Eur. J. Biochem. **268**, 2678–2686 (2001) © FEBS 2001

Purification and cloning of chloroplast 6-phosphogluconate dehydrogenase from spinach

Cyanobacterial genes for chloroplast and cytosolic isoenzymes encoded in eukaryotic chromosomes

Karsten Krepinsky[1], Marlies Plaumann[1], William Martin[2] and Claus Schnarrenberger[1]

[1]*Institute of Biology, Free University of Berlin, Germany;* [2]*Institute of Botany III, University of Düsseldorf, Germany*

Previous attempts to purify chloroplast 6-phosphogluconate dehydrogenase (cp6PGDH), a key enzyme of the oxidative pentose phosphate pathway, have been unsuccessful due to rapid activity loss. An efficient purification protocol was developed and the enzyme from spinach leaves was purified 1000-fold to apparent homogeneity with a specific activity of 60 U·mg^{-1}. The enzyme is a homodimer with subunits of 50 kDa. Antibodies raised against the purified cp6PGDH detected a 53-kDa protein from a crude extract, indicating alterations during purification. Purified cp6PGDH was microsequenced and the corresponding spinach cDNA was cloned using PCR techniques and degenerate primers. The cDNA for cytosolic 6PGDH from spinach was cloned for comparison. Phylogenetic analysis in the context of available homologues from eukaryotes and eubacteria revealed that animal and fungal cytosolic 6PGDH sequences are more similar to their homologues from γ-proteobacteria, whereas plant 6PGDH is more similar to its cyanobacterial homologues. The ancestral gene for higher plant 6PGDH was acquired from the antecedent of plastids through endosymbiosis and gene transfer to the nucleus. A subsequent gene duplication gave rise to higher plant cytosolic 6PGDH, which assumed the function of its pre-existing cytosolic homologue through endosymbiotic gene replacement. The protein phylogeny of both 6PGDH and of the first enzyme of the oxidative pentose phosphate pathway, glucose-6-phosphate dehydrogenase, indicate a surprisingly close relationship between the plant and *Trypanosoma brucei* lineages, suggesting that *T. brucei* (a relative of *Euglena gracilis*) may be secondarily nonphotosynthetic.

Keywords: antibodies; endosymbiosis; evolution; microsequencing; oxidative pentose phosphate pathway.

Much of plant metabolism is an inheritance from the cyanobacterial ancestors of plastids. Yet higher plant plastid genomes encode only 3–4% of the roughly 2300 proteins that are targeted to the organelle [1]. The vast majority of proteins are imported into plastids as transit peptide-bearing precursors that are encoded by nuclear genes, many of which were transferred from the ancestral plastid genome to the nucleus during the course of endosymbiosis [2,3]. Thoughts on this process, called endosymbiotic gene transfer, have long entailed the notion that the products of genes donated from organelles to the nucleus are specifically targeted back into the organelle from which the genes descended, an idea concisely formulated early on by Weeden [4]. But experimental scrutiny of this view has revealed that the products of nuclear genes that were acquired from organelles are very often rerouted to other cell compartments during evolution [2]. Comparative sequence analysis of chloroplast-cytosol isoenzymes for the pathways of central sugar phosphate metabolism in higher plants have played an important role in clarifying this issue, because they test the prediction that the gene for the plastid isoenzyme should reflect the history of the cyanobacterial symbiont, whereas the cytosolic enzyme should reflect the history of the host.

The evolutionary origins of the enzymes common to the Calvin cycle in chloroplast and glycolysis/gluconeogenesis in the cytosol have been mostly clarified for higher plants (reviewed in [2]). But the oxidative pentose phosphate pathway, which provides NADPH for reductive biosyntheses and for protection against oxidative stress, pentoses for synthesis of nucleotides [5] and sugar phosphates for the shikimate pathway, also involves several chloroplast-cytosol isoenzymes. The most important of these are those catalyzing the first two reactions of the pathway, glucose-6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGDH). Both enzymes are present in the chloroplast and the cytosol of green leaves of higher plants [6] and in nonphotosynthetic tissues such as etiolated radish cotyledons [7] and castor bean endosperm [8]. However, the subsequent enzymes of the oxidative pentose phosphate pathway are not always present as distinct

Correspondence to C. Schnarrenberger, Institut für Biologie, Königin-Luise-Str. 12–16a, 14195 Berlin, Germany.
Fax: + 49 30 838 54313, Tel.: + 49 30 838 53123,
E-mail: schnarre@zedat.fu-berlin.de

Abbreviations: cp, chloroplast; 6PGDH, 6-phosphogluconate dehydrogenase; G6PDH, glucose-6-phosphate dehydrogenase.
Enzymes: 6-phosphogluconate dehydrogenase (EC 1.1.1.44); glucose-6-phosphate dehydrogenase (EC 1.1.1.49).
Note: the sequences reported in this paper have been deposited in GenBank under the accession numbers AF 295670 (spinach chloroplast 6-phosphogluconate dehydrogenase) and AF 307144 (spinach cytosolic 6-phosphogluconate dehydrogenase).
(Received 19 December 2000, accepted 12 March 2001)

isoenzymes. For example, the enzymes of the regenerative part of the pathway are exclusively located in the chloroplasts of spinach leaves, so that the product of the two cytosolic dehydrogenases, ribulose-5-phosphate, has to be transported back into the chloroplast for further conversion in this pathway [9]. A specific translocator that transports cytosolic ribulose-5-phosphate into chloroplasts was recently identified [10].

The first enzyme of the oxidative pentose phosphate pathway is G6PDH. The cytosolic isoenzyme of G6PDH is very stable and is readily purified from potato [11] and pea leaves [12,13]. By contrast, the chloroplast enzyme is very unstable and is usually lost during the purification procedure. Genes for the cytosolic and chloroplast isoenzymes from potato have been isolated [11,14,15].

The second dehydrogenase of the oxidative pentose phosphate pathway is 6PGDH. It catalyses the $NADP^+$-dependent oxidative decarboxylation of 6-phosphogluconate to ribulose-5-phosphate. The cytosolic isoenzyme of 6PGDH has been studied in tomato [16] and maize [17], purified from corn roots [18] and beet leaves [19] and cloned from *Medicago sativa* [20].

But like G6PDH, the chloroplast isoenzyme of 6PGDH is extremely unstable and has resisted all efforts towards its purification to date; consequently its gene has not yet been identified. Here we report an efficient isolation procedure for active chloroplast 6PGDH, purification of the enzyme to apparent electrophoretic homogeneity, microsequencing of the protein and its cDNA cloning. A comparison of sequences for chloroplast and cytosolic 6PGDH to homologues from the cytosol of nonphotosynthetic eukaryotes and from available eubacteria indicates that the cyanobacterial imprint on plant metabolism extends beyond the confines of the chloroplast itself and into the cytosol as well. Furthermore, the phylogenies of both G6PDH and 6PGDH suggest that the imprint of plastids that were secondarily lost may be preserved in the human parasite *Trypanosoma brucei*.

