

Resurrecting the Ancestral Enzymatic Role of a Modulatory Subunit*[§] ◆

Received for publication, December 1, 2004, and in revised form, January 3, 2005
Published, JBC Papers in Press, January 4, 2005, DOI 10.1074/jbc.M413540200

Miguel A. Ballicora, Jennifer R. Dubay, Claire H. Devillers[‡], and Jack Preiss[§]

From the Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824

In the post-genomic era, functional prediction of genes is largely based on sequence similarity searches, but sometimes the homologues bear different roles because of evolutionary adaptations. For instance, the existence of enzyme and non-enzyme homologues poses a difficult case for function prediction and the extent of this phenomenon is just starting to be surveyed. Different evolutionary paths are theoretically possible for the loss or acquisition of enzyme function. Here we studied the ancestral role of a model non-catalytic modulatory subunit. With a rational approach, we “resurrected” enzymatic activity from that subunit to experimentally prove that it derived from a catalytic ancestor. We show that this protein (L subunit ADP-glucose pyrophosphorylase) evolved to have a regulatory role, losing catalytic residues more than 130 million years ago, but preserving, possibly as a by-product, the substrate site architecture. Inactivation of catalytic subunits could be the consequence of a general evolutionary strategy to explore new regulatory roles in hetero-oligomers.

An attractive model to study the evolution of allosteric control is ADP-glucose pyrophosphorylase (ADP-Glc PPase),¹ a regulatory enzyme for the synthesis of bacterial glycogen and starch in plants. It has evolved to be regulated by different key metabolites depending on the main carbon assimilation pathway of the organism (1, 2). The quaternary structure plays an important role in its allosteric regulation. Most of the bacterial ADP-Glc PPases are homotetramers (1), whereas the plant enzymes comprise two types of homologous but distinct subunits (2–5), the small (S subunit, 50–54 kDa), and the large (L subunit, 51–60 kDa). They form a heterotetrameric structure S_2L_2 where the S subunit is catalytic, but L subunit is modulatory without catalytic function (2, 6). The L subunit modifies the sensitivity of the S subunit toward allosteric effectors, possibly by protein-protein interactions. This is of remarkable physiological importance because expression of different L subunits in different plants tissues confers distinct allosteric prop-

erties to the tetramer according to the necessities for starch synthesis (7). The similarity between S and L subunits (~50% identity) suggests a common ancestry (5). Therefore, gene duplication, divergence leading to new roles (catalytic and modulatory), additional divergence of modulatory genes, and further establishment of differential tissue expression seems to be the strategy used by nature to control and regulate the synthesis of starch, one of the most important storage compounds on earth. To understand the evolutionary path that led to this divergence, it is fundamental to know the ancestral role of these subunits.

The ancestor of S and L subunits could have been 1) catalytic as the S subunit or 2) non-catalytic as the L subunit. The first scenario is supported by the similarity between plant subunits with many active bacterial ADP-Glc PPases (1). However, we cannot simply rule out the second possibility because of the existence of non-catalytic homologues such as the products of the *glgD* genes from Gram-positive bacteria (8). There is strong evidence that the L subunit from the potato (*Solanum tuberosum* L.) tuber ADP-Glc PPase binds substrates. The heterotetramer (S_2L_2), as well as bacterial homotetramers (α_4), binds four ADP-[¹⁴C]glucose molecules (9, 10). We hypothesized that the L subunit maintained the structure of the substrate site needed for binding, whereas the catalytic ability was severely reduced by mutations of essential residues. To test this hypothesis, we sought to design an L subunit with significant catalytic activity mutating as few residues as possible.

EXPERIMENTAL PROCEDURES

Materials

α -D-[U-¹⁴C]Glucose 1-phosphate (Glc1P) was purchased from Amer sham Biosciences. [³²P]PP_i was purchased from PerkinElmer Life Sciences. Glc1P, ATP, ADP-glucose (ADP-Glc), 3-phosphoglycerate (3-PGA), and inorganic pyrophosphatase were purchased from Sigma. *Pfu* DNA polymerase was purchased from Stratagene (La Jolla, CA, USA). Ampligase®, a thermostable DNA ligase, was purchased from Epicenter (Madison, WI). All other reagents were purchased at the highest quality available.

DNA Methods

The Macromolecular Structure, Sequencing, and Synthesis Facility (MS³F) at Michigan State University performed the synthesis of oligonucleotides and automated DNA sequencing.

Site-directed Mutagenesis

Mutant D145N on the S subunit was obtained as described previously (6). Combined chain reaction (CCR) was used to introduce other mutations (11). The plasmids encoding the wild-type S subunit (pML10), the S subunit mutant D145N (pML10-D145N), or the wild-type L subunit (pMON17336) were used as templates for the mutagenesis (12, 13). In the S subunit, the residues were changed with the following oligonucleotides: R33K, 5'-GGA GCT GGG ACC AAA CTT TAT CCT CTA-3'; K43T, 5'-AAA AAA AGA GCA ACC CCA GCT GTT CCA-3'; K198R, 5' T GAA TTT GCA GAG CGC CCG CAA GGA GAG C-3'. In the L subunit, the oligonucleotides used were: K44R, 5'-GGA GAA GGG ACC CGC TTA TTC CCA CTT A-3'; T54K, 5'-AGT AGA ACT

* This work was supported by a Department of Energy Research grant and a Northern Regional United States Department of Agriculture grant. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

◆ This article was selected as a Paper of the Week.

§ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Fig. 1 and Table I.

‡ Present address: Alcimed Co., 84 boulevard Vivier Merle, 69485 Lyon cedex 03, France.

§ To whom correspondence should be addressed. Tel.: 517-353-3137; Fax: 517-353-9334; E-mail: preiss@msu.edu.

¹ The abbreviations used are: ADP-Glc, ADP-glucose; PPase, pyrophosphorylase; Glc1P, glucose 1-phosphate; 3-PGA, 3-phosphoglycerate; S, small; L, large.

GCA AAG CCT GCT GTT CCG-3'; K213R, 5'-AG TTT GCT GAA CGT CCA AAA GGT TT-3'. L subunit mutants on the residue 160 were obtained as described previously (6). The mutated fragments were exchanged to the double mutant plasmid pMON17336-K44R/T54K using the restriction sites SspI and SacI to obtain the triple mutants L_{K44R/T54K/D160N} and L_{K44R/T54K/D160E}. The coding regions of the mutant plasmids were sequenced to confirm that only the intended mutations were introduced.

Protein Methods

Protein concentration during enzyme purification was measured by using bicinchoninic acid reagent (14) from Pierce, with bovine serum albumin as standard. Interfering substances were previously removed precipitating with 12% trichloroacetic acid and 0.025% sodium deoxycholate and dissolving in 5% SDS, 0.1 N NaOH. Protein concentration of the purified enzymes was determined by UV absorbance at 280 nm using an extinction coefficient of 1.0 (6). Electrophoresis (SDS-PAGE) and immunoblotting of the protein samples were performed as described previously (6). Samples were desalted with Bio-Rad 10 DG chromatography columns. Centricon-30 devices (Amicon Inc.) were used to concentrate the enzymes.

Enzyme Assays

Pyrophosphorolysis Direction—The formation of [³²P]ATP from [³²P]PP_i and ADP-Glc was measured as follows. The standard aqueous reaction mixture contained 50 mM Hepes buffer (pH 8.0), 7 mM MgCl₂, 2 mM dithiothreitol, 1 mM [³²P]PP_i (300–1500 cpm/nmol), 2 mM ADP-Glc, 10 mM NaF, 2 mM 3-PGA, and 0.2 mg/ml bovine serum albumin, plus enzyme in a total volume of 0.25 ml. After 10 min of incubation at 37 °C, the reaction was terminated by the addition of 3 ml of cold 5% trichloroacetic acid. The [³²P]ATP formed was measured as described previously (3).

Synthesis Direction—The synthesis of ADP-[¹⁴C]Glc from [¹⁴C]Glc1P and ATP was measured by the method of Yep *et al.* (15). The standard aqueous reaction mixture contained 50 mM Hepes buffer (pH 8.0), 7 mM MgCl₂, 2 mM dithiothreitol, 0.5 mM [¹⁴C]Glc1P (800–1000 cpm/nmol), 1.5 mM ATP, 2 mM 3-PGA, 1.5 units/ml inorganic pyrophosphatase, and 0.2 mg/ml bovine serum albumin, plus enzyme in a total volume of 0.2 ml. Reaction mixtures were incubated for 10 min at 37 °C and terminated by heating in a boiling water bath for 1 min. ADP-[¹⁴C]Glc was assayed as described previously (15).

Unit Definition—In the above assays, one unit of enzyme activity is equal to 1 μmol of product, either [³²P]ATP or [¹⁴C]ADP-glucose, formed per minute.

