Eur. J. Biochem. **271**, 4123–4131 (2004) © FEBS 2004 doi:10.1111/j.1432-1033.2004.04350.x

Chloroplast phosphoglycerate kinase from *Euglena gracilis*

Endosymbiotic gene replacement going against the tide

Ulrich Nowitzki[1], Gabriel Gelius-Dietrich[1], Maike Schwieger[2], Katrin Henze[1] and William Martin[1]

[1]*Institute of Botany III, Heinrich-Heine-University Düsseldorf, Germany;* [2]*Heinrich-Pette-Institute for Experimental Virology and Immunology at the University of Hamburg, Germany*

Two chloroplast phosphoglycerate kinase isoforms from the photosynthetic flagellate *Euglena gracilis* were purified to homogeneity, partially sequenced, and subsequently cDNAs encoding phosphoglycerate kinase isoenzymes from both the chloroplast and cytosol of *E. gracilis* were cloned and sequenced. Chloroplast phosphoglycerate kinase, a monomeric enzyme, was encoded as a polyprotein precursor of at least four mature subunits that were separated by conserved tetrapeptides. In a Neighbor-Net analysis of sequence similarity with homologues from numerous prokaryotes and eukaryotes, cytosolic phosphoglycerate kinase of *E. gracilis* showed the highest similarity to cytosolic and glycosomal homologues from the Kinetoplastida. The chloroplast isoenzyme of *E. gracilis* did not show a close relationship to sequences from other photosynthetic organisms but was most closely related to cytosolic homologues from animals and fungi.

Keywords: endosymbiotic gene replacement; *Euglena gracilis*; phosphoglycerate kinase; polyproteins.

The complex chloroplasts of the photosynthetic flagellate *Euglena gracilis* are surrounded by three membranes, evidence for their origin through secondary endosymbiosis [1]. The two partners involved in this endosymbiotic event are thought to be a relative of extant Kinetoplastida as host cell and a green alga as endosymbiont. *Euglena gracilis* is linked to the Kinetoplastida by a number of morphological homologies [2–7] and shares unique characters such as the kinetoplastid-specific redox enzyme trypanothione reductase [8] and the unusual base 'J', which is found only in the telomeric regions of Kinetoplastida and *Euglena* [9,10]. Phylogenetic analyses of nucleus-encoded genes for ribosomal RNA [11], tubulins [12], glycolytic glyceraldehyde dehydrogenase [13], the ER-specific protein calreticulin [14] and mitochondrial Hsp60 [15], as well as the mitochondrion-encoded *coxI* gene [15,16] strongly support this relationship. The endosymbiont that has developed into today's euglenid chloroplast was shown in cytological studies [1] and the comparative analysis of chloroplast genomes [17–20] to be derived from a eukaryotic green alga.

Essential to the compartmentation of sugar phosphate metabolism between chloroplast and cytosol in *Euglena* are glycolytic Calvin cycle isoenzyme pairs [21]. Glycolytic 3-phosphoglycerate kinase (PGK, EC 2.7.2.3) catalyses the ADP-dependent dephosphorylation of 1,3-bisphosphoglycerate to 3-bisphosophoglycerate. A chloroplast isoform in photosynthetic eukaryotes catalyses the reverse reaction as part of the Calvin cycle. In the Kinetoplastida, the closest relatives of *Euglena gracilis*, two glycolytic isoforms of PGK have been detected. One is located in the cytosol and the other in the glycosomes, specialized peroxisomes harbouring the first seven steps of glycolysis. Both isoforms are derived from a gene duplication and in phylogenetic analysis were shown to be monophyletic with, but highly divergent from, cytosolic orthologs in protozoa, fungi and animals [22]. In plants the cytosolic PGK was replaced by a copy of the chloroplast isoform, acquired from the cyanobacterial endosymbiont that gave rise to the plastids [23].

Here we report the purification and cloning of the chloroplast PGK (cpPGK) from *Euglena gracilis* which is translated as a polyprotein precursor, cloning of the cytosolic PGK isoenzyme (cPGK), and the histories of both PGK isoforms in the context of endosymbiotic gene acquisitions.

Correspondence to K. Henze, Institute of Botany III, Heinrich-Heine-University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany. Fax: +49 211 813554, Tel.: +49 211 8113983, E-mail: katrin.henze@uni-duesseldorf.de

Abbreviations: PGK, phosphoglycerate kinase; cPGK, cytosolic phosphoglycerate kinase; cpPGK, chloroplast phosphoglycerate kinase; LHCP, light harvesting complex protein; RbcS, ribulose-1,5-bisphosphate carboxylase/oxygenase.

Enzyme: 3-Phosphoglycerate kinase (PGK, EC 2.7.2.3).

(Received 6 July 2004, revised 23 August 2004, accepted 31 August 2004)

Materials and methods

Strain and culture conditions

Euglena gracilis strain SAG 1224–5/25 was grown in 5 L of *Euglena* medium with minerals [24] under continuous light and a constant flow of 2 L·min^{-1} air with 2% (v/v) CO_2. Cells were harvested 5 days after inoculation.

PGK purification from whole cells and chloroplasts

All steps were performed at 4 °C unless stated otherwise. *Euglena* cells (200 g) were homogenized in buffer 1 (10 mM Tris/HCl pH 7.5, 1 mM dithiothreitol) using a French-Press at 8000 p.s.i. and centrifuged for 30 min at 27 500 ***g***. The 30–80% ammonium sulfate fraction of the supernatant was

collected by centrifugation, dialysed against buffer 2 (10 mM Tris/HCl pH 8.5, 1 mM dithiothreitol) to < 2 mS·cm^{-1}, and loaded on a 2.6 × 13 cm DEAE-Sepharose (Amersham Biosciences, Uppsala, Sweden) column. The column was washed with 140 mL buffer 2 and proteins were eluted in a 70 mL 0–350 mM KCl gradient in buffer 2. Most of the PGK activity was detected in the wash fraction.

This fraction was pooled with the active fractions of the gradient, concentrated by ammonium sulfate precipitation, dialysed against buffer 1, and loaded on a 2.6 × 10 cm DEAE Fractogel 650 S (Merck, Darmstadt, Germany) column. The column was washed with 110 mL buffer 1 and proteins were eluted in a 125 mL 0–350 mM KCl gradient in buffer 1. Fractions containing PGK activity were pooled, dialysed against buffer 1 and loaded at 20 °C on a 1.6 × 10 cm Source 30Q (Amersham Biosciences) column. The column was washed with 40 mL buffer 1 and proteins were eluted in a 100 mL 0–300 mM KCl gradient in buffer 1.

