GLK gene pairs regulate chloroplast development in diverse plant species

David W. Fitter, David J. Martin, Martin J. Copley, Robert W. Scotland and Jane A. Langdale*

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

Received 9 April 2002; accepted 17 May 2002.
*For correspondence (fax +44 1865 275147; e-mail jane.langdale@plants.ox.ac.uk).

Summary

Chloroplast biogenesis is a complex process that requires close co-ordination between two genomes. Many of the proteins that accumulate in the chloroplast are encoded by the nuclear genome, and the developmental transition from proplastid to chloroplast is regulated by nuclear genes. Here we show that a pair of Golden 2-like (GLK) genes regulates chloroplast development in Arabidopsis. The GLK proteins are members of the GARP superfamily of transcription factors, and phylogenetic analysis demonstrates that the maize, rice and Arabidopsis GLK gene pairs comprise a distinct group within the GARP superfamily. Further phylogenetic analysis suggests that the gene pairs arose through separate duplication events in the monocot and dicot lineages. As in rice, AtGLK1 and AtGLK2 are expressed in partially overlapping domains in photosynthetic tissue. Insertion mutants demonstrate that this expression pattern reflects a degree of functional redundancy as single mutants display normal phenotypes in most photosynthetic tissues. However, double mutants are pale green in all photosynthetic tissues and chloroplasts exhibit a reduction in granal thylakoids. Products of several genes involved in light harvesting also accumulate at reduced levels in double mutant chloroplasts. GLK genes therefore regulate chloroplast development in diverse plant species.

Keywords: GARP genes, transcription factors, chloroplasts, gene redundancy.

Introduction

Chloroplast differentiation is an essential developmental process which takes place in all plant cells that ultimately fix carbon dioxide. In undifferentiated embryonic and meristematic cells, plastids are present in a proplastid form. These small organelles, which can develop into chloroplasts, etioplasts, amyloplasts, leucoplasts or chromoplasts, contain their own genome and are separated from the cytoplasm by a double envelope membrane (reviewed by Mullet, 1988). As cells in the shoot apical meristem divide to form leaf primordia, proplastids within the cells also divide and initiate a programme of chloroplast development. Within the chloroplast, the stroma, the thylakoid membrane and the lumen contained within the folds of the thylakoid membrane all differentiate (Mullet, 1988). Further development distinguishes different types of chloroplasts. For example, in plants such as rice and Arabidopsis, that utilize the C3 photosynthetic pathway, chloroplasts in all cell types exhibit granal (stacked) thylakoids and accumulate Calvin cycle enzymes (reviewed by Edwards and Walker, 1983). In these plants, CO2 is fixed by ribulose bisphosphate carboxylase (RuBPCase) in the mesophyll cells, whereas bundle sheath cells act primarily as supporting and conducting tissue. In contrast, C4 plants such as maize compartmentalize photosynthetic reactions between mesophyll and bundle sheath cells. RuBPCase activity is restricted to bundle sheath cells, whereas mesophyll cells accumulate a set of cell-specific enzymes ( phosphoenolpyruvate carboxylase, pyruvate phosphate dikinase and malate dehydrogenase) that act to fix and then shuttle carbon to the bundle sheath cells (reviewed by Edwards and Walker, 1983). Thus three types of chloroplasts develop in C4 plants (reviewed by Langdale and Nelson, 1991). In addition to the C3-type, which is present at certain developmental stages and in certain tissues, distinct chloroplasts differentiate in the bundle sheath and mesophyll cells of the C4 leaf. These dimorphic chloroplasts accumulate individual complements of photosynthetic enzymes and can exhibit distinct arrangements of thylakoid membranes.
In all chloroplasts the thylakoid membrane contains the components required for light harvesting and photophosphorylation (reviewed by Staehelin and Arntzen, 1986). The membrane is composed of a lipid bilayer that is interspersed with the proteins, pigments and other components of the five functional complexes that are vital for photosynthesis. These complexes are the mobile chlorophyll $a/b$ light-harvesting complex (LHC); photosystem II (PSII); cytochrome $f/b$ (cyt$f/b$); photosystem I (PSI); and adenosine triphosphate (ATP) synthase (Staehelin and DeWit, 1984). The photosystems are portioned in a characteristic fashion between the granal and stromal (unstacked) thylakoids. PSI is localized preferentially to the stromal thylakoids, PSII to the granal thylakoids (reviewed by Allen and Forsberg, 2001). Soluble photosynthetic proteins, including the carboxylating enzyme RuBPCase, accumulate within the stroma (Andrews and Lorimer, 1987).

Assembly of the photosynthetic apparatus requires coordination between the chloroplast and nucleus because complexes are comprised of proteins encoded by both genomes. As such, a large number of nuclear-encoded, cytosolically synthesized proteins have to be imported into the chloroplast. All these proteins are transported by a single pathway that mediates passage through the double envelope membrane (reviewed by Keegstra and Froehlich, 1999). Once inside the chloroplast, however, proteins are targeted to either the thylakoid membrane or the thylakoid lumen by one of four pathways (reviewed by Robinson et al., 1998). The SecA and $\Delta p$H pathways transport luminal proteins, whereas integral membrane proteins are imported by either the signal recognition pathway (SRP) or the spontaneous pathway (Robinson et al., 1998).

