Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1604 (2003) 180–189
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Cytochrome *f* and subunit IV, two essential components of the photosynthetic *bf* complex typically encoded in the chloroplast genome, are nucleus-encoded in *Euglena gracilis*

José Luis Santillán Torres[a], Ariane Atteia[a,1], M. Gonzalo Claros[b], Diego González-Halphen[a,*]

[a] *Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, 04510 Mexico City, Mexico*
[b] *Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, E-29071 Málaga, Spain*

Received 20 February 2003; received in revised form 2 May 2003; accepted 13 May 2003

Abstract

The photosynthetic protist *Euglena gracilis* contains chloroplasts surrounded by three membranes which arise from secondary endosymbiosis. The genes *petA* and *petD*, encoding cytochrome *f* and subunit IV of the cytochrome *bf* complex, normally present in chloroplast genomes, are lacking from the chloroplast DNA (cpDNA) of *E. gracilis*. The *bf* complex of *E. gracilis* was isolated, and the identities of cytochrome *f* and subunit IV were established immunochemically, by heme-specific staining, and by Edman degradation. Based on N-terminal and conserved internal protein sequences, primers were designed and used for PCR gene amplification and cDNA sequencing. The complete sequence of the *petA* cDNA and the partial sequence of the *petD* cDNA from *E. gracilis* are described. Evidence is provided that in this protist, the *petA* and *petD* genes have migrated from the chloroplast to the nucleus. Both genes exhibit a typical nuclear codon usage, clearly distinct from the usage of chloroplast genes. The *petA* gene encodes an atypical cytochrome *f*, with a unique insertion of 62 residues not present in other *f*-type cytochromes. The *petA* gene also acquired a region that encodes a large tripartite chloroplast transit peptide (CTP), which is thought to allow the import of apocytochrome *f* through the three-membrane envelope of *E. gracilis* chloroplasts. This is the first description of *petA* and *petD* genes that are nucleus-localized.

Keywords: Chloroplast DNA; *Euglena gracilis*; Photosynthetic complex

1. Introduction

Plastoquinol-cytochrome *c*/plastocyanin oxidoreductase, b_6f complex, or *bf* complex (EC 1.10.99.1) is an intrinsic membrane complex of chloroplasts and cyanobacteria that catalyzes electron transfer between Photosystem II and Photosystem I coupled to proton translocation across the photosynthetic membranes [1]. The *bf* complex comprises four major subunits, three of which contain redox-prosthetic groups, cytochrome *f*, a Rieske-type iron–sulfur protein, and cytochrome b_6. The fourth major constituent, subunit IV, lacks a prosthetic group but is required for the full catalytic activity of the complex [2]. Up to four additional small subunits of unknown function with apparent molecular masses between 3.4 and 4.0 kDa may also be present in isolated *bf* complex preparations [3,4]. Numerous sequence comparisons have shown that cytochrome b_6 and subunit IV are homologous to the N- and C-terminal portions, respectively, of bacterial and mitochondrial cytochromes *b*. Thus, these subunits are believed to have arisen from the fragmentation of an ancestral cyanobacterial cytochrome *b* gene. Cytochrome *b* of the *bc* complex from the photosynthetic bacterium *Rhodobacter capsulatus* was reengineered as a split form equivalent to subunits b_6 and IV, and shown to support photosynthetic growth [5]. This clearly shows a functional equivalence between cytochrome *b* and subunits b_6 plus IV. A low-resolution structure of the *bf* complex determined by electron crystallography is now available [6],

Abbreviations: BN-PAGE, blue native polyacrylamide gel electrophoresis; cpDNA, chloroplast DNA; CTP, chloroplast transit peptide; ER, endoplasmic reticulum; ⟨H⟩, local hydrophobicity; *mesoH*, mesohydrophobicity; nt, nucleotides; TMBZ, 3,3′,5,5′ tetramethylbenzidine

* Corresponding author. Tel.: +52-55-5622-5620; fax: 52-55-5622-5611.
E-mail address: dhalphen@ifisiol.unam.mx (D. González-Halphen).
[1] Present address: Division of Biology and Medicine, Brown University, Providence, RI 02912, USA.

0005-2728/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0005-2728(03)00058-6

as well as the partial architecture determined by X-ray crystallographic studies of cytochrome *f* and the Rieske-type iron–sulfur protein [7–9].

In the vast majority of the chloroplast genomes that have been sequenced, the genes encoding the three major subunits of the *bf* complex are invariably found to be present: *petA* encoding cytochrome *f*, *petB* encoding cytochrome b_6, and *petD* encoding subunit IV. In contrast, the *petC* gene, encoding the Rieske-type iron–sulfur protein, is always nucleus-localized. Four additional genes, *petG* (or *petE*), *petL*, *petM* and *petN* encoding small molecular mass subunits in the range of 3.0–4.0 kDa may be present or absent from chloroplast DNA (cpDNA) [3,4].

The early divergent protist *Euglena gracilis* contains chloroplasts that are surrounded by three membranes [10], which are thought to have arisen from a secondary endosymbiotic event, in which an *Euglena* ancestor engulfed an eukaryotic green alga [11,12]. The completely sequenced 143172 bp chloroplast genome of *E. gracilis* is unusual in the sense that it lacks two genes that encode essential components of the *bf* complex, i.e., the *petA* and *petD* genes [13,14]. Since the *bf* complex is an essential participant of photophosphorylation, and cytochrome *f* and subunit IV are indispensable for its function, we thought it was of interest to explore if the *petA* and *petD* genes had migrated from the chloroplast to the nucleus in *E. gracilis*.

In this work, we show that cytochrome *f* and subunit IV are constituents of the *bf* complex from *E. gracilis*, and we present the complete cDNA sequence of *petA* and the partial sequence of *petD* from this photosynthetic protist. Both genes exhibit typical nuclear codon usage. To our knowledge, this is the first description of *petA* and *petD* genes that are found in the nuclear genome. In addition, the *petA* gene encodes a preprotein exhibiting a relatively large tripartite chloroplastic transit peptide (CTP). This CTP is thought to direct and import cytochrome *f* to and into the thylakoid membrane through a unique series of membrane fusions and translocation events.

2. Materials and methods

2.1. Strain and culture conditions

E. gracilis strain Z was grown in Huntner medium under continuous light [15]. The cells were harvested at the late exponential phase of growth.

2.2. Isolation of E. gracilis chloroplasts and cytochrome bf complex solubilization

All solutions and buffers used were supplemented with 0.2 mM PMSF, 1 mM benzamidine, and 5 mM aminocaproic acid. Chloroplasts were obtained as described by Manning et al. [16], except that disruption of cells was done by hand-shaking for 4 min in the presence of half a volume of glass beads (0.5 mm in diameter). The *bf* complex from *E. gracilis* was partially purified following the procedure of Hurt and Hauska [17]. Isolated chloroplast membranes from *E. gracilis* (1 mg/ml of chlorophyll) were solubilized in the presence of 0.5% sodium cholate and 60 mM octylglucoside, fractionated by ammonium sulfate precipitation, and isolated by sucrose gradient centrifugation.

2.3. Protein analysis

SDS-PAGE was performed as described by Schägger and von Jagow [18], using 1.2-mm-thick slab gels (12% acrylamide). Gels were fixed and stained as described in the same work. Apparent molecular masses were calculated based on prestained molecular weight markers (BenchMark Prestained Protein Ladder, Gibco BRL). Protein concentrations were determined according to Markwell et al. [19]. The isolation of polypeptides for N-terminal sequencing was done as previously described [20]. N-terminal sequencing was carried out by Dr. J. d'Alayer on an Applied Biosystems Sequencer at the Laboratoire de Microséquençage des Protéines, Institut Pasteur, Paris (France). Western blots were carried out as previously described [21]. Antiserum was raised against a synthetic peptide derived from *Chlamydomonas reinhardtii* subunit IV [22]. Secondary antibodies were bound to alkaline phosphatase.