Expression and Purification of Mutant Enzymes

The *Escherichia coli* AC70R1-504 cells, which lack endogenous ADP-Glc PPase activity (12), were co-transformed with compatible plasmid pML10 (S subunit) and pMON17336 (L subunit) or their respective mutated derivatives for co-expression of the subunits. Transformed cells were grown in 2 liter of LB medium, induced, harvested, and sonicated in buffer A (50 mM Hepes (pH 8.0), 5 mM MgCl₂, 0.1 mM EDTA, 10% sucrose) as described previously (16). The purification was performed at 0–4 °C. The crude extract was applied onto a 12-ml DEAE-Sepharose column (Amersham Biosciences) equilibrated in buffer A, and eluted with a linear NaCl gradient (10 column volumes, 0–0.5 M). Purest fractions were pooled, desalted, concentrated, and applied onto a Mono Q HR 16/10 (FPLC, Amersham Biosciences) column equilibrated with buffer A and eluted with a linear NaCl gradient (16 bed volumes, 0.1–0.5 M). The post-Mono Q fractions were pooled, concentrated, resuspended in buffer B (buffer A plus 1.3 M ammonium sulfate), and applied onto a phenyl-Superose HR 5/5 (FPLC, Amersham Biosciences) column equilibrated with buffer B. The sample was eluted with a decreasing linear gradient (20 bed volumes) of ammonium sulfate (1.3 to 0 M ammonium sulfate). The purest fractions were pooled, concentrated, and desalted. After this procedure, enzymes were >95% pure. The enzymes were followed by pyrophosphorolysis activity, except S_{D145N}L_{WT}, S_{D145N}L_{K44R}, and S_{K43T}L_{WT}, which were followed by SDS-PAGE and immunoblotting.

Calculation of Kinetic Constants

Curves were performed varying the concentration of the effector studied while keeping a constant saturated concentration of the other components of the standard reaction mixture. The kinetic data were plotted as initial velocity in nmol min⁻¹ (*v*) versus effector concentration in μM ([S]). This plot was used to fit the parameters of a modified Hill equation (17), $v = V_{\max} [S]^n / (S_{0.5}^n + [S]^n)$, by the Levenberg-Marquardt

non-linear least squares algorithm provided by the computer program Origin® 5.0 (18). The parameter $S_{0.5}$ is the concentration of substrate needed to obtain 50% of the maximum activity (V_{\max}). The activation curves were fitted with the similar equation, $v = v_0 + (V_{\max} - v_0) [S]^n / (S_{0.5}^n + [S]^n)$. The parameter v_0 is the activity in absence of activator, and $A_{0.5}$ is the concentration of activator needed to obtain 50% of the maximal activation ($V_{\max} - v_0$). The standard deviations were obtained by the Levenberg-Marquardt method (18). Specific activity was determined at saturating concentrations of effectors and different dilutions of the enzyme to plot activity versus protein concentration. The specific activity was determined from the slope in the linear range. From this value, k_{cat} was calculated as number of molecules of ADP-Glc synthesized (consumed in the reverse direction) per molecule of enzyme per second. For this calculation, we used 202 kDa as molecular mass of the S₂L₂ enzyme (12).

Identification of Putative Essential Missing Residues in the L Subunit

The set of invariants from an alignment of catalytic ADP-Glc PPase subunits should contain most of, if not all, the essential residues for catalysis. Among those, the ones that are absent in the L subunits were candidates to be responsible for the lack of catalytic activity. In the comparison, to avoid missing important residues, we excluded sequences of modulatory or potentially inactive ADP-Glc PPase subunits. A first conservative alignment was performed with ADP-Glc PPases with experimentally proven catalytic activity (Fig. 1A). Those enzymes were from *Rhodobacter sphaeroides* (19), *Agrobacterium tumefaciens* (20), *E. coli* K12 (21), *E. coli* (SG14 mutant) (22), *Geobacillus stearothermophilus* (GlgC) (8), *Anabaena sp.* PCC 7120 (23), *Hordeum vulgare*, endosperm S subunit (24), *Arabidopsis thaliana* S subunit (7), and *S. tuberosum*, tuber S subunit (12). In that alignment, 64 residues were invariant in the putative catalytic domain of the potato tuber S subunit (residues 21–290 (2, 25)). From those, the only four absent in the potato tuber L subunit were Arg³³, Lys⁴³, Gln¹²⁶, and Ile²³² (S subunit numeration).

A second comparison with a broader and less strict selection of enzymes strongly suggested that Gln¹²⁶ and Ile²³² are not essential. The criteria to select enzymes with high probability of being catalytic in this second round were: 1) from bacterial genome projects that detected only one ADP-Glc PPase gene per organism and 2) genes identified as glgC (active) and not as glgD, known to produce non-catalytic polypeptides in *Bacillus* (8). Arg³³ and Lys⁴³ were still invariant in this second alignment, but Gln¹²⁶ and Ile²³² were replaced in 7 and 12 out of 58 cases, respectively. Among S subunits from plants, Arg³³ and Lys⁴³ were conserved except sequence 20 (homologous of Lys⁴³ to Asn) and sequence 21 (homologous of Arg³³ to Lys) (supplemental Table I). The existence of duplicated genes (paralogs) and pseudo-genes in plants raised the question whether these exceptions actually generate active enzymes.

Structure Prediction and Homology Modeling

Homology modeling of the putative pyrophosphorylase domain (residues 31–310) of the wild-type L subunit of the ADP-Glc PPase from potato tuber and the double mutant L_{K44R/T54K} was performed with the program Modeller 6v2 (26). The structure was modeled using two simultaneous templates; the known atomic coordinates of the *E. coli* dTDP-Glc PPase (RffH) complexed with dTTP and Mg²⁺ (Protein Data Bank code 1MC3), and human Agx2 (UDP-N-acetylglucosamine pyrophosphorylase; Protein Data Bank code 1JVG) (27, 28). In this way, we took advantage of the features of the program Modeller to use the best part of each template (29). We guided the initial alignment for the homology modeling with the secondary structure prediction performed with the PSI-PRED method (30) available on the PSI-PRED server (bioinf.cs.ucl.ac.uk/psipred/) (31). Alignment with the templates was based on homology, secondary structure, and hydrophobicity.

There were strong indications that ADP-Glc PPases share with the templates a pyrophosphorylase domain with the same fold despite the identity was ~20% (2). In addition, threading and fold recognition analysis with the program GenTHREADER (32) predicted a common fold with a very high confidence. Running the region from residues 31–320 from the L subunit from potato tuber, gave E values (expected number of false hits per sequence query) of 2×10^{-5} , and 4×10^{-5} for 1G0R (dTDP-Glc PPase) and 1JV1 (UDP-N-acetylglucosamine pyrophosphorylase), respectively. In the active site, we modeled a deoxyribose triphosphate ligand extracted from the coordinates of the dTTP bound to RffH. That is the shared moiety between the ATP and dTTP bound to the ADP-Glc PPase and dTTP-Glc PPase, respectively.

The model was checked with the programs Verify3D (33) and PRO-CHECK (34). Minor adjustments in the alignments with the templates were checked and modified iteratively to improve the three-dimensional-one-dimensional scores provided by Verify3D along the sequence. There was no structural change, in any of the iterations, in the area of the residues analyzed in this work. The region between residues 217–232, which corresponds to an insertion, was the only one with low 3D-1D score. This is expected from insertions of more than eight residues (35). However, this area is very far from the region analyzed in the paper, in the other lobe of the catalytic domain, exposed, and not facing the active site.

Recently, the crystal structure of the tetrameric form (S_4) of the S subunit from potato tuber was solved (36). This structure was in an inhibited conformation, which made it unsuitable as a template for modeling an active L subunit. However, it confirmed that the templates selected in this study (RffH and Agx2) were appropriate, because they share the same fold in the catalytic domain as previously predicted (2). RffH and Agx2 are not allosterically regulated and are better templates to model the active conformation of the substrate site of the L subunit.

Phylogenetic Analysis

ADP-Glc PPase sequences from Supplemental Table I were obtained from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). The expression data were taken from the literature. Multiple sequence alignment was performed by the Clustal program (37) available on-line (clustalw.genome.jp), and afterward, it was manually refined with the BioEdit v6.0.7 program (www.mbio.ncsu.edu/BioEdit/bioedit.html). For the phylogenetic analysis, the N terminus of all the sequences was discarded. That region is not homologous and corresponds to the leading peptide for subcellular targeting, which has been shown in the S subunits that belongs to an exon with a different evolutionary history (38). Alignment started in residue 2 of the mature recombinant S subunit from potato tuber ($A^2VSDSQNSQT\dots$) and residue 13 of the mature recombinant L subunit from potato tuber ($Q^{13}TVFVDMPR\dots$). The unrooted tree was built by the neighbor-joining method (39) as implemented in the PHYLIP package (evolution.genetics.washington.edu/phylip.html), version 3.6. The tree was drawn with the program Phylodraw (40). The outgroup chosen was the ADP-Glc PPase from *Anabaena* because cyanobacterial enzymes are the closest to plant enzymes (2). The topology of the tree of the L subunit gene family from plants was further investigated using the Maximum-likelihood method as implemented in the program TREE-PUZZLE (41). The tree was rooted with the sequence of the L subunit from the unicellular algae *Chlamydomonas reinhardtii* and plotted with the program TREEVIEW (42).

