Available online at www.sciencedirect.com

SCIENCE DIRECT

Biochimica et Biophysica Acta 1658 (2004) 212–224

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

Higher plant-like subunit composition of mitochondrial complex I from *Chlamydomonas reinhardtii*: 31 conserved components among eukaryotes

Pierre Cardol[a], Frank Vanrobaeys[b], Bart Devreese[b], Jozef Van Beeumen[b], René F. Matagne[a], Claire Remacle[a,*]

[a] *Genetics of Microorganisms, Department of Life Sciences, University of Liège, B22, Institute of Botany, B4000 Liege, Belgium*
[b] *Laboratory of Protein Biochemistry and Protein Engineering, Ghent University, Ghent, Belgium*

Received 13 April 2004; received in revised form 31 May 2004; accepted 8 June 2004
Available online 2 July 2004

Abstract

The rotenone-sensitive NADH:ubiquinone oxidoreductase (complex I) is the most intricate membrane-bound enzyme of the mitochondrial respiratory chain. Notably the bovine enzyme comprises up to 46 subunits, while 27 subunits could be considered as widely conserved among eukaryotic complex I. By combining proteomic and genomic approaches, we characterized the complex I composition from the unicellular green alga *Chlamydomonas reinhardtii*. After purification by blue-native polyacrylamide gel electrophoresis (BN-PAGE), constitutive subunits were analyzed by SDS-PAGE coupled to tandem mass spectrometry (MS) that allowed the identification of 30 proteins. We compared the known complex I components from higher plants, mammals, nematodes and fungi with this MS data set and the translated sequences from the algal genome project. This revealed that the *Chlamydomonas* complex I is likely composed of 42 proteins, for a total molecular mass of about 970 kDa. In addition to the 27 typical components, we have identified four new complex I subunit families (bovine ESSS, PFFD, B16.6, B12 homologues), extending the number of widely conserved eukaryote complex I components to 31. In parallel, our analysis showed that a variable number of subunits appears to be specific to each eukaryotic kingdom (animals, fungi or plants). Protein sequence divergence in these kingdom-specific sets is significant and currently we cannot exclude the possibility that homology between them exists, but has not yet been detected.

Keywords: Complex I; *Chlamydomonas* mitochondria; Proteomics; Bioinformatics

1. Introduction

The NADH:ubiquinone oxidoreductase (complex I; EC 1.6.5.3) is the most complicated energy-transducing enzyme complex of the mitochondrial inner membrane. To date, it is known to contain 46 subunits (one being not yet identified) in *Bos taurus* (980 kDa) [1,2] and 33 subunits in *Neurospora crassa* (>700 kDa) [3]. Because of this complexity, information on its structure and mechanism of action is still limited. A simpler membrane-associated enzyme has been characterized in detail in bacteria. It is made of 14 subunits, all being conserved in mammalian and fungal complex I. The bacterial enzyme is considered to be the minimal structure required to ensure electron transfer coupled to proton translocation [4,5]. Seven of these subunits represent homologues of the complex I hydrophobic ND1–6 and ND4L subunits encoded in the mitochondrial genome in fungi and mammals. Both the bacterial NADH:ubiquinone oxidoreductase and the eukaryotic complex I contain a non-covalently bound FMN molecule and eight to nine iron–sulfur clusters [6,7]. As revealed by electron microscopy studies, the enzyme has an L-shaped structure with a membrane-embedded arm perpendicular to an arm protruding in the matrix. The former is made of hydrophobic subunits within the lipid bilayer, whereas the latter consists of hydrophilic subunits that are responsible for dehydrogenase activity and bears all the known prosthetic groups [8,9].

Complex I has also been purified from a variety of higher plants [10–12]. Recently, the dissection of complex I protein components from *Arabidopsis* and rice mitochondria showed that the two higher plant enzymes contain the 14 subunits

* Corresponding author. Tel.: +32-324-3663812; fax: +32-324-366 3840.
E-mail addresses: pierre.cardol@ulg.ac.be (P. Cardol), c.remacle@ ulg.ac.be (C. Remacle).

0005-2728/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbabio.2004.06.001

characteristic of prokaryotes as well as a set of nine proteins widely found in other eukaryotes [13]. At least 10 proteins present in the bovine complex I and absent from the fungal enzyme are also missing in the complex I from plants and are not encoded in their genomes. Furthermore, a series of plant-specific complex I associated subunits, most of them of unknown function, have been identified as well [13].

Complex I from the unicellular green alga *Chlamydomonas reinhardtii* has also been purified and very partially characterized. With a molecular mass of approximately 850–1000 kDa [14–16], the enzyme is made of more than 25 subunits which can be separated by electrophoresis [16]. Four of these subunits, all present in other eukaryotic complexes I, have been identified by antibody cross reaction [15] or by N-terminal sequencing [16]. In *Chlamydomonas*, only five complex I subunits (ND1, 2, 4, 5 and 6) are encoded in the mitochondrial genome [17]. The presence in the algal nuclear genome of the proteins generally encoded in the mitochondrial DNA (ND3, 4L in mammals and fungi; ND3, 4L, 7, and 9 in higher plants) has never been demonstrated. The nearly complete sequencing of the *C. reinhardtii* nuclear genome now provides the opportunity for a more complete characterization of the nuclear-encoded components from the alga [18].

The blue-native polyacrylamide gel electrophoresis technique (BN-PAGE) allows a separation of membrane complexes under native form [19]. It has been successfully applied to detect complex I assembly of *Chlamydomonas* wild-type and mutant strains deficient in complex I activity [14–16]. This technique also constitutes a rapid procedure for purification of small amounts of complex I usable in mass spectrometry-based identification of proteins using standard gel electrophoresis approaches [13,20].

In the present study, complex I from *Chlamydomonas* has been purified by BN-PAGE and a number of subunits have been identified by mass spectrometry. In view to enlarge the comparison between the algal complex I and its eukaryotic and prokaryotic counterparts, we have completed this proteomic analysis by searching the *Chlamydomonas* nuclear genome for homologous sequences of complex I subunits identified in plants, mammals, fungi and bacteria. Our data show that *Chlamydomonas* complex I is a higher plant-like enzyme, probably composed of 42 subunits. These include a set a 31–32 components widely conserved among eukaryotes whereas six proteins found in the *Chlamydomonas* enzyme but missing in the mammalian or fungal enzyme have been also detected in the complex I from higher plants.

2. Material and methods

2.1. Isolation of mitochondria

The strain of *C. reinhardtii* used in this work is the cell-wall-less mutant *cw15 mt*$^+$. Cells were routinely grown under light (75 μmol m^{-2} s^{-1} PAR) in TAP liquid medium. Mitochondria were isolated by digitonin treatment according to a published procedure [21] with minor modifications [15]. The protein content was determined by the Bradford method [22]. The mitochondrial preparation from 2 l of culture (10^{10} cells) gave 2.0 ± 0.4 mg protein in the crude mitochondrial fraction; the chlorophyll content was < 0.1 μg mg^{-1} protein.

2.2. BN-PAGE and two-dimensional SDS-PAGE

Sample preparation and BN-PAGE were carried out according to a published procedure [19] with the following modifications. Mitochondria (500 μg) were solubilized by addition of 100 μl of 0.75 mM ε-amino-*n*-caproic acid, 0.5 mM EDTA, 50 mM Bis-Tris–HCl (pH 7.0) and *n*-dodecyl-β-D-maltoside (1.5% [w/v]). After electrophoresis, the NADH dehydrogenase activities were detected by incubating the gel in 100 mM MOPS-KOH (pH 8.0) containing 100 mM NADH and 1 mg ml^{-1} NBT [23].

To resolve complex I into its constitutive subunits, the gel slice corresponding to the 850–1000-kDa stained activity was cut out from the BN-gel, treated 15 min at 60 °C in standard BN-Buffer containing 1% SDS and 1% β-mercaptoethanol, washed in BN-buffer [24], sealed vertically on 8% (w/v) acrylamide SDS-PAGE stacking gel, and laid on a 15% (w/v) acrylamide SDS-PAGE separating gel (see Ref. [20] for details). Two different gels, using Tris–Tricine or Tris–Glycine buffer, were used to obtain an optimal separation of all protein bands [19,25].

For the detection of proteins, the gels were incubated 16 h in 50% (v/v) ethanol, 2% (v/v) phosphoric acid, washed in water, stained for 1 h in a solution containing 0.2% (w/v) Coomassie blue G-250, 17% (w/v) ammonium sulfate, 34% (v/v) methanol and 3% (v/v) phosphoric acid, washed twice in water and then destained in a 30% (v/v) methanol solution. The apparent molecular mass of proteins was estimated using Precision Plus Dual Color (Bio-Rad Laboratories, Hercules, CA) and Multi Mark Pre-Stained Protein Standards (Invitrogen, Carlsbad, CA).