2.4. Nucleic acid preparations

Total DNA and RNA from *E. gracilis* was obtained as described by Aldrich and Cullis [23], and Sokolovsky et al. [24], respectively. First-strand cDNA was obtained from total RNA using the SuperScript first-strand synthesis kit for RT-PCR (Gibco BRL). All other standard molecular biology techniques were as described [25].

2.5. Cloning and sequencing of cDNA of the gene petA from E. gracilis

Using cDNA obtained by RT-PCR as template, PCRs were performed using a set of eight specific oligodeoxynucleotides designed from the sequences obtained for the *petA* gene encoding cytochrome *f*: 5′-CGC(T) ATT GTG TGC(T) GCC AAC TGC(T) CA-3′ (F2, forward), 5′-AGC AAG GAC ACC CCG AAC ATT TTG-3′ (F3, forward), 5′-AAG AAG TAC GCT GAG ATG ACT G-3′ (F4, forward), 5′-CTC GAA C(T)TG CTT CTT CTT CAA A(C)AC-3′ (F5, reverse), 5′-A(G)TA AAT CTG GCC G(T)CG GCC G(T)CG A(G)TT-3′ (F6, reverse), and 5′-A(G)AC GGG ACC G(T)AT AAC AAG A(G)AT A(G)TT-3′ (F7, reverse), 5′-TTA CCT CTG AAG ATG TCG CCA TG-3′ (F8, forward), and 5′-ACA AGT CAT CAA TGC TCA CAA GG-3′ (F9, reverse). The primer 5′-CTG GAA TTC (T)24-3′ (T1, reverse) previously described [26] based on the poly-A sequence was used to amplify the 3′-end of the *petA* cDNA. In order to amplify the 5′-end of the *petA* cDNA, a primer based on the consensus

leader sequence of *Euglena* premessenger RNAs was used: 5′-CAC AAG CTT CTT TCT GAG TGT CTA TTT TTT TTC-3′ (F1, forward). These short leader sequences may be transferred to premessenger RNAs by transplicing, specifically to those premessenger RNAs that encode proteins of cellular compartments like the chloroplast [26]. For RT-PCRs, Taq DNA polymerase (Qiagen) was used. Samples were denatured by heating for 3 min at 94 °C, subjected to 35 cycles of 1 min denaturation at 94 °C, 1 min annealing at 58 °C, and 1 min extension at 72 °C, and a final cycle of 10 min at 72 °C. The RT-PCR products were cloned in a pGEM-T Easy vector (Promega) or a pMOSBlue vector (Amersham). Five amplicons encoded a fragment of the *petA* cDNA gene, and these clones were named F1–F6, F2–F6, F2–F7, F4–F5, F4–T1, in accordance with the pair of primers that were utilized to generate them. Sequencing was carried out by the Unidad de Biología Molecular (IFC-UNAM).

2.6. Cloning and sequencing of cDNA of the gene petD from E. gracilis

A set of five specific oligodeoxynucleotides was designed from the sequences obtained for subunit IV: 5′-CAC AAC(T) TAC(T) TAC GGC GAG(A) CCC(T) GC-3′ (S1, forward), 5′-CCC(T) CTG GAG(A) ATC CTG CCC(T) GAG TG-3′(S2, forward), 5′-G TAT TTC TTC CCC ACC TTC AAC-3′ (S3, forward), 5′-TC A(G)AT GAA A(G)GG A(C)AC GGT GAT G(C)AG-3′ (S4, reverse), and 5′-CTG GAA TTC (T)24-3′ (T1, reverse). Three RT-PCR products encoded a fragment of the *petD* cDNA gene and these clones were named S1–S4, S3–T1, and S2–T1, in accordance with the pair of primers that were used. Deoxyoligonucleotide synthesis, RT-PCRs, cloning, and sequencing were performed as described above for the *petA* cDNA.

2.7. Southern and Northern hybridization

Southern and Northern hybridization were as described [25]. Two probes were designed from the apocytochrome *f* cDNA sequence: one of 201 bp with the pair of deoxyoligonucleotides F8–F9 that encode the N-terminal extension domain of the cytochrome *f* pre-apoprotein, and a second of 165 bp with the pair of deoxyoligonucleotides F3–F6 that encodes the mature cytochrome *f* apoprotein. A probe of 131 bp encoding part of subunit IV was obtained with the pair of deoxyoligonucleotides S2–S5. The probes were labeled with [α-^{32}P] dCTP using the Random Primer labeling kit (Gibco BRL).

2.8. Database accession numbers

The nucleotide sequences discussed in this paper were deposited in the DDBJ/EMBL/GenBank nucleotide sequence database under the accession numbers AF443625 (*petA* cDNA sequence from *E. gracilis*), and AY216719 (*petD* cDNA sequence from *E. gracilis*).

2.9. Sequence analysis in silico

Protein sequence data were taken from GenBank [27] non-redundant protein sequence. Multiple protein sequence alignments were constructed with Clustal_X [28]. Local hydrophobicity (⟨H⟩) and mesohydrophobicity (*mesoH*) were calculated as in Funes et al. [29].

3. Results

3.1. Cytochrome f and subunit IV are constituents of the bf complex of E. gracilis

Chloroplasts were obtained from *E. gracilis* by disruption of cells with glass beads followed by differential centrifugation. The *bf* complex of *E. gracilis* was isolated by detergent-solubilization of chloroplasts, followed by ammonium sulfate fractionation and sucrose gradient centrifugation [17]. SDS-PAGE of the isolated *E. gracilis bf* complex revealed the presence of four main polypeptides and three minor protein components (Fig. 1). The *E. gracilis bf* complex exhibits a subunit composition similar to that of *bf* complexes from other chloroplast sources. To further identify the subunits of the *E. gracilis bf* complex, parallel gels were subjected to 3,3′,5,5′ tetramethylbenzidine (TMBZ) staining or transferred to nylon membranes for immunoblot analysis. TMBZ stained the two heme-containing bands, with apparent molecular masses of 30.9 (cytochrome *f*) and 16.6 kDa (cytochrome b_6) (Fig. 1). Heme b_H is covalently linked to cytochrome *b* in the *bc* complex of *Bacillus subtilis* [30], and

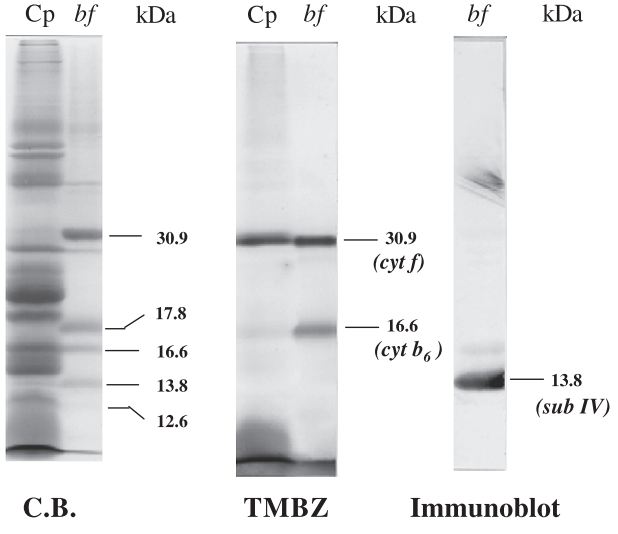

Fig. 1. Identification of the subunits of the isolated *bf* complex from *E. gracilis*. C.B.: Coomassie blue stained gel of *E. gracilis* total chloroplast proteins (Cp) and of the isolated *bf* complex (bf). TMBZ: heme-specific staining of *E. gracilis* total chloroplast proteins (Cp) and of the isolated *bf* complex (bf). Immunoblot: a gel with the *bf* complex from *E. gracilis* was transferred onto a nylon membrane (20 μg of protein) and immunostained with an antibody raised against subunit IV of the *C. reinhardtii bf* complex.