Fractions with PGK activity were pooled, dialysed against buffer 1, and loaded at 20 °C on a Mono Q HR 5/5 (Amersham Biosciences) column. The column was washed with 5 mL buffer 1, proteins were eluted in a 15 mL gradient of 0–70 mM KCl in buffer 1, and fractions of 0.4 mL were collected. Two peaks of PGK activity eluted at 40 mM KCl (PGK1) and 55 mM KCl (PGK2), respectively. After dialysis against buffer 2 both peak fractions were further purified separately, but under the same conditions, on a 1.6 × 5 cm Reactive Blue 72 (Sigma, Taufkirchen, Germany) column. The column was washed with 40 mL buffer 2, and proteins were eluted in a 50 mL gradient of 0–400 mM NaCl in buffer 2. Fractions containing PGK activity were pooled and concentrated by ultrafiltration (Millipore, Eschborn, Germany) to 30 μL, applied to a preparative 6.0 cm, 6% native polyacrylamide gel (Mini Prep Cell, Bio-Rad, München, Germany), and electrophoresed at 300 V and 20 °C. Fractions of 190 μL were collected at 100 μL·min^{-1} and assayed for PGK activity. Purified proteins were sequenced as described previously [25], both N-terminally and internally after endopeptidase LysC digestion.

cpPGK was partially purified from isolated *Euglena* chloroplasts. Chloroplasts isolated as described previously [26] were suspended in buffer 2 and lysed by sonication for 2 s. The lysate was centrifuged for 20 min at 30 000 ***g***, and the supernatant was diluted with buffer 2 to a final volume of 20 mL and applied to a 1.6 × 5 cm Reactive blue 72 column. Proteins were eluted as described above. Fractions with PGK activity were pooled, dialysed against buffer 1 and loaded onto a Mono Q HR 5/5 column (Amersham Biosciences). Proteins were eluted as described above.

Protein determination and PGK assay

Protein concentration was determined according to Bradford [27] using bovine serum albumin as a standard. Enzyme activity was measured photometrically at 20 °C in 1 mL of 50 mM HEPES pH 7.6, 4.5 mM $MgCl_2$, 4 mM dithioerythritol, 2 mM ATP, 200 μM NADH, 6 U·mL^{-1} glyceraldehyde-3-phosphate dehydrogenase, 6 U·mL^{-1} triose-phosphate isomerase, 4 mM 3-phosphoglycerate. One unit is the amount of enzyme that catalyses the oxidation of 1 μM NADH in one minute.

cDNA cloning and Northern blotting

RNA purification and cDNA library construction were performed as described previously [13,28]. A 1550 bp cDNA fragment coding for the glycosomal PGK (PGK-C) of *Trypanosoma brucei* [29] was radioactively labelled as a heterologous probe for cPGK and hybridized against 10^5 recombinant clones of the *Euglena* cDNA library [25]. Six independent clones encoding the same transcript were identified. The sequence of one full-length clone (pbP12.1) was determined.

A homologous hybridization probe for the cpPGK was generated by PCR. Primers 5′-GAYTTYAAYGTNCCN TTYGA-3′ and 5′-CCDATNGCCATRTTRTTNAR-3′ were designed against the sequenced peptides DFNVPFD and LNNMAIG, obtained from purified chloroplast PGK. Amplification conditions were 35 cycles of 1 min at 93 °C, 1 min at 50 °C, 1 min at 72 °C in 25 μL of 10 mM Tris/HCl (pH 8.3), 50 mM KCl, 1.0 mM $MgCl_2$, 0.05 mM of each dNTP, 0.02 U·μL^{-1} Ampli *Taq* polymerase (PerkinElmer, Norwalk, CT, USA), 2 ng·μL^{-1} *Euglena* cDNA, and 0.8 μM of each of the primers. The 720 bp amplification product was sequenced and used as a hybridization probe to screen 3×10^5 recombinant cDNA clones. Sixteen independent clones of sizes ranging from 1.0 to 3.2 kb were isolated and shown by sequencing to encode the same transcript. The sequence of the longest clone pcpPGK4 was determined by constructing nested deletions with exonuclease III and mung bean nuclease [25]. Northern blotting was performed as described previously [30]; the blot was probed with the cpPGK-specific 720 bp PCR fragment.

Phylogenetic analysis

PGK homologues were identified by a BLAST search of the nonredundant database at GenBank (http://www.ncbi.nlm.nih.gov/). Homologues were retrieved and aligned using CLUSTALW [31]. Gaps in the alignment were removed with the script RMGAPS. Protein LogDet distances, which are based on the determinant of a distance matrix comprising the relative frequencies of all amino acid pairs between two sequences [32], were calculated with the LDDIST program available at http://artedi.ebc.uu.se/molev/software/LDDist.html. Neighbor-Net networks [33] of protein LogDet distances [34] were constructed with NNET and visualized with SPLITSTREE [35]. Sequences were retrieved from GenBank under the accession numbers BAA79084 *Aeropyrum pernix*, NP_534233 *Agrobacterium tumefaciens*, O66519 *Aquifex aeolicus*, O29119 *Archaeoglobus fulgidus*, P41756 *Aspergillus oryzae*, Q8L1Z8 *Bartonella henselae*, P18912 *Bacillus stearothermophilus*, P40924 *Bacillus subtilis*, NP_879795 *Bordetella pertussis*, AAB53931 *Borrelia burgdorferi*, NP_768162 *Bradyrhizobium japonicum*, Q9L560 *Brucella melitensis*, NP_240262 *Buchnera aphidicola*, Q9A3F5 *Caulobacter vibrioides*, P94686 *Chlamydia trachomatis*, P41758 *Chlamydomonas reinhardtii*, Q01655 *Corynebacterium glutamicum*, P25055 *Crithidia fasciculata* glycosome, P08966 *Crithidia fasciculata* cytosol, P08967 *Crithidia fasciculata* glycosome, YP_011741 *Desulfovibrio vulgaris*, Q01604 *Drosophila melanogaster*, P11665 *Escherichia coli*, P51903 *Gallus gallus*, P43726 *Haemophilus influenzae*, P50315 *Haloarcula vallismortis*, P56154 *Helicobacter*