MATERIALS AND METHODS

Protein purification

All procedures were carried out at 4 °C unless otherwise indicated. The cp6PGDH was purified within two days. One hundred grams of 4–5-week-old spinach leaves (*Spinacia oleracea* L. var. Monnopa) was homogenized in 200 mL buffer A (20 mM Tris/HCl, pH 7.5, 10 mM 2-mercaptoethanol, 1 mM EDTA) with 10 g polyvinylpolypyrrolidone (Sigma, Deisenhofen, Germany) in a Waring blender. The homogenate was filtered through four layers of Miracloth (Calbiochem, Bad Soden, Germany) and centrifuged at 24 000 ***g*** for 20 min. The supernatant was precipitated with 0.8% (v/v) of a 10% (w/v) polyethyleneimine solution (M_r 60 000; Sigma, Deisenhofen, Germany) and centrifuged at 24 000 ***g*** for 20 min. The supernatant was applied to a Fractogel TSK-DEAE 650 S (Merck, Darmstadt, Germany) column (3 × 14 cm) equilibrated with buffer A. The column was washed with 200 mL buffer A and proteins were eluted with a linear 400-mL gradient of 0–200 mM KCl in buffer A at a flow rate of 2 mL·min^{-1}. Fractions of 4 mL were collected. Two peaks of 6PGDH activity were separated; the cytosolic 6PGDH eluted at 60 mM KCl and the cp6PGDH eluted at 160 mM KCl as previously described [6]. Fractions containing cp6PGDH were pooled and concentrated by ultrafiltration with Macrosep 50 K (Pall Gelman, Dreieich, Germany) at 4000 ***g*** to about 10 mL. This solution was stored overnight on ice. The next day the solution was concentrated by ultrafiltration to 2 mL and applied at room temperature to a Superdex G200 column (Amersham Pharmacia Biotech, Freiburg, Germany) equilibrated with buffer B (buffer A containing 200 mM KCl) at a flow rate of 1 mL·min^{-1}. Fractions of 1 mL were collected and those with cp6PGDH activity were pooled and applied to a 5-mL ADP-Agarose column (Sigma, Deisenhofen, Germany) equilibrated with buffer A. The column was washed with 10 mL buffer B, 10 mL buffer B containing 5 mM NAD^+, 10 mL 20 mM K_2HPO_4/KH_2PO_4 buffer, pH 5.6, with 10 mM 2-mercaptoethanol, 5 mL buffer B containing 0,5 mM $NADP^+$ and 5 mL buffer B containing 1 mM $NADP^+$. The cp6PGDH was eluted with 5 mL buffer B containing 2 mM $NADP^+$ and concentrated by ultrafiltration to 0.3–0.5 mL. The solution was directly used for microsequencing and immunization or stored at −20 °C in the presence of 50% (v/v) glycerol.

Other biochemical methods

6PGDH activity was measured photometrically (Uvikon 930, Bio-Tek Instruments, Neufahrn, Germany) at room temperature in 1 mL of 50 mM Tris/HCl, pH 8.0, 7.5 mM $MgCl_2$, 1 mM 6-phosphogluconate, and 0.18 mM $NADP^+$ at 340 nm. One unit is the amount of enzyme that catalyses the 6-phosphogluconate-dependent reduction of 1 μmol of $NADP^+$ in 1 min. The Michaelis–Menten constants for 6-phosphogluconate and $NADP^+$ were determined by varying the concentration of substrate or cofactor, keeping the other components of the reaction mixture constant. A Lineweaver-Burk plot was used to determine K_m values.

Proteins were determined according to Bradford [21] using bovine serum albumin as a standard. SDS/PAGE was performed as described by Laemmli [22]. Proteins in gels were detected by silver staining [23]. The apparent molecular mass of the cp6PGDH was estimated by gel filtration (Superdex G200) during protein purification with ferritin (450 kDa), catalase (240 kDa), aldolase (160 kDa), BSA (67 kDa) and chymotrypsinogen (25 kDa) as calibration standards. Native molecular masses were interpolated from the plot of elution volume versus log M_r.

Microsequencing

The purified protein of cp6PGDH was submitted to microsequencing. There were two attempts to determine the N-terminus of the purified protein. The native protein was directly applied onto a poly(vinylidene difluoride) membrane with the Prosorb sample preparation cartridge (PE Biosystems, Weiterstadt, Germany). In another attempt the protein was submitted to SDS/PAGE and semidry blotted onto a poly(vinylidene difluoride) membrane prior to sequencing.

To obtain internal sequence information 10 μg of the purified cp6PGDH was concentrated by SDS/PAGE, followed by in-gel digestion with the endoproteinase LysC, peptide elution and separation of the peptides with

RP-HPLC. Four peptides (K19, K24, K31 and K38) were sequenced by amino-terminal Edman degradation in an automatic Procise 492 sequencer (PE Biosystems, Weiterstadt, Germany) by a commercial service (Toplab, München, Germany).

Polyclonal antibodies and Western blot

A chicken was immunized with 60 μg of the purified cp6PGDH by a commercial service (Biogenes, Berlin, Germany). Polyclonal antibodies were purified from egg yolk and resuspended in storage buffer (20 mM Tris/glycine/HCl, pH 7.5, 250 mM NaCl, 0.02% thimerosal) at a final concentration of 9 $mg \cdot mL^{-1}$.

Proteins were separated by SDS/PAGE and then semidry blotted onto a nitrocellulose membrane [24]. Transfer efficiency was checked by using a prestained molecular mass standard (SeeBlue, Invitrogen, Groningen, the Netherlands). The following steps were performed at room temperature with shaking. After blocking the membrane in 1% casein in NaCl/Tris for 20 min, the blot was incubated with anti-cp6PGDH polyclonal antibodies diluted 1 : 1000 in NaCl/Tris (pH 8.0) overnight. The membrane was washed three times in NaCl/Tris containing Tween-20 (0.1%) and incubated with rabbit anti-(chicken IgG) IgG alkaline phosphatase conjugate (Biogenes, Berlin, Germany), diluted 1 : 1000 in NaCl/Tris, for 1–2 h. The membrane was washed three times in NaCl/Tris containing Tween-20 and twice in NaCl/Tris. The alkaline phosphatase was detected colorimetrically with Nitro Blue tetrazolium/5-bromo-4-chloroindol-2-yl phosphate tablets (Roche, Mannheim, Germany).