it has been proposed that heme is also covalently attached to the cytochrome b_6 subunit of *bf* complexes of chloroplasts and cyanobacteria. In addition, an immunochemical analysis was carried out with antibodies directed against subunit IV of the *C. reinhardtii bf* complex. This antibody recognized the 13.8 kDa band of the *E. gracilis bf* complex (Fig. 1), which was therefore identified as subunit IV. The main polypeptide bands of the isolated *bf* complex were also subjected to Edman degradation (Table 1), which revealed the presence of cytochrome *f*, cytochrome b_6, and subunit IV. *E. gracilis* subunit IV was susceptible to Edman degradation, contrary to what is found in other systems, as in the green alga *C. reinhardtii*, where subunit IV appears to be blocked [22]. The polypeptide band of 17.8 kDa was identified as a Photosystem II component, the so-called oxygen-evolving enhancer subunit (OEE2) [31]. It remains to be ascertained if this soluble polypeptide is a mere contaminant of the *E. gracilis bf* complex or if it is a legitimately associated component. Finally, the N-terminal sequence of the 12.6 kDa component was also obtained, but did not exhibit any significant similarity with proteins deposited in the databases. The Rieske-type iron–sulfur protein was not detected. This polypeptide may have a blocked N-terminal residue, may have dissociated from the preparation, or it may be represented by one of the less abundant bands that were not subjected to Edman degradation. Hence, we unambiguously identified three of the four main components of the *bf* complex from *E. gracilis*. The obtained N-terminal sequences allowed the design of primers for PCR amplification of the corresponding *petA* and *petD* cDNAs.

3.2. Characterization of the petA cDNA from E. gracilis

cDNA was synthesized from total *E. gracilis* RNA by RT-PCR. Then, based on the N-terminal sequence obtained for cytochrome *f* and on conserved internal sequences, deoxyoligonucleotides were designed and used for PCR amplification. PCRs gave a total of five overlapping amplicons. Each of them encoded a fragment of the *petA* cDNA,

Table 1
N-terminal sequences of the major polypeptide components of the isolated *E. gracilis bf* complex

Band (kDa)	N-terminal sequence	Subunit
30.9	(S/Y) P I F A Q Q A (X/Y) E N P R E A T G R I V C A N C H	cytochrome *f*
17.8	A Y G E G A N V F G K R K E T D Q F F E I S G D G W S G K L	Photosystem II component (OEE2)
16.6	S R V Y D W F E E R L E I Q A I A D D V S S K Y V P (P) H (V) N	cytochrome b_6
13.8	N V M K K P D L S D P K L R A (K L A K) G M G H N Y Y G E PAW P N D L L Y M F P V C I L G T F A (XA)	subunit IV
12.6	S G S G V R S A K K G G K A Q G G Q A G V G Y K G S T E P G	unidentified protein

Proteins were subjected to Edman degradation. Parentheses and X denote ambiguous identification of certain residues.

and exhibited similarity to chloroplast *petA* gene sequences in the databases. The complete sequence obtained of the *E. gracilis petA* cDNA contains an open reading frame of 1648 bp (accession number AF443625). The deduced amino acid sequence from this cDNA, starting from the first ATG codon, and ending at the stop codon UGA (position 1295–1297 bp), is a 496 amino acid cytochrome *f* pre-apoprotein. The *petA* cDNA has a 3′ UTR of 153 bp with a possible polyadenylation signal AATAAA at nucleotides 1574–1579. The only other *E. gracilis* nucleus-localized gene encoding a chloroplast protein that exhibits an identical polyadenylation site is the hydroxymethylbilane synthase gene (DDBJ/EMBL/GenBank accession number X15743). This putative polyadenylation site is localized 58 bp upstream of the poly(A) tail [32]. Other published *E. gracilis* nuclear gene sequences do not seem to carry a typical polyadenylation signal [33,34].

The pattern of codon utilization for the *petA* gene of *E. gracilis* was compared with the pattern of codon usage of known nuclear and chloroplast genes of this protist (Table 2). In other *E. gracilis* nucleus-localized genes, there is a significant bias in each codon family; this is because triplets that end in A are rarely used in the nuclear genome of this photosynthetic protist. In addition, the *Euglena* chloroplast genes also exhibit a biased codon usage, since they have less of a tendency to use those codons that end in C and G. The codon usage of the *petA* gene of *E. gracilis* is typically nuclear and clearly different from the one of chloroplast genes.

In plants and green algae, nucleus-encoded proteins that are transported through chloroplast membranes are synthesized in the cytoplasm as larger precursors containing a cleavable N-terminal sequence, the CTP [35]. Nevertheless, non-canonical transit peptides that are not processed have also been described [36]. Several *E. gracilis* nucleus-encoded chloroplast proteins are synthesized as precursors containing an N-terminal extension ranging from 134 to 141 residues [33,37]. The deduced cytochrome *f* pre-apoprotein of *E. gracilis* has a cleavable N-terminal sequence of 149 amino acids ($M_{[-149]}$ to $A_{[-1]}$) (Fig. 2), since the Edman degradation showed that the mature cytochrome *f* polypeptide starts with the sequence YPIFAQQA. The *E. gracilis* CTP of cytochrome *f* has three predicted domains (Fig. 2): the first one ($M_{[-149]}$ to $T_{[-107]}$) consists of a hydrophobic domain (residues $V_{[-142]}$ to $S_{[-122]}$) and ends at the recognition site for the signal peptidase between residues $T_{[-107]}$ and $Q_{[-106]}$. This domain localization and amino acid distribution suggest that it could act as a signal peptide that will direct the protein into the microsomal membranes. This is an extra domain not present in the CTPs of the preproteins from organisms that contain orthodox, double membrane chloroplasts [38]. The second domain ($Q_{[-106]}$ to $A_{[-46]}$) contains an amphiphilic transmembrane domain (residues $V_{[-66]}$ to $A_{[-46]}$) as well as the basic residues that are recognized by the chloroplast import machinery. It ends in a site for the stromal processing peptidase between

Table 2
Codon usage of the *petA* gene compared to nuclear and chloroplast genes of *E. gracilis*

	N	f	Cl		N	f	Cl		N	f	Cl		N	f	Cl
UUU	9.3	6.0	60.2	UCU	12.9	20.0	26.7	UAU	9.7	10.0	30.8	UGU	3.0	2.0	8.4
UUC	26.2	38.1	7.8	UCC	19.0	26.1	3.6	UAC	22.7	12.0	5.1	UGC	11.4	4.0	2.7
UUA	0.9	2.0	55.5	UCA	5.5	2.0	16.2	UAA	0.9	0.0	3.0	UGA	0.7	2.0	0.3
UUG	16.7	18.0	18.3	UCG	5.2	6.0	5.2	UAG	0.2	0.0	0.7	UGG	12.9	2.0	14.3
CUU	8.5	6.0	19.0	CCU	12.9	14.0	21.3	CAU	6.3	2.0	19.3	CGU	8.6	4.0	15.0
CUC	12.5	12.0	0.5	CCC	22.6	32.1	2.4	CAC	12.2	2.0	2.5	CGC	18.8	12.0	3.4
CUA	0.9	0.0	6.5	CCA	11.0	10.0	10.7	CAA	9.6	10.0	25.1	CGA	5.2	4.0	7.2
CUG	32.2	26.1	1.0	CCG	8.7	14.0	2.3	CAG	30.9	50.1	3.7	CGG	8.4	2.0	0.9
AUU	18.3	26.1	49.1	ACU	12.7	6.0	22.0	AAU	8.8	10.0	46.8	AGU	4.2	4.0	14.5
AUC	29.0	28.1	5.0	ACC	22.9	34.1	2.1	AAC	28.6	32.0	9.5	AGC	15.7	16.0	2.4
AUA	0.4	0.0	32.2	ACA	11.4	2.0	21.0	AAA	9.1	0.0	69.8	AGA	2.5	0.0	18.8
AUG	25.2	22.0	22.1	ACG	9.1	14.0	4.8	AAG	54.1	50.1	11.7	AGG	3.8	2.0	4.8
GUU	24.0	38.1	35.4	GCU	25.9	26.1	27.6	GAU	22.6	12.0	30.6	GGU	18.4	6.0	32.8
GUC	22.4	36.1	2.5	GCC	36.6	40.1	3.2	GAC	36.7	20.0	6.2	GGC	29.1	40.0	4.8
GUA	1.8	0.0	18.9	GCA	17.8	26.1	17.8	GAA	13.9	8.0	38.2	GGA	13.9	24.0	24.6
GUG	42.5	48.1	3.6	GCG	11.4	18.0	4.8	GAG	46.7	52.1	8.3	GGG	13.5	8.0	4.5