RESULTS

Identification of critical missing residues was the first important step to design an active L subunit. A sequence alignment of ADP-Glc PPases with reported activity (experimentally assayed) allowed us to focus on few candidates. The set of invariants should contain most of, if not all, the essential residues. The subset of the ones absent in the L subunit was of special interest. We chose Lys⁴⁴ and Thr⁵⁴ in the L subunit from potato tuber as the best candidates to study because the homologous residues (Arg³³ and Lys⁴³ in the S subunit, Fig. 1A) were differentially conserved in the active bacterial and S subunits (see a detailed analysis under "Experimental Procedures"). In addition, Lys⁴⁴ and Thr⁵⁴ are in a highly conserved region of ADP-Glc PPases (Fig. 1A).

Resurrection of ADP-Glc PPase Catalysis—To convert the modulatory L subunit into a catalytic one, we substituted Lys⁴⁴ and Thr⁵⁴ by Arg⁴⁴ and Lys⁵⁴, respectively. The mutant $L_{K44R/T54K}$ was expressed alone, but no activity was detected, most probably because the L subunit is unable to form a stable tetramer in absence of the S subunit (43). The activity of L subunit mutants cannot be readily tested with a co-expressed wild-type S subunit because of the intrinsic activity of the latter. We co-expressed the L subunit mutants with S_{D145N} , an inactive S subunit in which the catalytic Asp¹⁴⁵ was mutated (6). In this way, the activity deriving from the S subunit was reduced more than three orders of magnitude (Table I). Co-expression of the L subunit double mutant $L_{K44R/T54K}$ with

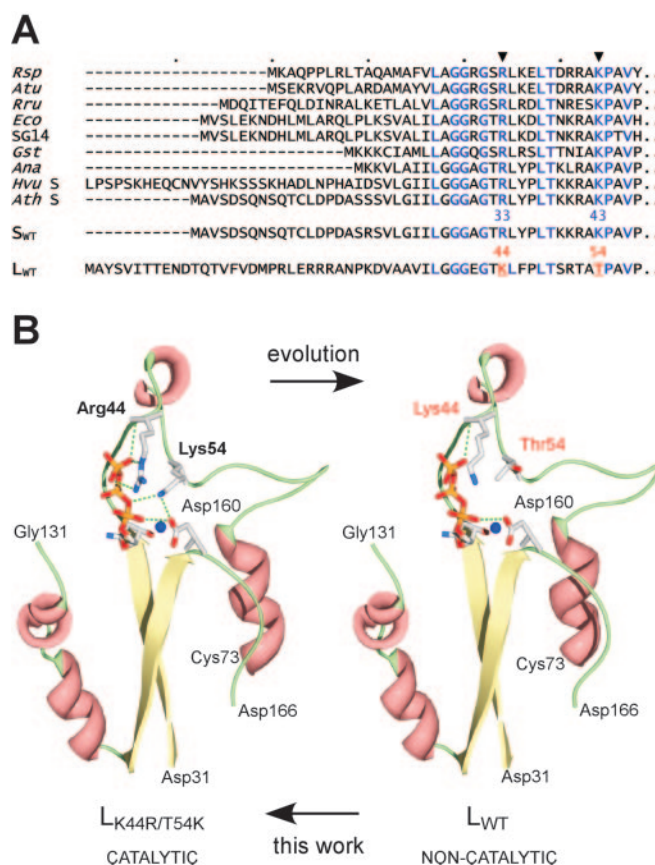


FIG. 1. Active site models of the wild-type and double mutant L subunit. A, sequence comparison of the potato (*S. tuberosum*) tuber L subunit (L_{WT}) with proteins with experimentally assayed ADP-Glc PPase activity. Sequences and their accession numbers are: Rsp, *R. sphaeroides*, Q9RNH7; Atu, *A. tumefaciens*, P39669; Rru, *Rhodospirillum rubrum*, Q9ZFN4; Eco, *E. coli* K12, P00584; SG14, *E. coli* SG14 mutant, P00584; Gst, *G. stearothermophilus* (GlcC), O08326; Ana, *Anabaena* sp. PCC 7120, S24991; Hvu, *H. vulgare*, endosperm S 5 subunit, CAA88449; Ath, *A. thaliana* S subunit, P55228; S_{WT} , potato tuber S subunit, P23509; L_{WT} , Q00081. Residues 100% conserved are in blue, and the ones conserved in the first group but absent in L_{WT} are in red. B, $L_{K44R/T54K}$ and L_{WT} were modeled based on a dTDP-Glc PPase and an UDP-GlcNAc PPase (see "Experimental Procedures"). Fragments from residues 31–73 and 131–166 are depicted. Deoxyribose triphosphate, common moiety between dTTP and ATP, was also modeled with Mg^{2+} (blue sphere). The nitrogen from which the adenine base (thymine in case of the dTDP-Glc PPase) is attached to the ribose is depicted in blue. Dotted green lines represent proper distances for hydrogen bonds.

S_{D145N} generated an enzyme with considerable activity, 10 and 18% of the wild-type enzyme ($S_{WT}L_{WT}$) in the forward and reverse direction, respectively (Table I). Single mutation K44R generated an enzyme ($S_{D145N}L_{K44R}$) with no significant activity over the control ($S_{D145N}L_{WT}$). Mutation T54K ($S_{D145N}L_{T54K}$) was more effective, but the combination of both mutations in the L subunit ($S_{D145N}L_{K44R/T54K}$) had the most dramatic effect (Table I). Therefore, we concluded that the two residues Arg⁴⁴ and Lys⁵⁴ are needed for restoring catalytic activity to the L subunit.

The resurrection of catalytic activity in the L subunit by incorporation of Arg⁴⁴ and Lys⁵⁴ predicted that the homologous residues in the S subunit are also critical. Replacement of those two residues by the homologous in the L subunits (mutations R33K and K43T) decreased the activity one and two orders of magnitude, respectively, in both directions, confirming the hypothesis (Table I). The mutant enzymes were still activated by 3-PGA and inhibited by orthophosphate (P_i) (data not shown), which are the main regulators of plant enzymes (2). The wild-type enzyme and $S_{D145N}L_{K44R/T54K}$ had very similar kinetic

TABLE I
Activity of ADP-Glc PPase mutants

The reaction in the forward (synthesis of ADP-Glc) direction uses ATP and Glc1P as substrates releasing PP_i and ADP-Glc as products. The reaction can also be assayed in the reverse (pyrophosphorolysis of ADP-Glc) direction. Enzymes were obtained by co-expression of the S and L subunits and purified to homogeneity, and the activity was assayed in both directions of the reaction as described under "Experimental Procedures." WT, wild type.

Subunits		k_{cat}	
S	L	Forward (synthesis) ^a	Reverse (pyrophosphorolysis) ^b
		s^{-1}	s^{-1}
WT + WT		108 ± 3	165 ± 7
D145N + WT		0.057 ± 0.003	0.125 ± 0.007
D145N + K44R		0.104 ± 0.003	0.111 ± 0.007
D145N + T54K		3.1 ± 0.3	1.9 ± 0.1
D145N + K44R/T54K		10.8 ± 0.7	30.3 ± 2.4
R33K + WT		12.1 ± 0.7	13.8 ± 0.3
K43T + WT		1.08 ± 0.03	0.94 ± 0.03

^a Activator (4 mM 3-PGA) and substrates (2 mM ATP and 0.5 mM Glc1P) were saturating.

^b Activator (4 mM 3-PGA) and substrate ADP-Glc (2 mM) were saturating. The concentration of PP_i (1.4 mM) was the highest to avoid precipitation in the assay mixture.

TABLE II
Kinetic properties of the mutant $S_{D145N}L_{K44R/T54K}$ and the wild-type ADP-Glc PPase from potato tuber

Saturation curves for different effectors were performed as indicated under "Experimental Procedures."