2.3. Mass spectrometric analyses

Individual bands of the 2D-gels were excised and collected in individual Eppendorf tubes. The gel slices were then digested with trypsin and the resultant peptides were extracted and dried as described elsewhere [26]. For mass spectrometric analyses, the peptide mixture was dissolved in 12 μl of 0.1% formic acid in water. One microliter of this digest mixture was added to an equal volume of matrix solution (100 mM alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% TFA in water) and 1 μl of the resultant mix was co-crystallized onto the target plate. Analysis was performed on a 4700 Proteomics Analyzer, a MALDI tandem time-of-flight mass spectrometer (Applied Biosystems, Framingham, CA, USA). The mass spectrometer was

externally calibrated with a standard peptide mixture in both the MS and MS/MS mode. The samples were first analyzed in the MS mode, and detected peptides were further submitted to MS/MS analysis. The acquired fingerprint data and the concomitant MS/MS data were then submitted to our local MASCOT server (www.matrixscience.com). If the protein band could not be identified by this method, we used a hyphenated strategy. We therefore coupled an automated nano-HPLC system (LC Packings, The Netherlands) to an electrospray ionisation-linear ion trap mass spectrometer (Q-TRAP LC-MS/MS system, Applied Biosystems). The experimental setup of the HPLC system has been described elsewhere [20]. Briefly, the peptide mixture was loaded onto a nano-column (Zorbax 3.5 μm, 150 mm × 75 μm I.D, Agilent Technologies) using an in-line pre-concentration step on a micro pre-column cartridge (2 mm × 800 μm I.D., LC Packings), and separated using a linear gradient (50 min). The instrument performed an MS scan as survey scan and selected the two most intense ions for an enhanced resolution scan. If their charge state was two or three, the peptides were automatically selected for an enhanced product ion scan (MS/MS). The eluted peptides were detected 'on-line' by the mass spectrometer. The total cycle time of the setup of the instrument was approximately 4.5 s. The acquired MS and MS/MS data were then submitted to the MASCOT server.

2.4. Bioinformatic analyses for gene identification

Protein sequences were obtained from the National Center for Biotechnology Information (NCBI) server (http://www.ncbi.nlm.nih.gov). *Chlamydomonas* homologous sequences were identified using the WU-TBLASTN 2.0 facility (http://www.biology.duke.edu/chlamy_genome/) with *N. crassa*, *B. taurus*, *A. thaliana*, *O. sativa* and *C. elegans* proteins against the "20021010" *Chlamydomonas* EST assembly (v3.0 released on December 6, 2002; distributed by the *Chlamydomonas* genetic resource center), a draft genome sequence of *Chlamydomonas* (v1.0 released on February 4, 2003; distributed by the DOE Joint Genome Institute) and the sequences of individual EST clones from NCBI.

Ipsort [27] (http://hypothesiscreator.net/iPSORT/), Mitopred [28] (http://mitopred.sdsc.edu/) and TargetP v1.0 [29] (http://www.cbs.dtu.dk/services/TargetP/) targeting predictions were determined using full-length predicted protein sequences. Eukaryotic homologous sequences were identified using BLASTP, TBLASTN, RPS-BLAST, PSI-BLAST, CDART tools available at the NCBI server. Calculations of the molecular masses and hydropathy profiles of amino acid sequences were done with the compute pI/MW tool [30] and with the Protscale tool using a window of seven residues [31], respectively, both from the ExPaSy molecular Biology Server (http://au.expasy.org/). Multiple sequence alignments were performed with the ClustalW tool (v1.8) [32] from the Baylor College of Medicine (BCM) (http://searchlauncher.bcm.tmc.edu/) and shaded with the BoxShade 3.21 tool (http://www.ch.embnet.org/software/BOX_form.html).

3. Results

3.1. Isolation of mitochondrial complex I of C. reinhardtii and identification of its constitutive subunits

Purification of mitochondria from green algae is complicated by the presence of a rigid cell wall that is hard to break, and by the low content of mitochondria per cell (less than 1% of the cell volume). In *C. reinhardtii*, however, this problem can be overcome by using mutant cells lacking a cell wall [21,33]. After disruption of cells using a digitonin treatment and purification of organelles, we obtained a mitochondrial fraction that was essentially free of chlorophyll-associated protein contaminants (see Material and methods). After solubilization by *n*-dodecyl-β-D-maltoside (1.5% final concentration), the mitochondrial protein complexes were separated by BN-PAGE. Following Coomassie blue staining, the protein profile exhibited four major bands (Fig. 1a) corresponding to the native complexes I, III (dimeric form), IV and V (dimeric form) [16], as well as several additional weaker bands representing unknown protein complexes. Recent data on the higher plant complex

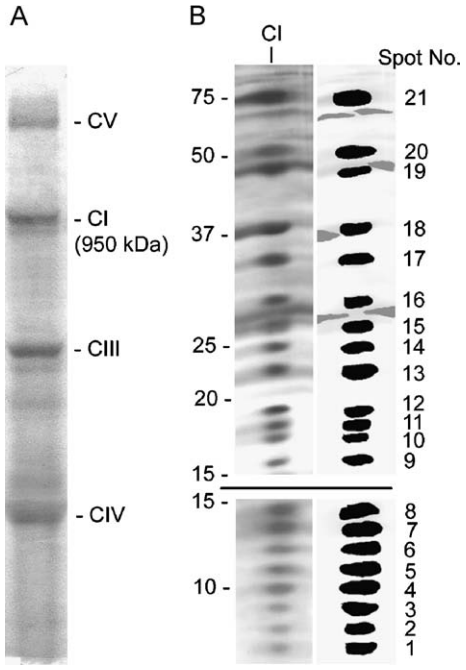

Fig. 1. (A) Coomassie-Blue stained image of the main respiratory-chain complexes from *Chlamydomonas* mitochondria separated by BN-PAGE. (B) Coomassie blue-stained image of native complex I (obtained by BN-PAGE) separated into its subunits by SDS-PAGE. Apparent molecular mass of markers is shown in kDa. The top panel shows a Tris–Glycine buffer separation. The lower panel shows proteins separated in Tris–Tricine buffer. Alongside the gel is a schematic representation of the protein spots analyzed by MS (Table 1).

I indicate a molecular mass of circa 1000 kDa [13]. This would modify the estimated molecular mass of the band on BN-PAGE corresponding to the *C. reinhardtii* complex I to 950 kDa instead of 850 kDa, as previously published [14,15]. We confirmed that the protein band localized at circa 950 kDa corresponded to complex I by incubating the gel with NADH and nitrobluetetrazolium (NBT) to provide evidence of NADH dehydrogenase activity (data not shown).

The band corresponding to complex I was excised from the BN-gel and its constitutive subunits were then separated by SDS-PAGE. To separate polypeptides in the ranges >15 kDa and < 15 kDa, we used Tris–Glycine and Tris–Tricine gel running buffers, respectively. A minimum of 21 bands were observed after Coomassie blue staining (Fig. 1b). The estimated molecular masses of the individual protein bands ranged between 7 and 75 kDa (Table 1). In order to systematically identify the proteins, the 21 bands were excised from the gel and digested with trypsin. The resultant peptides were analyzed by MALDI-MS using peptide mass fingerprint (PMF). In cases where PMF was not sufficient for unambiguous identification, MALDI-MS/MS and nano-flow LC-MS/MS on an electrospray Q-TOF instrument were performed to provide MS/MS data [20].

Thirty proteins representing a total molecular mass of about 700 kDa were identified from the 21 bands of *Chlamydomonas* complex I (Table 1). Database searching revealed that 18 of these proteins were clear orthologues of the following bovine complex I subunits : B8, B18, 13 kDa-A, ND3, B14, B16.6, PDSW, PSST, B17.2, B13, AQDQ, TYKY, 30 kDa, 24 kDa, 39 kDa, 49 kDa, 51 kDa and 75 kDa.