Values are shown in percentages for nuclear genes (N), the *petA* gene (f, gray boxes) and chloroplast (Cl) genes. Data to construct this table were obtained for *E. gracilis* from the Codon Usage Database (http://www.kazusa.or.jp/codon/) with accession numbers [gbpln]:42 for nuclear genes and [gbpln]:12 for chloroplast genes.

amino acids $A_{[-46]}$ and $A_{[-45]}$ and contains several hydroxylated and charged amino acids, similar to other tripartite CTPs of *E. gracilis* nucleus-encoded chloroplast precursor proteins [33,37,39–41]. The third domain ($A_{[-44]}$ to $A_{[-1]}$) contains a hydrophobic segment (residues $A_{[-17]}$ to $Y_{[+1]}$) and ends in a conserved cleavage site Ala-X-Ala (residues $A_{[-3]}$ to $A_{[-1]}$) for the thylakoid processing protease [39,42,43]. This third region operates as the thylakoid-targeting and translocation signal since it possesses the characteristic design of a charged region followed by a stretch of hydrophobic residues that ends with a cleavage site with small non-polar residues at $-3/-1$ [44]. In conclusion, the *Euglena* CTP is clearly different from plant CTPs and designed in accordance with a different import pathway. Indeed, algorithms like MitoProt [45,46] or ChloroP [47] do not predict *E. gracilis* CTPs correctly. In contrast, the algorithm SignalP [48] or TargetP [49] detect the first hydrophobic region as a domain that corresponds to a clear signal peptide.

The *petA* cDNA encodes a mature cytochrome *f* apoprotein of 347 amino acids that has all the common structural characteristics of the cytochrome *f* family: a heme binding domain $YX_{(19)}CANCH$ (residues Y_1 and from C_{21} to H_{25}); two highly acidic domains that contribute to the interaction with soluble plastocyanin (residues K_{120} to R_{127} and K_{249} to R_{251}); and the membrane anchor domain necessary for the functional assembly of the protein (residues V_{313} to V_{332}). A sequence comparison of the cytochrome *f* apoproteins from *E. gracilis*, *Chlamydomonas subcaudata*, *Chlorella vulgaris*, *Brassica rapa*, and *Phormidium laminosum* shows 52% or more identity between them, suggesting that the protein is highly conserved even in non-related species (Fig. 3). Furthermore, the alignment revealed that in the internal sequence of the apocytochrome *f* of *E. gracilis*, there is a

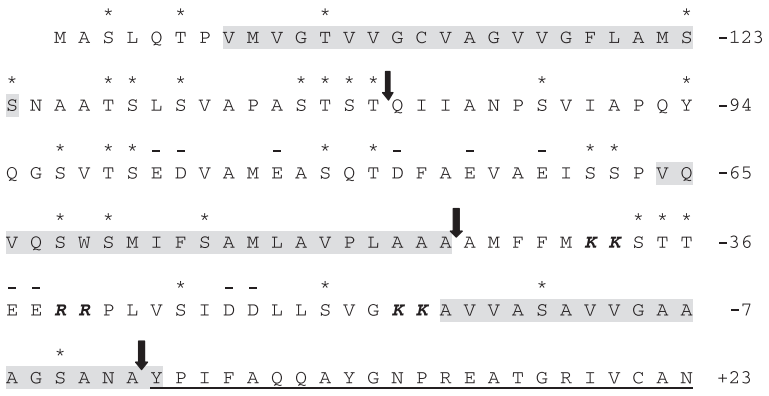

Fig. 2. Deduced amino acid sequence of the CTP of the apocytochrome *f* preprotein of *E. gracilis*. Black arrows indicate predicted peptidase cleavage sites. Gray-shadowed sequences represent the hydrophobic domains present in the CTP. Minus and asterisks show acid and hydroxylated amino acids residues, respectively. Basic residues are in bold and italic case. The sequence of the mature cytochrome *f* obtained by Edman degradation is underlined. The sequence is numbered using 1 as the first amino acid of the mature protein and negative numbers for the residues of the CTP.

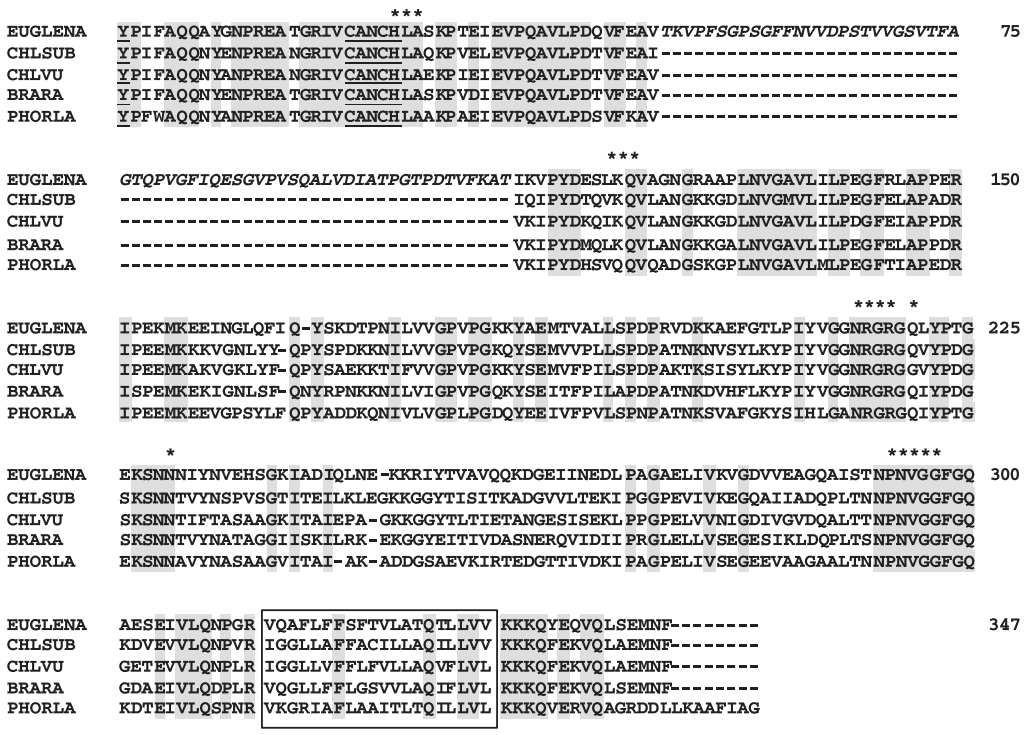

Fig. 3. Sequence comparison of different cytochromes *f*. Shown are the sequences of cytochrome *f* of *E. gracilis* (EUGLENA), *C. subcaudata* Q95AG0 (CHLSUB), *C. vulgaris* P56316 (CHLVU), *B. rapa* P36438 (BRARA), and *P. laminosum* P95522 (PHORLA). Amino acids involved in the H-bond of five-water chain are indicated by asterisks, and residues involved in heme-binding are underlined. Identical residues are shadowed, and the unique insertion of 62 residues in the apocytochrome *f* of *Euglena* is shown in italics. Residues that constitute the putative transmembrane stretch are boxed.

unique region of 62 residues (T_{49} to T_{110}) not found in other apocytochrome *f* sequences from a wide variety of plants, algal and cyanobacterial sources. Unexpectedly, this sequence exhibited 50% similarity to a protein related to molybdopterin-guanine dinucleotide biosynthesis in bacteria [50].