Parameter		$S_{WT}L_{WT}$	$S_{D145N}L_{K44R/T54K}$
Forward (synthesis) direction			
ATP	$S_{0.5}$ (μM) (n_H) ^a	97 ± 7 (1.6)	170 ± 13 (1.8)
Glc1P	$S_{0.5}$ (μM) (n_H)	27 ± 2 (1.1)	11 ± 1 (1.2)
3-PGA	Activation ^b (-fold)	60 ± 5	20 ± 2
	$A_{0.5}$ (μM) (n_H)	135 ± 11 (0.9)	13.6 ± 0.7 (1.3)
Reverse (pyrophosphorolysis) direction			
ADP-Glc	$S_{0.5}$ (μM) (n_H)	200 ± 20 (1.4)	70 ± 6 (1.5)
PP_i	$S_{0.5}$ (μM) (n_H)	37 ± 2 (1.0)	98 ± 10 (1.0)
3-PGA	Activation ^b (-fold)	3.1 ± 0.3	7.6 ± 1.0
	$A_{0.5}$ (μM) (n_H)	1.1 ± 0.3 (1.1)	4.0 ± 0.5 (0.9)
3-PGA (+ 2 mM P_i)	$A_{0.5}$ (μM) (n_H)	96 ± 19 (1.0)	101 ± 13 (1.1)

^a Coefficients of the Hill equation (n_H) are in parentheses.

^b Activation values (-fold) are the ratio between the activity at saturated concentration of activator (3-PGA) and the activity in absence of activator.

properties indicating that the architecture of the substrate site has been conserved. The apparent affinities for the substrates in both directions were in the same range and the allosteric properties of $S_{D145N}L_{K44R/T54K}$ resembled those of the wild type as well (Table II). The presence of 2 mM P_i decreased the apparent affinity of 3-PGA for both the wild type and $S_{D145N}L_{K44R/T54K}$ indicating that this new form has a similar sensitivity to P_i inhibition (Table II).

The fact that only two mutations in the L subunit restored enzyme activity is the ultimate evidence that the L subunit derives from a catalytic ancestor. To confirm that the catalysis occurs in the L subunit of $S_{D145N}L_{K44R/T54K}$, we disrupted the substrate site in each subunit and compared the kinetic properties. Previously, in the wild-type enzyme, replacement of Lys¹⁹⁸ in the S subunit decreased the substrate affinity (Glc1P), whereas disruption of the homologous residue Lys²¹³ in the L subunit did not have the same effect (9). In the case of $S_{D145N}L_{K44R/T54K}$, the mutation K213R on the L subunit severely decreased the apparent affinity for Glc1P, whereas K198R on the S subunit did not (Table III). This indicated that the L subunit double mutant, and not S_{D145N} , was catalytic. In the wild-type enzyme, Lys²¹³ does not seem to play any important role; but in $S_{D145N}L_{K44R/T54K}$, it recovered its ancestral ability to confer to the enzyme a high apparent affinity for Glc1P. Previous results showed that Asp¹⁴⁵ in the S_{WT} subunit is essential for catalysis, whereas the homologous Asp¹⁶⁰ in the L_{WT} subunit is not (6). Conversely, in this study, mutation D160N or D160E in the $L_{K44R/T54K}$ subunit abolished the activity (Table IV), which shows the ancestral essential role of this residue and confirms that the catalysis

of $S_{D145N}L_{K44R/T54K}$ occurs in the L subunit.

Structural Basis for the Ancestral Function—A comparative model of $L_{K44R/T54K}$ illustrates the predicted role of Arg⁴⁴ and Lys⁵⁴ (Fig. 1B). In the model, Asp¹⁶⁰, which is homologous to the catalytic Asp¹⁴⁵ in the S subunit and catalytic Asp¹⁴² in the *E. coli* ADP-Glc PPase (6, 25), interacts with Lys⁵⁴. This type of interaction (Lys⁵⁴-Asp¹⁶⁰) has also been observed in crystal structures of enzymes that catalyze similar reactions, such as dTDP-glucose pyrophosphorylase (dTDP-Glc PPase) and UDP-N-acetylglucosamine pyrophosphorylase (UDP-GlcNAc PPase), and postulated to be important for catalysis by correctly orienting the Asp (28, 44, 45). In addition, Lys⁵⁴ interacts with the oxygen bridging the α - and β -phosphates as it has been observed in the crystal structure of *E. coli* dTDP-Glc PPase (28). That contact may neutralize a negative charge density stabilizing the transition state and making the pyrophosphate a better leaving group. Arg⁴⁴ interacts in the model with the β - and γ -phosphates of ATP, which correspond to the PP_i product (Fig. 1). Similarly, Arg¹⁵ in the *E. coli* dTDP-Glc PPase was postulated to contribute to the departure of PP_i (44). The kinetic data agreed with the predicted interaction of PP_i with Arg⁴⁴. A Lys in that position, in both the catalytic L subunit mutant and the S subunit, decreased the apparent affinity for PP_i at least 20-fold (Fig. 2). In the model of the wild-type non-catalytic L subunit, Lys⁴⁴ and Thr⁵⁴ do not interact as Arg⁴⁴ and Lys⁵⁴ (Fig. 1).

Phylogenetic Analysis—A phylogenetic tree of the ADP-Glc PPases from photosynthetic eukaryotes provides information about the origin of the subunits. In multicellular organisms, S and L subunits form two and four distinct groups, respec-

TABLE III
Effect of the mutations on the Glc1P site

Reaction conditions were as in Table I. To determine specific activity, concentrations of effectors were saturating. The increase values were the relative values of $S_{0.5}$ compared to the enzyme $S_{D145N-L_{K44R/T54K}}$ (Table II).

Subunits		Glc1P, $S_{0.5}$		k_{cat}	
S	L	μM	Increase	Forward (synthesis) s^{-1}	Reverse (pyrophosphorolysis) s^{-1}
D145N + K44R/T54K/K213R		910 \pm 110	82-fold	4.4 \pm 0.3	14.1 \pm 0.3
D145N/K198R + K44R/T54K		13.2 \pm 0.5	1.2-fold	7.4 \pm 0.3	20.5 \pm 0.7

TABLE IV
Effect of mutations on the residue Asp¹⁶⁰ on the L subunit

Enzymes were obtained by co-expression of the L and S subunits as described under "Experimental Procedures." Pyrophosphorolysis activity was measured in crude extracts. Concentrations of the activator 3-PGA (4 mM) and substrate ADP-Glc (2 mM) were saturating. The saturating concentration of substrate PP_i (1.4 mM) was the highest concentration to avoid precipitation.

Subunits		Specific activity <i>unit/mg</i>
S	L	
D145N + WT		<0.001
D145N + K44R/T54K		0.35 \pm 0.05
D145N + K44R/T54K/D160N		<0.002
D145N + K44R/T54K/D160E		<0.002

DISCUSSION

The resurrection of the enzymatic activity of the L subunit ADP-Glc PPase by only two mutations reveals the catalytic ancestry. In addition, it points out the important role of the residues Arg⁴⁴ and Lys⁵⁴ for catalysis. A comparison with other pyrophosphorylases agrees with this data.

The importance of Arg⁴⁴ and Lys⁵⁴ in the Active L Subunit Mutant Is Possibly Widespread in Other Pyrophosphorylases—The identity between ADP-Glc PPases and other nucleotide-diphosphate-sugar pyrophosphorylases is very low (~20%). However, prediction of the secondary structure and further structure-based alignment of the sequences emphasized the similarities (1, 2, 25). Residues Arg⁴⁴ and Lys⁵⁴ of the active L subunit mutant are not only present in active ADP-Glc PPases but also in other pyrophosphorylases in structurally similar positions (25, 44, 45). In addition, in some cases these residues were shown to be important. For instance, Lys²⁵ in the UDP-GlcNAc PPase (GlmU), homologous to Lys⁵⁴ in the mutant $L_{K44R/T54K}$, was mutated to Ala and the specific activity decreased 8-fold (45). In this work, homology modeling predicted that Arg⁴⁴, an important residue for activity in the mutant $L_{K44R/T54K}$, interacts with PP_i making it a better leaving group in the forward (synthesis) direction. In the reverse (pyrophosphorolysis) direction, it is predicted to have a role in PP_i binding as a substrate. In other pyrophosphorylases, a non-conservative mutation from Arg to Ala, at the homologous position, reduced the specific activity in the nucleotide-sugar synthesis direction two to three orders of magnitude (45, 46). Unfortunately, the effect on the PP_i apparent affinity has not been examined. The presence of homologous residues to Arg⁴⁴ and Lys⁵⁴ suggest that they may be also important in other distantly related pyrophosphorylases, such as the enzyme 4-diphosphocytidyl-2-C-methylerythritol synthetase (47). In this case, despite the very low percentage of identity, the homology can also be identified with a structural alignment.

Evolutionary History of ADP-Glc PPase L Subunits—We cannot ascertain when the L subunit from potato tuber became modulatory only by phylogenetic analysis, which is based upon primary structure rather than function. However, with structure-function relationship information, we could theoretically trace function or the absence of it in the tree. Both groups III and IV lack the residues homologous to Arg³³ and Lys⁴³ (S_{WT} numeration, Fig. 1A). Group III has Lys, His, and Gln, and group IV has Gln, respectively, in place of a homologous Arg³³ (Fig. 3). A distinctive characteristic in both groups is that a Thr replaces the catalytic Lys⁴³ (Fig. 3). The most likely scenario supported by phylogenetic analysis (Fig. 3) is that the shared ancestor of both branches lacked a Lys in that position and, consequently, was already non-catalytic. Because one branch (III) is composed of dicot and the other (IV) of monocot plants, we can infer that the L subunit from groups III and IV ceased to be catalytic before monocots split from dicots among the angiosperms more than 130 million years ago (48, 49). Conceivably, they had already acquired adaptive modulatory properties at that point in evolution given that catalysis had been already lost (Fig. 3).