pylori, P00558 *Homo sapiens*, P20971 *Methanothermus fervidus*, Q58058 *Methanococcus jannaschii*, O27121 *Methanothermobacter thermoautotrophicus*, P47542 *Mycoplasma genitalium*, O06821 *Mycobacterium tuberculosis*, NP_840413 *Nitrosomonas europaea*, Q8YPR1 *Nostoc* sp., O02609 *Oxytricha nova*, NP_246799 *Pasteurella multocida*, P27362 *Plasmodium falciparum*, BAA33801 *Populus nigra* cytosol, BAA33803 *Populus nigra* chloroplast, NP_892316 *Prochlorococcus marinus*, O58965 *Pyrococcus horikoshii*, P29405 *Rhizopus niveus*, P00560 *Saccharomyces cerevisiae*, NP_457468 *Salmonella enterica*, P41759 *Schistosoma mansoni*, P74421 *Synechocystis* sp., NP_898418 *Synechococcus* sp., P50313 *Tetrahymena thermophila*, NP_683058 *Thermosynechococcus elongatus*, S54289 *Thermotoga maritima*, P09403 *Thermus thermophilus*, O83549 *Treponema pallidum*, P14228 *Trichoderma reesei*, P08891 *Trypanosoma brucei* A glycosome, P07378 *Trypanosoma brucei* C glycosome, P07377 *Trypanosoma brucei* B cytosol, P41762 *Trypanosoma congolense* glycosome, P41760 *Trypanosoma congolense*, cytosol, P12783 *Triticum aestivum* cytosol, P12782 *Triticum aestivum* chloroplast, NP_871308 *Wigglesworthia glossinidia*, NP_966880 *Wolbachia* sp., NP_907231 *Wolinella succinogenes*, P50314 *Xanthobacter flavus*, P29407 *Yarrowia lipolytica*, NP_994796 *Yersinia pestis*, P09404 *Zymomonas mobilis*. The *Cyanidioschyzon merolae* chloroplast PGK sequence was retrieved from http://merolae.biol.s.u-tokyo.ac.jp, accession number CMJ305C.

Results

Purification and cloning of *Euglena* chloroplast PGK

Two isoforms of PGK with a molecular mass of 60 kDa were purified to electrophoretic homogeneity (Fig. 1) from total *Euglena gracilis* cells. PGK1, eluting at 40 mM KCl from the Mono Q column, was purified 294-fold and had a specific activity of 1179 U·mg^{-1}. PGK2, eluting at 55 mM KCl from Mono Q, was purified 259-fold and had a specific activity of 1037 U·mg^{-1} (Table 1). Partial purification of cpPGK from isolated *Euglena* chloroplasts also yielded two peaks of PGK activity eluting at nearly the same salt concentrations from Reactive Blue 72 and Mono Q (data not shown). These findings strongly suggest that two very similar isoforms of the chloroplast PGK were purified from total *Euglena* cells, which can be separated on Mono Q. Both proteins had identical N-terminal amino acid sequences as determined by N-terminal protein sequencing (Table 2).

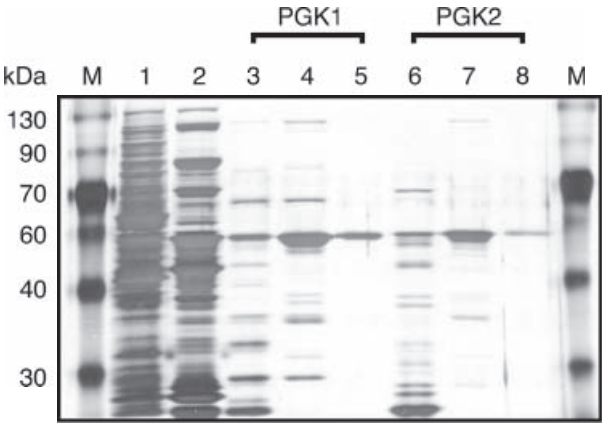

Fig. 1. SDS/PAGE of the purified chloroplast phosphoglycerate kinase isoenzymes of *E. gracilis*. M, Marker proteins; lane 1, crude extract; lane 2, active fractions from Source 30Q; lanes 3 and 6, first (PGK1) and second (PGK2) active peak eluting from Mono Q, peaks were treated separately from here; lanes 4 and 7, active fractions from Reactive Blue 72; lanes 5 and 8, active fractions from preprarative gel electrophoresis.

Table 1. Purification of phosphoglycerate kinases PGK1 and PGK2 from *Euglena*.

Purification step	Total activity (U)	Total Protein (mg)	Specific activity (U·mg^{-1})	Purification (fold)
Crude extract	35945	9875	4	–
AS precipitation	29583	6055	5	1
DEAE Sepharose	29522	2072	14	4
DEAE Fractogel	20460	1100	19	5
Source 30 Q	20295	297	68	17
PGK1				
Mono Q	5415	8.50	637	159
Reactive Blue 72	3570	4.50	793	198
Native PAGE	1014	0.86	1179	294
PGK2				
Mono Q	6336	9.60	660	165
Reactive Blue 72	5244	6.40	819	205
Native PAGE	1856	1.79	1037	259

Table 2. N-terminal and internal peptide sequences from purified phosphoglycerate kinases PGK1 and PGK2.

Peptide	Sequence
N-terminus	
PGK1	AVTGETSLNKLQLKDADV
	KGKRVFIRVDFNVPFDKK
PGK2	AVTGEXSLNKLQLKDADVKG
PGK2 internal peptides	
Peptide 1	VDFNVPFDKKD
Peptide 2	VLNNMAIGSS
Peptide 3	ADVXVND

The amino acid sequences of three internal proteolytic fragments from PGK2 were determined (Table 2). Using degenerate primers designed against the sequences of peptides 1 and 2, a PCR amplification product of 720 bp was obtained and used as a hybridization probe to isolate 16 cDNA clones coding for cpPGK. The longest cDNA clone, pcpPGK4, was completely sequenced. It contained an open reading frame (ORF) of 3000 bp which encoded three consecutive PGK proteins (Fig. 2). As the cDNA clone was not complete at the 5′-end, no transit peptide and only the C-terminal part of the first PGK segment were found. The two subsequent PGK proteins are complete. All three PGK proteins are separated by a conserved motif of four amino acids (SVAM). The two complete PGK segments encode

