

University of Groningen

## The atypical subunit composition of respiratory complexes I and IV is associated with original extra structural domains in *Euglena gracilis*

Miranda-Astudillo, H. V.; Yadav, K. N. S.; Colina-Tenorio, L.; Bouillenne, F.; Degand, H.; Morsomme, P.; Boekema, E. J.; Cardol, P.

*Published in:*  
Scientific Reports

*DOI:*  
[10.1038/s41598-018-28039-z](https://doi.org/10.1038/s41598-018-28039-z)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Miranda-Astudillo, H. V., Yadav, K. N. S., Colina-Tenorio, L., Bouillenne, F., Degand, H., Morsomme, P., Boekema, E. J., & Cardol, P. (2018). The atypical subunit composition of respiratory complexes I and IV is associated with original extra structural domains in *Euglena gracilis*. *Scientific Reports*, *8*(1), [9698]. <https://doi.org/10.1038/s41598-018-28039-z>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# SCIENTIFIC REPORTS



OPEN

## The atypical subunit composition of respiratory complexes I and IV is associated with original extra structural domains in *Euglena gracilis*

H. V. Miranda-Astudillo<sup>1</sup>, K. N. S. Yadav<sup>2</sup>, L. Colina-Tenorio<sup>3</sup>, F. Bouillenne<sup>4</sup>, H. Degand<sup>5</sup>, P. Morsomme<sup>5</sup>, E. J. Boekema<sup>2</sup> & P. Cardol<sup>1</sup> 

In mitochondrial oxidative phosphorylation, electron transfer from NADH or succinate to oxygen by a series of large protein complexes in the inner mitochondrial membrane (complexes I–IV) is coupled to the generation of an electrochemical proton gradient, the energy of which is utilized by complex V to generate ATP. In *Euglena gracilis*, a non-parasitic secondary green alga related to trypanosomes, these respiratory complexes totalize more than 40 *Euglenozoa*-specific subunits along with about 50 classical subunits described in other eukaryotes. In the present study the *Euglena* proton-pumping complexes I, III, and IV were purified from isolated mitochondria by a two-steps liquid chromatography approach. Their atypical subunit composition was further resolved and confirmed using a three-steps PAGE analysis coupled to mass spectrometry identification of peptides. The purified complexes were also observed by electron microscopy followed by single-particle analysis. Even if the overall structures of the three oxidases are similar to the structure of canonical enzymes (e.g. from mammals), additional atypical domains were observed in complexes I and IV: an extra domain located at the tip of the peripheral arm of complex I and a “helmet-like” domain on the top of the cytochrome c binding region in complex IV.

Mitochondria generate most of the energy in eukaryotic cells via oxidative phosphorylation (OXPHOS). This process can be separated in two parts: (i) the respiratory chain classically comprising four membrane protein complexes (complexes I to IV) and two mobile electron carriers, ubiquinone and cytochrome c, which together transfer electrons from reduced NADH or succinate to oxygen and establish an electrochemical proton gradient across the inner mitochondrial membrane (proton-motive force); (ii) the ATP synthase (complex V) which works like a molecular motor that utilizes the energy of the proton-motive force across the membrane to synthesize ATP from ADP and inorganic phosphate.

The NADH:ubiquinone oxidoreductase (complex I) is one of the largest membrane protein assemblies known to date and has a central role in energy production. With an apparent molecular mass of ca. 1000 kDa, it comprises 45 subunits in mammals<sup>1</sup>, and more than 40 components in ascomycete fungi and green plants<sup>2</sup>. The bacterial complex I is made up of 14 different subunits, also present in eukaryotes. This so-called “core” complex therefore represents the minimal structural form of complex I and its subunits sum up a molecular mass of approximately 530 kDa<sup>3</sup>. The solved structures of this complex show a characteristic “L-shape” which comprises two main arms, one membrane domain in charge of the proton pumping activity and a peripheral arm involved in the electron

<sup>1</sup>Laboratoire de Génétique et Physiologie des microalgues, InBioS/Phytosystems, Institut de Botanique, Université de Liège, Liège, Belgium. <sup>2</sup>Department of Electron Microscopy, Groningen Biological Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands. <sup>3</sup>Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico, Mexico. <sup>4</sup>InBioS/Center for Protein Engineering, Université de Liège, Liège, Belgium. <sup>5</sup>Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. Correspondence and requests for materials should be addressed to P.C. (email: [pierre.cardol@uliege.be](mailto:pierre.cardol@uliege.be))

transfer from NADH to ubiquinone<sup>4–6</sup>. The eukaryotic dimeric cytochrome *bc1* complex (complex III) is composed by ten different protein subunits per monomer. Three of these subunits (cytochrome *b*, cytochrome *c1* and the Rieske iron–sulfur protein) contain redox centers and participate in electron transfer from reduced quinone to cytochrome *c*. The seven supernumerary subunits are not embedded in the membrane, they are peripherally localized at the membrane surfaces and are not directly related with the catalytic activity of the complex in the electron transport chain<sup>7–9</sup>. The mammalian cytochrome *c* oxidase (complex IV) is a multimeric enzyme formed by 14 subunits. The core subunits (Cox1, Cox2, and Cox3) are highly hydrophobic integral membrane proteins without substantial extramembrane domains. Their structure is highly-conserved in  $\alpha$ -proteobacteria and eukaryotes. The remaining 11 mammalian accessory subunits are located surrounding the catalytic core<sup>10,11</sup>. This complex exists as a 420 kDa dimer in its crystalline state, although the fully active monomeric form can be also isolated<sup>12–14</sup>.

Euglenids are a group of free-living, single-celled flagellates that thrive predominantly in aquatic environments. Euglenids belong to the monophyletic group of Euglenozoa together with other heterotrophic flagellates like Symbiontida (free-living flagellates found in low-oxygen marine sediments), Diplonemea (free-living marine flagellates) and Kinetoplastea (free-living and parasitic flagellates, e.g. *Trypanosoma*)<sup>15,16</sup>. *Euglena gracilis*, a model organism among Euglenids, is a secondary photosynthetic unicellular eukaryote that arises from an endosymbiosis between a green alga and an ancient phagotroph euglenozoan species<sup>17,18</sup>. This flagellate has a remarkable adaptability to various environmental conditions. It can grow photoautotrophically, heterotrophically, and photoheterotrophically<sup>19</sup> and is able to metabolize various exogenous carbon sources such as sugars, alcohols and amino acids<sup>20</sup>. *E. gracilis* has a mitochondrial electron transfer system constituted by the conserved OXPHOS complexes (Complexes I–V) and an alternative oxidase (AOX) sensitive to diphenylamine, salicylhydroxamic acid (SHAM), *n*-propylgallate and disulfiram<sup>21–23</sup>.

The subunit composition of the OXPHOS complexes among the Euglenozoa includes additional and atypical subunits together with some conserved classical ones<sup>24–27</sup>. This divergent subunit composition derives notably in an atypical structure in the dimeric ATP synthase from euglenids<sup>28,29</sup>. Additionally, the *Euglena* complexes I and V present higher estimated molecular masses (1.5 and 2 MDa, respectively) when compared with other eukaryotic complexes<sup>27</sup>. The consequences of atypical subunits or structural domains in the function of the respiratory-chain pathway is still enigmatic and only hypotheses have been formulated so far, such as a role in the response of *E. gracilis* to stress<sup>30</sup>, resistance to inhibitors<sup>27,31,32</sup>, or any specific role in the parasitic lifestyle of trypanosomes<sup>33</sup>. Regarding the last point, our group has shown that many additional subunits originally described in kinetoplastids<sup>33–37</sup> are shared with *E. gracilis*. 19 of the 45 conventional subunits of the eukaryotic complex I are found in *E. gracilis* (ND1/4/5, NDUFS1–3/7/8, NDUFV1/2, NDUFA7/9/12/13, NDUFAB1, NDUFB7/11, NDUFC1, and CAG1/2) together with 14 trypanosomatid subunits (NDTB1/2/5/6/11/12/17/18/22/25/28/29/31/34). Amongst the 10 classical complex III subunits, eight were found in *E. gracilis* (QCR1/2/6/7/10, RIP1, COB, CYT1) together with two trypanosomatid subunits (QCRTB1/2). From the 13 mammalian CIV subunits, only seven subunits (COX1/2/3/5 A/5B/6B/8 A) could be identified in *E. gracilis* along with nine kinetoplastid subunits (COXTB1/2/4/5/6/8/10/12/16)<sup>27</sup>. So it is unlikely that these subunits play a specific role in the parasitic lifestyle but rather play a fundamental role in Euglenozoa.

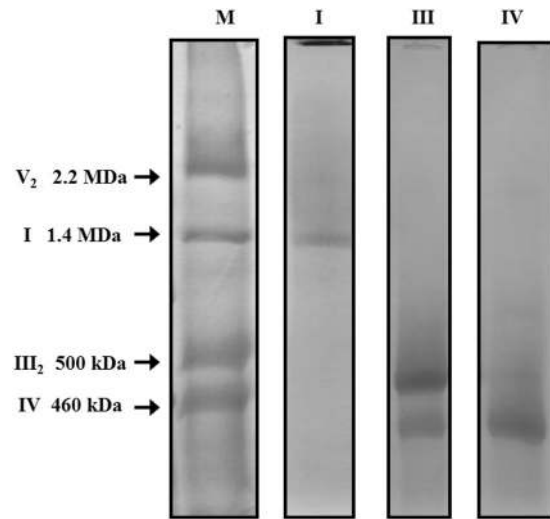
In the present work, we studied the unusual subunit composition of proton-pumping respiratory complexes (*i.e.* complexes I, III, and IV) from *Euglena gracilis* in more detail and investigated the consequences of these atypical subunit compositions on the structural level by single-particle electron microscopy.