Although many of the structural components of the import pathways and photosynthetic complexes are known (Allen and Forsberg, 2001; Robinson et al., 1998), very few genes that regulate chloroplast development have been characterized. One example is the maize Golden2 (G2) gene which acts as a general regulator of chloroplast development in C$_3$-type photosynthetic cells and as a specific regulator of bundle sheath cell chloroplast development in C$_4$ photosynthetic tissue (Hall et al., 1998; Langdale and Kidner, 1994; Rossini et al., 2001). Chloroplasts in g2 mutants are smaller than wild type and develop only rudimentary thylakoid lamellae that exhibit very few grana (Langdale and Kidner, 1994). A second G2-like (Glk) gene, ZmGlk1, is thought to regulate mesophyll cell chloroplast development in C$_4$ tissues (Rossini et al., 2001). Unlike G2, however, ZmGLK1 function has been postulated only on the basis of expression patterns and not on loss-of-function phenotypes.

The GLK proteins are members of the recently categorized GARP superfamily of transcription factors (Riechmann et al., 2000) defined by G2 in maize; the Arabidopsis RESPONSE REGULATOR-B (ARR-B) proteins (Imamura et al., 1999); and the PHOSPHATE STARVATION RESPONSE1 (PSR1) protein of Chlamydomonas (Wykoff et al., 1999). In the case of G2, three of the four defining features of most transcription factors have been verified experimentally in heterologous systems. G2 is nuclear-localized (Hall et al., 1998), is able to transactivate reporter gene expression, and can both homo-dimerize and hetero-dimerize with ZmGLK1 (Rossini et al., 2001). DNA-binding activity of GLK proteins has yet to be demonstrated, however, the putative DNA-binding domain is highly conserved with domains in other GARP proteins such as ARR1 and ARR2 (Riechmann et al., 2000). Notably, ARR1 and ARR2 have been shown to bind DNA (Sakai et al., 2000), thus it is likely that GLK proteins act as transcriptional regulators of chloroplast development.

The spatial compartmentalization of G2 and ZmGlk1 transcripts in C$_4$ maize tissue may represent a specialization required for the development of distinct bundle sheath and mesophyll chloroplasts (Rossini et al., 2001). This view is supported by the observation that the rice genes OsGlk1 and OsGlk2, which are orthologous to ZmGlk1 and G2, respectively, are expressed in overlapping domains in photosynthetic tissues (Rossini et al., 2001). Thus the two rice genes may act redundantly. Unfortunately, this suggestion has not been confirmed because loss-of-function mutations in OsGlk genes are not yet available. We have therefore examined GLK gene function in the C$_3$ dicot Arabidopsis with a view to testing the hypothesis that GLK gene pairs act redundantly to promote chloroplast development in C$_3$ species. In this report we demonstrate that two GLK genes are present in the Arabidopsis genome, and that these genes are structurally similar to the monocot Glk genes. A phylogenetic analysis confirms that the GARP gene superfamily is monophyletic in the context of putative relatives, and that the GLK gene pairs from Arabidopsis, maize and rice comprise a monophyletic group within the superfamily.

Through characterization of insertional mutants, we demonstrate that the GLK gene products function in a redundant manner as transcriptional regulators of C$_3$-type chloroplast development. We propose that this redundancy enabled one GLK gene to play a derived role in the evolution of distinct bundle sheath and mesophyll chloroplasts in C$_4$ plants.

**Results**

**GLK genes in Arabidopsis**

Two Arabidopsis GLK genes were identified when the ZmG2 sequence was used as a query in BLAST searches of the Arabidopsis genome database: http://www.arabidopsis-
The AtGLK1 gene encodes a predicted protein of 420 amino acids and the AtGLK2 gene encodes a protein of 386 amino acids. Figure 1 illustrates the sequence similarity between the Arabidopsis GLK proteins and the GLK proteins of maize and rice. Sequence conservation is particularly high across the putative DNA-binding domain and across the C-terminal domain, referred to as the GCT box, which is required to mediate homo- and hetero-dimerization of the maize proteins (Rossini et al., 2001). The DNA-binding domain has previously been postulated to fold as a helix-loop-helix (Rossini et al., 2001). However, using all six protein sequences, the consensus output from secondary structure prediction software is a domain that folds as three α-helices (Figure 1). Immediately downstream of the third helix is an AREAEAA motif that is conserved exactly in five of the GLK proteins, and is present as AREVEAA in ZmGLK1.