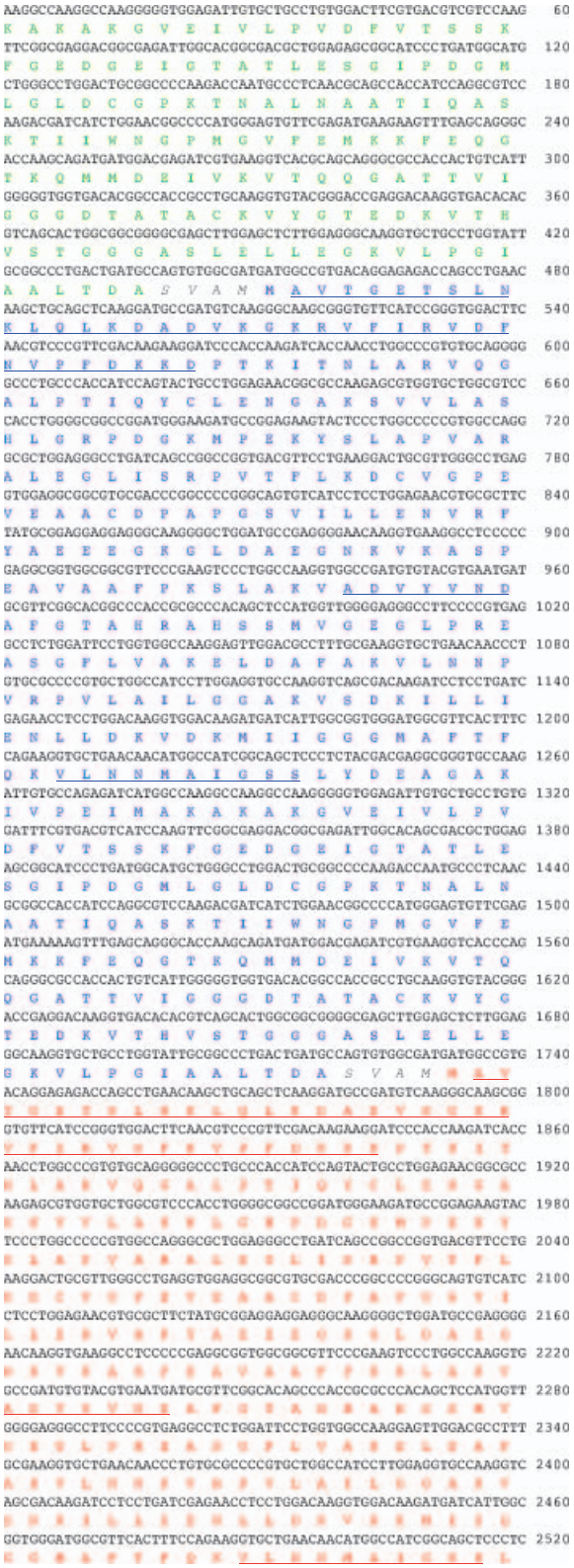

Fig. 2. cDNA sequence and conceptual translation of clone pcpPGK4. The three consecutive phosphoglycerate kinase proteins are printed in colour. N-terminal and internal peptide sequences generated from the purified proteins PGK1 and PGK2 (Table 2) are underlined. The SVAM tetrapeptides are shown in italic.

almost identical proteins of 423 amino acids that differ in only one residue. Asp422 of the second PGK protein (and also of the identical C-terminal fragment of the first unit) was replaced by Asn in the third PGK protein at the 3′ end. At the nucleotide level sequence identity of the PGK segments is 97–99%. The calculated M_r of the deduced amino acid sequence is 44 475 Da, which is in reasonably good agreement with the M_r of 48 kDa estimated from SDS/PAGE (Fig. 1). All three peptide sequences generated from the purified cpPGK were found in the two complete PGK segments of pcpPGK4, identifying the encoded proteins as chloroplast isoforms of PGK (Fig. 2).

A Northern blot of poly(A^+) mRNA was probed with the cpPGK-specific 720 bp PCR fragment and revealed two transcripts of 4.4 kb and 5.6 kb. Both transcripts are long enough to encode polyproteins of three and four consecutive PGK proteins of 423 amino acids, respectively, plus a putative transit peptide for chloroplast import (Fig. 3).

Cloning of *Euglena* cytosolic PGK

As the cytosolic PGK (cPGK) isoenzyme was not recovered by our purification procedure, a 1550 bp cDNA fragment coding for the glycosomal PGK (PGK-C) of *Trypanosoma brucei* was used to retrieve cPGK-specific clones from the *Euglena* cDNA library. The complete sequence of clone pbP12.1 revealed a 1391 bp cDNA which contained a 1245 bp ORF. The high homology of the encoded protein to other PGK sequences and the absence of a transit peptide identifies it as the cytosolic PGK from *E. gracilis*. Alignment of the cPGK amino acid sequence from *E. gracilis* with PGK sequences retrieved from GenBank revealed that it is a homologue of the cytosolic and glycosomal PGK isoenzymes of Kinetoplastida, with which it shares ≈ 55% amino acid identity.

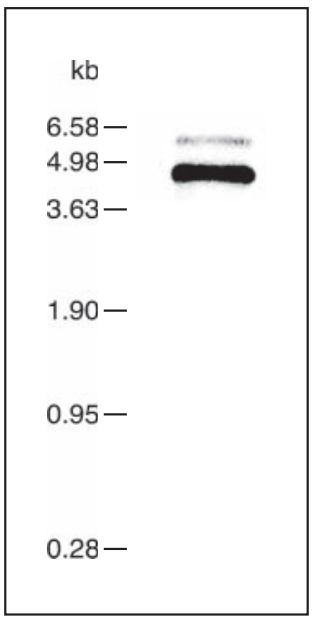

Fig. 3. Northern blot. Northern blot of 2 μg mRNA hybridized with a 720 bp probe specific for chloroplast PGK.

Neighbor-Net analysis

A Neighbor-Net sequence similarity network comparing the cytosolic and and chloroplast PGK protein sequences from *Euglena gracilis* with a representative sample of homologues from archaebacteria, eubacteria and eukaryotes was generated from LogDet distances based on a CLUSTALW alignment of the sequences (Fig. 4). As seen in many other analyses involving prokaryotic sequences, the branching order among PGK sequences from eubacteria is not resolved in the similarity network [36,37]. This could be due to extensive lateral gene transfer among prokaryotes [38,39] or to saturation at variable amino acid sites [40]. A strong split recovers the archaebacteria as a monophyletic group that is well separated from the eubacteria. All the eukaryotic groups appear among the eubacterial sequences.