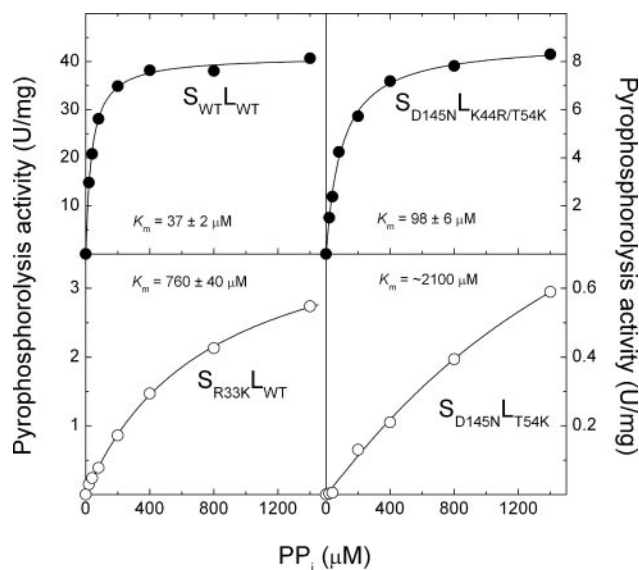
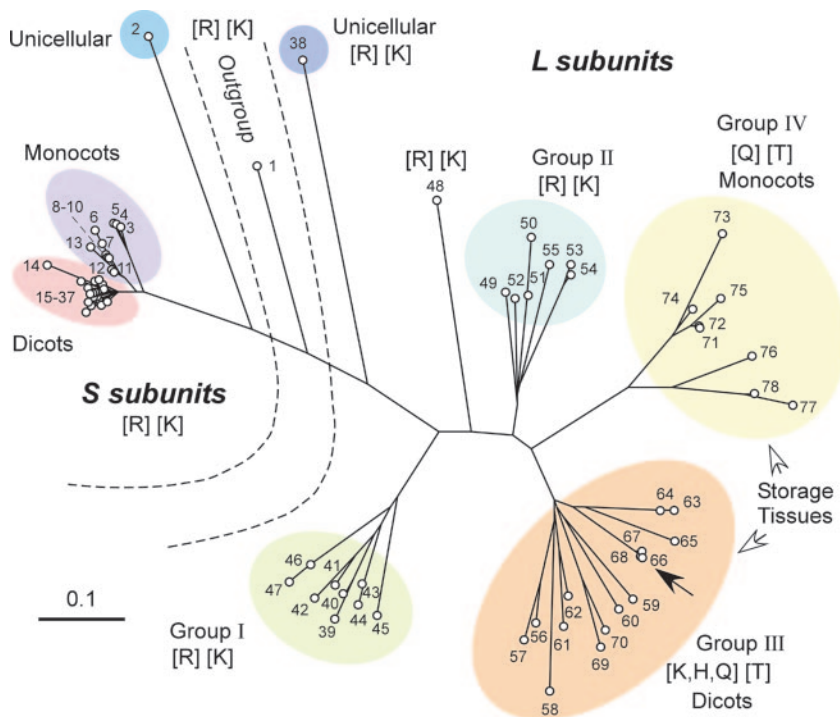


FIG. 2. Effect of the mutations R33K (S subunit) and K44R (L subunit) on the pyrophosphate saturation curves. Activity was analyzed in the pyrophosphorolysis direction as indicated under "Experimental Procedures" for the wild-type enzyme and the mutants $S_{R33K-L_{WT}}$, $S_{D145N-L_{K44R/T54K}}$, and $S_{D145N-L_{T54K}}$. Complete saturation could not be reached for the mutants $S_{R33K-L_{WT}}$ and $S_{D145N-L_{T54K}}$. Higher concentrations of PP_i lead to precipitation and cloudiness of the reaction mixture. The kinetic constants were calculated as described under "Experimental Procedures."

tively (Fig. 3). The two main groups of S subunits are from dicot and monocot plants, whereas L subunit groups correlate better with the reported tissue expression. The first L subunit group, generally expressed in photosynthetic tissues (supplemental Table I), comprises L subunits from dicots and monocots. Group II displays a broader expression pattern, whereas groups III and IV are expressed in storage organs (roots, stems, tubers, seeds). Subunits from group III are only from dicot plants, whereas group IV are seed-specific subunits from monocots. These last two groups stem from the same branch of the phylogenetic tree and split before monocot and dicot separation (Fig. 3).

FIG. 3. Phylogenetic relationship of ADP-Glc PPases from photosynthetic eukaryotes. The unrooted tree was built as described under “Experimental Procedures” and reference numbers are listed on Supplemental Table I. All the ADP-Glc PPases are from angiosperms because there are no complete sequences available from gymnosperms. Dotted lines separate L from S subunits, whereas the cyanobacterial ADP-Glc PPase from *Anabaena* was included as outgroup. The black arrow points to the potato tuber L subunit analyzed in this study (sequence 66). The general topology of the tree had a strong support by a maximum-likelihood method (Supplemental Fig. 1). Between brackets, it is indicated the residues conserved in each group in the positions that correspond to Lys⁴⁴ (first bracket) and Thr⁵⁴ (second bracket) in the potato tuber L subunit. The scale on the bottom left represents the number of substitutions per site.



Modulatory L subunits from branches I and II of the phylogenetic tree may be non-catalytic too (7), but the reasons for this lack of activity are not clear. They contain in their sequence the important residues analyzed in this work (Fig. 3), as well as all the others conserved in the catalytic subunits. This presence predicts a catalytic role; however, the only biochemical evidence gathered so far has not confirmed it. None of the L subunits from *A. thaliana* (groups I–III, supplemental Table I), when expressed in presence of an inactive S subunit (APS2), originated a form with detectable activity (7). Despite no apparent essential residue seems to be missing, other subtle ways of enzyme inactivation during the course of evolution can explain those results (50). At this time, we cannot discard the alternative hypothesis that *A. thaliana* L subunits from groups I and II are catalytic but could not interact properly with APS2. In the literature, no other L subunit from these two groups has been expressed and tested for intrinsic activity. Based on sequence, we cannot rule out the possibility of finding a catalytically competent subunit among groups I and II.

Phylogenetic trees based on protein sequences are very powerful tools to test evolutionary hypothesis, but they do not show the functional information that those sequences carry. Gene synthesis of hypothetical ancestors and characterization of their products may tackle this problem (51). As performed in this paper, analyses of structure-function relationships should complement phylogenetic analysis for tracing and understanding how and when certain former enzymes acquired new roles and lost catalytic function. The use of this information could potentially increase the accuracy of function prediction for these difficult cases (50, 52).

Enzyme Inactivation as a General Strategy—Divergence from an ancestral catalytic subunit to form hetero-oligomers with non-catalytic and catalytic subunits could have been a strategy used by nature to add complexity and explore new regulatory functions. There are phylogenetic evidences suggesting this could be a general evolutionary mechanism. In a superfamily of enzymes, if only one branch (“odd one out”) in the tree is non-catalytic, the common ancestor most probably was catalytic as the majority of the branches. Six cases of hetero-oligomers with solved three-dimensional structures that

meet this criterion have been observed (50). In this work, we experimentally tested this hypothesis for a plant ADP-Glc PPase and found that the common ancestor of catalytic and non-catalytic subunits was catalytic. This enzyme is a very attractive model because the “inactivated” subunit acquired differential modulatory properties in plant tissues, expanding the plasticity for regulation of a key synthetic pathway. After gene duplication, the ADP-Glc PPase subunits acquired asymmetric functions. The S subunit remained catalytic with defective allostereism whereas the L subunit became modulatory (*i.e.* modifying the properties of the S subunit) and lost catalytic function.

Considering the ancient inactivation of L subunits and their higher divergence (Fig. 3), it is remarkable that the general architecture of their substrate site has not changed much by evolution. The wild-type enzyme and S_{D145N}L_{K44R/T54K} had comparable apparent affinities for the substrates. This preservation occurred despite substrate binding does not seem to be an adaptive pressure because is not needed for the modulatory roles of the L subunit. It was shown previously that mutations in the L subunit of a critical residue for Glc1P binding (Lys²¹³) do not affect the modulatory properties of the potato tuber ADP-Glc PPase (9). In addition, in the wheat endosperm enzyme, Gln has replaced that Lys. Perhaps, the active-site architecture has been kept as a by-product of evolution or *spandrel* (53) because some disruptions could indirectly affect the modulatory properties. For instance, Asp¹⁶⁰ in the wild-type L subunit is conserved in all ADP-Glc and other nucleotide-sugar PPases (6, 25, 44). In the catalytic subunits, it has a catalytic role but replacement in the wild-type L subunit of Asp¹⁶⁰ altered the modulatory properties of the tetramer (6). This residue, Asp¹⁶⁰, “recovered” its critical role for activity in the L_{K44R/T54K} mutant (Table IV).