Both Arabidopsis GLK sequences contain recognizable nuclear localization signals (NLS). PSORT software predictions (Nakai and Kanehisa, 1992) indicate that AtGLK1 has an NLS of the Simian Virus 40 (SV40) large T antigen type (Kalderon et al., 1984) just upstream of the DNA-binding domain, at amino acid position 130–133. A similar type of NLS is found in OsGLK2. AtGLK2 is also predicted to have
an SV40 large T antigen-type NLS, downstream of the DNA-binding domain at position 223–226. In addition, however, like ZmG2 and ZmGLK1, AtGLK2 has an NLS upstream of the DNA-binding domain that is of the Xenopus nucleoplasmin type (Robbins et al., 1991).

GLK gene phylogeny

The GLK DNA-binding domain is characteristic of the newly defined GARP superfamily of transcription factors (Riechmann et al., 2000). GARP genes have thus far been identified only in plant genomes, however, the GARP DNA-binding domain sequence is similar to TEA-binding domain sequences (Bürglin, 1991) and to the single MYB-like domain in MYB-related proteins (Baranowskij et al., 1994). Both TEA- and MYB-related sequences are found in a range of eukaryotic genomes. To determine whether GARP gene sequences are phylogenetically distinct from TEA- and MYB-related sequences, phylogenetic analysis was carried out with 38 Arabidopsis GARP gene sequences; Psr1 from Chlamydomonas (Wykoff et al., 1999); maize and rice Glk gene sequences (Rossini et al., 2001); six TEA genes; and the MYB-related potato gene MybSt1. Figure 2(a) demonstrates that there is strong bootstrap support (86%) for the monophyly of the GARP superfamily relative to TEA genes and MybSt1. Thus GARP genes are phylogenetically distinct from both TEA- and MYB-related genes. Previous reports placing PSR1 and related Arabidopsis genes in the MYB family are therefore inaccurate (Rubio et al., 2001), as PSR1 is clearly nested within the GARP superfamily.

The GARP superfamily is reported to comprise 56 genes in Arabidopsis (Riechmann et al., 2000). Prior to analysis of AtGLK gene function, it was therefore important to determine how many of these GARP genes could be classified as GLK genes. Using the presence of both the DNA-binding domain and the GCT box as the defining criterion of GLK gene structure, AtGLK1, AtGLK2 and the Arabidopsis PSEUDO RESPONSE REGULATOR 2 (APRR2) gene (Makino et al., 2000) were identified as GLK genes. APRR2 encodes a protein that comprises a GARP DNA binding domain, a domain similar to the GCT box, and a receiver-like domain at the N-terminus which is similar to that encoded by the ARR-B genes (Imamura et al., 1999; Makino et al., 2000). Phylogenetic analysis of nucleotide sequence encoding only the DNA-binding domain demonstrated reasonable bootstrap support for the existence of a distinct GLK gene group nested within the GARP superfamily (Figure 2a). Notably, however, AtGLK1 and AtGLK2 belong to this group but APRR2 does not. APRR2 also lacks the AREAEAA motif. Thus only two GLK genes are present in the Arabidopsis genome.

Relationship between monocot and dicot GLK genes

Based on synteny analysis, gene sequence, intron/exon structure and expression patterns, relationships between

the maize and rice Glk genes have been determined (Rossini et al., 2001). ZmGlk1 is orthologous to OsGlk1 and G2 is orthologous to ZmG2. Thus the gene duplication occurred prior to the divergence of rice and maize. To assess the relationship between Arabidopsis GLK genes and the monocot genes, phylogenetic analysis was carried out using the complete cDNA sequence of each GLK gene and APRR2 cDNA sequence as an outgroup. Figure 2(b) demonstrates that the Arabidopsis GLK genes are more closely related to each other than they are to any of the monocot genes. Therefore the simplest hypothesis is that independent duplication events occurred in the monocot and dicot lineages. However, gene structure data are not entirely congruent with this suggestion. For example, AtGLK2 has five exons, like ZmG2, whereas AtGLK1 has six exons, like ZmGlk1 and OsGlk1. OsGlk2 has six exons, but the position of the extra intron is not the same as in the GLK1 genes (Figure 1). To resolve the discrepancy between the phylogenetic and gene structure data, further GLK gene sequences need to be analysed.

AtGLK gene transcript accumulation profiles

Consistent with a proposed role in chloroplast development, maize and rice Glk gene transcripts accumulate exclusively in photosynthetic tissue (Hall et al., 1998; Rossini et al., 2001). To determine the spatial expression profiles of the Arabidopsis GLK genes, RNA gel-blot analysis was carried out with RNA extracted from different tissues. Figure 3 demonstrates that both AtGLK1 (Figure 3a and AtGLK2 (Figure 3b) transcripts are predominantly detected in rosette and cauline leaves. In cotyledons, shoots and flowers, AtGLK1 transcript levels are lower than AtGLK2 levels. Only AtGLK2 transcripts accumulate in roots and siliques. Thus in the most photosynthetically
Figure 4. Atglk1.1 and Atglk2.1 mutant phenotypes.
(a,b) Schematic diagram showing the dSpm insertion site in Atglk1.1 (a) and Atglk2.1 (b) alleles. Horizontal lines represent untranslated regions. Boxes represent different domains of the coding region. Black triangles designate intron positions conserved in both genes; red triangle represents position of extra intron in AtGLK1.
(c,d) RNA gel-blot analysis of AtGLK1 (c) and AtGLK2 (d) transcript accumulation in wild-type (WT), Atglk1.1, Atglk2.1 and double mutant plants. The ethidium bromide-stained gel was used to assess RNA loading levels and relative hybridization intensity was determined as for Figure 3(a,b).
(e) Wild-type and mutant phenotypes 4 weeks after germination.
(f) Siliques of wild-type and mutant plants.
active tissues, both AtGLK1 and AtGLK2 transcripts accumulate. In other tissues, AtGLK2 transcripts predominate.

