure 6). This does not indicate a large gene family, and the observed patterns could arise from a single gene. A similar result (not shown) was obtained with bovine genomic DNA. A number of nuclear-encoded mitochondrial proteins have been found to have more than one expressed gene, for example, the subunit c (Gay & Walker, 1985b; Dyer et al., 1989; Dyer & Walker, 1991) and the α-subunit (Walker et al., 1989) of ATP synthase, and the ADP/ATP translocase (Battini et al., 1987; Walker et al., 1987b; Neckleman et al., 1987; Houldsworth & Attardi, 1988; Powell et al., 1989). In contrast to the 30-kDa subunit, in these cases extensive differences in sequence have been found in the 5′ and 3′ noncoding regions.

Added in Proof

The sequence of a 31-kDa protein from *N. crassa* complex I (Videira et al., 1990) is homologous to the bovine 30-kDa

subunit. Neither of the cysteine residues in the bovine protein is conserved.

ACKNOWLEDGMENTS

We are grateful to Dr. J. M. Cooper for the gift of a sample of the IP fraction of bovine complex I and to 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., & Young, I. G. (1981) *Nature (London) 290*, 457–465.

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

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

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

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

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

Bird, A. P. (1986) *Nature (London) 321*, 209–213.

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

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.

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

Dupuis, A., Skehel, J. M., & Walker, J. E. (1991) *Biochemistry* (in press).

Dyer, M. R., & Walker, J. E. (1991) *J. Mol. Biol.* (in press).

Dyer, M. R., Gay, N. J., & Walker, J. E. (1989) *Biochem. J. 260*, 249–258.

Earley, F. G. P., & Ragan, C. I. (1980) *Biochem. J. 191*, 429–436.

Earley, F. G. P., & Ragan, C. I. (1981) *FEBS Lett. 127*, 45–47.

Farrell, P. J., Deininger, P. L., Bankier, A., & Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1565–1569.

Fearnley, I. M., Runswick, M. J., & Walker, J. E. (1989) *EMBO J. 8*, 665–672.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

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

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

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

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

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

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

Houldsworth, J., & Attardi, G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 377–381.

Jacobs, H. T., Elliott, D. J., Math, V. B., & 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., & Ohyama, K. (1988) *J. Mol. Biol. 203*, 353–372.

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

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

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

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

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

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

Mayes, S. R., Cook, K. M., & Barber, J. (1990) *FEBS Lett. 262*, 49–54.

Neckelmann, N., Li, K., Wade, R. P., Shuster, R., & Wallace, D. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7580–7584.

Nicholson, D. W., & Neupert, W. (1988) in *Protein transfer and organelle biogenesis*, pp 677–746, Academic Press, New York.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., & Ozeki, H. (1986) *Nature (London) 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., & Ozeki, H. (1988) *J. Mol. Biol. 203*, 281–298.

Patel, S. D., & Ragan, C. I. (1988) *Biochem. J. 256*, 521–528.

Patel, S. D., Cleeter, M. W. J., & Ragan, C. I. (1988) *Biochem. J. 256*, 529–535.

Pilkington, S. J., & Walker, J. E. (1989) *Biochemistry 28*, 3257–3264.

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.

Roe, B. A., Ma, D.-P., Wilson, R. K., & Wong, J. F. H. (1985) *J. Biol. Chem. 260*, 9759–9774.

Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H., & Schatz, G. (1986) *EMBO J. 5*, 1327–1334.

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. (1990) *Biochem. J. 266*, 421–426.

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.

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., Tohodoh, N., Shimada, H., & Sugirua, M. (1986) *EMBO J. 5*, 2043–2049.

Simpson, L., Neckelman, N., de la Cruz, V. F., Simpson, A. M., Feagin, J. E., Jasmer, D. P., & Stuart, K. (1987) *J. Biol. Chem. 262*, 6182–6196.

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

Videira, A., Tropschug, M., & Werner, S. (1990) *Biochem. Biophys. Res. Commun. 171*, 1168–1174.

Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.
Viñas, O., Powell, S. J., Runswick, M. J., Iacobazzi, V., & Walker, J. E. (1990) *Biochem. J. 265*, 321–326.
von Heijne, G. (1986) *EMBO J. 5*, 1335–1342.
Walker, J. E., Gay, N. J., Powell, S. J., Kostina, M., & Dyer, M. R. (1987a) *Biochemistry 26*, 8613–8619.
Walker, J. E., Cozens, A. L., Dyer, M. R., Fearnley, I. M., Powell, S. J., & Runswick, M. J. (1987b) *Chem. Scr. 27B*, 97–105.
Walker, J. E., Powell, S. J., Viñas, O., & Runswick, M. J. (1989) *Biochemistry 28*, 4702–4708.
Wolstenholme, D. R., McFarlane, J. R., Okimoto, R., Clary, D. O., & Wahleithner, J. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1324–1328.