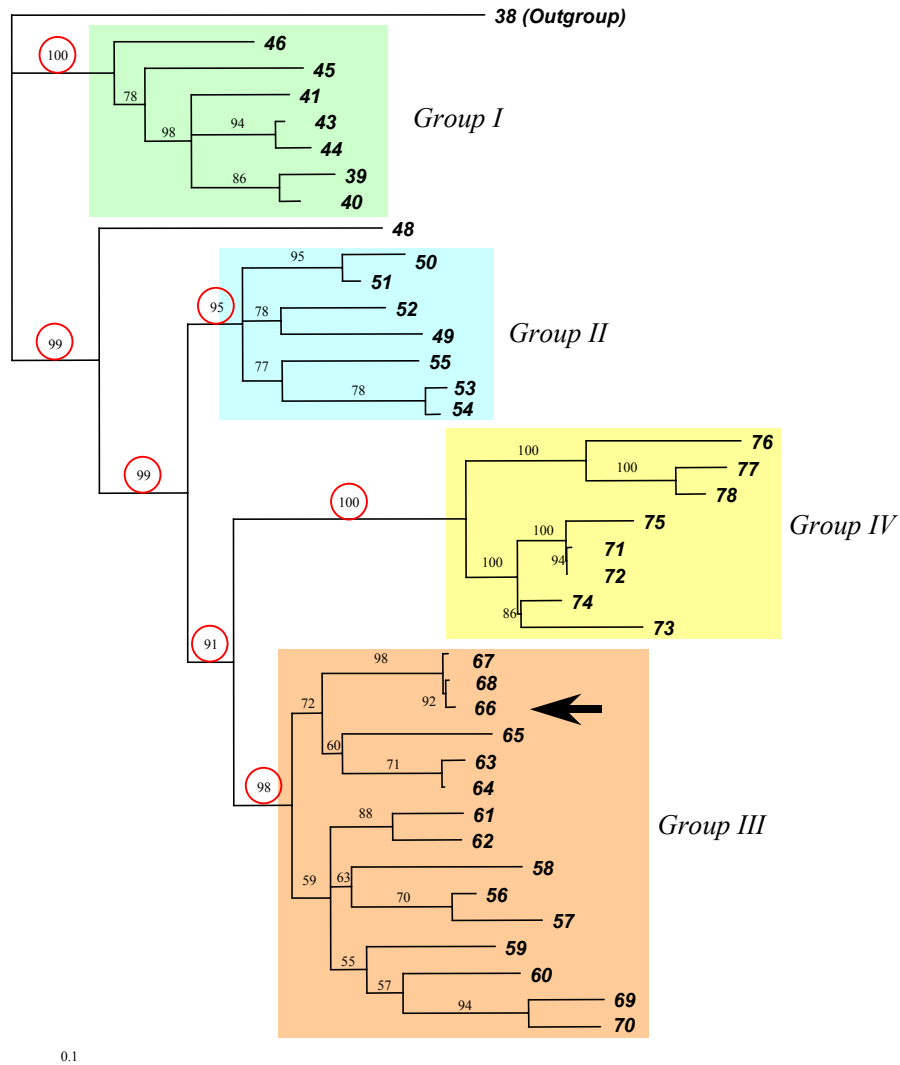
Adaptation from enzymes to non-enzymes and vice versa is a very exciting topic in protein evolution because it involves the generation of novel functions (50). Not only subunits from hetero-oligomers, but also many enzymes and non-enzymes homologues, are being detected. Based on the odd one out criterion, it has been reported some cases of non-enzymes that became enzymes, but the opposite scenario was observed more

often (50, 54). Some examples may be critical for the appearance of higher levels of complexity and control. For instance, a thrilling proposal is that many transcription regulators have an ancient catalytic past (55, 56). Supporting experimental evidence, as performed in the present study, will be enlightening to confirm these hypotheses and understand how key regulatory roles have evolved at a molecular level.

Acknowledgments—We thank C. M. Bejar and A. Yep for stimulating discussions and S. Wakao and Dr. E. S. Trombetta for helpful criticisms and revision of the manuscript.

REFERENCES

- Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2003) *Microbiol. Mol. Biol. Rev.* **67**, 213–225
- Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2004) *Photosynth. Res.* **79**, 1–24
- Morell, M. K., Bloom, M., Knowles, V., and Preiss, J. (1987) *Plant Physiol.* **85**, 182–187
- Okita, T., Nakata, P. A., Anderson, J. M., Sowokinos, J. R., Morell, M., and Preiss, J. (1990) *Plant Physiol.* **93**, 785–790
- Smith-White, B. J., and Preiss, J. (1992) *J. Mol. Evol.* **34**, 449–464
- Frueauf, J. B., Ballicora, M. A., and Preiss, J. (2003) *Plant J.* **33**, 503–511
- Crevillén, P., Ballicora, M. A., Mérida, A., Preiss, J., and Romero, J. M. (2003) *J. Biol. Chem.* **278**, 28508–28515
- Takata, H., Takaha, T., Okada, S., Takagi, M., and Imanaka, T. (1997) *J. Bacteriol.* **179**, 4689–4698
- Fu, Y., Ballicora, M. A., and Preiss, J. (1998) *Plant Physiol.* **117**, 989–996
- Haugen, T. H., and Preiss, J. (1979) *J. Biol. Chem.* **254**, 127–136
- Bi, W., and Stambrook, P. J. (1998) *Anal. Biochem.* **256**, 137–140
- Ballicora, M. A., Laughlin, M. J., Fu, Y., Okita, T. W., Barry, G. F., and Preiss, J. (1995) *Plant Physiol.* **109**, 245–251
- Iglesias, A. A., Barry, G. F., Meyer, C., Bloksberg, L., Nakata, P. A., Greene, T., Laughlin, M. J., Okita, T. W., Kishore, G. M., and Preiss, J. (1993) *J. Biol. Chem.* **268**, 1081–1086
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) *Anal. Biochem.* **150**, 76–85
- Yep, A., Bejar, C. M., Ballicora, M. A., Dubay, J. R., Iglesias, A. A., and Preiss, J. (2004) *Anal. Biochem.* **324**, 52–59
- Ballicora, M. A., Fu, Y., Nesbitt, N. M., and Preiss, J. (1998) *Plant Physiol.* **118**, 265–274
- Hill, A. V. (1910) *J. Physiol. (Lond.)* **40**, 4–7
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) *Numerical Recipes in C: The Art of Scientific Computing*, pp. 656–706, Cambridge University Press, New York
- Igarashi, R. Y., and Meyer, C. R. (2000) *Arch. Biochem. Biophys.* **376**, 47–58
- Uttaro, A. D., Ugalde, R. A., Preiss, J., and Iglesias, A. A. (1998) *Arch. Biochem. Biophys.* **357**, 13–21
- Ballicora, M. A., Sesma, J. I., Iglesias, A. A., and Preiss, J. (2002) *Biochemistry* **41**, 9431–9437
- Meyer, C. R., Ghosh, P., Nadler, S., and Preiss, J. (1993) *Arch. Biochem. Biophys.* **302**, 64–71
- Charng, Y. Y., Kakefuda, G., Iglesias, A. A., Buikema, W. J., and Preiss, J. (1992) *Plant Mol. Biol.* **20**, 37–47
- Doan, D. N., Rudi, H., and Olsen, O. A. (1999) *Plant Physiol.* **121**, 965–975
- Frueauf, J. B., Ballicora, M. A., and Preiss, J. (2001) *J. Biol. Chem.* **276**, 46319–46325
- Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol.* **234**, 779–815
- Peneff, C., Ferrari, P., Charrier, V., Taburet, Y., Monnier, C., Zamboni, V., Winter, J., Harnois, M., Fassy, F., and Bourne, Y. (2001) *EMBO J.* **20**, 6191–6202
- Sivaraman, J., Sauve, V., Matte, A., and Cygler, M. (2002) *J. Biol. Chem.* **277**, 44214–44219
- Sanchez, R., and Sali, A. (2000) in *Protein Structure Prediction, Methods and Protocols* (Webster, D. M., ed) pp. 97–129, Humana Press, Totowa, NJ
- Jones, D. T. (1999) *J. Mol. Biol.* **292**, 195–202
- McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) *Bioinformatics* **16**, 404–405
- Jones, D. T. (1999) *J. Mol. Biol.* **287**, 797–815
- Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) *Nature* **356**, 83–85
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Krieger, E., Nabuurs, S. B., and Vriend, G. (2003) in *Structural bioinformatics* (Bourne, P. E., and Weissig, H., eds) pp. 509–523, Wiley-Liss, Inc., Hoboken, NJ
- Jin, X., Ballicora, M. A., Preiss, J., and Geiger, J. H. (2005) *EMBO J.*, in press
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) *Trends Biochem. Sci.* **23**, 403–405
- Hannah, L. C., Shaw, J. R., Giroux, M. J., Reyss, A., Prioul, J.-L., Bae, J.-M., and Lee, J.-Y. (2001) *Plant Physiol.* **127**, 173–183
- Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425
- Choi, J. H., Jung, H. Y., Kim, H. S., and Cho, H. G. (2000) *Bioinformatics* **16**, 1056–1058
- Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002) *Bioinformatics* **18**, 502–504
- Page, R. D. M. (1996) *Comput. Appl. Biosci.* **12**, 357–358
- Wang, S.-M., Lue, W.-L., Yu, T.-S., Long, J.-H., Wang, C.-N., Eimert, K., and Chen, J. (1998) *Plant J.* **13**, 63–70
- Blankenfeldt, W., Asuncion, M., Lam, J. S., and Naismith, J. H. (2000) *EMBO J.* **19**, 6652–6663
- Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreux, D., Cambillau, C., and Bourne, Y. (1999) *EMBO J.* **18**, 4096–4107
- Gómez-Casati, D. F., Igarashi, R. Y., Berger, C. N., Brandt, M. E., Iglesias, A. A., and Meyer, C. R. (2001) *Biochemistry* **40**, 10169–10178
- Richard, S. B., Bowman, M. E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A. M., Cane, D. E., and Noel, J. P. (2001) *Nat. Struct. Biol.* **8**, 641–648
- Bremer, K. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4707–4711
- Mathews, S., and Donoghue, M. J. (1999) *Science* **286**, 947–950
- Todd, A. E., Orenge, C. A., and Thornton, J. M. (2002) *Structure (Lond.)* **10**, 1435–1451
- Thornton, J. W. (2004) *Nat. Rev. Genet.* **5**, 366–375
- Gerlt, J. A., and Babbitt, P. C. (2000) *Genome Biol.* **1**, reviews0005.1–0005.10
- Gould, S. J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10750–10755
- Pils, B., and Schultz, J. (2004) *J. Mol. Biol.* **340**, 399–404
- Grishin, N. V. (2001) *J. Mol. Biol.* **307**, 31–37
- Aravind, L., and Koonin, E. V. (1998) *Curr. Biol.* **8**, R111–R113