Table 1
Subunit identification in the purified *Chlamydomonas* complex I preparation

Spot	kDa	NP	*Chlamydomonas* entry	MM	Protein identification	TP	MP	IP
1	7	N.I.	–	–	–			
2	8	N.I.	–	–	–			
3	9	1	AAS58500	12	CI-AGGG ?	M	M	M
4	10	2	AAS58501	10	CI-6 kDa subunit (*A. thaliana*)	M	M	–
5	11	2	AAQ63699	11	CI-B8 subunit	M	M	M
6	12	3	AAQ73135	11	CI-B18 subunit	–	–	–
7	13	2	AAQ64641	13	CI-20.9 kDa subunit (*N. crassa*)	–	–	–
		1	AAQ64639	16	CI-13 kDa-A subunit	M	–	M
		3	AAS48193	16	CI-8 kDa subunit (*A. thaliana)*	M	M	M
8	14	1	AAQ55461	14	CI-ND3 subunit	M	–	M
		4	AAQ84469	16	CI-B14 subunit	M	M	–
9	16	4	AAS58503	15	unknown protein	–	–	–
		4	AAQ64637	17	CI-B16.6 subunit	–	–	–
10	17	4	AAQ55459	18	CI-PDSW subunit	–	–	–
		4	AAS48192	20	CI-13 kDa (*A. thaliana*)	M	M	M
11	18	4	AAQ63698	19	CI-PSST subunit	–	–	–
		8	AAQ64638	18	CI-B17.2 subunit	–	M	M
		2	AAQ73139	18	CI-B13 subunit	M	M	M
12	19	10	AAQ64640	22	CI-AQDQ subunit	M	M	M
		4	AAS58500	22	unknown protein	M	M	M
		3	AAS58502	22	like NP_566309 (*A. thaliana*)	M	M	M
13	23	7	AAQ63697	27	CI-TYKY subunit	M	M	M
		4	AAS58499	23	CI-B14.7 subunit?	M	–	–
14	25	4	AAQ55457	32	CI-30 kDa subunit (ND9)	–	M	M
15	27	6	AAQ63695	31	CI-24 kDa subunit	M	–	M
		2	AAS48195	25	FBP-like (*A. thaliana*)	–	M	–
16	29	5	AAS48196	31	FBP-like (*A. thaliana*)	M	M	M
17	32	11	AAS48197	33	FBP-like (*A. thaliana*)	–	–	–
18	38	12	AAQ55458	40	CI-39 kDa subunit	M	M	M
19	43	9	AAQ63700	53	CI-49 kDa subunit (ND7)	M	M	M
20	50	6	AAQ63696	53	CI-51 kDa subunit	M	–	M
21	75	12	AAQ73136	76	CI-75 kDa subunit	M	–	M

MS/MS spectra derived from trypsinated peptides of proteins were matched at local Mascot against a custom protein database and against a translated EST contig database (release October 10, 2002; available at ftp://ftp.biology.duke.edu/pub/chlamy_genome/). Deduced protein sequences from *Chlamydomonas* were blasted against protein or translated NCBI databases http://www.ncbi.nlm.nih.gov/).

Table headings: spot, protein spot number from Fig. 1; kDa, observed molecular mass in kDa (these values are somewhat different from those previously published [16], especially concerning the small molecular mass proteins. However, the use of different molecular mass markers may explain the slight differences in protein band mass estimations; NP, number of peptides matching to predicted protein sequence (N.I.: not identified); *Chlamydomonas* entries are accession numbers from GenBank; MM, predicted molecular mass of matched sequence in kDa; protein identification, the complex I subunits are named according to the bovine nomenclature, except for those where another species is mentioned. Protein sequences were analyzed by three targeting prediction programs: TargetP v1.01 (TP), MitoPred (MP) and IPsort (IP). M indicates mitochondrial targeting; a dash (–) indicates no prediction of mitochondria targeting.

Two additional proteins could also be orthologues of bovine complex I components. The 23-kDa AAS58499 protein sequence is 15–20% identical to the bovine B14.7 complex I subunit and its related component 21.3b in the fungus *N. crassa*. The alignment was performed at the exclusion of the first 41 N-terminal residues in the *Chlamydomonas* sequence which were not found in bovine and fungal sequences (see Fig. 2a for alignment and accession numbers). In spite of the differences in length between these subunits, the fact that their hydrophobicity profiles matched in the regions aligned with the bovine sequence (Fig. 2b) is additional evidence to consider these proteins as being orthologues. The 9-kDa AAS58498 polypeptide shares 15% identity and 25% similarity with the bovine complex I AGGG subunit (also 9 kDa) (data not shown). The protein is predicted to be targeted to mitochondria by the topology programs Mitopred, IPsort and TargetP 1.0 (Table 1). However, taking into account that the hydropathy profiles of these two proteins do not match (data not shown) and that we have not found any related sequences in fungi or higher plants genomes, it is hard to conclude that the *Chlamydomonas* AAS58498 protein belongs to the AGGG complex I protein family. Beside

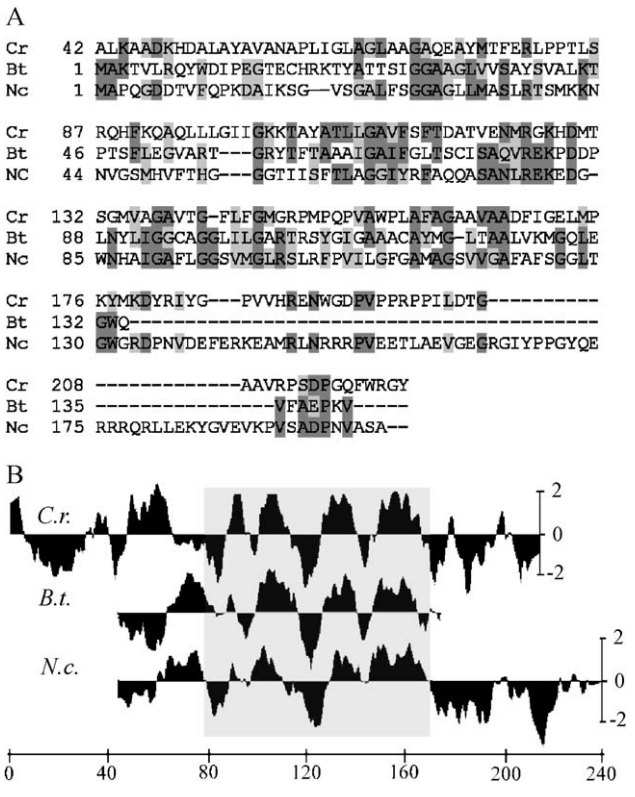

Fig. 2. Comparison of the sequences of *C. reinhardtii* 23-kDa (AAS58499), *B. taurus* B14.7 (NP_783649) and *N. crassa* 21.3b (S14277) complex I subunits. (A) Clustal W alignment of the protein sequences. Amino acids conserved in at least two sequences are shown on a grey background; similar residues are shown on a light-grey background. (B) Comparison of the hydrophobicity profiles of amino acid sequences, aligned vertically according to (A).

counterparts of complex I proteins from mammals, seven homologues to fungal or plant complex I components that were not before described as having counterparts in the bovine enzyme were identified in the *Chlamydomonas* preparation (Table 1). Three proteins (27, 29 and 32 kDa) belong to a gene family related to bacterial siderophores, known as ferripyochelin-binding proteins (FBP). Such FBP-like proteins have already been found in association with complex I from *Arabidopsis* and rice mitochondria [13], whereas no homologous gene has ever been detected in mammalian or fungal genomes.

A series of small molecular mass proteins (<20 kDa) were purified along with the *Chlamydomonas* complex I. The sequence of the 13-kDa AAQ64641 protein clearly matches to the 20.9-kDa complex I component in *N. crassa* and its newly discovered counterpart in higher plants (*Arabidopsis* 9-kDa At4g16450 gene product [13]). The 86-amino-acid AAS58501 protein is similar to the 6-kDa At4g20150 gene product from *A. thaliana* (25% identity) and to an 81-amino-acid polypeptide from rice (BAC79719). The 142-amino-acid AAS48193 sequence matches the *Arabidopsis* 8-kDa At2g31490 gene product (33% identity) and the rice BAB67906 protein (30% identity). Both these subunits are predicted to be addressed to mitochondria by three different prediction programs (Table 1) and their homologues have been found in preparations of complex I from *Arabidopsis* and rice [13]. Their presence in complex I of the green alga *Chlamydomonas* provides additional evidence for these two proteins to be part of complex I in photosynthetic eukaryotes.

The 17-kDa AAS48192 protein sequence shows 26–31% identity with members of a protein family found in complex I preparations from higher plants (*Arabidopsis* At2g42310/At3g57785 and rice AK108131 gene products; the rice sequence was previously described as 3605.m00182 TIGR entry [13]). The plant proteins and their homologue in *C. reinhardtii* moreover display significant similarity to an 11.7-kDa unknown protein of *N. crassa*. Interestingly, a position-specific iterative BLAST analysis (PSI-BLAST) suggests that the four proteins are distant relatives of the 14.5-kDa bovine complex I subunit ESSS and of its putative homologue in *C. elegans* (see Fig. 3a for alignment and accession numbers). Despite great differences in length between the sequences of these proteins, the correspondence of their hydrophobicity profiles supports the view that these proteins are homologous (Fig. 3b).