Cloning

The degenerate oligonucleotides k38forward 5′-GTKTAYA-AYMGRACMGC-3′ and k31reverse 5′-GCNARRTCRT-CYTTNAC-3′ were designed against the peptides K38 and K31, respectively, determined by microsequencing. These were used in a PCR reaction containing PCR buffer supplied with the polymerase, 1.5 mM $MgCl_2$, 250 μM of each dNTP, 2 μM of k38forward, 4 μM of k31reverse, 1 U *Taq* DNA polymerase (Roche, Mannheim, Germany), and 3 μg of M13 amplified spinach cDNA of a Lambda ZAP II cDNA library from young spinach leaves (Stratagene, Heidelberg, Germany) in a total volume of 50 μL in a T3 Thermocycler (Biometra, Göttingen, Germany). After an initial step of 3 min at 94 °C, PCR was performed for 35 cycles of 1 min at 48 °C, 30 s at 72 °C and 45 s at 94 °C, followed by a terminal step of 5 min at 72 °C. The resulting 653-bp amplification product was thymine/adenine (TA)-cloned into the pGEM-T Easy vector (Promega, Mannheim, Germany), transformed into *Escherichia coli* JM109 as recommended by the supplier, and verified by sequencing.

In addition, a 227-bp amplification product was obtained by PCR with conditions as above using the degenerate oligonucleotides k31forward(1/2) 5′-GTNAARGAYGAY-YT(R/Y)GC-3′ and k24reverse 5′-ATNGCRTTNACYT-CTTC-3′ (2 μM each) against the peptides K31 and K24, respectively. The sequence of the complete cDNA of cp6PGDH was determined with the oligonucleotides 653forward 5′-GCCTTGTGTTACATACATCGG-3′ and 653reverse 5′-CCGATGTATGTAACACAAGGC-3′against the 653-bp amplification product and the respective vector primers T7 and T3. PCR and cloning of the amplification product was performed as above with 2 ng of cDNA from the spinach library, 1 μM of each primer, and 35 cycles each of 1 min at 60 °C, 1 min with primer 653forward and 30 s with primer 653reverse at 72 °C and 45 s at 94 °C. Finally, the coding sequence was cloned with primers designed against the N- and C-termini of the cp6PGDH.

The cDNA of cytosolic 6PGDH from spinach was obtained by PCR as above using the primers cytforward (1/2) 5′-GARTGGTAY(G/C)ARAAYAC-3′ and cytreverse 5′-TGCATRTCNCCRTAYTC-3′ against the conserved regions EWY(E/Q)NT and EYGDMQ, respectively, of cytosolic 6PGDH from various plants. The resulting 275-bp amplification product was used to obtain the complete cDNA as above.

Standard molecular techniques were performed as described by Sambrook *et al.* [25]. Cloned PCR products were sequenced by a commercial service (Seqlab, Göttingen, Germany). Primers were also synthesized by a service (Biotez, Berlin, Germany).

Phylogenetic data analysis

6PGDH sequences from various taxa were extracted from GenBank. Standard sequence processing was performed with the GCG Package [26]. Sequence alignments were created with CLUSTALW and refined by eye. Distance between sequences for phylogenetic inference with the neighbor-joining method [27] was measured as numbers of amino-acid substitutions per site corrected for multiple substitutions with the Dayhoff option of PHYLIP [28]. Alternatively, sequence alignments were created with PILEUP, refined by eye and trees were inferred by protein maximum likelihood with PROTML using the JTT matrix starting from the neighbor-joining tree of ML distances [29]. Alignments are available from the authors upon request.

RESULTS AND DISCUSSION

Purification and characterization of spinach leaf cp6PGDH

Chloroplast 6PGDH from spinach leaves was purified 1000-fold from a crude extract by the following steps: polyethyleneimine precipitation, anion-exchange chromatography on DEAE–Fractogel, gel filtration in Superdex G200, and by affinity chromatography on ADP-Agarose (Table 1). Chloroplast and cytosolic isoenzymes of 6PGDH were separated by anion-exchange chromatography [6]. The main difficulty during purification was the stabilization of the enzyme. This could be achieved by the addition of 2-mercaptoethanol to the buffers while dithiothreitol, another reducing agent, was far less effective. In addition, the procedure had to be performed as quickly as possible (i.e. within 36 h after homogenization). The final specific activity of the enzyme preparation was 60 $U \cdot mg^{-1}$. In the final preparation the enzyme was stable in buffer and 50% glycerol for at least 6 months. The recovery was about 7% of the original total activity, however, considering that the chloroplast isoenzyme accounts only for about 30% of the total activity in the crude extract [6,9,30], an actual

Table 1 Purification of cp6PGDH from spinach leaves.

Purification step	Protein (mg)	Activity (U)	Specific activity ($U \cdot mg^{-1}$)	Purification factor	Yield (%)
Crude extract	920	53 [a]	0.06	1	100
PEI precipitation	710	52 [a]	0.07	1.2	98
DEAE–Fractogel	17.5	13	0.74	12	25
Superdex G200	3.3	9.1	2.8	47	17
ADP-Agarose	0.06	3.6	60	1000	7

[a] Sum of activity of cp6PGDH and cytosolic 6PGDH.

recovery of about 23% can be assumed. On SDS/PAGE (Fig. 1A) only a single protein band of 50 kDa was detected when silver stained. The native enzyme showed a mass of 107 kDa as determined by gel filtration (data not shown). The native protein is thus a homodimer.

The biochemical properties of purified cp6PGDH were determined. The enzyme showed classical Michaelis–Menten kinetics with both substrate and cofactor. The K_m (6-PG) was 40 μM and the K_m (NADP) was 6 μM, confirming previous data with a nonpurified enzyme preparation [6].

Polyclonal antibodies were raised in chicken and isolated from egg yolk. In Western blots the antibodies recognized a 53-kDa protein in a crude extract from spinach and a 50-kDa protein from the purified preparation (Fig. 1B), the same molecular mass as determined by SDS/PAGE and silver staining. Therefore, degradative changes must have altered the full-length protein. This is supported by the fact that two proteins of 53 and 50 kDa were detected by anti-cp6PGDH IgG in several-day-old crude extracts. The cytosolic isoenzyme was also recognized by the antibodies against the cp6PGDH in a dot-blot assay with the chloroplast and cytosolic 6PGDH after separation on DEAE–Fractogel (data not shown), probably due to the high sequence identity of 74% between the two proteins (see below).

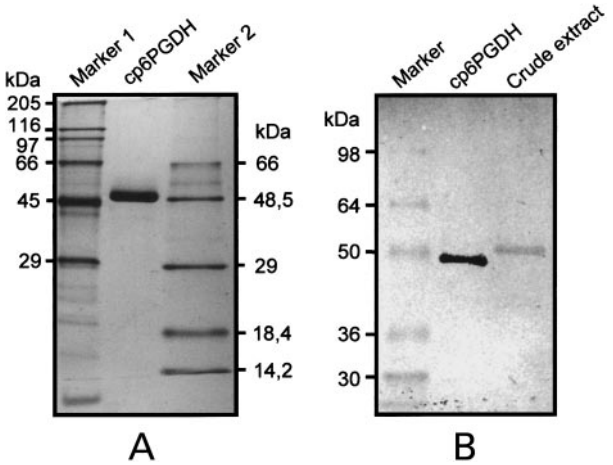

Fig. 1. Purification of cp6PGDH from spinach leaves. (A) SDS/PAGE of the purified cp6PGDH from spinach leaves after silver staining. Marker 1, high molecular mass standard; cp6PGDH, purified cp6PGDH (1 μg); marker 2, low molecular mass standard. (B) Western Blot of the purified cp6PGDH and proteins of a crude extract from spinach leaves. Marker, prestained molecular mass standard; cp6PGDH, purified cp6PGDH (300 ng); crude extract, crude extract from spinach leaves (30 μg).