## Results and Discussion

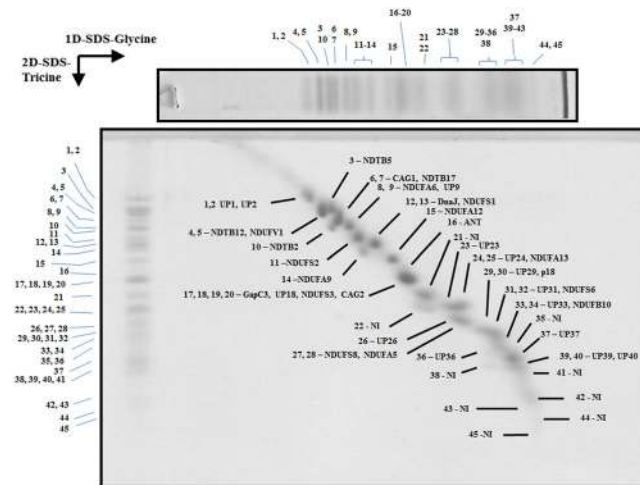
**Purification of the mitochondrial complexes I, III and IV from *E. gracilis*.** In a previous study we reported that the respiratory complexes of *E. gracilis* comprise many subunits which were also described in trypanosomatid respiratory complexes and, conversely, many subunits were lacking which are conserved among mammals and fungi<sup>27</sup>. In order to confirm this atypical subunit composition, we decided to further purify and characterize the subunit composition of complexes I, III, and IV. To this end, a purification protocol involving two chromatographic steps was developed (see Methods section for further details) to obtain enriched fractions of each of the three complexes. The fraction containing complex I was almost pure, as judged by the main 1.4 MDa complex observed after BN-PAGE analysis, while the fractions enriched in complex III (500 kDa) or complex IV (460 kDa) were slightly contaminated (Fig. 1 and Suppl. Information Fig. S1). It is to note the estimated molecular masses determined here are slightly different from the values previously reported<sup>27</sup>.

The enriched complexes were applied to a 1D-BN/2D-glycine-SDS/3D-tricine-SDS gel system. The 3D gels thus obtained for each complex are shown in Figs 2–4. Compared to 2D BN-SDS PAGE analysis, the 3D gels offer a better resolution of the polypeptide components. The Coomassie-stained spots were excised from the gel and analyzed by tandem mass spectrometry (MS/MS). All the obtained sequences (Suppl. Information) were then analyzed *in silico* to search for possible homologs in protein databases (NRPS, TriTrypDB). The predicted physico-chemical properties for each subunit, such as hydrophobicity (GRAVY) and iso-electric point (IP), predicted structural features like transmembrane helices (TMH) and conserved domains (CD) were also determined by *in silico* approaches (Tables 1–3 and Suppl. Information). Finally, in order to investigate how the subunit composition of these complexes affects their overall structure, electron microscopy in combination with single particle analysis was performed. A total of *ca.* 2500–3000 raw images from each complex were recorded and analysed by electron microscopy. The particle projections from each complex were classified into several groups, out of which 8 best class averages were selected per complex (an average of 3000–4000 particles per class) (Figs 5–7). The main findings for each complex will be described and discussed in the following sections.

**The atypical subunits build the extra module in the peripheral arm of complex I.** With the above described 3D gel system, a total of 31 spots were obtained for complex I (Fig. 2), compared to the 22 spots

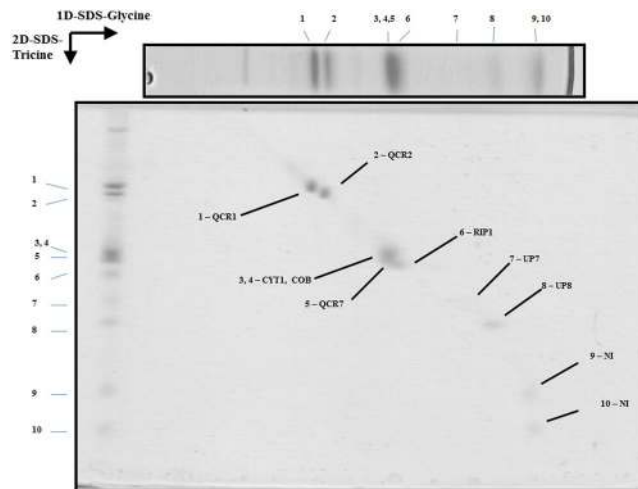


**Figure 1.** Electrophoretic patterns of purified complexes I, III and IV from *Euglena gracilis*. Isolated mitochondria were solubilized with n-dodecyl- $\beta$ -D-maltoside followed by a two-step chromatographic purification (Complete gels used are shown in Suppl Information Fig. S5). (M) Total mitochondrial proteins from *Euglena gracilis*. (I) purified NADH:ubiquinone oxidoreductase (complex I) fraction, (III) enriched cytochrome bc1 complex (complex III) fraction, remaining complex IV could be observed, (IV) enriched cytochrome c oxidase (complex IV) fraction. Estimated molecular masses and identities of Coomassie brilliant blue-stained protein complexes are indicated (see Suppl. Information Fig. S1 for more details).



**Figure 2.** 3D resolution of the polypeptides that constitute the *Euglena* mitochondrial NADH:ubiquinone oxidoreductase. Monomeric purified complex I was resolved by BN-PAGE. The band of interest was excised and resolved in a glycine-SDS-PAGE (12% acrylamide). The 2D gel was then subjected to 3D tricine-SDS-PAGE (14% acrylamide) and stained with Coomassie brilliant blue. The identified subunits are labeled. NI, not identified; UP, unnamed protein with no homolog in other species. Details about subunit identification are given in Table 1.

obtained in our previous study by 2D BN/SDS-PAGE analysis<sup>27</sup>. Ten of these spots matched for a single polypeptide, eleven matched for two peptides co-migrating, one matched for four proteins and nine spots did not lead to identification in our database. Globally, this analysis revealed the presence of at least 45 polypeptides associated to complex I, with molecular masses ranging from 7.6 to 58.5 kDa, and 36 of them were identified by MS analysis (Table 1). Six polypeptides correspond to complex I core subunits (NDUFS1/2/3/6/8, NDUFV1), eight to the  $\alpha$ -proteobacterial ancestor (NDUFA5/6/9/12/13, NDUFB10, CAG1/2), four are Euglenozoa-specific subunits (NDTB2/5/12/17), and fourteen proteins do not present homologs and thus remain as unnamed proteins (UP). This analysis also corroborated the association of two short domains related to DNAJ and the glyceraldehyde 3-phosphate dehydrogenase (GapC3) enzymes [erroneously annotated as glycerol-3-phosphate dehydrogenase (G3PD) in our previous study<sup>27</sup>]. The ANT transporter and the Euglenozoa-specific ATP synthase subunit p18 were also detected bound to complex I (Table 1, subunits 16 and 30, respectively). The ANT is rather abundant in the mitochondrial inner membrane (12% of total protein)<sup>38</sup>, and therefore we cannot exclude contamination due



**Figure 3.** 3D resolution of the polypeptides that constitute the *Euglena* mitochondrial cytochrome bc1 complex. Dimeric purified complex III was resolved by BN-PAGE. The band of interest was excised and resolved in a glycine-SDS-PAGE (12% acrylamide). The 2D gel was then subjected to 3D tricine-SDS-PAGE (14% acrylamide) and stained with Coomassie brilliant blue. The identified subunits are labeled. NI, not identified; UP, unnamed protein with no homolog in other species. Details about subunit identification are given in Table 2.

to hydrophobic interactions. The euglenozoan p18 subunit has been reported to be attached to the ATP synthase complex in many euglenozoan species<sup>25,27,28,39–42</sup> and, recently, its position in the complex attached externally to the  $F_1$  sector was determined<sup>43</sup>. The reason why subunit p18 is attached to complex I is unclear. The NDTB5 subunit (subunit 3, Table 1) was previously described as a paralogous of NDUFA9 subunit<sup>27</sup>. The euglenoid subunit presents a molecular mass higher (52.2 kDa) than the conserved subunit (38.7 kDa) (subunit 14, Table 1). This homology-pair characteristic is also present in the trypanomatid enzyme<sup>44</sup>, and thus suggests that euglenoids incorporated this new subunit by gene duplication.

Overall, these first results are similar to our previous study (Suppl. Information) and thus confirm the atypical subunit composition of *Euglena gracilis* complex I. The sum of the molecular masses from all the identified peptides (Table 1) is 1333 kDa, which is close to the value of 1.4 MDa estimated directly by BN-PAGE (Figs 1 and Suppl. Information S1). Additional subunits might also be present in the mature complex but have probably escaped from identification (see<sup>27</sup> for a discussion about the underlying reasons). Among those subunits not identified by mass spectrometry but previously identified at genomic level for *Euglena* complex<sup>27</sup>, there are subunits present in canonical complex I of mammals, fungi and plants, like highly hydrophobic core mitochondrion-encoded (NAD1, NAD4, NAD6) or nucleus-encoded (NDUFV2, NDUFA7, NDUFAB1, and NDUFB7) subunits (Suppl. Information), all inherited from the alpha-proteobacterial ancestor<sup>45</sup>. This value of 1.4 MDa for *Euglena* complex I contrasts with the *ca.* 1 MDa value described in many other organisms including mammals<sup>46,47</sup>, yeasts<sup>48,49</sup>, green algae<sup>50</sup> or flowering plants<sup>51</sup>, and prompted us to investigate the structure of *Euglena* complex I.