To assess whether light influences AtGLK transcript accumulation patterns, RNA was extracted at 2 h intervals through a 28 h diurnal light/dark cycle from rosette leaves of 4-week-old plants. Figure 3(c) demonstrates that AtGLK1 transcript levels are noticeably higher in samples harvested during the 16 h light period than in samples harvested during the dark period. Light also appeared to influence AtGLK2 transcript accumulation levels (Figure 3d), however, transcript levels increased in anticipation of the light period. This observation suggested that AtGLK2 transcript accumulation profiles are regulated by an endogenous (possibly circadian) mechanism. To investigate this suggestion further, plants were entrained to a light/dark cycle for 4 weeks, then kept in the dark for 28 h. As before, samples were harvested every 2 h through the 28 h period and RNA was extracted and analysed. Figure 3(e) demonstrates that AtGLK1 transcript levels remained low throughout the 28 h period, confirming the observation that accumulation levels are light-regulated. In contrast, however, the AtGLK2 transcript accumulation profile (Figure 3f) was similar to that seen previously (Figure 3d), although absolute levels of transcript were lower during the first half of the light period. This suggests that AtGLK2 transcript accumulation patterns are regulated both by endogenous mechanisms and by light.

Despite the fact that AtGLK1 transcript levels are noticeably higher in the light period of a diurnal cycle, only very low levels were detected during the conversion from etioplast to chloroplast (data not shown). In contrast, AtGLK2 transcript levels increased dramatically after dark-grown plants were exposed to light (data not shown). Thus it is likely that AtGLK2 functions to facilitate the etioplast-to-chloroplast conversion and that AtGLK1 plays only a limited role.

AtGLK1 and AtGLK2 exhibit functional redundancy

To elucidate GLK gene function, insertional mutants were identified in the Sainsbury Laboratory Arabidopsis Transposant collection (Tissier et al., 1999). The Atgk1.1 mutant contains a dSpm insertion 112 bp upstream of the AtGLK1 translation start site (Figure 4a) and the Atgk2.1 mutant contains a dSpm in exon I of AtGLK2, 333 bp downstream of the translation start site (Figure 4b). To determine whether the dSpm insertions altered the expression patterns of the respective genes, RNA gel-blot analysis was carried out. Figure 4c) demonstrates that AtGLK1 transcripts are barely detectable in Atgk1.1 mutants but accumulate to normal levels in Atgk2.1 mutants. Similarly, AtGLK2 transcripts do not accumulate in Atgk2.1 mutants but are present at normal levels in Atgk1.1 mutants (Figure 4d). One exception to this observation is the detection of a truncated AtGLK2 transcript in young tissue of Atgk2.1 mutants (data not shown). This truncated transcript lacks the 3' end and is therefore assumed to be non-functional.

Despite the fact that GLK gene transcripts do not accumulate, the single mutants are essentially indistinguishable from wild type throughout most of development (Figure 4e). This suggests that the AtGLK genes are functionally redundant. The exception to this redundancy is observed in siliques tissue. Atgk1.1 mutants exhibit pale green siliques (Figure 4f). Notably, AtGLK1 transcripts do not accumulate in siliques of wild-type plants (Figure 3a), suggesting that the absence of both gene products leads to a pale green phenotype. To confirm this suggestion, Atgk1.1,Atgk2.1 double-mutant plants were generated. Double mutants accumulated low levels of AtGLK1 transcripts (Figure 4c) and barely detectable levels of AtGLK2 transcripts (Figure 4d). The presence of AtGLK transcripts, albeit at low levels, suggests that there may be some residual GLK function in the double mutants. However, plants were pale green throughout development (Figure 4e,f). Although plants were smaller than wild type, no other phenotypic perturbations were observed. Thus AtGLK genes are likely to function specifically to promote chloroplast development.

AtGLK gene function is dosage dependent.

Mutant phenotypes suggested that AtGLK1 and AtGLK2 act in a redundant manner. To determine the effect of gene copy number on GLK function, chlorophyll accumulation levels were analysed in segregating F1 individuals from a selfed double heterozygous plant. Two wild-type and four double mutant plants were preselected, then 18 plants were selected randomly from the population. Figure 5(a) shows that chlorophyll levels were markedly reduced in double mutants as compared to the other 20 plants. In all cases, chlorophyll levels were higher in rosette leaves than cauline leaves.