Finally, the *Chlamydomonas* AAS58500, AAS58502 and AAS58503 proteins do not show any similarity to known complex I-associated proteins in eukaryotes. The AAS58502 polypeptide shares 25% identity with the *A. thaliana* At3g07480 gene product and both the proteins are addressed to mitochondria by multiple prediction programs. Concerning the AAS58500 and AAS58503 gene products, none of them are predicted to have a mitochondrial localization. In order to identify distant relatives or to detect conserved domains and architecture homologies, their sequences were

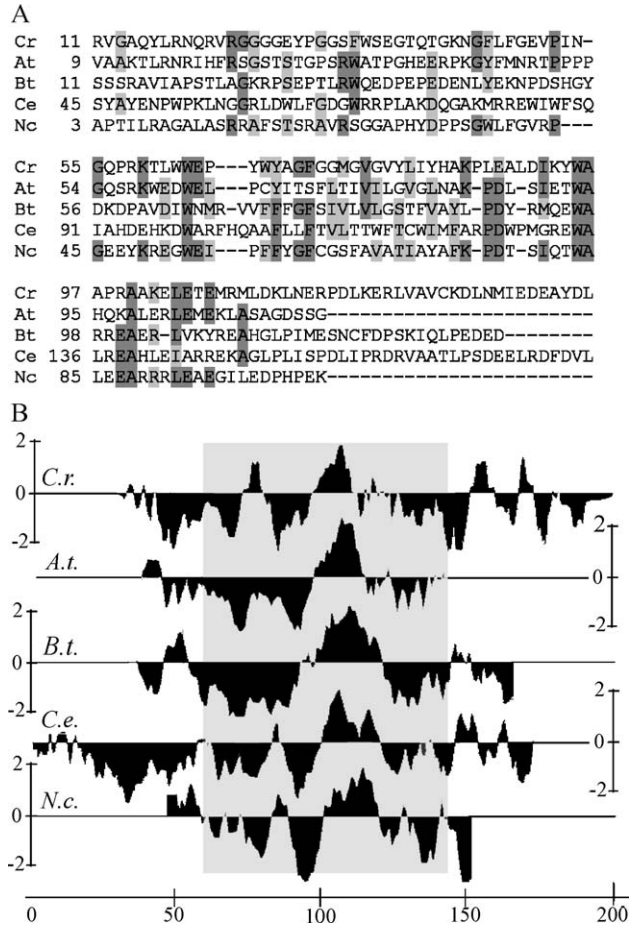

Fig. 3. Comparison of the ESSS-like proteins of *C. reinhardtii* (AAS48192), *A. thaliana.* (At3g57785), *B. taurus* (Q8HXG5), *C. elegans* (NP_501358) and *N. crassa* (XP_324110). (A) Clustal W alignment of the protein sequences. Amino acids conserved in at least three sequences are shown on a grey background; similar residues are shown on a light-grey background (B) Comparison of the hydrophobicity profiles of amino acid sequences, aligned vertically according to (A).

submitted to various BLAST analyses against sequences from other organisms but no match was ever obtained. These three proteins thus need to be studied further in order to confirm their association with complex I. While we cannot exclude the possibility that co-elution or nonspecific association of proteins occurred during the one-step purification process used in our work, it should be however mentioned that we did not find any other known mitochondrial proteins in our complex I sample.

3.2. Searching for homologues of complex I subunits from a genomic analysis using Chlamydomonas and eukaryotic databases

A number of complex I subunits identified in mammals, fungi and higher plants were not found in our proteomic analysis of *Chlamydomonas* complex I, maybe because some components, in particular small size or hydrophobic ones, have escaped detection using our experimental approach. In this context, a 12-kDa protein not found in the present mass spectrometry study (Table 1) was detected in *Chlamydomonas* mitochondria after immunological tests using a rabbit polyclonal antibody directed against the SDAP bovine complex I subunit (Gift from Dr. J. Hirst) (data not shown). Thus, to extend our proteomic study, we searched the *Chlamydomonas* genome for counterparts to known complex I subunits identified in unrelated organisms. In a first step, we examined the data of the literature concerning the subunit composition of the prokaryotic NADH dehydrogenase enzyme and of complex I from mammals (*B. taurus*), fungi (*N. crassa*) and higher plants (*A. thaliana*).

The bovine genes coding for the 38 nuclear-encoded proteins and the seven mitochondria-encoded subunits (ND proteins) of complex I have all been sequenced [2]. The bovine subunits are named according to their molecular mass and, in some cases, following their first four amino acids [34]. This nomenclature has been used as a helpful basis for Table 2.

The seven mitochondria-encoded ND subunits in mammals are orthologues of the bacterial NUOA, NUOH and NUOJ-N products. The five proteins (75 kDa, 51 kDa, 24 kDa, TYKY and PSST) responsible for binding iron–sulfur clusters and FMN along with the 49-kDa and 30-kDa subunits (encoded by the mitochondrial *nad7* and *nad9* genes in higher plants) are counterparts of the bacterial NUOB-NUOG and NUOI proteins [35]. These 14 proteins corresponding to the minimal bacterial enzyme are also detected both at genomic and proteomic levels in the fungus *N. crassa* [3], in the worm *C. elegans* [36] and in higher plants [13]. By compiling recent data on the complex I subunit conservation among eukaryotes [3,13,36], it appears that 13 other bovine complex I proteins, absent in the prokaryotic enzyme, are highly conserved among eukaryotes (MWFE, SDAP, B8, B13, 13 kDa-A, B14, B17.2, B18, AQDQ, PGIV, B22, PDSW and 39 kDa) (Table 2).

Though the 9.5-kDa protein (A44210) from fungal complex I was also proposed to be homologous to the B9 bovine subunit (Q02371) [37], various BLAST analyses performed at the NCBI website failed to find any significant similarity between the two polypeptide sequences. Furthermore, considering that their hydropathy profiles do not match at the exception of a 15-amino-acid hydrophobic stretch (data not shown), and that we have not found any related sequences in genomes of non-mammalian animals and in higher plants, we conclude that the 9.5-kDa and B9 subunits are probably not homologues. A set of three other *N. crassa* complex I components (17.8, 20.9, 21.3a) have also no known counterparts in the mammalian enzyme [3]. Finally, a set of 10 proteins have been recently proposed to be plant-specific: the 11- and 16-kDa NADH dehydrogenase subunits, at least four ferripyochelin-binding protein-like subunits and four non-related proteins of small molecular masses [13].

Table 2
Genomic analysis of complex I subunits from model prokaryotic and eukaryotic organisms and identification of homologous sequences

E. coli 14 subunits	Mammals (*B. taurus*) 45 subunits		Fungi (*N. crassa*) 38 subunits		Higher plants (*A. thaliana*) 41 subunits		Green algae (*C. reinhardtii*) 42 subunits		MS	% id
Bacterial NADH dehydrogenase orthologues (14 subunits)										
NUOB	PSST (20)	P42026	19.3	O47950	24	At5g11770	18	AAQ63698	*	76/72/83
NUOI	TYKY (23)	P42028	21.3c	Q12644	25.5	At1g16700	23	AAQ63697	*	75/70/78
NUOE	24	M22539	24	X78083	28.3	At4g02580	27	AAQ63695	*	54/46/53
NUOC	30	P23709	31	P23710	ND9 (22.6)	Q95748	25	AAQ55457	*	52/50/60
NUOD	49	S04104	49	X54508	ND7 (44.6)	P93306	43	AAQ63700	*	68/62/68
NUOF	51	P25708	51	P24917	53.5	At5g08530	50	AAQ63696	*	71/72/75
NUOG	75	J02877	78	X57602	81.5	At5g37510	75	AAQ73136	*	51/52/50
NUOH	ND1 (36)	P03887	ND1 (42)	P08774	ND1 (36)	NP_085565	ND1 (31.6)	AAB93446		45/41/50
NUON	ND2 (39)	P03892	ND2 (66)	A25096	ND2 (55)	NP_085584	ND2 (42.4)	AAB93444		14/24/26
NUOA	ND3 (13)	P03898	ND3	Q35141	ND3 (14)	NP_085553	14	AAQ55461	*	41/26/43
NUOM	ND4 (52)	P03910	ND4	S02153[1]	ND4 (55)	NP_085518	ND4 (48.7)	AAB93441		34/39/38
NUOK	ND4L (11)	P03902	ND4L (10)	P05509	ND4L (11)	NP_051111	24.2[3]	AAO61142		27/32/31
NUOL	ND5 (67)	P03920	ND5 (80)	P05510	ND5 (74)	NP_085478	ND5 (59)	AAB93442		27/35/41
NUOJ	ND6 (19)	P03924	ND6	S02156[1]	ND6 (23.5)	NP_085495	ND6 (17.7)	AAB93445		15/27/28
Eukaryote-specific subunits (17 subunits)										
	MWFE (7.5)	Q02377	9.8	CAE85571	7.5	At3g08610	7.5[3]	AAS48198		16/20/21
	SDAP (8)	P52505	9.6	S17647	14	At1g65290	14[3]	AAQ73138		49/52/55
	B8	X63219	10.5	Q07842	10.8	At5g47890	11	AAQ63699	*	52/37/49
	B12	Q02365	10.6[3]	XP_331394	7	AK059007[2]	6.5[3]	AAS48194		33/30/37
	B13	P23935	29.9	P24919	19.2	At5g52840	18	AAQ73139	*	24/20/39
	13a	S28238	18.4[3]	EAA26933	12.2	At3g03070	13[3]	AAQ64639	*	28/31/53
	B14	X63211	14.8	CAA53963	15	At3g12260	14	AAQ84469	*	41/24/31
	ESSS (14.5)	Q8HXG5	11.7[3]	XP_324110	13[3]	At3g57785	17	AAS48192	*	10/28/31
	PFFD (15)	Q02379	11.5[3]	EAA31476	14	At3g62790	11	AAQ98888		28/37/36
	B16.6	Q95KV7	13.5[3]	EAA29209	16.1	At1g04630	16	AAQ64637	*	42/37/45
	B17.2	O97725	13.4[3]	EAA31813	18	At3g03100	18	AAQ64638	*	25/23/34
	B18	Q02368	89.7[3]	EAA28195	12	At2g02050	12	AAQ73135	*	36/45/44
	AQDQ (18)	X63215	21	P25711	17.1	At5g67590	19	AAQ64640	*	48/34/44
	PGIV (19)	P42029	20.8	EAA35830	12	At5g18800	12.9[3]	AAQ55460		28/18/31
	B22	S28256	18[3]	CAD60692[1]	13.6	At4g34700	13.9[3]	AAQ73134		32/36/35
	PDSW (22)	Q9DCS9	12.3	X68965	12.5	At4g16450	17	AAQ55459	*	14/16/20
	39	X59418	40	P25284	44	At2g20360	38	AAQ55458	*	40/35/41
Subunits identified in at least two lineages (4 subunits)										
	B14.7	NP_783649	21.3b	S14277	N.I.[4]		23	AAS58499	*	15/20/–
	B15	P48305	7[3]	XP_322246	N.I.		N.I.			
	ASHI (19)	S28242	20.1[3]	XP_332152	N.I.[4]		N.I.[4]			
	N.I.		20.9	Q02854	9	At4g16450	13	AAQ64641	*	–/24/29
Mammal-specific subunits (11 subunits)										
	KFYI (6)[6]	Q02376	N.I.		N.I.		N.I			
	MNLL (7)	Q02378	N.I.		N.I.		N.I.			
	AGGG (8)	Q02374	N.I.		N.I.		N.I.[4]			
	B9[5,6]	Q02371	N.I.[5]		N.I.		N.I.			
	MLRQ (9)	Q01321	N.I.		N.I.		N.I.			
	10[6]	P25712	N.I.		N.I.		N.I.			
	B14.5a	Q05752	N.I.		N.I.		N.I.			
	B14.5b	Q02827	N.I.		N.I.		N.I.			
	SGDH (16)	Q02380	N.I.		N.I.		N.I.			
	B17	Q02367	N.I.		N.I.		N.I.			
	42	P34942	N.I.		N.I.		N.I.			
Fungus-specific subunits (3 subunits)										
	N.I.[5]		9.5[5]	A44210	N.I.		N.I.			
	N.I.		17.8	X71414	N.I.		N.I.			
	N.I.		21.3a	P19968	N.I.		N.I.			