Negatively stained complex I adopts mainly three positions on the carbon support film (Fig. 5). The first four side-view projections (panels A–D) show that complex I has an “L shape” conformation, which is in agreement with the bacterial and mitochondrial complex I structures [e.g.<sup>6,52</sup>]. The membrane arm is best seen in projections A and B (green arrow heads). The other projections correspond to the upper view (panel E) and to tilted side views (panels F–H). The comparison with the structure of bovine respiratory complex I (pdb: 5LDW<sup>52</sup>) revealed two main differences: (i) a matrix-exposed protuberance attached to the membrane arm at a central position (visible in projections B and C, purple arrow heads) and (ii) an extra domain located at the tip of the peripheral arm (panels A–D and F–H, red arrow heads). A third putative extra region is located in the distal part of the hydrophobic domain (Panels E and F, orange arrow heads). However, we cannot exclude that this density is due to the presence of detergent micelles surrounding the membrane arm. All of these features are also highlighted upon comparison with the yeast Complex I (pdf: 4WZ7<sup>53</sup>) (Suppl. Information Fig. S2). The first domain (purple arrow heads) probably corresponds to the gamma-carbonic anhydrase (CAG) domain, first described in complex I from flowering plant mitochondria<sup>54</sup> and later described in other organisms like the chlorophycean alga *Polytomella* sp.<sup>55</sup> or the amoeboid protozoon *Acanthamoeba castellanii*<sup>56</sup>. Members of the CAG family have been identified in all eukaryotic lineages but not in Opisthokonts (*i.e.* mammals and fungi)<sup>56,57</sup>. In this respect, two members (CAG1–2) have been found by proteomic approach in *Euglena* complex I (Fig. 2 and Table 1). In contrast, the extra domain located at the tip of the peripheral arm (Fig. 5, red arrow heads) has no counterpart in the complex I structures described in representatives of other lineages. Interestingly, based on the presence of some additional subunits in trypanosomes, a fatty acid synthase (FAS) domain has been proposed as an additional complex I structural domain. Among the non-canonical subunits identified in trypanosomes<sup>35,44</sup>, NDTB2 and NDTB17 were also confirmed here as components of *Euglena* complex I. Neither NDTB2 nor NDTB17 subunits possess any putative transmembrane helices but comprise a functional domain related to fatty acid

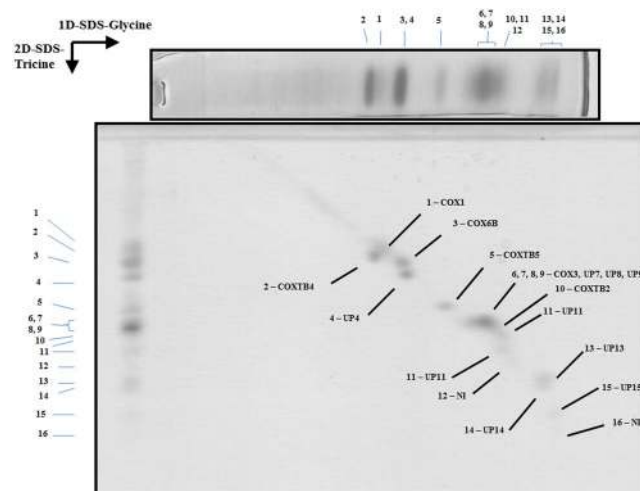


I	Name	Accession number	MW	MW (calc)	TMH (putative)	Conserved Domain
1	unknown	comp60085_c0_seq. 3	58.5	44.2	0	— <sup>a,b</sup>
2	unknown	comp54683_c0_seq. 1	58.5	18.1	0	— <sup>a,b</sup>
3	NDTB5	comp54702_c0_seq. 3	52.2	54.9	1 (C)	NDUFA9_like_SDR_a (M) <sup>a,b</sup>
4	NDTB12	>gn Egra Contig2023_c	51.0	51.8	0	— <sup>a,b</sup>
5	NDUFV1	comp62912_c0_seq. 6	51.0	57.8	0	NADH dehydrogenase I subunit F (A) <sup>a,b</sup>
6	CAG1	comp62122_c0_seq. 3	48.5	48.7	0	LbH_gamma_CA_like (M) <sup>a,b</sup>
7	NDTB17	comp63840_c0_seq. 1	48.5	53.3	0	Adenylate forming domain, Class I (C) <sup>a,b</sup>
8	NDUFA6	comp61975_c0_seq. 5	45.7	48.7	0	— <sup>a,b</sup>
9	unknown	comp46408_c0_seq. 1	45.7	8.5	0	— <sup>a,b</sup>
10	NDTB2	comp63125_c0_seq. 1	44.8	46.4	0	2-enoyl thioester reductase (N,M) <sup>a,b</sup>
11	NDUFS2	comp55715_c0_seq. 3_c	41.9	48.6	1 (M)	NADH dehydrogenase subunit D; NADH:ubiquinone oxidoreductase 49 kD subunit 7 (A) <sup>a,b</sup>
12	DnaJ	comp60945_c0_seq. 5_c	40.0	43.6	1 (C)	DnaJ molecular chaperone homology domain(N) <sup>a,b</sup>
13	NDUFS1	comp61469_c0_seq. 1_c	40.0	43.8	0	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) (A) <sup>a,b</sup>
14	NDUFA9	comp59654_c0_seq. 4	38.7	41.3	0	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, subunit 9, 39 kDa (A) <sup>a,b</sup>
15	NDUFA12	gn Egra Contig149	35.7	23.2	0	NADH ubiquinone oxidoreductase subunit NDUFA12 <sup>a,b</sup>
16	ADP/ATP	comp62013_c0_seq. 2_cut_gi 109788168	32.4	28.3	2(C)	Mitochondrial carrier protein x 2(N,M) <sup>a,b</sup>
17	GapC3	comp52123_c0_seq. 2_gi 125990644	31.0	37.1	0	Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain (C) <sup>a,b</sup>
18	unknown	comp58177_c0_seq. 3	31.0	32.6	3(M)	— <sup>a,b</sup>
19	NDUFS3	comp62960_c0_seq. 11	31.0	32.4	0	NADH dehydrogenase subunit C (M) <sup>a,b</sup>
20	CAG2	comp48089_c0_seq. 3_c	31.0	28.8	0	Gamma carbonic anhydrase-like (M) <sup>a,b</sup>
21	NI	—	27.7	—	—	—
22	NI	—	26.1	—	—	—
23	unknown	comp47716_c0_seq. 2_c	25.5	24.0	0	— <sup>a,b</sup>
24	unknown	comp51732_c0_seq. 5_gi 125990596	25.1	22.6	0	— <sup>a,b</sup>
25	NDUFA13	comp59406_c0_seq. 3	25.1	22.7	1 (M)	GRIM-19 protein (M) <sup>a,b</sup>
26	unknown	comp52747_c0_seq. 1_c	22.9	22.0	2 (M)	— <sup>a,b</sup>
27	NDUFS8	comp54309_c0_seq. 4_c	22.4	24.3	0	NADH dehydrogenase subunit I (M) <sup>a,b</sup>
28	NDUFA5	comp49716_c0_seq. 2	22.4	20.8	0	ETC_C1_NDUFA5 (M) <sup>a,b</sup>
29	unknown	comp49825_c0_seq. 3_c	21.3	19.5	1 (C)	— <sup>a,b</sup>
30	p18	comp56597_c0_seq. 3	21.3	21.0	0	— <sup>a,b</sup>
31	unknown	comp53986_c0_seq. 1_c	20.9	18.3	0	— <sup>a,b</sup>
32	NDUFS6	comp55416_c0_seq. 4	20.9	17.1	0	— <sup>a,b</sup>
33	unknown	comp41364_c0_seq. 3	18.8	17.7	1 (C)	— <sup>a,b</sup>
34	NDUFB10	comp54566_c0_seq. 2	18.8	17.4	0	— <sup>a,b</sup>
35	NI	—	18.3	—	—	—
36	unknown	comp54442_c0_seq. 1	17.5	17.2	1 (M)	— <sup>a,b</sup>
37	unknown	comp54200_c0_seq. 1	17.1	15.6	1 (N)	— <sup>a,b</sup>
38	NI	—	16.8	—	—	—
39	unknown	comp51611_c0_seq. 1_c	16.3	14.9	1 (M)	— <sup>a,b</sup>
40	unknown	comp51117_c0_seq. 1	16.3	9.9	0	— <sup>a,b</sup>
41	NI	—	15.5	—	—	—
42	NI	—	11.7	—	—	—
43	NI	—	10.7	—	—	—
44	NI	—	8.7	—	—	—
45	NI	—	7.6	—	—	—
		Total	1333	1097		

**Table 1.** Euglena Complex I subunits identified by mass spectrometry. NI not identified. When accession number ends with \_c, the first Methionine was used for sequence analysis. (N) Amino terminus, (C) Carboxy terminus, (M) Middle region, (A) Almost all sequence. <sup>a</sup>CCD. <sup>b</sup>DELTA-BLAST. <sup>c</sup>The e-value threshold for the blast results was  $10^{-5}$ .