Oligomerization and Intracellular Protein Transport: Dimerization of Intestinal Dipeptidylpeptidase IV Occurs in the Golgi Apparatus†

Thomas Jascur, Karl Matter,‡ and Hans-Peter Hauri*
Department of Pharmacology, Biocenter of the University of Basel, CH-4056 Basel, Switzerland
Received July 16, 1990; Revised Manuscript Received October 18, 1990

ABSTRACT: It was postulated that newly synthesized membrane proteins need to be assembled into oligomers in the endoplasmic reticulum in order to be transported to the Golgi apparatus. By use of the differentiated human adenocarcinoma cell line Caco-2, the general validity of this proposal was studied for small intestinal brush border enzymes which are dimers in most mammalian species. Chemical cross-linking experiments and sucrose gradient rate–zonal centrifugation revealed that dipeptidylpeptidase IV is present as a dimer in the brush border membrane of Caco-2 cells whereas the disaccharidase sucrase–isomaltase appears to be a monomer. Dipeptidylpeptidase IV was found to dimerize immediately after complex glycosylation, an event associated with the Golgi apparatus. Dimerization of this enzyme was inhibited by CCCP but did not depend on complex glycosylation of N-linked carbohydrates as assessed by the use of the trimming inhibitor 1-deoxymannojirimycin. It is concluded that dimerization of dipeptidylpeptidase IV occurs in a late Golgi compartment and therefore cannot be a prerequisite for its export from the endoplasmic reticulum.

The quaternary structure of integral plasma membrane proteins has become of central interest in the study of protein transport to the cell surface. From work with viral envelope proteins, the concept has emerged that assembly of these proteins into oligomers takes place in the endoplasmic reticulum (ER) and may be a prerequisite for their export from the ER and subsequent transport to the Golgi apparatus (Copeland et al., 1986, 1988; Gething et al., 1986; Kreis & Lodish, 1986; Doms et al., 1987, 1988; Boulay et al., 1988). Similar results have been obtained for endogenous membrane proteins of complex heterooligomeric structure [for a review, see Hurtley and Helenius (1989)]. However, one study is at variance with these general conclusions (Yewdell et al., 1988).

The biosynthesis of small intestinal brush border enzymes has been the subject of numerous investigations over the past years (Kenny & Maroux, 1982; Semenza, 1986; Hauri, 1988). These proteins are major constituents of the intestinal microvillus membrane and are involved in the terminal digestion of dietary carbohydrates and peptides. Some of them, including dipeptidylpeptidase IV (DPPIV)[1] and sucrase–isomaltase (SIM), are expressed in the differentiated human intestinal epithelial cell line Caco-2 (Pinto et al., 1983; Hauri et al., 1985). The enzymes are cotranslationally N-glycosylated to give the high-mannose precursors of 110 and 114 kDa for DPPIV, and 210 kDa for SIM. In the Golgi apparatus, they are processed to the complex glycosylated forms of 124 and 217 kDa, respectively. The intracellular transport of newly synthesized DPPIV and SIM in Caco-2 cells has been extensively studied. DPPIV is transported to the brush border membrane considerably faster than SIM, two rate-limiting steps being export from the ER and transit through the Golgi (Stieger et al., 1988).

Most brush border enzymes of rabbit, rat, and pig are known to exist as noncovalently linked homodimers in the microvillus membrane (Kenny & Maroux, 1982), but the intracellular site of their assembly into dimers is unknown. Therefore, the Caco-2 cell line expressing a number of human brush border hydrolases should provide a valuable system to assess the role of oligomerization for the intracellular transport of these endogenous plasma membrane proteins.

In the present paper, we provide evidence that in Caco-2 cells DPPIV is assembled into dimers in a late Golgi compartment and that transport through the Golgi apparatus, but not complex glycosylation, is required for dimerization of DPPIV. SIM, however, was found to be a monomer throughout its transport from the ER to the brush border

† Supported by the Swiss National Science Foundation.
* To whom correspondence should be addressed at the Department of Pharmacology, Biocenter of the University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.
‡ Present address: Department of Cell Biology, School of Medicine, Yale University, New Haven, CT 06510.

[1] Abbreviations: ApN, aminopeptidase N; DPPIV, dipeptidylpeptidase IV; SIM, sucrase–isomaltase; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; DMN, 1-deoxymannojirimycin; CCCP, carbonyl cyanide *m*-chlorophenylhydrazone; DSS, suberic acid bis(*N*-hydroxysuccinimide ester); DSP, 3,3′-dithiobis(propionic acid *N*-hydroxysuccinimide ester); EGS, ethylene glycol bis(succinimidyl succininate); MBS, *m*-maleimidobenzoic acid *N*-hydroxysuccinimide ester; DMS, dimethyl suberimidate.

0006-2960/91/0430-1908$02.50/0