After chlorophyll levels had been determined, individual plants were genotyped to assess the number of functional GLK gene copies in each. Figure 5(b) shows that wild-type and mutant AtGLK1 alleles were distinguished with a HindIII digest. Wild-type alleles yielded a 6.5 kb fragment and mutant alleles a 4.3 kb fragment. Similarly, AtGLK2 alleles were distinguished with an EcoRV digest (AtGLK2, 6.5 kb; Atgk2.1, 6.8 kb). Of the 18 plants, five had one, eight had two and five had three GLK gene copies.

Having determined gene copy number, ANOVA was carried out to determine whether copy number influenced chlorophyll levels. Figure 5(c) shows that both AtGLK1 and AtGLK2 copy numbers affect chlorophyll levels. Notably, plants with only one AtGLK copy (plants 1 and 13) are phenotypically closer to wild type in appearance than they
are to the double mutant. Thus a threshold level of GLK protein (equivalent to or less than one wild-type gene copy) is required for plants to appear normal (green). As the interaction term (GLK1*GLK2) is significant, the data further suggest that loss of both genes has a greater effect than would be expected were the effects of the Atglk1.1 and Atglk2.1 mutant alleles simply additive (Figure 5c). Therefore, in addition to functioning on their own, the two gene products may interact. This suggestion is supported by the observation that AtGLK1 and AtGLK2 both homo- and hetero-dimerize in the yeast two-hybrid system (data not shown).

**Chloroplast development is perturbed in Atglk double mutants**

Loss of Glk gene function in maize leads to perturbed chloroplast development – plastids are small and thylakoids are rudimentary (Langdale and Kidner, 1994). To

---

**Figure 5. Chlorophyll concentration in plants segregating Atglk1.1 and Atglk2.1 alleles.**

(a) Mean chlorophyll levels in cauline (white) and rosette (black) leaves. Two wild-type (WT) and four double mutant (D) plants were preselected, but the remaining 18 plants were chosen randomly.

(b) DNA gel-blot analysis of HindIII (AtGLK1)- and EcoRV (AtGLK2)- digested DNA extracted from individual plants assayed in (a). Blots were hybridized with AtGLK1 and AtGLK2 gene-specific fragments.

(c) Output from ANOVA. The error term 31.53 is a measure of within-subject variation (the variation within plants). The error term 42.97 is a measure of the between-subject variation (variation between plants of the same genotype), and is the result of nesting variables ‘Plant number’ with ‘Copies of GLK gene’ which is specified as a random variable. Importantly, the between subject error is not significant (1.36) and therefore the variation between plants of the same genotype is not significantly greater than that between leaves within a plant. Thus plants with the same genotype can be treated as identical. The between-plants error term is used as the denominator to calculate the F values for the main effect variable ‘Copies GLK1’, ‘Copies GLK2’, and the interaction term ‘Copies GLK1*GLK2’. The within-plant error is used as the denominator to calculate the F value for the ‘Leaf Type’ factor.

df = degrees of freedom; SeqSS = sequential sums of squares; AdjSS = adjusted sums of squares; AdjMS = adjusted mean squares.

---

**Figure 6. Chloroplast ultrastructure in wild-type and Atglk mutant plants.** TEM of bundle sheath (a,c,e,g) and mesophyll (b,d,f,h) chloroplasts in wild-type (a,b); Atglk1.1 (c,d); Atglk2.1 (e,f); and double mutant (g,h) plants. Solid arrows point to granal lamellae; open arrows to stromal lamellae. Scale bar = 1 μm.
determine whether the AtGLK genes influence chloroplast ultrastructure, transmission electron microscopy was carried out on wild-type and mutant rosette leaf tissue. Figure 6 shows that in both mesophyll and bundle sheath cells of double mutants, chloroplasts are ~50% smaller in cross-sectional area than the wild type. As chloroplasts divide by binary fission of existing organelles (reviewed by Osteryoung, 2000), a negative correlation exists between chloroplast number and chloroplast size. This is clear from chloroplast division mutants, such as accumulation and replication of chloroplasts 6 (arc6), where only two chloroplasts exist per cell, but chloroplasts are greatly enlarged as compared to the wild type (Pyke et al., 1994; Robertson et al., 1995). The arc mutants suggest that plastid division is distinct from plastid differentiation, as proliferation is affected, but the chloroplasts appear to develop normally (Pyke and Leech, 1994; Pyke et al., 1994). Thus the ARC genes must specifically regulate chloroplast division. In contrast, the GLK genes appear to effect plastid development and differentiation but not plastid division, as chloroplast numbers per cell are normal in double mutants (data not shown).

In addition to a reduction in size, chloroplasts in double mutants exhibit rudimentary thylakoid lamellae (Figure 6g,h). The absence of grana is striking in that chloroplasts contain almost entirely stromal lamellae. Notably, this chloroplast phenotype is very similar to that seen in chloroplasts of the g2 mutant of maize. This suggests that the GLK genes in maize and Arabidopsis have similar functions.