III	Name	Accession number	MW	MW (calc)	TMH (putative)	Conserved Domain <sup>c</sup>
1	QCR1	comp63646_c0_seq. 8_c	50.9	53.5	0	Peptidase_M16 (N), Peptidase_M16C (C) <sup>a,b</sup>
2	QCR2/TB1	comp60854_c0_seq. 1_c	47.9	51.1	0	Insulinase (Peptidase family M16) (N) <sup>a,b</sup>
3	CYT1	comp49373_c0_seq. 3_c	31.3	28.0	1 (C)	Cytochrome C1 family (A) <sup>a,b</sup>
4	COB	ALQ28773.1	31.3	43.9	9(A)	cytochrome b (A) <sup>a,b</sup>
5	QCR7	comp51517_c0_seq. 3_c	30.3	27.4	1 (N)	Ubiquinol-cytochrome C reductase complex 14kD subunit (C) <sup>a,b</sup>
6	RIP1	comp57996_c0_seq. 3_c	29.5	27.8	1 (M)	Rieske_cytochrome_bc1 (C) <sup>a,b</sup>
7	unknown	comp47102_c0_seq. 2_c	22.4	20.3	1 (C)	— <sup>a,b</sup>
8	unknown	comp57617_c0_seq. 1_c	18.8	17.6	0	— <sup>a,b</sup>
9	NI	—	14.7	—	—	—
10	NI	—	10.5	—	—	—
		Total	288	270		

**Table 2.** *Euglena* Complex III subunits identified by mass spectrometry. NI not identified. When accession number ends with \_c, the first Methionine was used for sequence analysis. (N) Amino terminus, (C) Carboxy terminus, (M) Middle region, (A) Almost all sequence. <sup>a</sup>CCD. <sup>b</sup>DELTA-BLAST. <sup>c</sup>The e-value threshold for the blast results was 10<sup>−5</sup>.



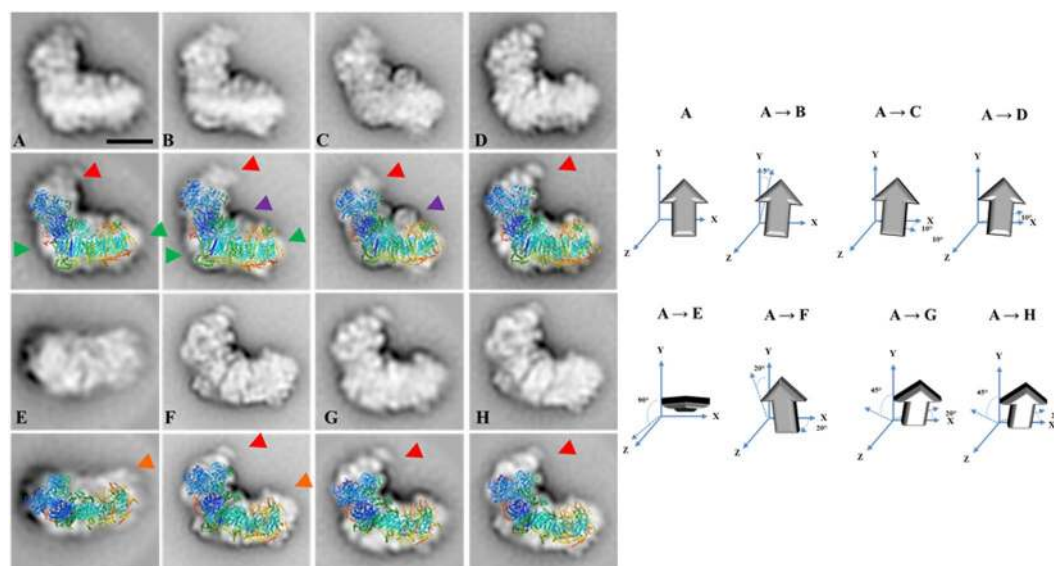
**Figure 4.** 3D resolution of the polypeptides that constitute the *Euglena* mitochondrial cytochrome c oxidase. Monomeric purified complex IV was resolved by BN-PAGE. The band of interest was excised and resolved in a glycine-SDS-PAGE (12% acrylamide). The 2D gel was then subjected to 3D tricine-SDS-PAGE (14% acrylamide) and stained with Coomassie brilliant blue. The identified subunits are labeled. NI, not identified; UP, unnamed protein with no homolog in other species. Details about subunit identification are given in Table 3.

synthesis (FAS) (Table 1). The NDTB17 subunit presents an adenylate forming domain, a domain which has been associated with the fatty acid activation necessary for their subsequent metabolism (e.g.  $\beta$ -oxidation)<sup>58,59</sup>. The NDTB2 subunit presents a 2-enoyl thioester reductase domain. This enzymatic activity has been associated with the NADPH-dependent conversion of trans-2-enoyl acyl carrier protein/coenzyme A (ACP/CoA) to acyl-(ACP/CoA) in fatty acid synthesis<sup>60</sup>. NDTB2 and NDTB17 are thus good candidates to participate in a FAS domain. In trypanosomes, it was suggested that such domain is located at the interface between the membrane and matrix arms<sup>44</sup>, but no extra density can be distinguished in this area in *Euglena* complex I. We propose that the FAS domain is therefore located at the tip of the matrix domain in Euglenozoa (Fig. 5, and Suppl. Information Fig. S2). It should be noted that a mitochondrial small acyl carrier protein NDUFAB1/ACPM is also found associated to complex I in mammals and fungi<sup>61</sup> but not in flowering plants<sup>51</sup>. This subunit has been found in *Euglena* at genome level (Suppl. Information). Altogether, these findings may indicate an ancient function related to fatty acid metabolism associated to complex I in Eukaryotes.

**Overall conserved subunit composition and structure of *Euglena* Complex III.** *Euglena* mitochondrial complex III was resolved into 10 protein spots of molecular masses ranging between 10.5 and 50.9 kDa (Fig. 3). Six corresponded to a single polypeptide in our database, one was found to comprise two different polypeptides, and two were not identified. Six of these polypeptides corresponded to canonical complex III subunits QCR1, QCR2, QCR7, CYT1, COB, and RIP1, whereas two polypeptides of about 20 kDa do not have homologs

IV	Name	Accession number	MW	MW (calc)	TMH (putative)	Conserved Domain
1	COX1	comp60817_c0_seq. 2_c	38.5	45.5	10(A)	Cytochrome C oxidase subunit I (A) <sup>a,b</sup>
2	COXTB4	comp55436_c0_seq. 3_c	36.4	35.6	0	— <sup>a,b</sup>
3	COX6B	comp54364_c0_seq. 4	35.1	33.0	0	— <sup>a,b</sup>
4	unknown	comp57506_c0_seq. 2_c	32.6	31.7	0	— <sup>a,b</sup>
5	COXTB5	gi 109781798_c	25.9	24.5	0	— <sup>a,b</sup>
6	COX3	comp54737_c0_seq. 2	22.9	27.3	3 (M, C)	Heme-copper oxidase subunit III (C) <sup>a,b</sup>
7	unknown	comp53543_c0_seq. 2	22.9	20.4	1 (M)	— <sup>a,b</sup>
8	unknown	comp51338_c0_seq. 1_c	22.9	19.8	0	— <sup>a,b</sup>
9	unknown	comp47102_c0_seq. 2_c	22.0	20.3	1 (C)	— <sup>a,b</sup>
10	COXTB2	comp54722_c0_seq. 3_c	22.0	19.7	0	— <sup>a,b</sup>
11	unknown	comp55710_c0_seq. 1_c	20.9	19.9	1 (M)	— <sup>a,b</sup>
12	NI	—	16.7	—	—	—
13	unknown	comp50120_c1_seq. 1_c	14.9	16.4	1 (N)	— <sup>a,b</sup>
14	unknown	comp48005_c0_seq. 1_c	13.6	13.1	1 (M)	— <sup>a,b</sup>
15	unknown	comp53374_c0_seq. 4_c	10.6	9.8	0	— <sup>a,b</sup>
16	NI	—	7.2	—	—	—
		Total	365	337		

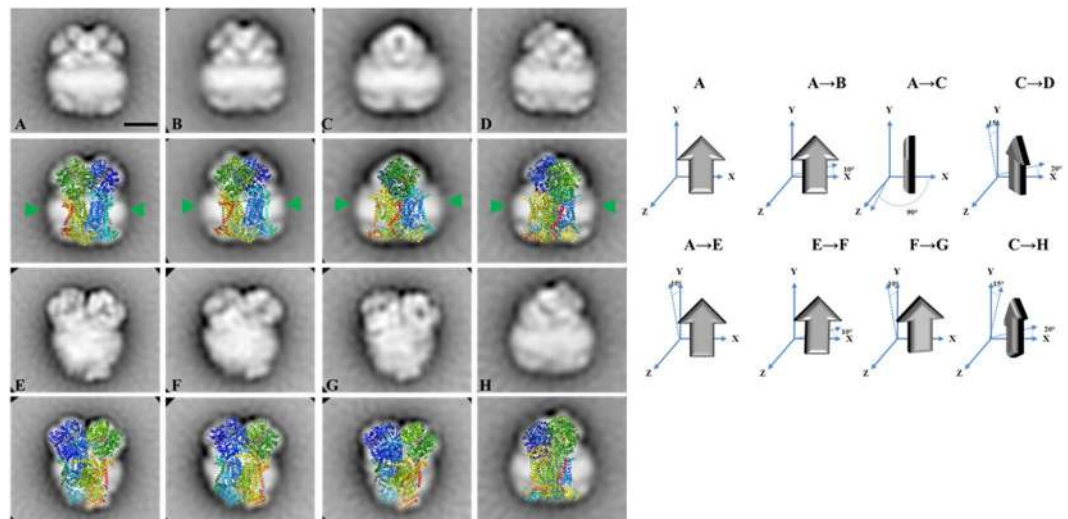
**Table 3.** *Euglena* Complex IV subunits identified by mass spectrometry. NI not identified. When accession number ends with \_c, the first Methionine was used for sequence analysis. (N) Amino terminus, (C) Carboxi terminus, (M) Middle region, (A) Almost all sequence. <sup>a</sup>CCD. <sup>b</sup>DELTA-BLAST. <sup>c</sup>The e-value threshold for the blast results was 10<sup>-5</sup>.