Supplemental Figure 1 Phylogenetic tree of the L subunits from plants. The tree was calculated as described in Supplementary Methods by the Maximum-likelihood method with the program TREE-PUZZLE. The reference numbers in bold and italic correspond to the sequences listed in Supplemental Table I. Sequences 42 and 47 (Supplemental Table I) were not included because they are not 100% complete. The small numbers are the support provided by the program TREE-PUZZLE for the respective branches of the tree. The arrow indicates the L subunit from potato tuber used in this work. The circled support numbers refer to the branches that determine the topology of the main groups. Numbers between 90 and 100 are of very high confidence. The scale on the bottom left represents the number of substitutions per site.

SUPPLEMENTAL TABLE I.

ADP-Glc PPases from oxygenic photosynthesizers

#	Accession	Species	Source	Reference
1	S24991	Anabaena sp. PCC 7120	prokaryote	(1)
Small subunits				
2	AAF75832	<i>Chlamydomonas reinhardtii</i>	unicellular eukaryote	(2)
Monocots				
3	AAK27313	<i>Oryza sativa</i>	endosperm plastidial	(3)
4	AAO16183	<i>Hordeum vulgare</i> 'Vulgare'	endosperm plastidial	(4)
5	AAK39640	<i>Zea mays</i>	embryo	(5)
6	P15280	<i>Oryza sativa</i>	endosperm	(6)
7	BAC75439	<i>Oryza sativa</i> 'Japonica'	NR*	NR
8	CAA88449	<i>Hordeum vulgare</i> 'Vulgare'	endosperm, cytosolic	(4,7)
9	P30523	<i>Triticum aestivum</i>	grains	(8)
10	AAF61173	<i>Triticum aestivum</i>	endosperm	‡
11	CAA88450	<i>Hordeum vulgare</i> 'Vulgare'	leaf, endosperm, putative plastidial	(4,7)
12	AAK69628	<i>Zea mays</i>	leaf	(5)
13	AAK69627	<i>Zea mays</i>	endosperm	(9)
Dicots				
14	AAF66435	<i>Perilla frutescens</i>	stem, root, seed, leaf	(10)
15	CAA58473	<i>Ipomoea batatas</i>	tuber, leaf (inducible by sucrose)	(11)
16	Q42882	<i>Lycopersicon esculentum</i>	leaf, fruit	(12)
17	P23509	<i>Solanum tuberosum</i>	tuber, leaf	(13,14)
18	P55228	<i>Arabidopsis thaliana</i>	leaf	(15,16)
19	CAB89863	<i>Brassica napus</i>	seed, leaf	‡
20	AAK27684	<i>Brassica rapa</i> 'Pekinensis'	extracted from leaf	(17)
21	AAF66434	<i>Perilla frutescens</i>	stem, root, seed, leaf	(10)
22	CAB01912	<i>Ipomoea batatas</i>	tuber, stem, leaf (inducible by sucrose)	(18)
23	CAB01911	<i>Ipomoea batatas</i>	tuber, stem, leaf	(18)
24	CAA65539	<i>Pisum sativum</i>	seed, leaf	(19)
25	P52417	<i>Vicia faba</i>	seed, leaf	(20)
26	AAK27720	<i>Cicer arietinum</i>	seed, weak in leaf	(21)
27	BAC66693	<i>Phaseolus vulgaris</i>	library from seed	‡
28	P52416	<i>Vicia faba</i>	seed	(20)
29	CAA65540	<i>Pisum sativum</i>	seed plastidial, weak in leaf, stem and root	(19)
30	AAK27721	<i>Cicer arietinum</i>	seed, weak in leaf	(21)
31	AAS00541	<i>Fragaria ananassa</i>	NR	NR
32	P55232	<i>Beta vulgaris</i>	sink and source leaf, weaker in hypocotyls, tap root, root	(22)
33	CAA58475	<i>Spinacia oleracea</i>	leaf	‡
34	AAB91462	<i>Cucumis melo</i>	leaf, early fruit, weaker in root and stem	(23)
35	AAB91466	<i>Citrullus lanatus</i>	extracted from fruit	(24)
36	AAD56041	<i>Citrus unshiu</i>	fruit, leaf	(25)
37	Literature [†]	<i>Manihot esculenta</i> 'Crantz'	tuber, leaf	(26)

Large subunits				
38	Literature [†]	<i>Chlamydomonas reinhardtii</i>	unicellular eukaryote	(2)
Group I				
39	AAK27685	<i>Brassica rapa</i> 'Pekinensis'	extracted from leaf	(17)
40	P55229	<i>Arabidopsis thaliana</i>	leaf	(27)
41	AAK27718	<i>Cicer arietinum</i>	leaf	(21)
42	AAS00543	<i>Fragaria ananassa</i>	NR	NR
43	AAC49943	<i>Lycopersicon esculentum</i>	leaf	(28)
44	P55243	<i>Solanum tuberosum</i>	tuber	(29)
45	AAM95945	<i>Oncidium</i> 'Goldiana'	leaf, flower	(30)
46	AAC49729	<i>Hordeum vulgare</i>	leaf, weak in endosperm	(31)
47	P12298	<i>Triticum aestivum</i>	leaf, endosperm, embryo	(32,33)
48	NP_911710	<i>Oryza sativa</i> 'Japonica'	NR	NR
Group II				
49	P55230	<i>Arabidopsis thaliana</i>	NR	NR
50	AAB91464	<i>Cucumis melo</i>	stem, root, early fruit weaker in leaf	(23)
51	AAB91468	<i>Citrullus lanatus</i>	extracted from fruit	(24)
52	AAK27719	<i>Cicer arietinum</i>	seeds, weak in leaf, stem and root	(21)
53	P55242	<i>Solanum tuberosum</i>	tuber, leaf	(29)
54	AAC49942	<i>Lycopersicon esculentum</i>	fruit, root	(28)
55	P55233	<i>Beta vulgaris</i>	sink and source leaf, weaker in hypocotyls, tap root, root	(22)
Group III				
56	AAB91467	<i>Citrullus lanatus</i>	extracted from fruit	(24)
57	AAB91463	<i>Cucumis melo</i>	stem, root, early fruit, weaker in leaf	(23)
58	AAS00542	<i>Fragaria ananassa</i>	NR	NR
59	AAD56042	<i>Citrus unshiu</i>	fruit, leaf	(25)
60	Literature [†]	<i>Manihot esculenta</i> 'Crantz'	tuber, stem, leaf	(26)
61	BAC66692	<i>Phaseolus vulgaris</i>	library from seed	‡
62	CAA65541	<i>Pisum sativum</i>	sink tissues: seed, pod, seed coat	(19)
63	AAC21562	<i>Ipomoea batatas</i>	tuber, stem (induced by sucrose)	(34)
64	CAB55495	<i>Ipomoea batatas</i>	tuber, stem (induced by sucrose)	(34)
65	AAF66436	<i>Perilla frutescens</i>	stem, cotyledon	(10)
66	Q00081	<i>Solanum tuberosum</i>	tuber, weaker in leaf	(35)
67	AAD56405	<i>Lycopersicon hirsutum</i>	fruit	(36)
68	AAC49941	<i>Lycopersicon esculentum</i>	stems, roots, early fruit	(28)
69	P55231	<i>Arabidopsis thaliana</i>	sink tissues, inducible in mesophyll by sucrose	(37)
70	Q9SIK1	<i>Arabidopsis thaliana</i>	NR	NR
Group IV				
71	P12299	<i>Triticum aestivum</i>	endosperm	(33,38)
72	P30524	<i>Hordeum vulgare</i>	endosperm	(31,39)
73	P55234	<i>Zea mays</i>	embryo	(40)
74	BAA23490	<i>Oryza sativa</i> 'Japonica'	endosperm	‡
75	P12300	<i>Triticum aestivum</i>	endosperm	(32)
76	P55241	<i>Zea mays</i>	endosperm	(41)
77	AAB38781	<i>Oryza sativa</i>	endosperm	‡
78	AAK27727	<i>Oryza sativa</i>	endosperm	‡