To further characterize the defects in chloroplast development that are conditioned by loss of AtGLK gene function, components of the photosynthetic machinery were examined by RNA and protein gel-blot analysis. To assess whether the accumulation of Calvin cycle enzymes was perturbed in mutant plants, RuBPCase accumulation patterns were examined. Both the large (LSU) and small (SSU) subunit of RuBPCase accumulated to normal levels in both single and double mutant plants (Figure 7a). Thus it is likely that mutant plants have the capacity to fix carbon.

The most obvious feature of chloroplast ultrastructure in double mutant plants is a lack of granal thylakoids. To test whether Atglk double mutants are fundamentally impaired in thylakoid membrane formation, levels of the VESICLE-INDUCING PROTEIN IN PLASTIDS 1 (VIPP1) protein were compared in wild-type and mutant plants. VIPP1 is a hydrophilic, integral thylakoid membrane protein that has been shown to be essential for vesicle formation (Kroll et al., 2001). Figure 7(a) demonstrates that VIPP1 levels are normal in double mutants and thus suggests that membrane assembly per se is unaffected by loss of GLK function.

The reduced levels of chlorophyll seen in double mutant plants suggested that components of the light-harvesting

![Figure 7. Accumulation patterns of photosynthetic gene products in wild-type and Atglk mutant plants.](image-url)
chlorophyll binding proteins that hold chlorophyll molecules within the thylakoid membranes. Figure 7(a) demonstrates that LHCP-II protein levels are reduced in double mutant plants but not in single mutant plants, in accord with macroscopic observations (Figure 4e). LHCP1 and LHCP6 gene transcript levels are also greatly reduced in the double mutant (Figure 7b).

The LHCP proteins are inserted into the thylakoid membrane through the SRP pathway (Schuenemann et al., 1998) which requires two stromal factors, cpSRP43 and cpSRP54. cpSRP43 dimerizes with cpSRP54 to bind LHCP (Groves et al., 2001; Jonas-Straube et al., 2001), but then transport across the membrane is facilitated by cpSRP43 alone (Klimyuk et al., 1999). SRP43 transcript levels are reduced in the double mutant, whereas SRP54 transcript levels are unaffected (Figure 7b). However, SRP43 protein levels are normal in mutant plants (data not shown), suggesting that the reduced levels of SRP43 transcripts are a consequence of reduced LHCP, rather than a cause. Other components of the photosynthetic apparatus that are transported across the thylakoid membrane to the lumen utilize the SecA or Sec pathways (Mould and Robinson, 1991; Nielsen et al., 1994; Yuan et al., 1994). Preliminary analysis suggests that other thylakoid-import pathways are unaffected by loss of GLK function (data not shown).

Although the reduced chlorophyll levels observed in double mutant plants may result simply from a reduced ability to anchor chlorophyll in the membrane, there may also be a defect in chlorophyll biosynthesis. Levels of transcripts encoding both subunit H of magnesium chelatase (CHLH; an enzyme that functions early in chlorophyll biosynthesis) and chlorophyll a/b oxygenase (CAO; a key enzyme in the latter stages of chlorophyll biosynthesis) are reduced in double mutant plants (Figure 7c).

To assess whether other complexes that accumulate in the thylakoid membrane are perturbed in double mutants, representative proteins from PSI (PsaD), PSII (D1 polypeptide), ATP synthase (CF1α), cyt f/b6 (Cytf) and the NAD(P)H dehydrogenase (NDHH) complexes were examined by protein gel-blot analysis. Previous studies have demonstrated that the stability of ATP synthase and cytochrome f/b6 complexes is dependent on the accumulation of all the subunits within a complex (Barkan et al., 1995; Rochaix, 1992). Thus the level of any individual subunit reflects the integrity of the complex as a whole. Consistent with the observation that LHCP-II levels are lower in double mutants (Figure 7a), Figure 7(d) shows that D1 polypeptide levels are also lower than in wild-type and single mutant plants. Thus PSII assembly is impaired. In contrast, all other proteins assayed are present at normal levels, suggesting that most of the complexes are unaffected by loss of GLK function.

Discussion

Chloroplasts are thought to have derived from a cyanobacterium ancestor by endosymbiosis (Margulis, 1975). During evolution, most of the prokaryotic genome of the ancestral cyanobacterium has been lost or transferred to the euukaryotic nuclear genome of the host cell (Douglas, 1994). Depending on the plant species, chloroplast genomes now contain between 87 and 183 known genes, around half of which encode components of the chloroplast translational machinery (Sugiura et al., 1998). The remaining chloroplast genes encode components of the ATP-synthesizing complex, the cytochrome b/f complex, PSI, PSII, various subunits of the respiratory enzyme NAD(P)H dehydrogenase, and also the LSU of RuBPCase (Lawlor, 1993). The nucleus therefore encodes many chloroplast-localized photosynthetic proteins including five subunits of PSI, four subunits of PSII, the SSU of RuBPCase, and proteins involved in import mechanisms such as the SecA pathway. Perhaps more importantly, the nucleus encodes genes that regulate the developmental programme of chloroplast biogenesis. Of these regulatory genes, the GLK genes are now among the best characterized. There are no GLK gene sequences in the genome of the cyanobacterium Synechocystis; however, all chloroplast-containing organisms examined, from algae to higher plants, appear to contain GLK sequences in their genomes. These observations suggest first, that the GLK genes did not originate in the chloroplast genome, and second, that GLK gene function is associated specifically with chloroplast development rather than with the ability to photosynthesize.