(continued on next page)

Table 2 (*continued*)

E. coli 14 subunits	Mammals (*B. taurus*) 45 subunits	Fungi (*N. crassa*) 38 subunits	Higher plants (*A. thaliana*) 41 subunits		Green algae (*C. reinhardtii*) 42 subunits		MS	% id
Plant-specific subunits (12 subunits)								
	N.I.	N.I.	6	At4g20150	10	AAS58501	*	–/–/25
	N.I.	N.I.	8	At2g31490	13	AAS48193	*	–/–/33
	N.I.	N.I.	17[3]	At3g07480	19	AAS58502	*	–/–/25
	N.I.	N.I.	25 (FBP-like)	At5g63510	27	AAS48195	*	–/–/47
	N.I.	N.I.	27 (FBP-like)	At3g48680	N.I.			
	N.I.	N.I.	30 (FBP-like)	At1g47260	32	AAS48197	*	–/–/42
	N.I.	N.I.	32 (FBP-like)	At5g66510	29	AAS48196	*	–/–/45
	N.I.	N.I.	11 (NDH11)	At1g67350	N.I.			
	N.I.	N.I.	16 (NDH16)	At2g27730	N.I.			
	N.I.[4]	N.I.	N.I.		9	AAS58498	*	–/–/–
	N.I.	N.I.	N.I.		16	AAS58503	*	–/–/–
	N.I.	N.I.	N.I.		19	AAS58500	*	–/–/–

Mammal, fungal and plant protein sequences were blasted against *C. reinhardtii* databases and against the NCBI protein bank. Names and molecular masses of complex I components are based on published reviews [2,3,13]. The percentage of identity (%id) between *Chlamydomonas* proteins and its eukaryotic homologues are given when applicable. *N.I.* (not identified) indicates that no sequence of significant similarity was found. Asterisks in the column (MS) indicate the presence of the protein in the mass spectrometry identification records of Table 1.

ND proteins are mitochondria-encoded subunits. (1)*Podospora anserina* sequences. (2)*Oryza sativa* gene product identified in complex I [13]. (3)Molecular masses predicted from sequence analyses only (these subunits were not identified by proteomic analysis). (4)No clear homologous sequence was identified (see text for details). (5)Previously described as orthologues [37] (see text for details). (6)No related sequences in non-mammalian species have been identified for these three proteins (see text for details).

Taking these data into account and in order to identify algal sequences, the known complex I protein sequences were used to search mitochondrial, genomic and EST databases from the *Chlamydomonas* genome project (http://www.biology.duke.edu/chlamy_genome/). In parallel, all available complex I protein sequences were used to search for homologous sequences among eukaryotes, using BLAST facilities from the NCBI server. In *Chlamydomonas*, orthologues of the 27 conserved subunits were identified (Table 2). These proteins include the five mitochondria-encoded subunits ND1, 2, 4, 5 and 6. Among the 22 conserved nuclear genes, we found in particular the genes encoding the hydrophobic ND3 and ND4L proteins, which are located in the mitochondrial genome of all non-related organisms. In addition, we identified four new widely conserved complex I component families. The PFFD subunit (15 kDa) and the newly described B16.6 protein of bovine complex I [38], which have counterparts identified in complex I preparations from plants [13], are homologous to predicted proteins of *N. crassa* (11.5 and 13.5 kDa, respectively) and *C. reinhardtii* (11 and 16 kDa, respectively). Homologous genes are also present in the worm *C. elegans* genome [36] (see Fig. 4a and b for alignments and accession numbers). The 7-kDa AK059007 gene product identified as a complex I component in rice (described as 2468.m001222 TIGR entry [13]) shows similarity with the predicted 6.5-kDa AAS48194 polypeptide of *Chlamydomonas*. Both sequences also match with the B12 subunit of mammal complex I and with *N. crassa* 10.6-kDa and *C. elegans* 12.1-kDa unidentified proteins (see Fig. 4c for alignments and accession numbers). Finally, homologues of the bovine ESSS subunit are also present in all genomes investigated and orthologues were described above (Fig. 3). Taken together, our findings indicate that the PFFD, B16.6, B12 and ESSS subunits are conserved in all eukaryotes examined, which extends the number of conserved complex I components to 31.

Besides these widely conserved complex I components, we also examined the remaining subunits presented in Table 2 as subunits previously identified in at least two lineages (mammals and fungi or fungi and plants) and lineage-specific subunits. Carroll et al. [1] pointed out the similarity between the bovine B14.7 subunit, the *Neurospora* 21.3b subunit and proteins of the mitochondrial inner membrane translocase (TIM17, 22 and 23). The *Chlamydomonas* 23-kDa complex I component (AAS58499) identified through our proteomic analysis also shares similarity with TIM proteins. When looking for homologous sequences to these proteins, we indeed found putative homologues of TIM17, 22 and 23 in plant genomes but also found the *Chlamydomonas* BI72446 EST product and the *Arabidopsis* At2g42210 gene product whose identification as B14.7 orthologues or as TIM family members was not possible on the sole basis of the sequences. Conversely, the two proteins identified by BLAST analysis in the databases of *C. elegans* (NP_501427, NP_500627) are clearly related to TIM17/22 proteins. While no plant sequence was found to be similar to the B15 bovine subunit or to its homologue in *C. elegans* (NP_492001), the *Neurospora* 7-kDa XP_322246 protein lined up with the carboxyl-terminal part of the animal complex I subunit. Bovine and fungal polypeptides share 24% identity and 36% similarity and their hydrophobicity profiles match closely in the region aligned with the fungal protein sequence (data not shown). The *Neurospora* XP_332152 unknown protein (173 aa) is homologous to the mammalian complex I ASHI subunit (19 kDa) and to the *C. elegans* CAA21668 polypeptide: their