**Figure 5.** 2D Projection maps of purified monomeric complex I from *E. gracilis* obtained by single particle averaging. The monomeric complex I was purified by a two-step chromatographic procedure in the presence of  $\beta$ -dodecyl-*n*-maltoside and analyzed by EM. Overlap of the bovine respiratory complex I (pdb: 5LDW<sup>52</sup>) over the EM images was performed. (A–D) side view projections showing the characteristic “L shape” conformation. (E) upper view, (F–H) almost non-tilted from a side situation. Three unusual densities are observed: (i) an extra domain in the tip of the peripheral arm (red arrow heads), (ii) matrix-exposed protuberance attached to the membrane arm at a central position (purple arrow heads) and (iii) an extra region located in the extreme of the hydrophobic domain (orange arrow heads). At the right: schematic representation of the rotation of the complex in the left panels. The approximate rotation angles over the axis are indicated. The scale bar is 10 nm.

in the databases and thus remained as unnamed proteins (UP) (Table 2 and Suppl. Information). Two additional complex III canonical subunits (QCR6/10) that were previously identified at genomic level in *Euglena* could not be identified here (Suppl. Information). No CD could be confidently identified for the two UP proteins and only one putative additional TMH is predicted for UP7 protein (e-value threshold = 10<sup>-5</sup>).



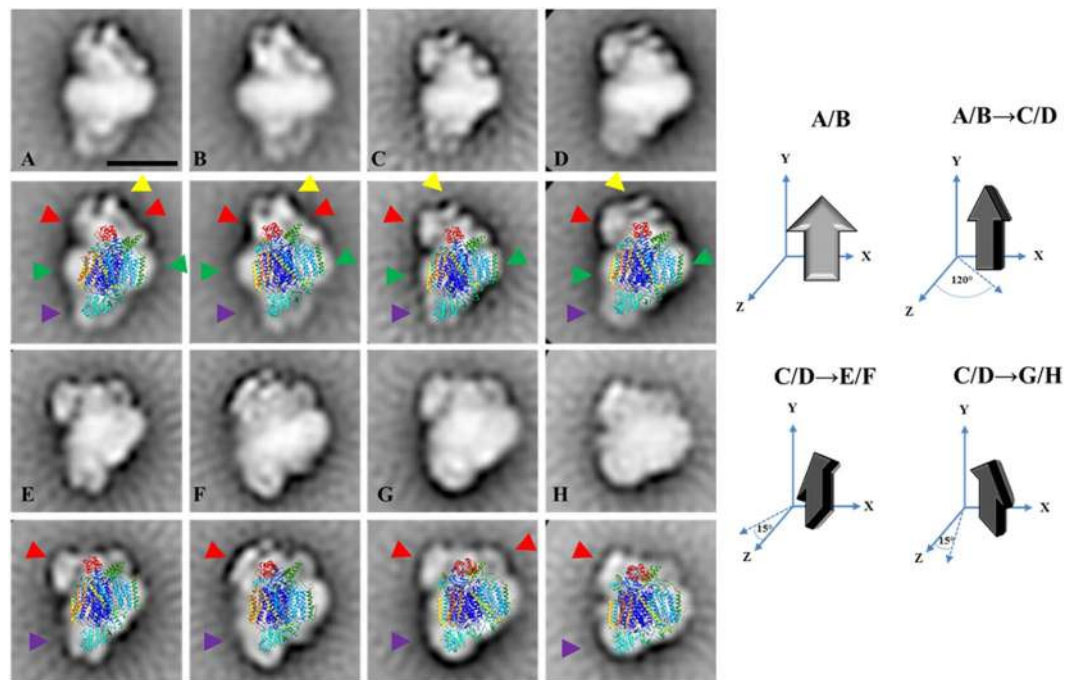


**Figure 6.** 2D Projection maps of purified dimeric complex III from *E. gracilis* obtained by single particle averaging. The dimeric complex III was purified by a two-step chromatographic procedure in the presence of  $\beta$ -dodecyl-*n*-maltoside and analyzed by EM. Overlap of the chicken dimeric complex III (pdb: 4U3F<sup>63</sup>) over the EM images was performed. (A–D) side view projections, (E–H) slightly tilted side-view. The membrane region is indicated by the green arrow heads. At the right: schematic representation of the rotation of the complex in the left panels. The approximate rotation angles over the axis are indicated. The scale bar is 10 nm.

The reported structures for the eukaryotic dimeric complex III are composed of 10 to 11 subunits per monomer<sup>7,62</sup>. In *Euglena*, the sum of all the determined peptides (288 kDa per monomer, Table 2) is in good agreement with the size of a native dimeric complex (ca. 500 kDa, Fig. 1). The *Euglena* complex III particles classify into two principal groups on the carbon film (Fig. 6): Side-view projections (panels A–D) and slightly tilted side-view projections (panels E–H). The membrane region of the complex is evident in the side-view projections (Panels A–D green arrow heads). Panels A and B represent the side-by-side monomer placement, panel C represents the 90° rotation (one monomer in front of the other). The comparison with the structure of chicken dimeric complex III [pdb: 4U3F<sup>63</sup>] explains all the projections obtained and corroborates the dimeric oligomeric state of this complex. The absence of additional electronic densities from any projection concurs with the absence of extra atypical subunits. In conclusion, the overall structure of *E. gracilis* complex III is highly similar to the one reported for other species. However, the lack of QCR8 coding sequence in *Euglena* nucleotide database<sup>27</sup> questions the complex III biogenesis in *Euglena*. QCR8, which has one TMH<sup>63</sup>, is required to form an early core subcomplex with COB and QCR7 during complex III biogenesis in *S. cerevisiae*<sup>62,64</sup>. The QCR7 subunit usually shields the quinone pocket of the CYT1 subunit from exposure to the aqueous environment on the matrix side<sup>65</sup>. Interestingly, *Euglena* QCR7 subunit presents an atypical N-terminal extension (~78 residues; Suppl. Information Fig. S3) with a putative TMH that may take the role of the missing QCR8 subunit. Another role for this QCR7 extension in *Euglena* could be to maintain the structural integrity of the quinone reduction site. Such a function has already been described for the atypical extension in the CYT1 subunit from *Rhodobacter sphaeroides bc1* complex<sup>66</sup>. A subunit named QCR9 was also found in *Euglena* complex III by N-terminal sequencing<sup>67</sup> even though this subunit does not correspond to the canonical QCR9 subunit described in other species<sup>27,67</sup>. Altogether, this analysis confirmed the dimeric state of the purified complex and showed no atypical domains within the overall structure.

#### Atypical subunit composition of *Euglena* complex IV is associated to the presence of an extra domain in the cytochrome *c* binding region.