3.3. Characterization of the petD cDNA from E. gracilis

Based on the amino acid sequence obtained for subunit IV and on conserved internal regions of the protein, oligodeoxynucleotides were designed and used for PCR-amplification. Three overlapping amplicons were obtained using cDNA from *E. gracilis* as template (data not shown). The sequences of the amplified products encoded fragments of subunit IV, and exhibited similarity to chloroplast *petD* gene sequences in the databases. A partial sequence of 697 bp of the *petD* cDNA of *E. gracilis* was constructed (accession number AY216719). The *petD* cDNA does not exhibit a known polyadenylation signal. The pattern of codon utilization for the *petD* gene of *E. gracilis* was also found to be nuclear, and unambiguously distinct from the chloroplast codon usage (data not shown). The deduced amino acid sequence predicts a partial subunit IV preprotein of 173 residues that is homologous to subunits IV from other sources. Determination of the N-terminal sequence of the mature protein by Edman degradation allowed derivation of the complete 159-residue sequence of the mature *Euglena* subunit IV. The sequence of the putative CTP is clearly truncated, since only 15 residues could be obtained. In the subunit IV predicted sequence, there is an inconsistent residue with respect to the sequence of the mature protein obtained by Edman degradation (R_{23} instead of H_{23}). We assume that this difference may be due to errors either in RT-PCR amplification or in the automated amino acid sequence analysis, since there is no previous evidence suggesting RNA editing in *E. gracilis*. In addition, the partial sequence of the subunit IV preprotein has a predicted peptidase cleavage site in N_{15} [49,51]. This result is supported by the N-terminal sequence data obtained with the mature protein. The sequence of the entire CTP of the subunit IV precursor remains to be established.

3.4. DNA–DNA and DNA–RNA hybridization experiments

Southern blot analysis was carried out to ascertain that the *petA* and *petD* genes were present in the nuclear genome. Total *E. gracilis* DNA was restricted with different enzymes and hybridized at high and low stringency and temperatures ranging from 55 to 65 °C with different *petA*- and *petD*-derived probes. In all the assayed conditions, no single-band hybridization signal could be obtained. Instead, several bands that produce a smear along the gel were observed (data not shown). This methodological problem has been previously observed in hybridization experiments carried out with restricted mitochondrial DNA from *E.*

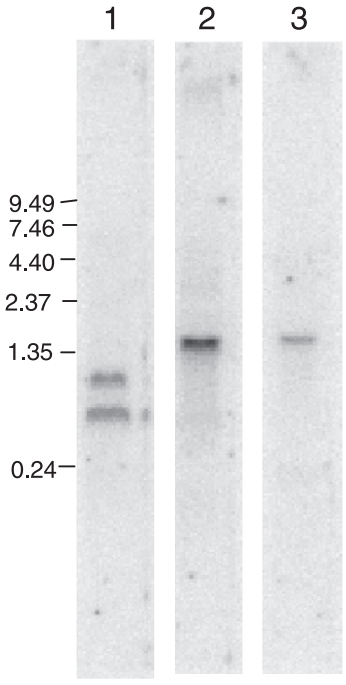

Fig. 4. The *petA* and *petD* genes are expressed in *E. gracilis*. Fifteen micrograms of *E. gracilis* total RNA were fractionated on denaturing 1% agarose gel containing 0.66% of formaldehyde, transferred to a nylon membrane, and decorated with the following probes (see Materials and methods): a 131 bp *petD* cDNA fragment (S2–S5) (lane 1), a 201 bp *petA* cDNA fragment (F8–F9) (lane 2), and a 165 bp *petA* cDNA fragment (F3–F6) (lane 3). Size markers are indicated on the left of the figure (values in kb).

gracilis [52]. In contrast, DNA–RNA hybridization experiments carried out with total RNA preparations exhibited discrete bands (Fig. 4). Using as probes either a 201 bp fragment of the *petA* cDNA encoding the N-terminal extension of cytochrome *f*, or a 165 bp fragment encoding an internal region of the mature cytochrome *f* apoprotein, a single 1.7 kb mRNA band was observed. This result is consistent with the size of 1648 bp obtained for the *petA* cDNA of *E. gracilis*. In the case of subunit IV, RNA–DNA hybridization analysis shows two mRNAs, one of 0.6 kb and the second of 1.0 kb (for an expected minimum size of the *petD* mRNA of 0.8 kb). The presence of two bands may be due to the presence of isoforms or of different alternative-splicing products. We conclude that both nucleus-localized genes, *petA* and *petD*, are functionally expressed, since the corresponding mRNAs can be detected and the corresponding mature polypeptides are present in the *E. gracilis bf* complex.

4. Discussion

4.1. Nucleus-encoded petA and petD genes

To date, the predicted primary structure of several nucleus-encoded *E. gracilis* chloroplast proteins have been analyzed. Here, we report the complete cDNA of the nucleus-encoded cytochrome *f* gene from this euglenoid (the *petA* gene) and the partial sequence of the cDNA of the nucleus-localized gene encoding subunit IV (the *petD* gene). *E. gracilis petA* gene exhibits several characteristics of a chloroplast gene that migrated to the nucleus: a typical nuclear codon usage, the presence of a putative polyadenylation signal, and the presence of DNA sequences encoding CTPs. The *petD* gene seems to share some of these characteristics, although only the partial sequence of its CTP could be obtained in this work and no putative polyadenylation signal could be identified. To our knowledge, this is the first report of *petA* and *petD* genes in the nuclear genome. The nucleus-localized *E. gracilis petA* and *petD* genes are both expressed, as demonstrated by Northern blot hybridization. In addition, the corresponding subunits cytochrome *f* and subunit IV are present in the mature, isolated *bf* complex from *E. gracilis*. Therefore, the migration of *petA* and *petD* genes from the chloroplast to the nucleus resulted in the functional relocalization of these genes, which eventually led to the inactivation and subsequent obliteration of the original chloroplast *petA* and *petD* genes. The causes that may intensify the migration of genes from the chloroplast to the nucleus could be the mutational pressure on the chloroplast genome due to oxidative cpDNA damage, lack of recombination, or the accumulation of deleterious mutations [53]. *Astasia longa*, a colorless and heterotrophic flagellate protist, is closely related to *E. gracilis* [54]. *A. longa* is thought to have lost photosynthesis through a secondary evolutionary process [55]. Its circular highly reduced 73 kb plastid DNA lacks the chloroplast genes encoding photosynthesis-related proteins [56]. Therefore, all the *pet* genes are also absent from this reduced plastid genome. It would be of interest to explore the presence of *petA* and *petD* genes (or pseudogenes) in the nuclear genome of *A. longa*, and to be able to infer if the migration of these genes occurred before or after the divergence of *A. longa* from *E. gracilis*.