Among the eukaryotes, the cytosolic and chloroplast homologues from plants and red and green algae form a separate cluster that also includes the cyanobacterial sequences, implying a cyanobacterial, i.e. chloroplast, origin of both isoenzymes in this group. All other eukaryotic sequences form a monophyletic group that again is separated into two distinct subgroups. One contains the highly divergent cytosolic and glycosomal PGK sequences from Kinetoplastida and the cytosolic isoform of *E. gracilis*, showing that cPGK of *E. gracilis* is orthologous to both isoforms in the Kinetoplastida. The second subgroup comprises the cytosolic PGKs of protozoa, fungi and animals together with the chloroplast isoform of *Euglena*. Accordingly, cpPGK from *E. gracilis* has a different origin than its homologues in algae and plants and, although all nonplant eukaryotic PGKs in the network appear to share a common eubacterial ancestry, even if the precise donor lineage is not revealed, it also has a different phylogenetic history than the cytosolic isoform.

Discussion

The chloroplast PGK of *Euglena gracilis* is synthesized as a polyprotein precursor

CpPGK from *Euglena gracilis* was purified to homogeneity (Fig. 1) and the protein microsequenced. A partial cDNA was cloned that encoded at least three consecutive copies of the enzyme. The mature protein units were separated by a conserved SVAM tetrapeptide (Fig. 2). These findings suggest that cpPGK from *Euglena* is synthesized as a polyprotein precursor from which the mature proteins are processed after import into the plastid. Three other nucleus-encoded chloroplast proteins were previously found to be expressed as polyprotein precursors with a single bipartite transit sequence in *Euglena*; light harvesting complex protein (LHCP) I [41], LHCP II [42,43] and ribulose-1,5-bisphosphate carboxylase/oxygenase (RbcS) [44]. These precursors comprise up to eight mature protein units that are separated by decapeptides with the consensus sequence XMXAXXGXKX [45]. Proteolytic processing of the precursors at the decapeptides takes place in the chloroplast [46,47] and was shown to be carried out by a sequence-specific thiol protease, which is localized in the chloroplast stroma [48]. In contrast, the segments of the PGK polyprotein are separated by a tetrapeptide (SVAM). A very similar topology was found in the dinoflagellate *Amphidinium carterae*, another organism with secondary plastids, where the segments of a putative polyprotein precursor of the chlorophyll *a*-*c*-binding protein are also separated by a tetrapeptide (SPLR) [49]. The protease that processes the PGK precursor remains to be identified. The short tetrapeptide spacers suggest that it may be different from the one acting on the decapeptide spacers [48].

Notably, only a subset of nucleus-encoded plastid proteins is encoded as polyprotein precursors in *E. gracilis*. Several other nuclear genes for plastid proteins have been shown to encode single proteins, e.g. enolase [28], fructose-1,6-bisphosphate aldolase [50], glyceraldehyde-3-phosphate dehydrogenase [13] and the extrinsic 30 kDa protein of photosystem II [51]. The question is why some proteins are expressed as polyproteins in *Euglena*, and probably also in the dinoflagellate *Amphidinium*, while others are not. The LHCPs and RbcS are among the most abundant proteins in algae and plants. Multigene families guarantee their synthesis in adequate amounts in these organisms [52–54]. In analogy the synthesis of polyproteins in *E. gracilis* was assumed to be a means to supply sufficient amounts of these proteins without the necessity of maintaining large multigene families [45]. In chloroplast PGK, a protein expressed as a polyprotein precursor has been found that functions as a monomer and is not organized into a higher plant multigene family. Thus, substitution for multigene families alone cannot explain the existence of polyprotein precursors in *E. gracilis* and other possible explanations have to be considered. Firstly, the processing of polyproteins is an additional step in gene expression that might be post-translationally regulated through the expression-level of the processing protease [45]. Secondly, although single protein precursors such as glyceraldehyde-3-phosphate dehydrogenase [13] are efficiently transferred into the chloroplast, it cannot be excluded that import across three membranes as polyprotein precursors might be more efficient for some proteins. LHCP II and RbcS polyprotein precursors are inserted into the ER membrane and transferred as integral membrane proteins to the Golgi apparatus before import into the chloroplast [46,47,55]. Because no single-protein precursors have yet been analyzed, it remains to be seen whether this pathway is restricted to polyproteins or whether it is the general chloroplast protein import pathway in *E. gracilis*. Thirdly, expression of polyproteins might be of no advantage whatsoever, but simply a chance occurrence whose fixation is made possible by the existence of the chloroplast polyprotein processing protease. Identification of more polyproteins and comparison of expression patterns with single precursors may help to better understand why some chloroplast proteins are expressed in this unique fashion in *E. gracilis*.

Kinetoplastid PGK in the cytosol of *E. gracilis*

PGK phylogeny has been previously analysed for a broad spectrum of organisms by Brinkmann and Martin [23]. The results of our Neighbor-Net analysis (Fig. 4) are congruent with that distinct overall picture of PGK gene phylogeny. All nonplant eukaryotic PGKs form