Microsequencing and cloning of cp6PGDH

Attempts to sequence the undigested protein failed, probably due to a blocked N-terminus. When the enzyme was digested with endoproteinase LysC, the sequences of four of the resulting peptides were obtained as follows: K19 (K)ILEAAGMK, K24 (K)EEVNAIRGGVDK, K31 (K)VKDDLADGGLVDK, and K38 (K)GFPISVYNRTASK. These peptides were used to design degenerate primers and to isolate gene-specific fragments by PCR from a cDNA library of spinach. The complete cDNA of cp6PGDH was determined by PCR with gene-specific primers and subsequent cloning and sequencing of the amplification products.

The cDNA of cp6PGDH comprised an open reading frame encoding 537 amino acids. A molecular mass of the protein of 58 301 Da was calculated. With a methionine at the theoretical N-terminus of the mature protein, a serine-rich transit peptide for chloroplast import of 45 amino acids was encountered. Without the transit sequence a molecular mass of 53 529 Da for the 492-amino acid large subunit was calculated. A theoretical isoelectric point of 5.25 was determined. All four peptide sequences determined were found in the deduced amino-acid sequence of the cp6PGDH cDNA (Fig. 2).

The cytosolic 6PGDH was also cloned with the same method as used for the cp6PGDH except with degenerate primers against conserved regions of known plant cytosolic sequences. The gene encodes a protein of 53 252 Da with 483 amino acids and a hypothetical pI of 5.96. The chloroplast and cytosolic isoenzymes are therefore of almost identical size and can hardly be separated on SDS/PAGE, only on native PAGE because of differing isoelectric points [6].

Cell compartment-specific 6PGDH isoenzymes

Chloroplast 6PGDH has previously been recalcitrant to purification. Here the enzyme from spinach was purified to homogeneity. It is a homodimer with subunits of 50 kDa. This is in accordance with cytosolic 6PGDHs purified from maize and beet with subunits of 55 and 52 kDa, respectively [18,19]. All four peptide sequences determined directly from the purified chloroplast enzyme by microsequencing were encountered in the deduced amino-acid

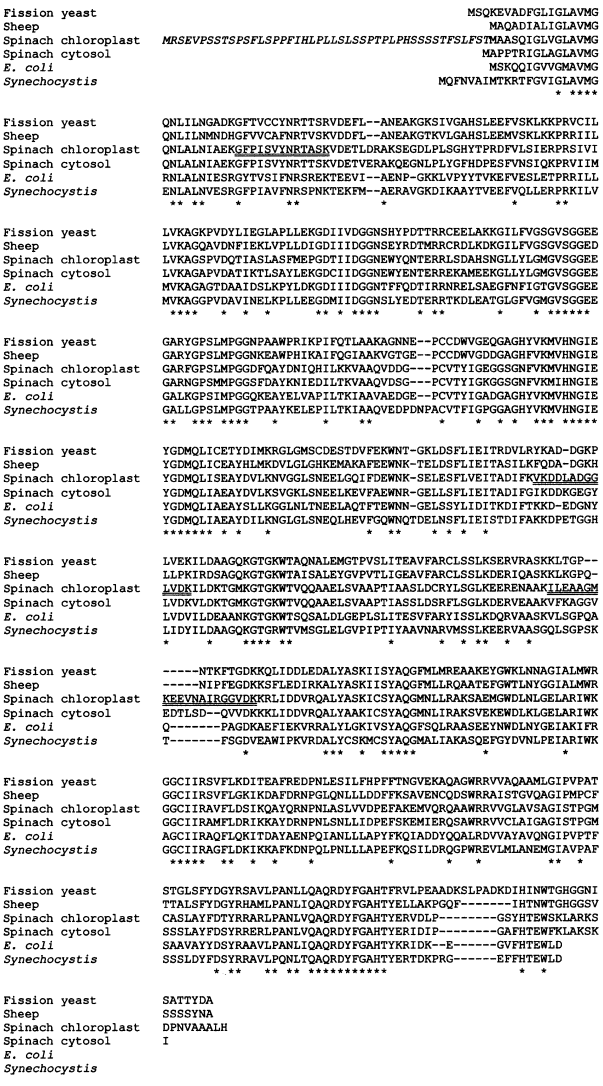

Fig. 2. Comparison of spinach chloroplast and cytosolic 6PGDHs deduced from cDNA sequences. Double-underlined regions indicate peptide sequences determined by microsequencing from cp6PGDH purified from spinach. The transit peptide of the cp6PGDH is indicated in italics, with methionine as the theoretical N-terminus of the mature protein. Conserved residues are indicated by an asterisk, and gaps by a dash. The other sequences are representative of eubacteria, animals and fungi (see legend to Fig. 3 for accession numbers).

sequence of the cDNA, indicating that it encodes the chloroplast isoenzyme of 6PGDH. Purified cp6PGDH from spinach has a blocked N-terminus, as is also the case for the cytosolic G6PDH of potato [11]. So although the open reading frame encodes an N-terminal serine-rich transit peptide [31] for chloroplast import, the precise processing site could not be directly determined. The sequence of spinach cp6PGDH is easily alignable to 6PGDHs from prokaryotes and eukaryotes, sharing its greatest similarity with plant and cyanobacterial homologues.

For direct comparison, a spinach homologue of cytosolic 6PGDH from higher plants was also cloned and sequenced. Although it is generally accepted that both of the two dehydrogenases of the oxidative pentose phosphate pathway in plants are compartmentalized in the plastids and the cytosol, respectively, there is newer evidence that an isoenzyme may also exist in peroxisomes of pea leaves as well [32]. However, the sequence of this gene and its targeting signal have not been determined yet.

Phylogeny and prokaryote-eukaryote transitions of 6PGDH genes

6PGDHs are highly conserved enzymes 467–492 amino acids in length with sequence identities of 40% and higher (*T. brucei* only 35%) in prokaryote-eukaryote comparisons (Fig. 3). Plant cytosol and chloroplast isoenzymes of 6PGDH are more similar to their cyanobacterial homologues than they are to homologues from any other prokaryote or from the cytosol of nonphotosynthetic eukaryotes. This rather straightforwardly indicates that the genes for plant 6PGDH were acquired by the nucleus via endosymbiotic gene transfer from the cyanobacterial antecedents of chloroplasts.