The deduced primary structure of cytochrome *f* is similar to the one from other species. In fact, the amino acids involved in the binding of the heme redox group are identical [7]. Moreover, the sequence shows the presence of residues involved in the H-bond of a five-water chain that are conserved in sequences of plant, algal, and cyanobacterial *f*-type cytochromes [9,57]. This water chain is thought to act as a structural frame [9] or as a proton wire as an exit port for protons pumped from the cytochrome *bf* complex [57].

Amino acid sequence comparison established the presence of 62 unique residues in *E. gracilis* cytochrome *f*, which are absent in the cytochrome *f* sequences from plant and algal sources. This extra region is rich in valines, prolines, and phenylalanines. When the *E. gracilis* cytochrome *f* primary structure was modeled upon the three-dimensional structure of *C. reinhardtii* cytochrome *f* (data not shown), it was observed that this 62-residue additional loop is located far away from the binding site of plastocyanin. Therefore, it is expected that this extra loop will not

interfere with the activity of the *E. gracilis bf* complex. The presence of this unique loop could also be responsible of a diminished association with the Rieske-type iron–sulfur protein, and could account for the apparent loss of this subunit during the isolation procedure. It is also tempting to speculate that the 62 amino acid insertion in the *Euglena* cytochrome *f* may be responsible for the binding of Photosystem II component OEE2, that co-migrated with the *bf* complex of this photosynthetic protist.

Cytochrome *f* is a protein whose rate of synthesis is regulated by the availability of the other subunits of the *bf* complex [58]. One of the main signals for this translational regulation is localized in the C-terminal part of cytochrome *f*, at the end of the transmembrane stretch, in the highly conserved region QVLLVLKKKQFEKV (residues Q_{297} to V_{310} of *C. reinhardtii* cytochrome *f*). The three residues Q_{297}, K_{303}, and F_{307}, are crucial for the integrity of the regulatory motif [59]. This conserved sequence corresponds to the region QTLLVVKKKQYEQV (residues Q_{327} to V_{340}) in *E. gracilis* cytochrome *f*. One of the critical residues, the equivalent of *Chlamydomonas* F_{307}, is replaced by a tyrosine (Y_{307}) in *Euglena.* It would be of interest to explore in more detail if this conserved substitution does not alter the assembly-mediated regulation of the *bf* complex, or if this regulation is achieved through different mechanisms in *E. gracilis*.

4.2. Analysis of the CTPs encoded by petA and petD genes

The N-terminal analysis of the corresponding mature proteins cytochrome *f* and subunit IV obtained by Edman degradation established the presence of a CTP in the cytochrome *f* pre-apoprotein. The size of this CTP is relatively large when compared to CTPs from other organisms, which exhibit a mean length of 140 amino acids [39,60,61]. This complex CTP may have been acquired through exon shuffling [61]. The CTP of the *E. gracilis* cytochrome *f* precursor contains all the structural and targeting features previously identified for the tripartite presequences of chloroplast precursors of *E. gracilis* [40]: a signal peptide, a stromal targeting region, and a thylakoid targeting sequence (Fig. 3). The last hydrophobic segment ($A_{[-17]}$ to $Y_{[+1]}$) is believed to function as a stop-transfer membrane anchor sequence during the membrane-fusion processes that lead to protein import in *E. gracilis* chloroplasts [40].

Why have some genes that are typically localized in the chloroplast genome been successfully transferred to the nucleus in *E. gracilis*? Two features may account for this phenomenon. Firstly, it has been previously argued that highly hydrophobic proteins, containing four or more helices that span the membrane, cannot be imported into organelles. Therefore, these polytopic membrane proteins must be synthesized in situ in order to be properly inserted and assembled into their organellar compartment [62–64]. These studies suggested that the highest average hydrophobicity over 60 to 80 amino acids of a polypeptide chain (*mesoH*), along with the maximum hydrophobicity of the putative transmembrane segments, are useful indicators of the likelihood that a protein could be imported into organelles. We carried out similar calculations with the sequence of *E. gracilis* cytochrome *f*, and compared its highest average hydrophobicity to that of orthodox, chloroplast-encoded cytochrome *f* sequences (Fig. 5). The *E. gracilis* cytochrome *f* exhibited a notable reduced highest average hydrophobicity, suggesting that this parameter is important in facilitating the migration of genes encoding membrane-bound proteins from organelles to the nucleus. The *Euglena* subunit IV protein behaved in the same way (results not shown). Secondly, *Euglena* chloroplasts are surrounded by three membranes, in contrast to the orthodox two-membrane containing chloroplasts found in green algae, red algae, and plants. This third membrane in *Euglena*, the so-called perichloroplast membrane, seems to be closely related to the endoplasmic reticulum (ER). In addition, the euglenoids and chlorophytes appear as closely related species in a phylogenetic analysis based on the inferred amino acid sequences from the genes *psbA* (Photosystem II reaction center protein D1), *rbcL*, and *rbcS* (the large and small

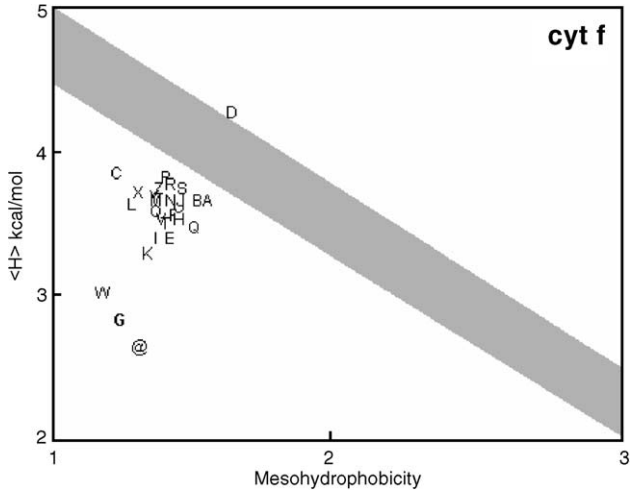

Fig. 5. Plot of mesohydrophobicity versus maximal local hydrophobicity for cytochrome *f* proteins. The MitoProtII program was used to calculate local hydrophobicity values (⟨H⟩) and mesohydrophobicity values for cytochrome *f* proteins using the PRIFT scale. Proteins are distributed on the abscissa according to their maximum mesohydrophobicity value and on the ordinate according to the hydrophobicity of the most hydrophobic 17-residue segment. The hypothetical boundary between importable and nonimportable proteins, indicated by a broad, gray diagonal, was derived from Claros et al. [63] and Pérez-Martínez et al. [64]. The following proteins with their symbols were analyzed in the left panel for cytochrome *f*: *B. rapa* P36438 (B), *Arabidopsis thaliana* P56771 (A), *C. reinhardtii* P23577 (C), *C. vulgaris* P56316 (D), *Cyanidium caldarium* Q9TLS4 (E), *Cyanophora paradoxa* P48123 (F), *E. gracilis* AF443625 (G), *Guillardia theta* O78494 (H), maize P46617 (I), *Marchantia polymorpha* P06246 (J), *Mesostigma viride* Q9MUN6 (K), *Nostoc* sp. P13626 (L), *Odontella sinensis* P49476 (M), *Oenothera hookeri* P04658 (N), rice P07888 (O), pea P00155 (P), *Phormidium laminosum* P95522 (Q), *Picea abies* O47042 (R), *Pinus thunbergii* P41619 (S), *Porphyra purpurea* P51265 (T), *Glycine max* P49161 (U), *Spinacia oleracea* P16013 (V), *Synechococcus* sp. P26293 (W), *Synechocystis* sp. P26287 (X), *Nicotiana tabacum* P06449 (Y), *Vicia faba* P06669 (Z), and wheat P05151 (@).

subunits of the rubisco holoenzyme) that participate in the photosynthetic process [12]. Therefore, it appears that a euglenoid ancestor engulfed a chlorophyte alga, and subsequently enslaved its chloroplasts, inside the boundaries of an ER-related membrane. Nowadays, *Euglena* chloroplast precursors are synthesized on membrane-bound polysomes, co-translationally inserted into the ER, and transported by vesicles as integral membrane proteins from the ER to the Golgi apparatus and from there to the chloroplast. A detailed model for the transport of precursor proteins through the *Euglena* three-chloroplast membranes has been previously proposed [40,61]. We hypothesize that this transport model may facilitate the import of membrane protein products into chloroplasts. This machinery requires atypical CTPs that cannot be easily analyzed nor clearly predicted with current algorithms.