With the current 3D gel system a total of 14 spots were obtained for *Euglena* mitochondrial complex IV (Fig. 4), while it was previously resolved into 10 protein spots<sup>27</sup>. A single polypeptide could be identified in our database for eleven spots, one spot matched to four polypeptides of similar molecular masses and two spots remained unidentified. Three polypeptides correspond to classical complex IV subunits (COX1, COX3, and COX6B), three to Euglenozoa-specific subunits (COXTB2/4/5), six peptides do not present any similarity with other existing proteins and therefore remain as unnamed proteins (UP) (Table 3 and Suppl. Information). Here again, no CD could be confidently identified for non-classical subunits (e-value threshold =  $10^{-5}$ ). Globally, our analysis allowed the identification of at least 16 polypeptides associated with *Euglena* complex IV with molecular masses ranging from 7.2 to 38.5 kDa. The stoichiometry of one subunit per monomer was proposed based on Coomassie staining<sup>68</sup>. Accordingly, the sum of all the associated peptides (365 kDa, Table 3) is lower than the molecular mass of 460 kDa determined for the whole complex by BN-PAGE (Fig. 1). Additional subunits might have escaped identification, such as previously identified core subunit COX2 or subunits identified at genomic level in *Euglena* like classical subunits (COX5A/5B/8A)<sup>27</sup>, or even some of the originally proposed kinetoplastid complex IV subunits (COXTB1/6/8/10/12/16)<sup>26</sup>. This impressive shift in molecular mass when compared to the mammalian enzyme [200 kDa<sup>46</sup>] is therefore due to additional subunits rather than to dimerization. In this respect, classification of the 460 kDa cytochrome *c* oxidase images obtained



**Figure 7.** 2D Projection maps of purified monomeric complex IV from *E. gracilis* obtained by single particle averaging. The monomeric complex IV was purified by a two-step chromatographic procedure in the presence of  $\beta$ -dodecyl-n-maltoside and analyzed by EM. Overlap of the monomeric bovine complex together with the cytochrome *c* (pdb: 5IY5<sup>13</sup>) over the EM images was performed. (A and B) side view projections; (C and D)  $\sim 120^\circ$  rotated views from A projection along a perpendicular axis to the membrane; (E–H) slightly tilted views: the lower region is tilted by  $\sim 10$ – $15^\circ$  to the front from D projection (E, F) and the upper region is tilted by  $\sim 10$ – $15^\circ$  to the front from D projection (G, H). The membrane region of the complex is indicated by the green arrow heads, the “helmet-like” domain (see text for details) is indicated with red and yellow arrow heads, Cox5A and Cox5B positions in bovine complex IV (purple arrow heads). At the right: schematic representation of the rotation of the complex in the left panels. The approximate rotation angles over the axis are indicated. The scale bar is 10 nm.

by single-particle analysis led to four principal groups (Fig. 7). All pictures depict asymmetric structures unlikely to be dimers: side-views (panels A and B),  $\sim 120^\circ$  rotated views from A projection along a perpendicular axis to the membrane (panels C and D), and slightly tilted views (panels E–H). The membrane region of the complex is highly recognizable in the non-tilted projections (panels A–D green arrow heads). The similarity with the structure of the monomeric bovine complex together with the cytochrome *c* (pdb: 5IY5<sup>13</sup>) was not obvious and the resulting overlays shown in Fig. 7 are subject to caution. In this regard, the opposite orientation of the complex with respect to domains located into the matrix and the intermembrane space led to more non-explained electronic densities at both side of the complex (Suppl. Information Fig. S2). Essentially, these comparisons brought two important pieces of information. They confirmed that the 460 kDa *Euglena* complex IV cannot be a dimer and is thus in a monomeric form. The overlays shown in Fig. 7 also highlighted a novel 5 nm extra density in the intermembrane space (red and yellow arrow heads). This extra density forms a “helmet-like” domain over the cyt *c* binding region. Because the projection of this extra domain varies across all the classes, we cannot exclude the possibility that in some of them this domain is not complete. For instance, the tip-like densities at the top of the domain (Fig. 7, yellow arrows) cannot be visualized on all pictures. Previous attempts to measure the *in vitro* oxidase activity using exogenous cyt *c* as electron donor revealed a specific requirement of *Euglena* complex IV for its endogenous cyt *c*<sup>68,69</sup>. This specificity was first explained by the atypical features found in the purified *Euglena* cyt *c*<sup>70</sup> even though other species could use the *Euglena* cyt *c* as electron donor *in vitro*<sup>69</sup>. Our results allow us to propose that the structure of the *Euglena* complex IV possesses a specific cavity for its endogenous cytochrome *c*.

Among the supernumerary subunits described in other eukaryotic species only Cox6b, the mammalian homolog of yeast Cox12, was identified. Cox6b/Cox12 is a structural but non-essential subunit in close contact with Cox2 and Cox3 subunits in the intermembrane space region<sup>71</sup>. Other classical subunits, such as Cox5A and Cox5B that have been identified in the previous genomic survey, are located in the matrix side of the complex with no TMH (Fig. 7, purple arrow heads), meanwhile Cox8A, an isoform of the ubiquitous Cox8B<sup>72</sup>, is located in the membrane in close contact with Cox1 subunit<sup>13</sup>. Six additional TMHs can be observed in the high resolution structures of this complex<sup>13,14</sup>, they correspond to subunits Cox4I1/6A2/6C/7A1/7B/7C which have not been identified at genomic or proteomic level in *Euglena* (27 and Suppl. Information). One putative TMH is predicted in five UP sequences (UP 7, 9, 11, 13, 15). These subunits could take the place of the classical subunits inside the membrane region to maintain the overall structure of the complex. The absence of putative TMHs among the three trypanosomatid subunits (CoxTB2/4/5) and the UP4/8/15 subunits may indicate that they participate to the

150 kDa “helmet-like” domain. Altogether, our analyses indicate that the atypical subunits of the *Euglena* complex IV may be involved in the construction of an atypical structure whose specific role remains to be elucidated.

**Concluding remarks.** The analysis of the subunit composition and the structure of proton-pumping respiratory oxidases (complexes I, III, and IV) of *Euglena gracilis* showed that this highly divergent organism (when compared to canonical yeast and mammal models) shares many atypical subunits described in respiratory complexes of trypanosomes<sup>27</sup>. While the canonical subunits and the lineage-specific subunits maintain the overall architecture of the respiratory complexes described in yeast and mammals, the lineage-specific subunits are presumably responsible for the atypical extra domains observed in the structure of monomeric complexes I and IV. Incidentally, the presence of these extra subunits/domains probably explains the shift in the molecular mass of complexes I and IV. One protein is also shared between complexes III and IV (UP7-CIII corresponds to UP9-CIV). Although we observed minor cross-contamination after the purification steps, the fact that this polypeptide is the only found in common between both complexes may suggest that it is involved in interactions that lead to the formation of the supercomplex previously observed<sup>27</sup>. Overall, the roles of these atypical subunits/domains in enzyme activities or supramolecular associations remain to be elucidated.

## Material and Methods

**Algal strain, growth conditions and mitochondria isolation.** *Euglena gracilis* (SAG 1224-5/25) was obtained from the University of Göttingen (Sammlung von Algenkulturen, Germany). Cells were grown in the dark under orbital agitation at 25 °C. Ethanol 1% was used as carbon source. The liquid mineral Tris-minimum-phosphate medium (TMP) pH 7.0 was supplemented with a mix of vitamins (biotin 10<sup>-7</sup>%, B12 vitamin 10<sup>-7</sup>% and B1 vitamin 2 × 10<sup>-5</sup>% (w/v)). The cells were collected at the middle of the logarithmic phase by a 10-min centrifugation step at 7000 × g and stored at -70 °C until use. Mitochondria were obtained by differential centrifugation following the procedure previously described<sup>28</sup> and stored at -70 °C until use. Protein concentration was determined by the Bradford method (Biorad).

**Purification of respiratory complexes.** All steps were performed at 4 °C. Seventy five milligrams of mitochondrial proteins were solubilized with n-dodecyl-β-D-maltoside (DDM, 4 g detergent per g protein) in buffer A containing Tris-HCl 50 mM, amino caproic acid 50 mM, MgSO<sub>4</sub> 1 mM, NaCl 50 mM, glycerol 10%, phenyl-methylsulfonyl fluoride (PMSF) 1 mM, tosyl-lysyl-chloromethylketone (TLCK) 50 μg/ml (pH 8.4). The mixture was incubated with gentle stirring for 30 min, and centrifuged at 38,000 × g for 30 min. The supernatant was diluted three times in buffer A without NaCl and supplemented with DDM 0.01%. After a filtration step (0.22 μm) the sample was loaded on an anion exchange column (Mono Q HR 5/5, 1 mL) connected to an ÄKTA explorer 100 (GE Healthcare Life Sciences) equilibrated with the same buffer and washed until a base line was obtained. The column was washed with 50 mM NaCl in the same buffer (10 VC) and eluted with a 50–500 mM NaCl linear gradient (40 VC). Two milliliter fractions were collected and visualized by BN-PAGE.

The samples enriched with Complex I were pooled and concentrated with an Amicon Ultra-15 Centrifugal Filter 100 kDa (EMD Millipore) to a final volume of 500 μL and injected to a Superose 6 10/300 (GE Healthcare Life Sciences) previously equilibrated with buffer A containing 200 mM NaCl and DDM 0.01%. The elution was carried out at 0.3 mL/min, 0.5 mL fractions were collected and visualized by BN-PAGE. The samples enriched with Complex I were pooled separately and stored at -70 °C until use.

The enriched fractions with Complex III and IV from the anion exchange column were pooled separately and concentrated to a final volume of 500 μL. Each fraction was injected separately to the Superose column equilibrated with buffer A containing 300 NaCl and DDM 0.01%. The elution of each injection was carried out as described above, 0.5 mL fractions were collected and visualized by BN-PAGE, the fractions enriched with each complex were stored at -70 °C until use.

**Non-denaturing and denaturing protein electrophoresis.** Each complex was resolved by BN-PAGE using a 3–10% acrylamide gradient gel at 4 °C. The first dimension band was then excised from the gel and submitted to a 2D/3D Glycine/Tricine SDS-PAGE carried out as in<sup>73</sup> at room temperature. Briefly, the subunits of each complex were separated in a Glycine-SDS-PAGE (12% acrylamide) and then each 2D lane was excised and separated individually in a Tricine-SDS-PAGE (14% acrylamide).