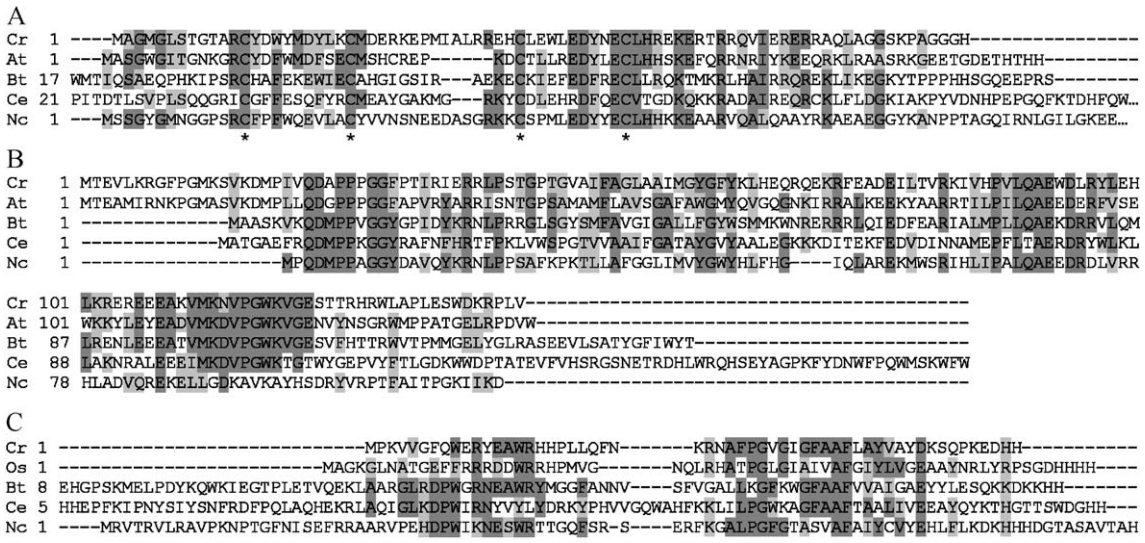

Fig. 4. Clustal W multiple sequence alignments of complex I PFFD-like, B16.6-like and B12-like subunits. (A) PFFD-like proteins: *C. reinhardtii* AAQ98888, *A. thaliana* At3g62790, *B. taurus* Q02379, *C. elegans* NP_491085, *N. crassa* EAA31476. (B) B16.6-like proteins: *C.r.* AAQ64637, *A.t.* At1g04630, *B.t.* Q95KV7, *C.e.* NP_492799, *N.c.* EAA29209. (C) B12-like proteins: *C.r.* AAS48194, *O. sativa* AK059007, *B.t.* Q02365, *C.e.* NP_495907, *N.c.* XP_331394. Amino acids conserved in at least three sequences are shown on a grey background; similar residues are shown on a light-grey background and conserved cysteine residues from PFFD subunits are marked by asterisks.

protein sequences share 25% identity and similar hydrophobicity profiles (data not shown). Interestingly, the amino-terminal part of the fungal protein sequence aligns with the amino-terminal part of the *Arabidopsis* At1g02870 gene product (193 aa) and the *Chlamydomonas* BI718025 EST product (196 aa). However, the hydropathy profiles of the plant proteins are different from those of fungal or mammal ASHI proteins and the consensus is not sufficient to conclude that the *Arabidopsis* and *Chlamydomonas* proteins belong to the complex I ASHI subunit family. Finally, the amino-terminal part of the 132 amino acid BU650459 EST product in *Chlamydomonas* was also found to share 29% identity and 39% similarity with the bovine KFYI subunit (76 aa). However, the bovine and algal protein sequences have quite different hydrophobicity profiles (data not shown), which suggests that the similarity might be fortuitous.

Our searches in the genomes of plants, algae and fungi failed to identify homologous sequences for 11 mammalian-encoded complex I subunits (Table 2). However, with the exception of bovine KFYI, B9 and 10-kDa subunits, we have been able to identify homologous sequences in some non-mammalian animal model species (insects *D. melanogaster* or *A. gambiae*, zebrafish *D. rerio* and worm *C. elegans*). The B17, AGGG and 42-kDa bovine complex I subunits have clear homologues in *A. gambiae* (XP_318629, XP_312296, XP_309517), *C. elegans* (NP_723927, NP_495919, NP_741215) and *D. rerio* (AAH59684, AAH59649, NP_955872) and their sequences share 10–15% identity and 20–24% similarity. Mammal SGDH subunit family encompasses the *D. melanogaster* NP_652042 and *C. elegans* NP_741000 proteins and the three proteins are 13% identical and 23% similar. The bovine complex I B14.5a and B14.5b components have homologues in *A. gambiae* (XP_308132, XP_312024) and in *C. elegans* (CAE73341, AAK29975) and their sequences share only 10% identity and about 20% similarity. The putative BX008826 and NM_168988 proteins from *A. gambiae* are 22–47% identical and 34–60% similar to the bovine MNLL and MLRQ subunits, respectively. These data reveal that the degree of conservation is very weak between the members of most of these animal complex I subunit families. Hence, we cannot exclude that homologous components do exist in higher plants, algae and fungi, but they were not identified in our analyses due to great divergence.

In this context, we have not identified in plant, algal or animal genomes sequences similar to the *N. crassa* 9.5-, 17.8- and 21.3a-kDa complex I subunits, whereas homologues were identified in the protein databases from the two fungi *Magnaporthe grisea* (EAA47579, EAA47033, EAA53227) and *Aspergillus nidulans* (EAA60724, EAA64574, EAA60040). The degree of conservation of the fungal subunits is highly variable, the 17.8-kDa family members being the less conserved (19% identity and 27% similarity) and the 9.5-kDa family proteins being the most conserved (45% identity and 63% similarity). Concerning hypothetic lineage-specific subunits in plants, and with respect to the findings of the MS identifications (Table 1), there are only three genes coding for FBP-like proteins in *Chlamydomonas*, whereas four homologous genes have been identified in *Arabidopsis* [13]. Finally, no sequence related to higher plant 11- and 16-kDa subunits was identified, whereas all other plant-specific subunits identified by Heazlewood et al. [13] have counterparts in *Chlamydomonas* complex I (see also the proteomic results).

4. Discussion

4.1. The multiprotein complex I of C. reinhardtii

By combined proteomic and genomic approaches, we have here analyzed the subunit composition of the mitochondrial complex I from the unicellular green alga *C. reinhardtii*. After isolation from a Blue-Native gel, the algal complex I was resolved into 21 bands by SDS-PAGE. While the two smallest polypeptide bands were not identified by mass spectrometric analysis, 30 coding sequences were identified out of the 19 remaining bands, for an approximate total molecular mass of 700 kDa. The finding of our genomic approach led to the identification of 42 protein sequences (Table 2), the sum of the individual molecular masses (about 970 kDa) being consistent with the native molecular mass of 950 kDa for *Chlamydomonas* complex I. All predicted subunits that have not been found in the electrophoresis gel either have a predicted molecular mass less than 14 kDa or are extremely hydrophobic, which is very similar to the results obtained by similar approach on human complex I [20]. While some low molecular mass subunits may not be visualised by the Coomassie-blue staining, hydrophobic subunits may have been lost during the second electrophoresis. In particular, none of the five mitochondria-encoded subunits (ND1, 2, 4, 5, 6) was detected. Moreover, such proteins contain few tryptic cleavage sites, generating only a limited number of peptides in the mass range that can be used for PMF and peptide sequencing (1–3.5 kDa).

4.2. Identification of ND3, 4L, 7 and 9 in the Chlamydomonas genome

C. reinhardtii as other related Chlamydomonad algae has a very low number of mitochondria-encoded proteins [39]. From the typical set of seven complex I subunits encoded by the mitochondrial genome of mammals and fungi (ND1-6, 4L), the ND3 and ND4L polypeptides are no longer encoded by the organelle genome in *Chlamydomonas*. In addition, the ND7 and ND9 encoding genes present in the mitochondrial genome of higher plants are also missing in the mitochondrial genome of the alga. The question of whether these four genes were present in the nuclear genome of *Chlamydomonas* has been previously addressed [40]. The four genes were detected in the nucleic acid databases, and, with the exception of *ND4L*, the corresponding subunits were also identified at the protein level. It was proposed that ND3 and ND4L subunits, which classically possess three transmembrane domains [34], are too hydrophobic to be imported across the two membranes surrounding mitochondria [41]. Hydropathy profile analysis of *C. reinhardtii* ND3 and ND4L protein sequences also predicted three transmembrane stretches but with a slightly reduced hydrophobicity (data not shown). This satisfies one of the criteria necessary for a protein whose gene has been translocated from the mitochondria to the nucleus. The presence of putative mitochondria targeting presequences is another criterion. When compared to ND3, ND4L, ND7 and ND9 sequences from other organisms, all the *Chlamydomonas* proteins exhibit long N-terminal extensions (up to 120 amino acids) with putative targeting sequences (data not shown).