Acknowledgements

The authors wish to thank Dr. J. d'Alayer (Institut Pasteur) for his expert help in sequencing peptides, Dr. D.W. Krogmann (Purdue University) for his comments to the manuscript, Dr. J.-L. Popot (I.B.P.C., Paris) for the gift of the anti-subunit IV antiserum, and Dr. Yves Choquet (I.B.P.C., Paris) for pointing out the differences in the regulatory motif of cytochrome *f*. This work was supported by grants TW01176-01 from NIH/Fogarty, USA; 27754N from CONACyT, Mexico; and IN202598 from DGAPA, UNAM, Mexico.

References

[1] W.A. Cramer, G.M. Soriano, M. Ponamarev, D. Huang, H. Zhang, S.E. Martínez, J.L. Smith, Some new structural aspects and old controversies concerning the cytochrome b_6f complex of oxygenic photosynthesis, Annu. Rev. Plant Physiol. Plant Mol. Biol. 47 (1996) 477–508.

[2] L.-B. Li, Y.-P. Zou, L. Yu, C.-A. Yu, The catalytic role of subunit IV of the cytochrome b_6f complex from spinach chloroplast, Biochim. Biophys. Acta 1057 (1991) 215–222.

[3] G.M. Soriano, M.V. Ponamarev, C.J. Carrell, D. Xia, J.L. Smith, W.A. Cramer, Comparison of the cytochrome bc_1 complex with the anticipated structure of the cytochrome b_6f complex: Le plus ca change le plus c'est la meme chose, J. Bioenerg. Biomembranes 31 (1999) 201–213.

[4] C. Breyton, The cytochrome b_6f complex: structural studies and comparison with the bc_1 complex, Biochim. Biophys. Acta 1459 (2000) 467–474.

[5] A.S. Saribas, S. Mandaci, F. Daldal, An engineered cytochrome b_6c_1 complex with a split cytochrome *b* is able to support photosynthetic growth of *Rhodobacter capsulatus*, J. Bacteriol. 181 (1999) 5365–5372.

[6] P. Bron, J.J. Lacapere, C. Breyton, G. Mosser, The 9 Å projection structure of cytochrome b_6f complex determined by electron crystallography, J. Mol. Biol. 287 (1999) 117–126.

[7] S.E. Martinez, D. Huang, A. Szczepaniak, W.A. Cramer, J.L. Smith, Crystal structure of chloroplast cytochrome *f* reveals a novel cytochrome fold and unexpected heme ligation, Structure 2 (1994) 95–105.

[8] H. Zhang, C.J. Carrell, D. Huang, V. Sled, T. Ohnishi, J.L. Smith, W.A. Cramer, Characterization and crystallization of the lumen side domain of the chloroplast Rieske iron–sulfur protein, J. Biol. Chem. 271 (1996) 31360–31366.

[9] Y.I. Chi, L.S. Huang, Z. Zhang, J.G. Fernández-Velasco, E.A. Berry, X-ray structure of a truncated form of cytochrome *f* from *Chlamydomonas reinhardtii*, Biochemistry 39 (2000) 7689–7701.

[10] M. Lefort-Tran, M. Pouphile, G. Freyssinet, B. Pineau, Structural and functional significance of the chloroplast envelope of *Euglena*: immunocytological and freeze fracture study, J. Ultrastruct. Res. 73 (1980) 44–63.

[11] S.P. Gibbs, The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae, Ann. N.Y. Acad. Sci. 361 (1981) 193–208.

[12] C.W. Morden, C.F. Delwiche, M. Kuhsel, J.D. Palmer, Gene phylogenies and the endosymbiotic origin of plastids, Biosystems 28 (1992) 75–90.

[13] B. Schlunegger, E. Stutz, The *Euglena gracilis* chloroplast genome: structural features of a DNA region possibly carrying the single origin of DNA replication, Curr. Genet. 8 (1984) 629–634.

[14] R.B. Hallick, L. Hong, R.G. Drager, M.R. Favreau, A. Monfort, B. Orsat, A. Spielmann, E. Stutz, Complete sequence of *Euglena gracilis* chloroplast DNA, Nucleic Acids Res. 21 (1993) 3537–3544.

[15] R. Moreno-Sánchez, J.C. Raya, Preparation of coupled mitochondria from *Euglena* by sonication, Plant Sci. 48 (1987) 151–157.

[16] J.E. Manning, D.R. Wolstenholme, R.S. Ryan, J.A. Hunter, O.C. Richards, Circular chloroplast DNA from *Euglena gracilis*, Proc. Natl. Acad. Sci. U. S. A. 68 (1971) 1169–1173.

[17] E. Hurt, G. Hauska, A cytochrome f/b_6 complex of five polypeptides with plastoquinolplastocyanin-oxidoreductase activity from spinach chloroplasts, Eur. J. Biochem. 117 (1981) 591–595.

[18] H. Schägger, G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Anal. Biochem. 166 (1987) 368–379.

[19] M.A.K. Markwell, S.M. Haas, L.L. Bieber, N.E. Tolbert, A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples, Anal. Biochem. 87 (1978) 206–210.

[20] A. Atteia, G. Dreyfus, D. González-Halphen, Cha racterization of the a and β subunits of the F_0F_1-ATPase from the alga *Polytomella* spp., a close relative of *Chlamydomonas reinhardtii*, Biochim. Biophys. Acta 1320 (1997) 275–284.

[21] A. Reyes-Prieto, M. El-Hafidi, R. Moreno-Sánchez, D. González-Halphen, Characterization of oxidative phosphorylation in the colorless chlorophyte *Polytomella* sp. Its mitochondrial respiratory chain lacks a plant-like alternative oxidase, Biochim. Biophys. Acta 1554 (2002) 170–179.

[22] Y. Pierre, C. Breyton, D. Kramer, J.L. Popot, Purification and characterization of the cytochrome b_6f complex from *Chlamydomonas reinhardtii*, J. Biol. Chem. 270 (1995) 29342–29349.

[23] J. Aldrich, A. Cullis, RAPD analysis in Flax: optimization of yield and reproducibility using KlenTaq1 DNA Polymerase, Chelex 100, and gel purification of genomic DNA, Plant Mol. Biol. Rep. 11 (1993) 128–141.

[24] V. Sokolovsky, R. Kaldenhoff, M. Ricci, V.E.A. Russo, Fast and reliable mini-prep RNA extraction from *Neurospora crassa*, Fungal Genet. Newslett. 37 (1990) 41–43.

[25] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

[26] L.H. Tessier, M. Keller, R.L. Chan, R. Fournier, J.H. Weil, P. Imbault, Short leader sequences may be transferred from small RNAs to premature mRNAs by trans-splicing in *Euglena*, EMBO J. 10 (1991) 2621–2625.

[27] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, B.A. Rapp, D.L. Wheeler, GenBank, Nucleic Acids Res. 28 (2000) 15–18.

[28] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G.

Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[29] S. Funes, E. Davidson, M.G. Claros, R. van Lis, X. Pérez-Martínez, M. Vázquez-Acevedo, M.P. King, D. González-Halphen, The typically mitochondrial DNA-encoded ATP6 subunit of the F_1F_0-ATPase is encoded by a nuclear gene in *Chlamydomonas reinhardtii*, J. Biol. Chem. 277 (2002) 6051–6058.
[30] J. Yu, N.E. Le Brun, Studies of the cytochrome subunits of menaquinone:cytochrome *c* reductase (*bc* complex) of *Bacillus subtilis*. Evidence for the covalent attachment of heme to the cytochrome *b* subunit, J. Biol. Chem. 273 (1998) 8860–8866.
[31] S.P. Mayfield, M. Rahire, G. Frank, H. Zuber, J.D. Rochaix, Expression of the nuclear gene encoding oxygen-evolving enhancer protein 2 is required for high levels of photosynthetic oxygen evolution in *Chlamydomonas reinhardtii*, Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 749–753.
[32] A.L. Sharif, A.G. Smith, C. Abell, Isolation and characterisation of a cDNA clone for a chlorophyll synthesis enzyme from *Euglena gracilis*. The chloroplast enzyme hydroxymethylbilane synthase (porphobilinogen deaminase) is synthesized with a very long transit peptide in *Euglena*, Eur. J. Biochem. 184 (1989) 353–359.
[33] R.L. Chan, M. Keller, J. Canaday, J.H. Weil, P. Imbault, Eight small subunits of *Euglena* ribulose 1–5 bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein, EMBO J. 9 (1990) 333–338.
[34] G. Houlne, R. Schantz, Molecular analysis of the transcripts encoding the lightharvesting chlorophyll *a*/*b* protein in *Euglena gracilis*: unusual size of the mRNA, Curr. Genet. 12 (1987) 611–616.
[35] M. Küchler, J. Soll, From nuclear genes to chloroplast localized proteins, Plant Sci. 161 (2001) 379–389.
[36] S. Miras, D. Salvi, M. Ferro, D. Grunwald, J. Garin, J. Joyard, N. Rolland, Noncanonical transit peptide for import into the chloroplast, J. Biol. Chem. 277 (2002) 47770–47778.
[37] R. Kishore, U.S. Muchhal, S.D. Schwartzbach, The presequence of *Euglena* LHCPII, a cytoplasmically synthesized chloroplast protein, contains a functional endoplasmic reticulum-targeting domain, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 11845–11849.
[38] C. Robinson, R.B. Klösgen, Targeting of proteins into and across the thylakoid membrane: a multitude mechanisms, Plant Mol. Biol. 26 (1994) 15–24.
[39] J. Inagaki, Y. Fujita, T. Hase, Y. Yamamoto, Protein translocation within chloroplast is similar in *Euglena* and higher plants, Biochem. Biophys. Res. Commun. 277 (2000) 436–442.
[40] C. Sulli, Z. Fang, U. Muchhal, S.D. Schwartzbach, Topology of *Euglena* protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles, J. Biol. Chem. 274 (1999) 457–463.
[41] K. Henze, A. Badr, M. Wettern, R. Cerff, W. Martin, A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 9122–9126.
[42] S. Smeekens, C. Bauerle, J. Hageman, K. Keegstra, P. Weisbeek, The role of the transit peptide in the routing of precursors toward different chloroplast compartments, Cell 46 (1986) 365–375.
[43] G. von Heijne, J. Steppuhn, R.G. Hermann, Domain structure of mitochondrial and chloroplast targeting peptides, Eur. J. Biochem. 180 (1989) 535–545.
[44] S. Clausmeyer, R.B. Klosgen, R.G. Herrmann, Protein import into chloroplasts—the hydrophilic lumenal proteins exhibit unexpected import and sorting specificities in spite of structurally conserved transit peptides, J. Biol. Chem. 268 (1993) 13869–13876.
[45] M.G. Claros, MitoProt, a Macintosh application for studying mitochondrial proteins, Comput. Appl. Biosci. 11 (1995) 441–447.
[46] M.G. Claros, P. Vincens, Computational method to predict mitochondrially imported proteins and their targeting sequences, Eur. J. Biochem. 241 (1996) 778–779.
[47] O. Emanuelsson, H. Nielsen, G. von Heijne, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites, Protein Sci. 8 (1999) 978–984.
[48] H. Nielsen, A. Krogh, Prediction of signal peptides and signal anchors by a hidden Markov model, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB 6), AAAI Press, Menlo Park, CA, 1998, pp. 122–130.
[49] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, J. Mol. Biol. 300 (2000) 1005–1016.
[50] J.F. Heidelberg, J.A. Eisen, W.C. Nelson, R.A. Clayton, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, L.A. Umayam, S.R. Gill, K.E. Nelson, T.D. Read, H. Tettelin, D. Richardson, M.D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R.D. Fleischmann, W.C. Nierman, O. White, S.L. Salzberg, H.O. Smith, R.R. Colwell, J.J. Mekalanos, J.C. Venter, C.M. Fraser, DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*, Nature 406 (2000) 477–483.
[51] H. Nielsen, J. Engelbrech, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1–6.
[52] L.H. Tessier, H. van der Speck, J.M. Gualberto, J.M. Grienenberger, The *cox1* gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications, Curr. Genet. 31 (1997) 208–213.
[53] W. Martin, W.R.G. Hermann, Gene transfer from organelles to the nucleus: how much, what happens, an why? Plant Physiol. 118 (1998) 9–17.
[54] A.N. Müllner, D.G. Angeler, R. Samuel, E.W. Linton, R.E. Triemer, Phylogenetic analysis of phagotrophic, phototropic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence, Int. J. Syst. Evol. Microbiol. 51 (2001) 783–791.
[55] F.E. Round, The evolution of pigmented and unpigmented unicells: a consideration of the protista, Biosystems 12 (1980) 61–69.
[56] G. Gockel, W. Hachtel, Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*, Protist 151 (2000) 347–351.
[57] G. Sainz, C.J. Carrell, M.V. Ponamarev, G.M. Soriano, W.A. Cramer, J.L. Smith, Interruption of the internal water chain of cytochrome *f* impairs photosynthetic function, Biochemistry 39 (2000) 9164–9173.
[58] Y. Choquet, F.-A. Wollman, Translational regulations as specific traits of chloroplast gene expression, FEBS Lett. 529 (2002) 39–42.
[59] Y. Choquet, K. Wostrikoff, B. Rimbault, F. Zito, J. Girard-Bascou, D. Drapier, F.-A. Wollman, Assembly-controlled regulation of chloroplast gene translation, Biochem. Soc. Trans. 29 (2001) 421–426.
[60] S.D. Schwartzbach, T. Osafune, W. Löffelhardt, Protein import into cyanelles and complex chloroplasts, Plant Mol. Biol. 38 (1998) 247–263.
[61] G.G. van Dooren, S.D. Schwartzbach, T. Osafune, G.I. McFadden, Translocation of proteins across the multiple membranes of complex plastids, Biochim. Biophys. Acta 1541 (2001) 34–53.
[62] J.-L. Popot, C. de Vitry, On the microassembly of integral membrane proteins, Annu. Res. Biophys. Chem. 19 (1990) 369–403.
[63] M.G. Claros, J. Perea, Y. Shu, F.A. Samatey, J.L. Popot, C. Jacq, Limitations to in vivo import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome *b*, Eur. J. Biochem. 228 (1995) 762–771.
[64] X. Pérez-Martínez, M. Vázquez-Acevedo, E. Tolkunova, S. Funes, M.G. Claros, E. Davidson, M.P. King, D. González-Halphen, Unusual location of a mitochondrial gene. Subunit III of cytochrome *c* oxidase is encoded in the nucleus of Chlamydomonad algae, J. Biol. Chem. 275 (2000) 30144–30152.