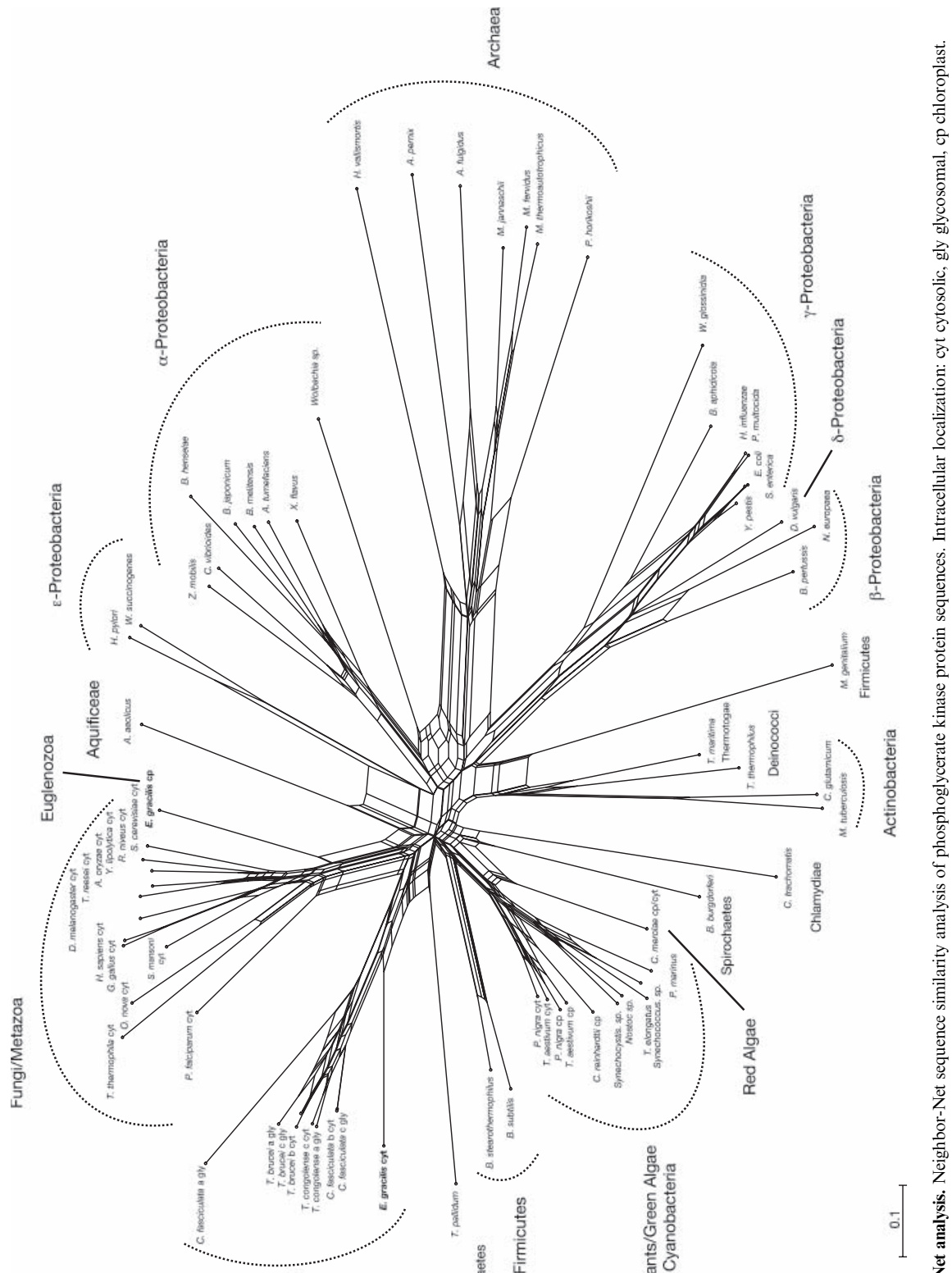

Fig. 4. Neighbor-Net analysis. Neighbor-Net sequence similarity analysis of phosphoglycerate kinase protein sequences. Intracellular localization: cyt cytosolic, gly glycosomal, cp chloroplast.

a monophyletic group, which is rooted among the eubacterial homologues. The archaebacterial homologues are monophyletic and are well separated from all other sequences analysed. This situation suggests a eubacterial origin of eukaryotic PGKs. Although a specific eubacterial donor cannot be identifed from the sequence similarity analysis in Fig. 4, the ancestor of mitochondria appears to be the most likely source. Endosymbiotic gene transfer from mitochondria and chloroplasts to the nucleus, and the subsequent retargeting of gene products to cytosolic pathways such as glycolysis, have been amply demonstrated in eukaryotes [56]. Furthermore, several other cytosolic proteins from *E. gracilis*, glycolytic glyceraldehyde-3-phosphate dehydrogenase [13] and fructose-1,6-bisphosphate aldolase [50], tubulin [12] and calretculin [14] have previously been reported to be of mitochondrial origin. It should be mentioned, however, that cytosolic PGKs from eukaryotes do not branch specifically with α-proteobacterial homologues in the Neighbor-Net analysis, and thus these enzymes fail to meet a criterion set forth for eukaryotic genes inferred to be of mitochondrial origin [57]. However, about half of the 63 proteins encoded in the *Reclinomonas americana* mitochondrial genome also fail to branch with α-proteobacterial homologues [58], indicating that there is a considerable degree of inherent uncertainty involved in phylogenetic analysis [59]. Furthermore due to frequent lateral gene transfer among bacteria contemporary α-proteobacteria cannot reasonably be expected to contain exactly the same set of orthologous genes as the ancestral mitochondrial endosymbiont [60]. Accordingly, the lack of a specific association between eukaryotic and α-proteobacterial PGK sequences does not constitute clear evidence against a mitochondrial origin of eukaryotic PGK.

The PGK sequences from the Kinetoplastida are highly divergent from all other eukaryotic cytosolic PGKs and form a separate subgroup. In *Trypanosoma brucei* and *Crithidia fasciculata* gene duplications have led to the emergence of cytosolic and glycosomal isoforms. Cytosolic PGK from *E. gracilis* is an orthologue of cytosolic and glycosomal PGKs in the Kinetoplastida. Thus it appears that after the kinetoplastid host cell engulfed a chlorophytic alga, and at the emergence of the euglenid lineage, no endosymbiotic gene replacement occurred in the *E. gracilis* cPGK.

Chloroplast PGK in *E. gracilis*, a molecular relic from the nucleus of the secondary endosymbiont

Acquisition of endosymbiotic organelles was, and probably still is, accompanied by extensive endosymbiotic gene transfer from the genome of the endosymbiont to the nucleus of the host cell, followed in many instances by recompartmentation of the encoded gene products, and thus resulting in chimaeric nuclear genomes and hybrid compartment proteomes [56]. In secondary endosymbiosis an additional level of complexity is added to the endosymbiotic gene transfer and gene replacement scenario with the nucleus of the eukaryotic endosymbiont. Therefore, in any phylogenetic analyses of *E. gracilis* nucleus-encoded chloroplast proteins, three different origins of genes have to be considered: the chloroplast genome of the endosymbiotic green alga, the now lost nucleus of that green alga, and the nucleus of the euglenozoan host cell.