4.3. Thirty-one widely conserved subunits exist within eukaryotic complex I

As already pointed out by Heazlewood et al. [13], the achievement of many genome projects has considerably increased the number of complex I conserved subunits. In 1998, a total of 23 components common to the bovine and *Neurospora* enzymes were described [42]. In 2002–2003, data from several studies revealed 23 to 27 conserved components between the eukaryotic kingdoms [3,13,36]. Here we have now extended the number of complex I conserved proteins to 31 (Table 2). Besides the 14 highly conserved subunits homologous to the prokaryotic NADH dehydrogenase enzyme components, all the species examined (mainly the mammal *B. taurus*, the fungus *N. crassa*, the worm *C. elegans*, the higher plant *A. thaliana* and the green alga *C. reinhardtii*) clearly contain 17 additional conserved components. An 18th subunit could be widely conserved: present in mammals, fungi and algae, B14.7 homologue is not clearly identified in the worm *C. elegans* and in higher plants.

Generally considered as accessory proteins, since they do not have counterparts in the bacterial enzyme, the role of these subunits is still largely unknown. It was speculated that at least some of them form a frame or a stabilizer to keep the redox groups in the right position so as to prevent the electrons from escaping and forming reactive oxygen species [43]. In the bovine complex I, 11 of these proteins belong to the Iα subcomplex (MWFE, B8, 13 kDa-A, B13, B14, B14.7, B16.6, B17.2, PFFD, PGIV, AQDQ, 39 kDa), five subunits are part of the Iβ subcomplex (ESSS, B12, PDSW, B18, B22) and the acyl-carrier SDAP protein is associated to both the subcomplexes, which basically correspond to peripheral and membrane parts of the enzyme [1,2]. During the last decade, efforts were made to determine the impact of subunits SDAP, 39 kDa, AQDQ, MWFE, B13, PGIV, PDSW on the enzyme activity or assembly, either in mammals or in *N. crassa* (see discussion by Heazlewood et al. [13]).

The B16.6 and B14.7 subunits were identified only recently as complex I components in mammals [1,38]. B16.6 is known as GRIM-19 protein, the product of a cell death regulatory gene induced by interferon-b and retinoic acid [38]. Its presumed presence in all eukaryotic cells provides a new key to investigate a general link between mitochondrial electron transport chain and cell death.

Bovine B14.7 [1], *Neurospora* 21.3b [44] and *Chlamydomonas* 23-kDa proteins share 15–20% identity and similar hydropathy profiles with the TIM22 and TIM23

complex components involved in protein translocation across the inner mitochondrial membrane. A similar import function of complex I subunits that have no cleavable mitochondrial import sequence has been suggested for bovine protein B14.7 [1]. On the other hand, disruption of the 21.3b corresponding gene in *N. crassa* results in the absence of complex I assembly and in the formation of two major subcomplexes, corresponding to the membrane and matrix domains [44]. These protein homologues might thus play a central role in the eukaryotic complex I assembly pathway.

To our knowledge, there are no deletion phenotypes or roles attributed to ESSS, B8, B12, B14, B17.2, B18, B22, 13 kDa-A and PFFD proteins. B14 and B22 subunits have been placed in the same protein family (pfam05347) on the basis of a conserved tripeptide (LYR or LYK) located near to the N-terminus of these proteins. However, we failed to obtain a significant consensus from Clustal W alignment between the B14 and B22 sequences. Moreover, the so-called conserved motif is degenerated in many species (MYR in the *Chlamydomonas* B22 protein, LFK in the mice BAB24512 B14 homologue, FFR in the higher plants B14 polypeptides), which perhaps makes irrelevant the classification of B14 and B22 in the same protein family.

4.4. The importance of a four regularly spaced cysteine residue motif

In the PGIV bovine protein and in its homologue in *N. crassa* (20.8 kDa), eight cysteine residues separated from each other by an average of nine amino acids were suggested to be involved in disulfide bridges or to provide ligands for binding Fe–S clusters [42]. The eight Cys residues are also present in higher plant PGIV homologue. However, Cys residues 1, 4, 5, 8 and 2, 3 are absent in the *C. reinhardtii* and *C. elegans* (NP_497574) homologues, respectively, which reduces to cysteines 6 and 7 the conserved residues (data not shown).

Looking for cysteine residues in other widely conserved accessory subunits revealed us that PSDW family members also possess two conserved cysteine residues 11 amino acids apart (data not shown). In PFFD (Fig. 4) and B18 (data not shown) protein families, a motif of four regularly spaced cysteine residues was also identified ($C\text{-}X_{(9)}\text{-}C\text{-}X_{(9\text{-}15)}\text{-}C\text{-}X_{(9)}\text{-}C$) which aligns with four cysteines present in the two complex IV chaperones COX17 and COX19. COX17 is involved in copper import into mitochondria and replacement of its C-terminal cysteine residue in yeast leads to the absence of the protein and of complex IV assembly [45]. On the basis of the four conserved cysteine residues, COX19 is also suggested to perform a similar function [46]. Although most of the cysteine residues are probably involved in formation of disulfide bridges, it can be hypothesized that PFFD, B18, PDSW or PGIV subunits play a role in Fe–S cluster binding, or even in metal transport to mitochondria.

4.5. Non-conserved complex I subunits: lineage-specific components?

While the number of conserved subunits in eukaryotes has increased, the number of lineage-specific complex I subunits has almost remained constant because new subunits continue to be described [1,13]. Eleven proteins are specific to mammals, three of which having no identified counterparts in the other animal species, three are specific to fungi and at least six are typical of the photosynthetic eukaryotes, including *C. reinhardtii* (Table 2). There are also two components conserved between fungi and animals which do not have counterparts in photosynthetic eukaryotes, and finally there is one subunit common to fungi and plants which is not found in animals (Table 2). In this context, the composition of *Chlamydomonas* complex I is very similar to that of the higher plant enzyme. Nevertheless, the degree of conservation of most of these subunits within their class is rather low, and in the absence of comprehensive studies on complex I from various species, it is hard to know whether some of these proteins represent weakly conserved orthologues, structural counterparts, or whether they are real lineage-specific components that would have been acquired or lost during the evolution of specific groups.

Based on current knowledge, it is tempting to speculate that a major part of these components are not required for electron transport in complex I but rather perform other functions, such as enzyme stability, or function against reactive oxygen species. The presence of ferripyochelin-binding protein-like subunits in photosynthetic eukaryotes is quite intriguing. Their putative function as sensor of Fe in the mitochondrion or their ability to bind and store iron for later assembly of Fe–S clusters [13,47] awaits further investigation. Interestingly, the NUO-20.9 mutant of *Neurospora* fails to assemble a complete membrane arm while the matricial arm is present [48]. This subunit is also the only one that has no animal counterpart while it is present in photosynthetic eukaryote complex I, pointing out the central role of a non-conserved ‘‘accessory’’ subunit in the assembly of the membrane arm the in non-mammalian enzymes.

Acknowledgements

We are grateful to Dr. H. Millar (University of Western Australia) for critical comments to the manuscript. We thank Dr. J. Hirst (Cambridge University, UK) for the generous gift of the antibody against SDAP bovine subunit. This research was supported by grants from the Belgian Fonds National de la Recherche Scientifique (2.4.587.04.F). P.C. is a Research Fellow of the F.N.R.S. and F.V. is a grantee from the ‘Bijzonder Onderzoekfonds’ of Ghent University. J.V.B. and B.D. are indebted to the Fund for Scientific Research-Flanders for project G.0312.02, providing the mass spectrometers. We also thank Dr. D. Baurain for helpful hints on

BLAST facilities and E. Lecocq and E. Schmetz for technical assistance.

References

[1] J. Carroll, R.J. Shannon, I.M. Fearnley, J.E. Walker, J. Hirst, Definition of the nuclear encoded protein composition of bovine heart mitochondrial complex I Identification two new subunits, J. Biol. Chem. 277 (2002) 50311–50317.

[2] J. Hirst, J. Carroll, I.M. Fearnley, R.J. Shannon, J.E. Walker, The nuclear encoded subunits of complex I from bovine heart mitochondria, Biochim. Biophys. Acta 1604 (2003) 135–150.

[3] A. Videira, M. Duarte, From NADH to ubiquinone in *Neurospora* mitochondria, Biochim. Biophys. Acta 1555 (2002) 187–191.

[4] T. Friedrich, The NADH:ubiquinone oxidoreductase (complex I) from *Escherichia coli*, Biochim. Biophys. Acta 1364 (1998) 134–146.

[5] A. Dupuis, M. Chevallet, E. Darrouzet, H. Duborjal, J. Lunardi, J.P. Issartel, The complex I from *Rhodobacter capsulatus*, Biochim. Biophys. Acta 1364 (1998) 147–165.

[6] T. Ohnishi, Iron–sulfur clusters/semiquinones in complex I, Biochim. Biophys. Acta 1364 (1998) 186–206.

[7] M. Chevallet, A. Dupuis, J.P. Issartel, J. Lunardi, R. van Belzen, S.P. Albracht, Two EPR-detectable [4Fe–4S] clusters, N2a and N2b, are bound to the NuoI (TYKY) subunit of NADH: ubiquinone oxidoreductase (Complex I) from *Rhodobacter capsulatus*, Biochim. Biophys. Acta 1557 (2003) 51–66.