Higher plant 6PGDHs form two adjacent clusters in the tree, one comprising only cytosolic sequences from various plants including a cytosolic sequence from *Arabidopsis thaliana*, the other comprising the cp6PGDH from spinach and two additional sequences from *A. thaliana* genome sequencing data designated here as *A. thaliana* 1 and *A. thaliana* 2 [33]. *A. thaliana* 1 and 2 are not annotated as possessing chloroplast transit peptides, but those might have been missed in annotation and the similarity to the spinach chloroplast enzyme is suggestive that they might encode chloroplast isoenzymes as well.

Clearly, the genes for chloroplast and cytosolic 6PGDH in higher plants arose from a duplication of the gene that was acquired from the ancestral plastid genome. The host that engulfed the cyanobacterial symbiont must also have possessed a cytosolic 6PGDH, as eukaryotes studied to date do so, but this enzyme was clearly replaced by the cyanobacterial intruder through endosymbiotic gene replacement. In previous evolutionary studies of enzymes of carbohydrate metabolism, one other case has been observed where an intruding cyanobacterial gene has replaced its preexisting cytosolic homologue, chloroplast and cytosolic 3-phosphoglycerate kinase [2,34].

The pre-existing cytosolic 6PGDH that was replaced by the cyanobacterial acquisition in plants, were it still around, probably would have branched with its homologues from the cytosol of animals and fungi (Fig. 3). But notably, on the basis of the available sequences sampled, cytosolic 6PGDH from animals and fungi also appear to be acquisitions from eubacteria. Unfortunately, there are currently no 6PGDH sequences available from α-proteobacteria, from which mitochondria are thought to have arisen [35], for comparison. The closest relatives in the current sample to animal and fungal 6PGDH are the sequences from the γ-proteobacteria *Haemophilus* and *Actinobacillus* (Fig. 3), whereby it should be noted that there are also γ-proteobacterial sequences (*E. coli* and relatives) that branch in a very different portion of the tree, next to Gram-positive bacteria (Fig. 3). Overall, this pattern of similarity indicates that, like plants, the animal and fungal lineages have also acquired their nuclear genes for 6PGDH through endosymbiotic gene transfer, the most reasonable scenario being that the donor was the antecedent of mitochondria. If so, one prediction would be that α-proteobacteria will eventually

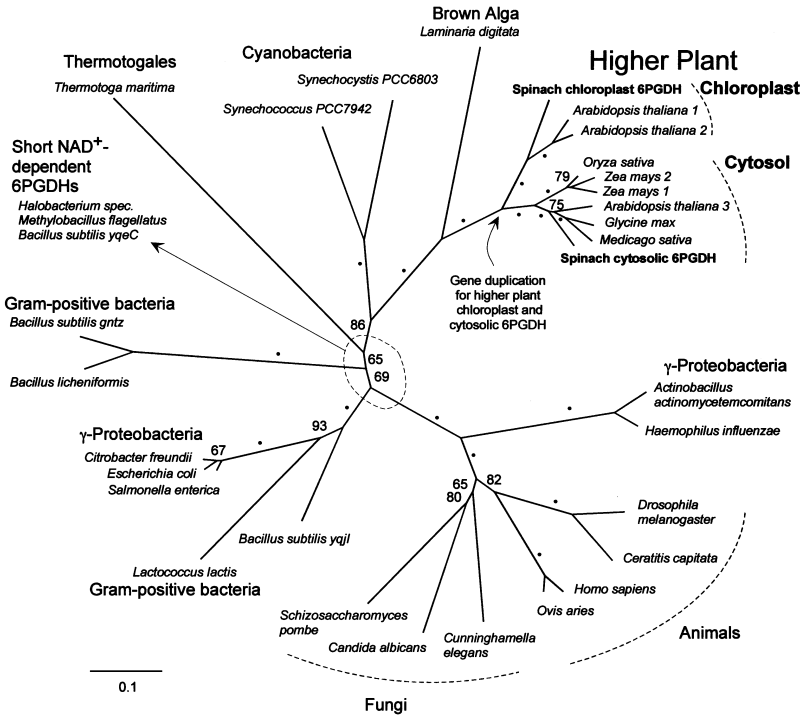

Fig. 3. 6PGDH protein phylogeny based on an alignment of 482 positions. The tree was constructed by the neighbor-joining method from a matrix of estimated numbers of amino-acid substitutions per site calculated with the Dayhoff option of PHYLIP. Numbers near branches indicate the bootstrap proportion for 100 replicas using the same method. The scale bar indicates 0.1 substitutions per site. Branches with > 98% RELL bootstrap support are indicated by a dot. Taxonomic affinities and subcellular localization are indicated. The approximate position of the branch with the short 6PGDHs of *Halobacterium sp.*, *B. subtilis yqeC* and *M. flagellatus* is indicated by an arrow (see text). Sequences were extracted from GenBank. Accession numbers to sequences shown are *Actinobacillus actinomycetemcomitans*, P70718; *A. thaliana* 1, AC007764; *A. thaliana* 2, AB005233; *A. thaliana* 3, AC068900; *B. licheniformis (gntz)*, JC2306; *B. subtilis gntz*, D26190; *B. subtilis yqjI*, A69964; *Candida albicans*, AB006102; *Ceratitis capitata*, P41570; *Citrobacter freundii*, U14466; *Cunninghamella elegans*, Y17297; *Drosophila melanogaster*, P41572; *E. coli*, P00350; *Glycine max*, AB007907; *Haemophilus influenzae*, U32737; *Homo sapiens*, P52209; *Lactococcus lactis*, U74322; *Laminaria digitata*, AJ130772; *Medicago sativa*, U18239; *Oryza sativa*, AP001552; *Ovis aries*, P00349; *Salmonella enterica*, U14509; *Schizosaccharomyces pombe*, D89161; *S. oleracea* chloroplast and cytosolic 6PGDH, this paper; *Synechococcus PCC7942*, P21577; *Synechocystis PCC6803*, P52208; *Thermotoga maritima*, AE001722; *Z. mays* 1, AF061837; *Z. mays* 2, AF061838. Accession numbers to short 6PGDHs are *B. subtilis yqeC*, G69950; *Halobacterium sp*. NRC-1, AE005131; *M. flagellatus*, AF167580.

be found that possess an even closer homologue of animal and fungal 6PGDH. Yet it should be kept in mind that the issue of whence eukaryotic genes arose is complicated to no small extent by lateral gene transfer among free-living prokaryotes (for a discussion see [2,36]), a process that is now very well documented [37,38].