On the basis of AtGLK gene transcript accumulation patterns (Figure 3) and mutant phenotypes (Figure 4), it is proposed that AtGLK2 functions in photosynthetic tissues throughout development. The spatial pattern of transcript accumulation (Figure 3a,b) suggests a role in chloroplast maintenance, whereas the profile during the etioplast to chloroplast transition (data not shown) suggests a pivotal role during early chloroplast biogenesis. During vegetative growth, AtGLK2 transcript profiles are regulated by endogenous mechanisms, but accumulation levels are enhanced by light. In contrast, AtGLK1 accumulation profiles are regulated by light – levels are highest during the light period of the diurnal cycle. It is therefore proposed that AtGLK1 acts to boost photosynthetic capability in tissues with a high photosynthetic workload. This suggestion is supported by the accumulation of AtGLK1 transcripts exclusively in leaf tissue, and predominantly in rosette and cauline leaves where the majority of photosynthesis takes place.

The defects in Atglk double mutant chloroplasts are limited to a reduction in granal thylakoids and lower chlorophyll, LHCP-II and D1 polypeptide levels. As VIPP1
levels are normal in mutant plants (Figure 7), we propose that membrane biosynthesis per se is unaffected by loss of GLK gene function. It is therefore proposed that thylakoids are perturbed due to loss of PSII complexes. LHCP gene transcript levels are reduced in mutant plants (Figure 7), whereas psbA (which encodes D1) levels are normal (data not shown). Thus it is likely that the primary defect is loss of LHCP or chlorophyll, and that D1 protein is degraded due to defective PSII complex assembly. As GLK proteins have a proposed DNA-binding domain that is similar in structure to ARR-B proteins, it is possible that they bind to a consensus DNA sequence that is similar to the one recognized by ARR-B proteins. The expected frequency of any particular 5 bp sequence is one in 1024 bp. Interestingly, in a region of DNA 1 kb upstream of the LHCP1 and LHCP6 promoter regions there are three and six copies of the ARR-B binding site (AGATT), respectively. In the region upstream of the chlorophyll biosynthetic genes CAO and CHLH, there are one and two sites, respectively.

Data obtained in Arabidopsis and rice have demonstrated that in C3 plants the GLK genes play a partially redundant role in regulating transcription of genes required for chloroplast development. Redundancy is widespread in the genomes of higher organisms (Thomas, 1993), and is thought to arise from gene duplication events as a result of either tandem duplication of specific chromosomal sections, or whole-genome tetraploidization. Three types of genetic redundancy have been defined, and each is subject to different selection pressures (Nowak et al., 1997). The first is true or complete redundancy, where the relative fitness of an organism with only one functional gene is identical to an organism where both genes are functional. Only individuals without either gene suffer a reduction in fitness. In the second example, redundancy is not complete in that two genes perform the same function, but at slightly different efficiencies. In such a case, the less efficient gene is fully redundant in that there is no perturbed phenotype in null mutants, but the other gene is partially redundant, with mutants exhibiting a partial loss of fitness. An example of this is observed with the Arabidopsis TRYPOTOPHAN SYNTHASE B genes (TSB1 and TSB2) (Last et al., 1991). Both genes are transcribed but TSB1 produces >90% of the tryptophan synthase B protein. Despite this observation, under certain conditions TSB2 is capable of supporting growth in the absence of TSB1 (Last et al., 1991). The third type of redundancy is where genes are redundant for one function, but one or both may be selected for because of another independent function. The Arabidopsis meristem identity genes APETALA1 (AP1; Mandel et al., 1992) and CAULIFLOWER (CAL; Kempin et al., 1995) exhibit redundancy of this type. Both encode MADS-box transcription factors. The ap1 mutants produce flowers that retain partial inflorescence meristem identity, whereas cal mutants do not exhibit a mutant phenotype. However, ap1 cal double mutants produce presumptive floral meristems that are completely converted into inflorescence meristems (Bowman et al., 1993). Hence, CAL activity appears to be entirely redundant with AP1 activity, but AP1 has some activities that are not redundant with CAL (Pickett and Meeks-Wagner, 1995).