[8] V. Guenebaut, R. Vincentelli, D. Mills, H. Weiss, K.R. Leonard, Three-dimensional structure of NADH-dehydrogenase from *Neurospora crassa* by electron microscopy and conical tilt reconstruction, J. Mol. Biol. 265 (1997) 409–418.

[9] G. Hofhaus, H. Weiss, K. Leonard, Electron microscopic analysis of the peripheral and membrane parts of mitochondrial NADH dehydrogenase (complex I), J. Mol. Biol. 221 (1991) 1027–1043.

[10] B. Combettes, J.M. Grienenberger, Analysis of wheat mitochondrial complex I purified by a one-step immunoaffinity chromatography, Biochimie 81 (1999) 645–653.

[11] A.G. Rasmusson, J. Mendel-Hartvig, I.M. Moller, J.T. Wiskich, Isolation of the rotenone-sensitive NADH-ubiquinone reductase (complex I) from red beet mitochondria, Physiol. Plant. 90 (1994) 607–615.

[12] S. Leterme, M. Boutry, Purification and preliminary characterization of mitochondrial complex I (NADH: ubiquinone reductase) from broad bean (*Vicia faba* L.), Plant Physiol. 102 (1993) 435–443.

[13] J.L. Heazlewood, K.A. Howell, A.H. Millar, Mitochondrial complex I from *Arabidopsis* and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits, Biochim. Biophys. Acta 1604 (2003) 159–169.

[14] F. Duby, P. Cardol, R.F. Matagne, C. Remacle, Structure of the telomeric ends of mt DNA, transcriptional analysis and complex I assembly in the dum24 mitochondrial mutant of *Chlamydomonas reinhardtii*, Mol. Genet. Genomics 266 (2001) 109–114.

[15] P. Cardol, R.F. Matagne, C. Remacle, Impact of mutations affecting ND mitochondria-encoded subunits on the activity and assembly of complex I in *Chlamydomonas*. Implication for the structural organization of the enzyme, J. Mol. Biol. 319 (2002) 1211–1221.

[16] R. van Lis, A. Atteia, G. Mendoza-Hernandez, D. Gonzalez-Halphen, Identification of novel mitochondrial protein components of *Chlamydomonas* reinhardtii. A proteomic approach, Plant Physiol. 132 (2003) 318–330.

[17] C. Vahrenholz, G. Riemen, E. Pratje, B. Dujon, G. Michaelis, Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication, Curr. Genet. 24 (1993) 241–247.

[18] J. Shrager, C. Hauser, C.W. Chang, E.H. Harris, J. Davies, J. McDermott, R. Tamse, Z. Zhang, A.R. Grossman, *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information, Plant Physiol. 131 (2003) 401–408.

[19] H. Schägger, G. Von Jagow, Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form, Anal. Biochem. 199 (1991) 223–231.

[20] B. Devreese, F. Vanrobaeys, J. Smet, J. Van Beeumen, R. Van Coster, Mass spectrometric identification of mitochondrial oxidative phosphorylation subunits separated by two-dimensional blue-native polyacrylamide gel electrophoresis, Electrophoresis 23 (2002) 2525–2533.

[21] U. Klein, C. Chen, M. Gibbs, M.K.A. Platt-Aloia, Cellular fractionation of *Chlamydomonas reinhardtii* with emphasis on the isolation of the chloroplast, Plant Physiol. 72 (1983) 481–487.

[22] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.

[23] D.R. Kuonen, P.J. Roberts, I.R. Cottingham, Purification and analysis of mitochondrial membrane proteins on nondenaturing gradient polyacrylamide gels, Anal. Biochem. 153 (1986) 221–226.

[24] I.R. Cottingham, A.L. Moore, Analysis of NADH dehydrogenases from plant [mung bean (*Phaseolus aureus*)] mitochondrial membranes on non-denaturing polyacrylamide gels and purification of complex I by band excision, Biochem. J. 254 (1988) 303–305.

[25] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.

[26] F. Vanrobaeys, B. Devreese, E. Lecocq, L. Rychlewski, L. De Smet, J. Van Beeumen, Proteomics of the dissimilatory iron-reducing bacterium *Shewanella oneidensis* MR-1, using a matrix-assisted laser desorption/ionization-tandem-time of flight mass spectrometer, Proteomics 3 (2003) 2249–2257.

[27] H. Bannai, Y. Tamada, O. Maruyama, K. Nakai, S. Miyano, Extensive feature detection of N-terminal protein sorting signals, Bioinformatics 18 (2002) 298–305.

[28] C. Guda, E. Fahy, S. Subramaniam, MITOPRED: a genome-scale method for prediction of nuclear-encoded mitochondrial proteins, Bioinformatics (2004) (in press).

[29] 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.

[30] B. Bjellqvist, G.J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J.C. Sanchez, S. Frutiger, D. Hochstrasser, The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences, Electrophoresis 14 (1993) 1023–1031.

[31] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.

[32] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[33] M. Eriksson, P. Gardestrom, G. Samuelsson, Isolation, purification and characterization of mitochondria from *Chlamydomonas reinhardtii*, Plant Physiol. 107 (1995) 479–483.

[34] J.E. Walker, The NADH: ubiquinone oxidoreductase (complex I) of respiratory chains, Q. Rev. Biophys. 25 (1992) 253–324.

[35] T. Friedrich, A. Abelmann, B. Brors, V. Guenebaut, L. Kintscher, K. Leonard, T. Rasmussen, D. Scheide, A. Schlitt, U. Schulte, H. Weiss, Redox components and structure of the respiratory NADH: ubiquinone oxidoreductase (complex I), Biochim. Biophys. Acta 1365 (1998) 215–219.

[36] W.Y. Tsang, B.D. Lemire, The role of mitochondria in the life of the nematode, *Caenorhabditis elegans*, Biochim. Biophys. Acta 1638 (2003) 91–105.

[37] H. Heinrich, J.E. Azevedo, S. Werner, Characterization of the 9.5-kDa ubiquinone-binding protein of NADH:ubiquinone oxidoreductase (complex I) from *Neurospora crassa*, Biochemistry 31 (1992) 11420–11424.

[38] I.M. Fearnley, J. Carroll, R.J. Shannon, M.J. Runswick, J.E. Walker, J. Hirst, GRIM-19, a cell death regulatory gene product, is a subunit

of bovine mitochondrial NADH:ubiquinone oxidoreductase (complex I), J. Biol. Chem. 276 (2001) 38345–38348.
[39] A.M. Nedelcu, R.W. Lee, Modes and tempos of mitochondrial and chloroplast genome evolution in *Chlamydomonas*: a comparative analysis, in: J.D. Rochaix, M. Goldschmidt-Clermont, S. Merchant (Eds.), The Molecular Biology of Chloroplast and Mitochondria in *Chlamydomonas*, Kluwer Academic Publishing, The Netherlands, 1998, pp. 63–91.
[40] X. Perez-Martinez, M. Vazquez-Acevedo, E. Tolkunova, S. Funes, M.G. Claros, E. Davidson, M.P. King, D. Gonzalez-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.
[41] H.L. Race, R.G. Herrmann, W. Martin, Why have organelles retained genomes? Trends Genet. 15 (1999) 364–370.
[42] A. Videira, Complex I from the fungus *Neurospora crassa*, Biochim. Biophys. Acta 1364 (1998) 89–100.
[43] T. Friedrich, H. Weiss, in: H. Baltscheffsky (Ed.), Origine and Evolution of the Proton-Pumping NADH:ubiquinone Oxidoreductase (Complex I) in Origin and Evolution of Biological Conversion, VCH, Stockholm, 1996, pp. 205–220.
[44] U. Nehls, T. Friedrich, A. Schmiede, T. Ohnishi, H. Weiss, Characterization of assembly intermediates of NADH: ubiquinone oxidoreductase (complex I) accumulated in *Neurospora* mitochondria by gene disruption, J. Mol. Biol. 227 (1992) 1032–1042.
[45] F.A. Punter, D.M. Glerum, Mutagenesis reveals a specific role for Cox17p in copper transport to cytochrome oxidase, J Biol Chem. 278 (2003) 30875–30880.
[46] M.P. Nobrega, S.C. Bandeira, J. Beers, A. Tzagoloff, Characterization of COX19, a widely distributed gene required for expression of mitochondrial cytochrome oxidase, J. Biol. Chem. 277 (2002) 40206–40211.
[47] R. Lill, G. Kispal, Maturation of cellular Fe–S proteins: an essential function of mitochondria, Trends Biochem. Sci. 25 (2000) 352–356.
[48] U. Schulte, H. Weiss, Generation and characterization of NADH:ubiquinone oxidoreductase mutants in *Neurospora crassa*, Methods Enzymol. 260 (1995) 3–14.