T. brucei 6PGDH and G6PDH: evidence of plastids past?

Very few sequences for 6PGDH from protists are available for comparison. An exception is the 6PGDH gene from the trypanosomatid human parasite *T. brucei*, a functional enzyme for which the structure is known [39]. But the sequence of *T. brucei* is highly divergent from other eukaryotic homologues. This divergence was manifested in phylogenetic analyses as an extremely long branch for the sequence, which was positioned unstably in protein phylogenies. This was apparently founded in many highly variable stretches in the *T. brucei* sequence at regions that were highly conserved in the other sequences of the alignment (data not shown). When these regions (that is, the ones where the *T. brucei* sequence was highly deviant) were excluded from the data, 240 well-conserved positions from the original 500 positions were left in the alignment, and the *T. brucei* sequence showed good conservation with the other sequences at those positions. Using that data set, the *T. brucei* homologue branched with its homologues from plants, not robustly but at a position above the cyanobacterial homologues, as shown in Fig. 4A.

The possession of an otherwise apparently plant-specific 6PGDH homologue in this nonphotosynthetic and nonplastid-bearing human parasite was intriguing, because some human parasites, most notably the apicomplexans, are very well known to be secondarily nonphotosynthetic although they still possess a plastid [40]. To see if other enzymes from this pathway in *T. brucei* showed a similar affinity to plant homologues, we examined the phylogeny of the other key enzyme of the oxidative pentose phosphate pathway, G6PDH. A previous report had indicated that eukaryotic G6PDH sequences are of eubacterial origin [41] and more recently, the *T. brucei* sequence has been reported [42]. A ProtML phylogeny of G6PDH revealed that also in the case of G6PDH enzyme, the *Trypanosoma* homologue branched with its counterparts from plants (Fig. 4B). At face value, these data would tend to suggest a common ancestry of the kinetoplastid and higher plant lineages and

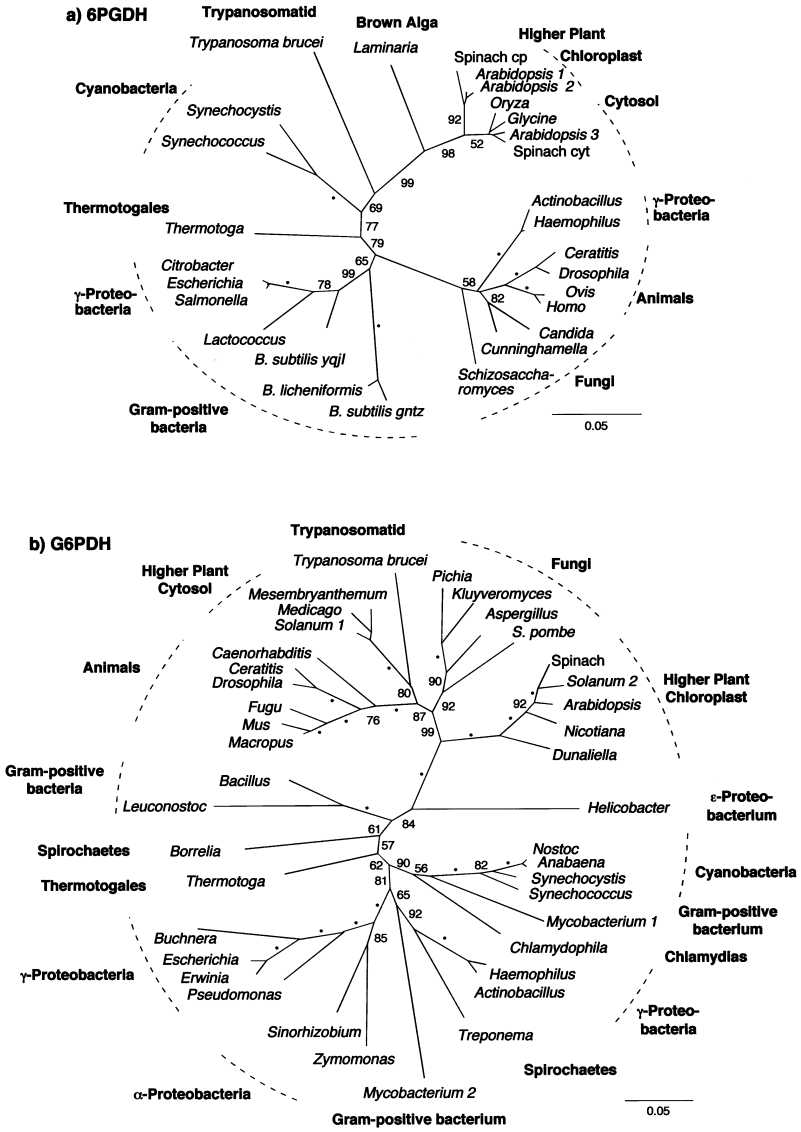

Fig. 4. Comparison of protein maximum likelihood phylogenies of 6PGDH and G6PDH. Taxonomic affinities and subcellular localization are indicated. Branches with > 98% RELL bootstrap support are indicated by a dot. The scale bar indicates 0.05 substitutions per site. (A) Tree of 6PGDH sequences including the highly divergent sequence from *T. brucei* (accession number P31072), based on an alignment of 240 highly conserved positions. (B) Tree of G6PDH sequences, based on an alignment of 663 positions and extending the analysis of Wendt *et al.* [41] by inclusion of the recently cloned *T. brucei* sequence [42]. Accession numbers to G6PDH sequences shown are *A. actinomycetemcomitans,* P77809; *Anabaena PCC7120*, P48992; *A. thaliana*, AJ001359; *Aspergillus niger*, P48826; *B. subtilis*, P54547; *Borrelia burgdorferi*, O51581; *Buchnera aphidicola*, P57405; *Caenorhabditis elegans*, Q27464; *C. capitata*, P41571; *Chlamydophila pneumoniae*, Q9Z8U6; *D. melanogaster*, P12646; *Dunaliella bioculata*, CAB52685; *Erwinia chrysanthemi*, P37986; *E. coli*, P22992; *Fugu rubripes*, P54996; *H. influenzae*, P44311; *Helicobacter pylori*, P56110; *Kluyveromyces lactis*, P48828; *Leuconostoc mesenteroides*, P11411; *Macropus robustus*, Q29492; *M. sativa*, Q42919; *Mesembryanthemum crystallinum*, AAD11426; *Mus musculus*, P97324; *Mycobacterium tuberculosis*1, O06573; *Mycobacterium tuberculosis*2, O08407; *Nicotiana tabacum*, Q43793; *Nostoc punctiforme*, P48848; *Pichia jadinii*, P11410; *Pseudomonas aeruginosa*, O68282; *S. pombe*, O00091; *Sinorhizobium meliloti*, Q9Z3S2; *Solanum tuberosum1*, P37830; *Solanum tuberosum2*, Q43839; *S. leracea* O24357; *Synechococcus PCC7942*, P29686; *Synechocystis PCC6803*, P73411; *T. maritima* Q9X0N9; *Treponema pallidum*, O83491; *T. brucei*, CAC07816; *Zymomonas mobilis*, P21907.

furthermore, a cyanobacterial origin for the *T. brucei* 6PGDH gene. Would this suggest a photosynthetic ancestry of the trypanosomatid parasite?