The cytosolic and chloroplast PGK homologues from plants, as well as red and green algae, are clearly distinct from all other eukaryotic homologues. They form a separate cluster in the sequence similarity network (Fig. 4) that also includes the sequences from cyanobacteria. This topology indicates that in the algae/plant lineage, when chloroplasts arose the PGK gene from the endosymbiotic cyanobacterium was transferred to the nucleus of the eukaryotic host cell. After gene duplication a copy of the cyanobacterial PGK also replaced the endogenous eukaryotic, cytosolic PGK that is still found in animals, fungi and euglenozoa (Fig. 4). In *E. gracilis*, gene replacement in the wake of secondary endosymbiosis went against the tide. In contrast to plants and algae, the cytosolic PGK of the kinetoplastid host cell been retained as the glycolytic isoform. The strong similarity of cpPGK from *E. gracilis* with cytosolic homologues from protists, animals and fungi (Fig. 4) shows that the cyanobacterial Calvin cycle isoenzyme of the euglenid chloroplast was replaced by a cytosolic isoform, probably retargeted from the nucleus of the green algal endosymbiont. Accordingly, cpPGK from *E. gracilis* is most probably a molecular relic, the only repesentative of the original cytosolic PGK found among photosynthetic eukaryotes to date.

Acknowledgements

We thank Eva Walla for excellent technical assistance and Stephan Zangers and Sven Schünke for the gap-removal script rmgaps. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

1. Gibbs, S.P. (1978) The chloroplast of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.* **56**, 2883–2889.
2. Kivic, P.A. & Walne, P.L. (1984) An evaluation of a possible phylogenetic relationship between the Euglenophyceae and Kinetoplastida. *Origins Life* **13**, 269–288.
3. Surek, B. & Melkonian, M. (1986) A cryptic cytostome is present in *Euglena*. *Protoplasma* **133**, 39–49.
4. Walne, P.L. & Kivic, P.A. (1989) Phylum Euglenida. In *Handbook of Protoctista*, Vol. 1, (Margulis, L., Corliss, J.O., Melkonian, M. & Chapman, D.J., eds), pp. 270–287. Jones and Bartlett, Boston, MA.
5. Vickermann, K. (1990) Phylum Zoomastigina class Kinetoplastida. In *Handbook of Protoctista*, Vol. 1, (Margulis, L., Corliss, J.O., Melkonian, M. & Chapman, D.J., eds), pp. 215–238. Jones and Bartlett, Boston, MA.
6. Cavalier-Smith, T. (1993) Kingdom Protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.
7. Corliss, J.O. (1994) An interim utilitarian ('user friendly') hierarchial classification and characterisation of the protists. *Acta Protozool.* **33**, 1–51.
8. Dumas, C., Ouelette, M., Tovar, J., Cunningham, M.L., Fairlamb, A.H., Tamar, S., Olivier, M. & Papadopoulou, B. (1997) Disruption of the trypanothione reductase gene of *Leishmania* decreases its ability to survive oxidative stress in macrophages. *EMBO J.* **16**, 2590–2598.
9. Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van Der Marel, G.A., van Boom, J.H., van Leeuwen, F. & Borst, P. (1999)

The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J.* **18**, 6573–6581.
10. Dooijes, D., Chaves, I., Kieft, R., Dirks-Mulder, A., Martin, W. & Borst, P. (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. *Nucleic Acids Res.* **28**, 3017–3021.
11. Sogin, M., Gunderson, J., Elwood, H., Alonso, R. & Peattie, D. (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**, 75–77.
12. Levasseur, P.J., Meng, Q. & Bouck, B. (1994) Tubulin genes in the algal protist *Euglena gracilis*. *J. Euk. Microbiol.* **41**, 468–477.
13. Henze, K., Badr, A., Wettern, M., Cerff, R. & Martin, W. (1995) A nuclear gene of eubacterial origin in *Euglena* reflects cryptic endosymbioses during protist evolution. *Proc. Natl Acad. Sci. USA* **92**, 9122–9126.
14. Navazio, L., Nardi, C., Baldan, B., Dianese, P., Fitchette, A.C., Martin, W. & Mariani, P. (1998) Functional conservation of calreticulin from *Euglena gracilis*. *J. Euk. Microbiol.* **45**, 307–313.
15. Yasuhira, S. & Simpson, L. (1997) Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. *J. Mol. Evol.* **44**, 341–347.
16. Tessier, L.H., van der Speck, H., Gualberto, J.M. & Grienenberger, J.M. (1997) The *cox1* gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. *Curr. Genet.* **31**, 208–213.
17. Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M. & Kowallik, K.V. (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165.
18. Leitsch, C.I.W., Kowallik, K.V. & Douglas, S. (1999) The atpA gene cluster of *Guillardia theta* (Cryptophyta): a piece in the puzzle of chloroplast genome evolution. *J. Phycol.* **35**, 115–122.
19. Lockhart, P.J., Howe, C.J., Barbrook, A.C., Larkum, A.W.D. & Penny, D. (1999) Spectral analysis, systematic bias and the evolution of chloroplasts. *Mol. Biol. Evol.* **16**, 573–576.
20. Stöbe, B. & Kowallik, K.V. (1999) Gene-cluster analysis in chloroplast genomics. *Trends Genet.* **15**, 344–347.
21. Martin, W. & Schnarrenberger, C. (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32**, 1–18.
22. Adjé, C.A., Opperdoes, F.R. & Michels, P.A.M. (1998) Molecular analysis of phosphoglycerate kinase in *Trypanoplasma borreli* and the evolution of this enzyme in Kinetoplastida. *Gene* **217**, 91–99.
23. Brinkmann, H. & Martin, W. (1996) Higher-plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. *Plant Mol. Biol.* **30**, 65–75.
24. Schlösser, U.G. (1997) SAG-Sammlung für Algenkulturen at the University of Göttingen. *Bot. Acta* **107**, 111–186.
25. Henze, K., Schnarrenberger, C., Kellermann, J. & Martin, W. (1994) Chloroplast and cytosolic triosephosphate isomerase from spinach: Purification, microsequencing and cDNA sequence of the chloroplast enzyme. *Plant Mol. Biol.* **26**, 1961–1973.
26. Price, C.A., Hadjeb, N., Newman, L. & Reardon, E.M. (1994) Isolation of chloroplasts and chloroplast DNA. In *Plant Molecular Biology Manual* (Gelvin, S. & Schilperoort, R., eds), pp. 1–15. Kluwer Academic Publishers, Dordrecht, the Netherlands.
27. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the prinziple of protein-dye-binding. *Anal. Biochem.* **72**, 248–254.
28. Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J.A., Albert, M.-A., Sensen, C.W., Gaasterland, T., Müller, M., Michels, P. & Martin, W. (2000) Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and the cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.* **17**, 989–1000.
29. Zomer, A.W., Allert, S., Chevalier, N., Callens, M., Opperdoes, F.R. & Michels, P.A. (1998) Purification and characterisation of the phosphoglycerate kinase isoenzymes of *Trypanosoma brucei* expressed in *Escherichia coli*. *Biochim. Biophys. Acta* **1386**, 179–188.
30. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
31. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) 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**, 4673–4680.
32. Lockhart, P.J., Steel, M.A., Hendy, M.D. & Penny, D. (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605–612.
33. Bryant, D. & Moulton, V. (2004) Neighbor-Net: an agglomerative method for the construction of planar phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265.
34. Thollesson, M. (2004) LDDist: a Perl module for calculating LogDet pair-wise distances for protein and nucleotide sequences. *Bioinformatics* **20**, 416–418.
35. Huson, D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.
36. Atteia, A., van Lis, R., Mendoza-Hernández, G., Henze, K., Riveros-Rosas, H. & González-Halphen, D. (2003) Bifunctional aldehyde/alcohol dehydrogenase (ADHE/AAD) in chlorophyte algal mitochondria. *Plant Mol. Biol.* **53**, 175–188.
37. Gelius-Dietrich, G. & Henze, K. (2004) Pyruvate Formate Lyase (PFL) and PFL activating enzyme in the chytrid fungus *Neocallimastix frontalis*: a free-radical enzyme system conserved across divergent eukaryotic lineages. *J. Euk. Microbiol.* **51**, 456–463.
38. Eisen, J. (2000) Horizontal gene transfer among microbial genomes: insights from complete genome analysis. *Curr. Opin. Genet. Dev.* **10**, 606–611.
39. Jain, R., Rivera, M.C., Moore, J.E. & Lake, J.A. (2002) Horizontal gene transfer in microbial genome evolution. *Theor. Pop. Biol.* **61**, 489–495.
40. Horner, D.S. & Pesole, G. (2003) The estimation of relative site variability among aligned homologous protein sequences. *Bioinformatics* **19**, 600–606.
41. Houlné, G. & Schantz, R. (1988) Characterization of cDNA sequences for LHCI apoproteins in *Euglena gracilis*: the mRNA encodes a large precursor containing several consecutive divergent polypeptides. *Mol. Gen. Genet.* **213**, 479–486.
42. Rikin, A. & Schwartzbach, S.D. (1988) Extremly large and slowly processed precursors to the Euglena light-harvesting chlorophyll a/b binding proteins of photosystem II. *Proc. Natl Acad. Sci. USA* **85**, 5117–5121.
43. Muchal, U.S. & Schwartzbach, S.D. (1992) Characterization of a Euglena gene encoding a polyprotein precursor to the light-harvesting chlorophyll a/b protein of photosystem II. *Plant Mol. Biol.* **18**, 287–299.
44. Chan, R., Keller, M., Canaday, S., Weil, J. & Imbault, P. (1990) Eight small subunits of *Euglena* ribulose-1,5-bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein. *EMBO J.* **9**, 333–338.
45. Houlné, G. & Schantz, R. (1993) Expression of polyproteins in *Euglena*. *Crit. Rev. Plant Sci.* **12**, 1–17.
46. Sulli, C. & Schwartzbach, S.D. (1995) The polyprotein precursor to the *Euglena* light harvesting chlorophyll a/b-binding protein is transported to the Golgi apparatus prior to chloroplast import and polyprotein processing. *J. Biol. Chem.* **270**, 13084–13090.
47. Sulli, C. & Schwartzbach, S.D. (1996) A soluble protein is imported into *Euglena* chloroplasts as a membrane-bound precursor. *Plant Cell* **8**, 43–53.