* NR: Expression data not reported

† Sequence is not available in databases and was taken from literature

‡ Information of tissue expression was extracted from the National Center for Biotechnology Information (NCBI) database, <http://www.ncbi.nlm.nih.gov/>

References for Supplemental Table I

1. Charng, Y. Y., Kakefuda, G., Iglesias, A. A., Buikema, W. J., and Preiss, J. (1992) *Plant Mol. Biol.* **20**, 37-47
2. Zabawinski, C., Van den Koornhuysse, N., D'Hulst, C., Schlichting, R., Giersch, C., Delrue, B., Lacroix, J. M., Preiss, J., and Ball, S. (2001) *J. Bacteriol.* **183**, 1069-1077
3. Sikka, V. K., Choi, S.-B., Kavakli, I. H., Sakulsingharoj, C., Gupta, S., Ito, H., and Okita, T. W. (2001) *Plant Sci.* **161**, 461-468
4. Johnson, P. E., Patron, N. J., Bottrill, A. R., Dinges, J. R., Fahy, B. F., Parker, M. L., Waite, D. N., and Denyer, K. (2003) *Plant Physiol.* **131**, 684-696.
5. Hannah, L. C., Shaw, J. R., Giroux, M. J., Reyss, A., Prioul, J.-L., Bae, J.-M., and Lee, J.-Y. (2001) *Plant Physiol.* **127**, 173-183
6. Anderson, J. M., Larsen, R., Laudencia, D., Kim, W. T., Morrow, D., Okita, T. W., and Preiss, J. (1991) *Gene* **97**, 199-205.
7. Thorbjornsen, T., Villand, P., Kleczkowski, L. A., and Olsen, O. A. (1996) *Biochem. J.* **313**, 149-154.
8. Ainsworth, C., Tarvis, M., and Clark, J. (1993) *Plant Mol. Biol.* **23**, 23-33.
9. Bae, J. M., Giroux, M., and Hannah, L. (1990) *Maydica* **35**, 317-322
10. Choi, S. B., Kim, K. H., Kavakli, I. H., Lee, S. K., and Okita, T. W. (2001) *Plant Cell Physiol* **42**, 146-153.
11. Kim, Y. S., Lee, M. H., Shin, H. S., Lee, D. S., Son, C. B., Song, H. W., Yoo, O. J., and Liu, J. R. (1995) *Mol. Cells* **5**, 605-610
12. Chen, B.-Y., Janes, H. W., and Gianfagna, T. (1998) *Plant Sci.* **136**, 59-67
13. Nakata, P. A., and Okita, T. W. (1996) *Mol. Gen. Genet.* **250**, 581-592
14. Nakata, P. A., Anderson, J. M., and Okita, T. W. (1994) *J. Biol. Chem.* **269**, 30798-30807.
15. Villand, P., Olsen, O. A., and Kleczkowski, L. A. (1993) *Plant Mol. Biol.* **23**, 1279-1284.
16. Wang, S.-M., Lue, W.-L., Yu, T.-S., Long, J.-H., Wang, C.-N., Eimert, K., and Chen, J. (1998) *Plant J.* **13**, 63-70
17. Kim, I.-J., Park, J. Y., Lee, Y.-W., Chung, W.-I., and Lim, Y. P. (2002) *J. Plant Biotechnology* **4**, 59-65
18. Bae, J. M., and Liu, J. R. (1997) *Mol. Gen. Genet.* **254**, 179-185
19. Burgess, D., Penton, A., Dunsmuir, P., and Dooner, H. (1997) *Plant Mol. Biol.* **33**, 431-444.
20. Weber, H., Heim, U., Borisjuk, L., and Wobus, U. (1995) *Planta* **195**, 352-361
21. Singh, S., Choi, S. B., Modi, M. K., and Okita, T. W. (2002) *Phytochemistry* **59**, 261-268
22. Müller-Röber, B., Nast, G., and Willmitzer, L. (1995) *Plant Mol. Biol.* **27**, 191-197
23. Park, S. W., Kahng, H. Y., Kim, I. J., Park, J. O., and Chung, W. I. (1998) *J. Plant Res.* **111**, 59-63
24. Kim, I. J., Kahng, H. Y., and Chung, W. I. (1998) *Biosci. Biotechnol. Biochem.* **62**, 550-555

25. Kim, I.-J., Noh, S.-J., Lee, B.-H., Jo, J., Kim, Y.-S., and Chung, W.-I. (2001) *Biochim. Biophys. Acta* **1518**, 324-328
26. Munyikwa, T. R. I., Kreuze, J., Fregene, M., Suurs, L., Jacobsen, E., and Visser, R. G. F. (2001) *Euphytica* **120**, 71-83
27. Kavakli, I. H., Kato, C., Choi, S. B., Kim, K. H., Salamone, P. R., Ito, H., and Okita, T. W. (2002) *Planta* **215**, 430-439.
28. Park, S. W., and Chung, W. I. (1998) *Gene* **206**, 215-221.
29. Lacognata, U., Willmitzer, L., and Müller-Röber, B. (1995) *Mol. Gen. Genet.* **246**, 538-548
30. Li, C. R., Zhang, X. B., and Hew, C. S. (2003) *Biol. Plantarum* **47**, 613-615
31. Eimert, K., Luo, C., Dejardin, A., Villand, P., Thorbjornsen, T., and Kleczkowski, L. A. (1997) *Gene* **189**, 79-82
32. Olive, M. R., Ellis, R. J., and Schuch, W. W. (1989) *Plant Mol. Biol.* **12**, 525-538
33. Burton, R. A., Johnson, P. E., Beckles, D. M., Fincher, G. B., Jenner, H. L., Naldrett, M. J., and Denyer, K. (2002) *Plant Physiol.* **130**, 1464-1475.
34. Harn, C. H., Bae, J. M., Lee, S. S., Min, S. R., and Liu, J. R. (2000) *Plant Cell Physiol.* **41**, 1235-1242
35. Nakata, P. A., and Okita, T. W. (1995) *Plant Physiol.* **108**, 361-368
36. Schaffer, A. A., Levin, I., Oguz, I., Petreikov, M., Cincarevsky, F., Yeselson, Y., Shen, S., Gilboa, N., and Bar, M. (2000) *Plant Sci.* **152**, 135-144
37. Rook, F., Corke, F., Card, R., Munz, G., Smith, C., and Bevan, M. W. (2001) *Plant J.* **26**, 421-433
38. Ainsworth, C., Hosein, F., Tarvis, M., Weir, F., Burrell, M., Devos, K. M., and Gale, M. D. (1995) *Planta* **197**, 1-10
39. Villand, P., Aalen, R., Olsen, O. A., Luthi, E., Lonneborg, A., and Kleczkowski, L. A. (1992) *Plant Mol. Biol.* **19**, 381-389
40. Giroux, M., Smith-White, B., Gilmore, V., Hannah, L. C., and Preiss, J. (1995) *Plant Physiol.* **108**, 1333-1334
41. Shaw, J. R., and Hannah, L. C. (1992) *Plant Physiol.* **98**, 1214-1216