This possibility deserves consideration. Euglenids, which are photosynthetic relatives of the trypanosomatids [43], possess plastids of secondary symbiotic origin and several nuclear genes of cyanobacterial origin that were acquired from the plastid genome [44], although it has generally been assumed that the trypanosome lineage diverged from the euglenid lineage prior to the acquisition of plastids in the latter. Several groups of eukaryotes are believed to have secondarily lost their plastids [45], perhaps most notably the oomycetes. These are plastidless, fungal-like organisms with a cellulose cell wall which encode nuclear genes such as the elongation factor EF-1α that reflect the past presence of a cyanobacterial symbiont [46]. Clear evidence for genes of plant or cyanobacterial origin in trypanosomatids is lacking, although it should be noted that a recent phylogeny involving several genes uncovered a weak phylogenetic link between the euglenozoan lineage, which includes euglenids plus trypanosomatids, and the plant lineage [47]. That link could conceivably reflect a phylogenetic signal stemming from the past presence of a plastid in the trypanosomatid evolutionary past, congruent with the data presented here. If so, as more data accumulate from trypanosome genome sequencing efforts, additional *T. brucei* genes other than 6PGDH and G6PDH should reflect a photosynthetic history of this parasite.

Lack of classical 6PGDH and G6PDH in archaea

To our knowledge, no archaeal sequence has yet been reported for either 6PGDH or G6PDH (see also [41]), nor was G6PDH activity detected in methanogenic archaea

[48]. Instead, several archaea are known to possess a glucose dehydrogenase as part of the nonphosphorylated Entner–Doudoroff pathway [49]. However, shorter putative 6PGDH-like sequences (about 300 amino acids instead of about 480 amino acids) have been identified through sequence similarity among various prokaryotes, including the methylotrophic β-proteobacterium *Methylobacillus flagellatus*, the Gram-positive *Bacillus subtilis vyeC*, and one halophilic archaeon, *Halobacterium sp*. The short sequence of *M. flagellatus* is considered to be a candidate for the NAD^+-specific 6PGDH of the cyclic formaldehyde oxidation [50]. The short 6PGDH sequences are 31–38% identical to the N-terminal region of classical 6PGDHs. They lack the C-terminus with its key residues for substrate binding [51] and have altered cofactor binding sites. Their long branches to a common ancestor with classical 6PGDHs are schematically indicated in Fig. 3 (long arrow). The classical-type 6PGDHs from *B. subtilis gntz* and *B. licheniformis* are also atypical, because they lack arginine 34, proposed to play an important role in binding the 2′-phosphate of $NADP^+$ [51], implicating a different cofactor specificity, probably NAD^+, with only low affinity to $NADP^+$.

That archaea studied or sequenced to date lack a 6PGDH homologue of the type found in eubacteria and eukaryotes is in agreement with the view that eukaryotes acquired their 6PGDH genes from eubacterial donors. But 6PGDH is by no means unique in this respect because the general pattern observed here, namely that animals and fungi acquired their 6PGDH from a different eubacterial donor than plants, has been found among four other genes of sugar phosphate metabolism: glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, transketolase [2] and glucose-6-phosphate isomerase [52]. Indeed, all of the enzymes of glycolysis, gluconeogenesis, the Calvin cycle and the oxidative pentose phosphate pathway studied to date in eukaryotes seem to be acquisitions from eubacteria [2,52,53] with the single exception of enolase, which is somewhat more similar to the archaeal enzyme [54].

ACKNOWLEDGEMENTS

We thank R. Grübnau for excellent technical assistance and the Deutsche Forschungsgemeinschaft for financial support.

REFERENCES

1. Abdallah, F., Salamini, F. & Leister, D. (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci.* **5**, 141–142.
2. 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.
3. 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.
4. Weeden, N.F. (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J. Mol. Evol.* **17**, 133–139.
5. Wood, T. (1986) Distribution of the pentose phosphate pathway in living organisms. *Cell Biochem. Funct.* **4**, 235–240.
6. Schnarrenberger, C., Oeser, A. & Tolbert, N.E. (1973) Two isoenzymes each of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase in spinach leaves. *Arch. Biochem. Biophys.* **154**, 438–448.
7. Schnarrenberger, C., Tetour, M. & Herbert, M. (1975) Development and intracellular distribution of enzymes of the oxidative pentose phosphate cycle in radish cotyledons. *Plant Physiol.* **56**, 836–840.
8. Nishimura, M. & Beevers, H. (1981) Isoenzymes of sugar phosphate metabolism in endosperm of germinating castor beans. *Plant Physiol.* **67**, 1255–1258.
9. Schnarrenberger, C., Flechner, A. & Martin, W. (1995) Enzymatic evidence for a complete oxidative pentose phosphate pathway in chloroplasts and an incomplete pathway in the cytosol of spinach leaves. *Plant Physiol.* **108**, 609–614.
10. Eicks, M., Deterding, S., Flügge, U.-I., Schubert, S. & Fischer, K. (2000) Transport of pentose phosphates across the inner envelope membrane of plastids and characterization of GPT-homologous phosphate translocator from *Arabidopsis thaliana*. In *Botanikertagung 2000* (Oelmüller, R. & Bergmann, M., eds) p. 53. Finkdruck Apolda, Finland.
11. Graeve, K., von Schaewen, A. & Scheibe, R. (1994) Purification, characterization and cDNA sequence of glucose-6-phosphate dehydrogenase from potato (*Solanum tuberosum* L.). *Plant J.* **5**, 353–361.
12. Srivastava, D.K. & Anderson, L.E. (1983) Isolation and characterization of light- and dithiothreitol-modulatable glucose-6-phosphate dehydrogenase from pea chloroplasts. *Biochim. Biophys. Acta* **724**, 359–369.
13. Fickenscher, K. & Scheibe, R. (1986) Purification and properties of the cytoplasmic glucose-6-phosphate dehydrogenase from pea leaves. *Arch. Biochem. Biophys.* **247**, 393–402.
14. von Schaewen, A., Langenkämper, G., Graeve, K., Wenderoth, I. & Scheibe, R. (1995) Molecular characterization of the plastidic glucose-6-phosphate dehydrogenase from potato in comparison to its cytosolic counterpart. *Plant Physiol.* **109**, 1327–1335.
15. Wendt, U.K., Wenderoth, I., Tegeler, A. & von Schaewen, A. (2000) Molecular characterization of a novel glucose-6-phosphate dehydrogenase from potato (*Solanum tuberosum* L.). *Plant J.* **23**, 723–733.
16. Tanksley, S.D. & Kühn, G.D. (1985) Genetics, subcellular localization, and molecular characterization of 6-phosphogluconate dehydrogenase isozymes in tomato. *Biochem. Genet.* **23**, 441–454.
17. Bailey-Serres, J., Tom, J. & Freeling, M. (1992) Expression and distribution of cytosolic 6-phosphogluconate dehydrogenase isozymes in maize. *Biochem. Genet.* **30**, 233–246.
18. Bailey-Serres, J. & Nguyen, M.T. (1992) Purification and characterization of cytosolic 6-phosphogluconate dehydrogenase isozymes from maize. *Plant Physiol.* **100**, 1580–1583.
19. Signorini, M., Bregoli, A.M., Caselli, L. & Bergamini, C.M. (1995) Purification and properties of 6-phosphogluconate dehydrogenase from beet leaves. *Biochem. Mol. Biol. Int.* **35**, 669–675.
20. Fahrendorf, T., Weiting, N., Shorrosh, B.S. & Dixon, R.A. (1995) Stress responses in alfalfa (*Medicago sativa* L.). XIX. Transcriptional activation of oxidative pentose phosphate pathway genes after the onset of isoflavonoid phytoalexin response. *Plant Mol. Biol.* **28**, 885–900.
21. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
22. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
23. Blum, H., Beier, H. & Gross, H.J. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**, 93–99.
24. Kyhse-Andersen, J. (1984) Electroblotting of multiple gels: a simple apparatus without tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J. Biochem. Biophys. Methods* **10**, 203–209.