Coomassie blue-stained protein spots were manually excised from the 3D gels and analyzed by mass spectrometry (MS) as described in<sup>27</sup>. Briefly, the Matrix Laser Desorption Ionization Analyses (MALDI) was performed on a 4800 MALDI time of flight (TOF/TOF) system (Applied Biosystems). Internal digested fragments of trypsin as well as TOF/TOF Cal Mix 5 (AB Sciex) were used as internal calibration. For direct protein identification with MASCOT against our homemade database (available in <https://figshare.com/s/57d2ba4ebfb-b472ae3de>), protein scores greater than 60 were considered as significant (P < 0.05). To determine the molecular mass of each subunit, PageRuler plus pre-stained protein ladder (Thermo Scientific) were used as size molecular markers. The molecular mass of each protein spot was calculated from its migration distance by comparing it with the migration of molecular markers (the linear regression between the logarithm of the migration distance and the molecular mass (kDa) of the molecular markers has a R<sup>2</sup> of 0.989).

**In silico analysis of the identified subunits from the respiratory complexes.** To further characterize the sequences identified by MS, each protein sequence was submitted to a tBLASTn analysis against the expressed sequence tags (EST) database from *Euglena gracilis* (taxid: 3039) available in the NCBI server. The obtained translated nucleotide sequences were manually assembled to generate the longest possible polypeptide. The resulting polypeptides were submitted to similarity searches by BLASTp against the non-redundant protein sequences database (NRPS) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), and against the Kinetoplastid



genomic resource database (TriTrypDB) (<http://www.tritrypdb.org/tritrypdb/>). The first methionine codon of each sequence was arbitrarily chosen as the putative start codon. The theoretical isoelectric point, molecular mass and grand average of hydropathicity (GRAVY) were determined using the algorithm ProtParam (<http://web.expasy.org/protparam/>), transmembrane helices (TMH) were predicted using Phobius server (<http://phobius.sbc.su.se/>) and TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), local conserved domains (CD) were searched using the NRPS Blastp, DELTA-Blast and Conserved Domain Blast (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), the e-value threshold for the *in silico* analysis results was  $10^{-5}$ , when a TMH or CD was found the topological localization inside the peptide was annotated. Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used for direct alignments of protein sequences.

**Visualization of isolated complexes I, III and IV by transmission electron microscopy.** 5  $\mu$ l of each purified complex solution was absorbed onto freshly glow-discharged carbon-coated copper grids, the excess amount of sample was blotted with filter paper and subsequently stained with 2% uranyl acetate for contrast. Imaging was performed on a Tecnai T20 equipped with a LaB6 tip operating at 200 kV. The “GRACE” system for semi-automated specimen selection and data acquisition<sup>74</sup> was used to record 2048  $\times$  2048 pixel images at 133,000  $\times$  magnifications using a Gatan 4000 SP 4K slow-scan CCD camera with a pixel size of 0.224 nm. Single particles were analyzed with the Xmipp software (including multi-reference and non-reference alignments, multivariate statistical analysis and classification, as in<sup>75</sup> and RELION software<sup>76</sup>). The best of the class members were taken for the final class-sums.

## References

- Carroll, J. *et al.* Bovine complex I is a complex of 45 different subunits. *J. Biol. Chem.* **281**, 32724–32727 (2006).
- Cardol, P. Mitochondrial NADH:Ubiquinone oxidoreductase (complex I) in eukaryotes: A highly conserved subunit composition highlighted by mining of protein databases. *Biochim. Biophys. Acta - Bioenerg.* **1807**, 1390–1397 (2011).
- Friedrich, T. *et al.* Redox components and structure of the respiratory NADH:ubiquinone oxidoreductase (complex I). *Biochim. Biophys. Acta - Bioenerg.* **1365**, 215–219 (1998).
- Di Luca, A., Gamiz-Hernandez, A. P. & Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci.* 201706278 <https://doi.org/10.1073/pnas.1706278114> (2017).
- Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–8 (2013).
- Efremov, R. *g.*, Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465**, 441–447 (2010).
- Berry, E. A., Guergova-Kuras, M., Huan, L. & Crofts, A. R. Structure and Function of Cytochrome bc Complexes. *Annu. Rev. Biochem.* **69**, 1005–1075 (2000).
- Zara, V., Conte, L. & Trumpower, B. L. Biogenesis of the yeast cytochrome bc1 complex. *Biochim. Biophys. Acta - Mol. Cell Res.* **1793**, 89–96 (2009).
- Xia, D. *et al.* Structural analysis of cytochrome bc1 complexes: Implications to the mechanism of function. *Biochim. Biophys. Acta - Bioenerg.* **1827**, 1278–1294 (2013).
- Timón-Gómez, A. *et al.* Mitochondrial cytochrome c oxidase biogenesis: Recent developments. *Semin. Cell Dev. Biol.* 1–16 <https://doi.org/10.1016/j.semcdb.2017.08.055> (2017).
- Vidoni, S. *et al.* MR-1S Interacts with PET100 and PET117 in Module-Based Assembly of Human Cytochrome c Oxidase. *Cell Rep.* **18**, 1727–1738 (2017).
- Denis, M. Structure and function of cytochrome-c oxidase. *Biochimie* **68**, 459–470 (1986).
- Shimada, S. *et al.* Complex structure of cytochrome *c*–cytochrome *c* oxidase reveals a novel protein–protein interaction mode. *EMBO J.* **36**, 291–300 (2017).
- Luo, F. *et al.* Structure of bovine cytochrome c oxidase in the ligand-free reduced state at neutral pH. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **74**, 92–98 (2018).
- Zakrys, B., Rafal, M. & Karnkowska, A. In *Euglena: Biochemistry, Cell and Molecular Biology* (eds Schwartzbach, S. & Shingeoka, S.) 3–28 (2017).
- Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, 1–18 (2014).
- Gibbs, S. P. The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann. New York Acad. Sci.* **81**, 193–208 (1981).
- Turmel, M., Gagnon, M. C., O’Kelly, C. J., Otis, C. & Lemieux, C. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* **26**, 631–648 (2009).
- Fujita, T., Aoyagi, H., Ogbonna, J. C. & Tanaka, H. Effect of mixed organic substrate on  $\alpha$ -tocopherol production by *Euglena gracilis* in photoheterotrophic culture. *Appl. Microbiol. Biotechnol.* **79**, 371–378 (2008).
- Tani, Y. & Tsumura, H. Screening for Tocopherol-producing Microorganisms and  $\alpha$ -Tocopherol Production by *Euglena gracilis* Z. *Agric. Biol. Chem.* **53**, 305–312 (1989).
- Moreno-Sánchez, R. *et al.* Oxidative phosphorylation supported by an alternative respiratory pathway in mitochondria from *Euglena*. *Biochim. Biophys. Acta - Bioenerg.* **1457**, 200–210 (2000).
- Sharpless, T. K. & Butow, R. A. An inducible alternate terminal oxidase in *Euglena gracilis* mitochondria. *J. Biol. Chem.* **245**, 58–70 (1970).
- Benichou, P., Calvayrac, R. & Claisse, M. Induction by antimycin A of cyanide-resistant respiration in heterotrophic *Euglena gracilis*: Effects on growth, respiration and protein biosynthesis. *Planta* **175**, 23–32 (1988).
- Morales, J. *et al.* Novel mitochondrial complex II isolated from *Trypanosoma cruzi* is composed of 12 peptides including a heterodimeric ip subunit. *J. Biol. Chem.* **284**, 7255–7263 (2009).
- Speijer, D. *et al.* Characterization of the respiratory chain from cultured *Crithidia fasciculata*. *Mol. Biochem. Parasitol.* **85**, 171–186 (1997).
- Verner, Z. *et al.* Malleable Mitochondrion of *Trypanosoma brucei*. *Int. Rev. Cell Mol. Biol.* **315**, 73–151 (2015).
- Perez, E. *et al.* The mitochondrial respiratory chain of the secondary green alga *Euglena gracilis* shares many additional subunits with parasitic *Trypanosomatidae*. *Mitochondrion* **19**, 338–349 (2014).
- Yadav, K. N. S. *et al.* Atypical composition and structure of the mitochondrial dimeric ATP synthase from *Euglena gracilis*. *Biochim. Biophys. Acta - Bioenerg.* **1858**, 267–275 (2017).
- Mühlleip, A. W., E. Dewar, C., Schnauffer, A., Kühlbrandt, W. & Davies, K. M. *In-situ* structure of trypanosomal ATP synthase dimer reveals unique arrangement of catalytic subunits. *PNAS* <https://doi.org/10.1073/pnas.1612386114> (2017).
- Castro-Guerrero, N. A., Krab, K. & Moreno-Sánchez, R. The alternative respiratory pathway of *Euglena* mitochondria. *J. Bioenerg. Biomembr.* **36**, 459–469 (2004).