48. Enomoto. T., Sulli, C. & Schwartzbach, S.D. (1997) A soluble chloroplast proteaseprocesses the *Euglena* polyprotein precursor to the light harvesting chlorophyll a/b binding protein of photosystem II. *Plant Cell Physiol*. **38**, 743–746.
49. Hiller, R.G., Wrench, P.M. & Sharples, F.P. (1995) The light harvesting chlorophyll a-c-binding protein of dinoflagellates: a putative polyprotein. *FEBS Lett*. **363**, 175–178.
50. Plaumann, M., Pelzer-Reith, B., Martin, W. & Schnarrenberger, C. (1997) Multiple recruitment of class-I aldolase to chloroplasts and eubacterial origin of eukaryotic class-II aldolases revealed by cDNAs from *Euglena gracilis*. *Curr. Genet*. **31**, 430–438.
51. Kuroda, I., Inagaki, J. & Yamamoto, Y. (1993) Precursor of the nuclear-encoded extrinsic 30 kDa protein in photosystem II of *Euglena gracilis* Z is not a polyprotein. *Plant Mol. Biol*. **21**, 171–176.
52. Montané, M. & Kloppstech, K. (2000) The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? *Gene* **258**, 1–8.
53. Dean, C., Pichersky, E. & Dunsmuir, P. (1989) Structure, evolution, and regulation of *RbcS* genes in higher plants. *Annu. Rev. Plant Physiol*. **40**, 415–439.
54. Durnford, D.G., Deane, J.A., McFadden, G.I., Gantt, E. & Green, B.R. (1999) A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J. Mol. Evol*. **48**, 59–68.
55. Van Dooren, G.G., Schwartzbach, S.D., Osafune, D. & McFadden, G. (2001) Translocation of proteins across the multiple membranes of complex plastids. *Biochim. Biophy. Acta* **1541**, 34–53.
56. Timmis, J.N., Ayliffe, M.A., Huang, C.Y. & Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Gen*. **5**, 123–136.
57. Canback, B., Anderson, S.G. & Kurland, C.G. (2002) The global phylogeny of glycolytic enzymes. *Proc. Natl Acad. Sci. USA* **99**, 6097–6102.
58. Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., Bryant, D., Steel, M.A., Lockhart, P.J., Penny, D. & Martin, W. (2004) A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol*. **21**, 1643–1660.
59. Penny, D., Foulds, L.R. & Hendy, M.D. (1982) Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* **297**, 197–200.
60. Theissen, U., Hoffmeister, M., Grieshaber, M. & Martin, W. (2003) Single eubacterial origin of eukaroytic sulfide: quinone oxidoreductase, a mitochondrial enzyme conserved from the early evolution of eukaryotes during anoxic and sulfidic times. *Mol. Biol. Evol*. **20**, 1564–1574.