25. Sambrook, J., Fritsch, E. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
26. Devereux, J., Haeberli, P. & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
27. Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for the construction of phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
28. Felsenstein, J. (1989) PHYLIP: phylogeny inference package (Version 3.2). *Cladistics* **5**, 164–166.
29. Kishino, H. & Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from the DNA sequence data and the branching order in Hominoidea. *J. Mol. Evol.* **38**, 179–183.
30. Averill, R.H., Bailley-Serres, J. & Kruger, N.J. (1998) Co-operation between cytosolic and plastidic oxidative pentose phosphate pathways revealed by 6-phosphogluconate dehydrogenase-deficient genotypes of maize. *Plant J.* **14**, 449–457.
31. Cline, K. & Henry, R. (1996) Import and routing of nucleus-encoded chloroplast proteins. *Annu. Rev. Cell Dev. Biol.* **12**, 1–26.
32. Corpas, F.J., Barroso, J.B., Sandalio, L.M., Distefano, S., Palma, J.M., Lupianez, J.A. & Del Rio, L.A. (1998) A dehydrogenase-mediated recycling system of NADPH in plant peroxisomes. *Biochem. J.* **330**, 777–784.
33. Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, E., Fukami, M., Miyajima, N. & Tabata, S. (1997) Structural analysis of *Arabidopsis thaliana* chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones. *DNA Res.* **4**, 215–230.
34. 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.
35. Gray, M.W., Burger, G. & Lang, B.F. (1999) Mitochondrial evolution. *Science* **283**, 1476–1481.
36. Martin, W. (1999) Mosaic bacterial chromosomes - a challenge en route to a tree of genomes. *Bioessays* **21**, 99–104.
37. Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128.
38. Ochman, H., Lawrence, J.G. & Groisman, E.S. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
39. Phillips, C., Dohnalek, J., Gover, S., Barrett, M.P. & Adams, M.J. (1998) A 2.8 Å resolution structure of 6-phosphogluconate dehydrogenase from the protozoan parasite *Trypanosoma brucei*: comparison with the sheep enzyme accounts for differences in activity with coenzyme and substrate analogues. *J. Mol. Biol.* **282**, 667–681.
40. McFadden, G.I. & Roos, D.S. (1999) Apicomplexan plastids as drug targets. *Trends Microbiol.* **7**, 328–333.
41. Wendt, U.K., Hauschild, R., Lange, C., Pietersma, M., Wenderoth, I. & von Schaewen, A. (1999) Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants. *Plant Mol. Biol.* **40**, 487–494.
42. Duffieux, F., Van Roy, J., Michels, P.A. & Opperdoes, F.R. (2000) Molecular characterization of the first two enzymes of the pentose-phosphate pathway of *Trypanosoma brucei*: glucose-6-phosphate dehydrogenase and 6-phosphogluconolactonase. *J. Biol. Chem.* **275**, 27559–27565.
43. Dooijes, D., Chaves, I., Kieft, R., Martin, W. & Borst, P. (2000) Conservation outside the order Kinetoplastida of base J as a constituent of nuclear but not nucleolar DNA in *Euglena gracilis*. *Nucleic Acids Res.* **28**, 3017–3021.
44. Henze, K., Badr, A., Wettern, M., Cerff, R. & Martin, W. (1995) A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl Acad. Sci. USA* **92**, 9122–9126.
45. Cavalier-Smith, T. (2000) Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**, 174–182.
46. van't Klooster, J.W., van den Berg-Velthuis, G., van West, P. & Govers, F. (2000) tef1, a *Phytophthora infestans* gene encoding translation elongation factor 1 alpha. *Gene* **249**, 145–151.
47. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I. & Doolittle, W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.
48. Purwantini, E., Gillis, T.P. & Daniels, L. (1997) Presence of F_{420}-dependent glucose-6-phosphate dehydrogenase in *Mycobacterium* and *Nocardia* species, but absence from *Streptomyces* and *Corynebacterium* species and methanogenic archaea. *FEMS Microbiol. Lett.* **146**, 129–134.
49. Schönheit, P. & Schäfer, T. (1995) Metabolism of hyperthermophiles. *World J. Microbiol. Biotechnol.* **11**, 26–57.
50. Chistoserdova, L., Gomelsky, L., Vorholt, J.A., Gomelsky, M., Tsygankov, Y.D. & Lidstrom, M.E. (2000) Analysis of two formaldehyde oxidation pathways in *Methylobacillus flagellatus* KT, a ribulose monophosphate cycle methylotroph. *Microbiol.* **146**, 233–238.
51. Tetaud, E., Hanau, S., Wells, J.M., Le Page, R.W.F., Adams, M.J., Arkison, S. & Barrett, M.P. (1999) 6-phosphogluconate dehydrogenase from *Lactococcus lactis*: a role for arginine residues in binding substrate and coenzyme. *Biochem. J.* **338**, 55–60.
52. Nowitzki, U., Flechner, A., Kellermann, J., Hasegawa, M., Schnarrenberger, C. & Martin, M. (1998) Eubacterial origin of nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase from spinach: sampling eukaryotic gene diversity in eukaryotic chromosomes through symbiosis. *Gene* **214**, 205–213.
53. Siebers, B., Klenk, H.P. & Hensel, R. (1998) PP_i-dependent phosphofructokinase from *Thermoproteus tenax*, an archaeal descendant of an ancient line in phosphofructokinase evolution. *J. Bacteriol.* **180**, 2137–2143.
54. 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 cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.* **17**, 989–1000.