31. Neumann, D. & Parthier, B. Effects of Nalidixic Acid, Chloramphenicol, Cycloheximide and Anisomycin on Structure and Development of Plastids and Mitochondria in Greening *Euglena gracilis*. *Exp. Cell Res.* **81**, 255–268 (1973).
32. Calvayrac, R., Van Lente, F. & Buetow, R. A. *Euglena gracilis*: Formation of Giant Mitochondria. *Science* (80). **173**, 252–254 (1971).
33. Panigrahi, A. K. *et al.* Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. *Mol. Cell. Proteomics* **7**, 534–545 (2008).
34. Panigrahi, A. K. *et al.* A comprehensive analysis of *trypanosoma brucei* mitochondrial proteome. *Proteomics* **9**, 434–450 (2009).
35. Acestor, N. *et al.* *Trypanosoma brucei* Mitochondrial Respiratorome: Composition and Organization in Procytic Form. *Mol. Cell. Proteomics* **10**, M110.006908 (2011).
36. Ziková, A., Schnauer, A., Dalley, R. A., Panigrahi, A. K. & Stuart, K. D. The F0F1-ATP synthase complex contains novel subunits and is essential for procyclic *Trypanosoma brucei*. *PLoS Pathog.* **5** (2009).
37. Ziková, A. *et al.* Structural and functional association of *Trypanosoma brucei* MIX protein with cytochrome c oxidase complex. *Eukaryot. Cell* **7**, 1994–2003 (2008).
38. Boxer, D. H. The location of the major polypeptide of the ox heart mitochondrial inner membrane. *FEBS Lett.* **59**, 149–152 (1975).
39. Detke, S. & Elsabrouty, R. Identification of a mitochondrial ATP synthase-adenine nucleotide translocator complex in *Leishmania*. *Acta Trop.* **105**, 16–20 (2008).
40. Gnipová, A. *et al.* The ADP/ATP carrier and its relationship to oxidative phosphorylation in ancestral protist *trypanosoma brucei*. *Eukaryot. Cell* **14**, 297–310 (2015).
41. Hashimi, H. *et al.* The assembly of F1F0-ATP synthase is disrupted upon interference of RNA editing in *Trypanosoma brucei*. *Int. J. Parasitol.* **40**, 45–54 (2010).
42. Nelson, R. E., Aphasizheva, I., Falick, A. M., Nebohacova, M. & Simpson, L. The I-complex in *Leishmania tarentolae* is a uniquely-structured F1-ATPase. *Mol. Biochem. Parasitol.* **135**, 219–222 (2004).
43. Montgomery, M. G., Leslie, A. G. W., Ziková, A. & Walker, J. E. *ATP synthase from Trypanosoma brucei has an elaborated canonical F1 domain and conventional catalytic sites*. <https://doi.org/10.1073/pnas.1720940115> (2017).
44. Duarte, M. & Tomás, A. M. The mitochondrial complex I of trypanosomatids - An overview of current knowledge. *J. Bioenerg. Biomembr.* **46**, 299–311 (2014).
45. Gabaldón, T. & Huynen, M. A. Lineage-specific gene loss following mitochondrial endosymbiosis and its potential for function prediction in eukaryotes. *Bioinformatics* **21**, 144–150 (2005).
46. Wittig, I. & Schägger, H. Advantages and limitations of clear-native PAGE. *Proteomics* **5**, 4338–4346 (2005).
47. Wittig, I. & Schägger, H. Features and applications of blue-native and clear-native electrophoresis. *Proteomics* **8**, 3974–3990 (2008).
48. Guerrero-Castillo, S., Vázquez-Acevedo, M., González-Halphen, D. & Uribe-Carvajal, S. In *Yarrowia lipolytica* mitochondria, the alternative NADH dehydrogenase interacts specifically with the cytochrome complexes of the classic respiratory pathway. *Biochim. Biophys. Acta - Bioenerg.* **1787**, 75–85 (2009).
49. Nubel, E., Wittig, I., Kerscher, S., Brandt, U. & Schägger, H. Two-dimensional native electrophoretic analysis of respiratory supercomplexes from *Yarrowia lipolytica*. *Proteomics* **9**, 2408–2418 (2009).
50. Cardol, P. *et al.* Higher plant-like subunit composition of mitochondrial complex I from *Chlamydomonas reinhardtii*: 31 Conserved components among eukaryotes. *Biochim. Biophys. Acta - Bioenerg.* **1658**, 212–224 (2004).
51. Klodmann, J., Sunderhaus, S., Nimtz, M., Jansch, L. & Braun, H. P. Internal Architecture of Mitochondrial Complex I from *Arabidopsis thaliana*. *Plant Cell* **22**, 797–810 (2010).
52. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536**, 354–358 (2016).
53. Zickermann, V. *et al.* Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* (80). **347**, 44–49 (2015).
54. Dudkina, N. V., Eubel, H., Keegstra, W., Boekema, E. J. & Braun, H.-P. Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. *Proc. Natl. Acad. Sci.* **102**, 3225–3229 (2005).
55. Sunderhaus, S. *et al.* Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. *J. Biol. Chem.* **281**, 6482–6488 (2006).
56. Gawryluk, R. M. R. & Gray, M. W. Evidence for an early evolutionary emergence of gamma-type carbonic anhydrases as components of mitochondrial respiratory complex I. *BMC Evol. Biol.* **10**, 176 (2010).
57. Cardol, P., Forti, G. & Finazzi, G. Regulation of electron transport in microalgae. *Biochim. Biophys. Acta - Bioenerg.* **1807**, 912–918 (2011).
58. Schmelz, S. & Naismith, J. H. Adenylate-forming enzymes. *Curr. Opin. Struct. Biol.* **19**, 666–671 (2009).
59. Watkins, P. A. Fatty acid activation. *Prog. Lipid Res.* **36**, 55–83 (1997).
60. Airenne, T. T. *et al.* Structure-function analysis of enoyl thioester reductase involved in mitochondrial maintenance. *J. Mol. Biol.* **327**, 47–59 (2003).
61. Carroll, J., Fearnley, I. M., Shannon, R. J., Hirst, J. & Walker, J. E. Analysis of the Subunit Composition of Complex I from Bovine Heart Mitochondria. *Mol. Cell. Proteomics* **2**, 117–126 (2003).
62. Smith, P. M., Fox, J. L. & Winge, D. R. Biogenesis of the cytochrome bc1 complex and role of assembly factors. *Biochim. Biophys. Acta - Bioenerg.* **1817**, 276–286 (2012).
63. Hao, G.-F. *et al.* Rational Design of Highly Potent and Slow-Binding Cytochrome bc1 Inhibitor as Fungicide by Computational Substitution Optimization. *Sci. Rep.* **5**, 13471 (2015).
64. Zara, V., Palmisano, I., Conte, L. & Trumppower, B. L. Further insights into the assembly of the yeast cytochrome bc1 complex based on analysis of single and double deletion mutants lacking supernumerary subunits and cytochrome b. *Eur. J. Biochem.* **271**, 1209–1218 (2004).
65. Cocco, T. *et al.* Structural and Functional-Characteristics of Polypeptide Subunits of the Bovine Heart Ubiquinol - Cytochrome-C Reductase Complex. *Eur. J. Biochem.* **195**, 731–734 (1991).
66. Esser, L. *et al.* Inhibitor-complexed structures of the cytochrome bc1 from the photosynthetic bacterium *Rhodobacter sphaeroides*. *J. Biol. Chem.* **283**, 2846–2857 (2008).
67. Cui, J.-Y., Mukai, K., Saeki, K. & Matsubara, H. Molecular Cloning and Nucleotide Sequences of cDNAs Encoding Subunits I, II, and IX of *Euglena gracilis* Mitochondrial Complex III. *J. Biochem.* **115**, 98–107 (1994).
68. Brönstrup, U. & Hachtel, W. Cytochrome c oxidase of *Euglena gracilis*: Purification, characterization, and identification of mitochondrially synthesized subunits. *J. Bioenerg. Biomembr.* **21**, 359–373 (1989).
69. Collins, N., Brown, R. H. & Merrett, M. J. Oxidative phosphorylation during glycolate metabolism in mitochondria from phototrophic *Euglena gracilis*. *Biochem J* **150**, 373–377 (1975).
70. Pettigrew, G. W., Leaver, J. L., Meyer, T. E. & Ryle, aP. Purification, properties and amino acid sequence of atypical cytochrome c from two protozoa, *Euglena gracilis* and *Crithidia oncopelti*. *Biochem. J.* **147**, 291–302 (1975).
71. Ghosh, A. *et al.* Mitochondrial disease genes COA6, COX6B and SCO2 have overlapping roles in COX2 biogenesis. *Hum. Mol. Genet.* **25**, 660–671 (2016).
72. Goldberg, A. *et al.* Adaptive evolution of cytochrome c oxidase subunit VIII in anthropoid primates. *Proc. Natl. Acad. Sci. USA* **100**, 5873–5878 (2003).
73. Colina-Tenorio, L. *et al.* Subunit Asa1 spans all the peripheral stalk of the mitochondrial ATP synthase of the chlorophycean alga *Polytomella* sp. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 359–369 (2016).



74. Oostergetel, G. T., Keegstra, W. & Brisson, A. Automation of specimen selection and data acquisition for protein electron crystallography. *Ultramicroscopy* **74**, 47–59 (1998).
75. Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* **3**, 977–990 (2008).
76. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

### Acknowledgements

P.C. acknowledges financial support from the Belgian Fonds de la Recherche Scientifique F.R.S.-F.N.R.S. (FRFC 2.4597, CDR J.0079) and European Research Council (H2020-EU BEAL project 682580). P.C. is a Senior Research Associate from F.R.S.-FNRS.

### Author Contributions

H.V.M.A., P.C. and E.J.B. conceived the research, H.V.M.A., L.C.T., K.N.S.Y., F.B., H.D. and P.M. performed the experiments, H.V.M.A., K.N.S.Y., P.C. and E.J.B. analysed the data, H.V.M.A. and P.C. wrote the manuscript, all authors reviewed the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-28039-z>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018