The AtGLK genes exhibit partial redundancy as there is a phenotype specific to the Atglk2.1 mutant allele (pale green siliques) but no phenotype specific to the Atglk1.1 allele. This observation may simply reflect the fact that the two genes have different spatial and temporal expression domains (AtGLK1 is not expressed in siliques), rather than different functions. Alternatively, the two genes may be functionally distinct, in which case AtGLK1 is redundant but AtGLK2 is not. This may be due to weaker efficiency of AtGLK1, as in the case of TSB genes, or may reflect the fact that, like the AP1/CAL interaction, AtGLK2 has some functions that are independent of AtGLK1. Notably, at certain stages of development AtGLK1 transcript levels are higher in Atglk2.1 mutants than in corresponding wild-type tissues (data not shown). Therefore one of the additional functions of AtGLK2 may be to regulate AtGLK1 accumulation levels. A similar regulatory mechanism is seen in Antirrhinum with products of the DEFICIENS (DEF) and GLOBOSA (GLO) genes. In this case, GLO transcription is independent of a functional DEF protein during early development; however, after initiation of organ primordia, GLO and DEF transcription is regulated by a DEF/GLO interaction (Schwarz-Sommer et al., 1992; Trobner et al., 1992). Perhaps significantly, DEF and GLO bind DNA as heterodimers (Zachgo et al., 1995).

The functional redundancy (albeit partial) of GLK gene action in Arabidopsis is a striking contrast to the apparent specialization of gene function in maize. Compartmentalization of Glk gene products between bundle sheath and mesophyll cells of maize may reflect the fact that ZmG2 and ZmGlk1 have evolved specific functions. Alternatively, the two proteins may be functionally interchangeable but restricted to spatial domains of action. Either way, it is possible that compartmentalization of Glk gene products coincided with the transition to C3 photosynthesis. However, as ZmG2 promotes the development of C3M and C4BS type chloroplasts, which are both structurally and functionally distinct, it is likely that other cell-specific factors were also involved.

Experimental procedures

Sequence analysis

BLAST searches were performed using appropriate algorithms with and without low-complexity filters on a range of nucleotide and protein databases at http://www.ncbi.nlm.nih.gov/BLAST/and
http://www.arabidopsis.org/BLAST/. Secondary structure prediction was performed using three software programs: SSPro (promoter.ics.uci.edu/BRNN-PRED/ Baldi et al., 1999), PSIPRED (insulin.brunel.ac.uk/psipred/ Jones, 1999) and PROF PREDICTION (http://www.aber.ac.uk/~phiww/Prof/ Ouali and King, 2000). PHD protein-folding predictions were made through the Predict Protein e-mail server (maple.bioc.columbia.edu/predictprotein/). NLS prediction was performed using PSORT software (psort.nib-b.ac.jp). Simple pairwise alignments were performed using GENEJOCKEY software (Biosoft, Cambridge, UK), and the Genetics Computer Group (GCC, Madison, WI) PILEUP package was used for multiple sequence alignments. The GCC PRETTYBOX function was used for highlighting conserved residues and the CANVAS (Deneba Systems, Miami, FL, USA) graphics programme was used to manipulate and present the alignment.

**Determination of gene structure**

ESTs were ordered from the Arabidopsis Biological Resource Centre (ABRC, Ohio State University, Columbus, OH). EST 122A10T7 corresponded to AtGLK1. EST 190N23T7 corresponded to AtGLK2. Both ESTs were sequenced and aligned to genomic sequence from BAC, Accession number AC007048(AtGLK1) and P1clone MLN1, Accession number AB005239(AtGLK2). Gene structure was determined by aligning EST sequence and genomic sequences.

**Phylogenetic analysis**

Phylogenetic analyses were carried out using PAUP* 4.0 (Swofford, 1998). For the GARP superfamily phylogenetic analysis, 36 Arabidopsis GARP genes, six GLK genes, the Psr1 gene from Chlamydomonas, six TEA genes, and the potato MYB-related gene MybS1 were used. The GARP genes were identified following BLAST searches of the Arabidopsis genome database (http://www.arabidopsis.org/BLAST/) using exon II of the maize G2 gene as the query sequence. Individual genes were identified from within BACs by searching for the highly conserved sequence motif ASHLQ (last five amino acids encoded by exon II of G2). For each gene, nucleotide sequences equivalent to the second exon of G2, plus 12 bp either side of the exon, were aligned using CLUSTALW (Thompson et al., 1994). This analysis was conducted using parsimony with all characters unordered and equally weighted. These results are presented as a majority rule bootstrap consensus tree (Figure 2a) estimated from 1000 bootstrap replicates and a heuristic search strategy comprising 1000 random addition replicates with MULPARS on.

For the GLK gene analysis, full-length cDNAs of each GLK gene and APRR2 were aligned by hand according to the amino acid alignment shown in Figure 1. As the number of taxa is relatively small, these data were analysed using the exhaustive search option. The analysis found one most parsimonious tree (Figure 2b). Bootstrap values were estimated from 1000 bootstrap replicates analysed using the exhaustive search option. These data were also analysed using the maximum likelihood method. MODELTEST (Posada and Crandall, 1998) was used to select a model of DNA substitution, GTR + G + I. The results of the maximum likelihood analysis yielded an identical topology to the parsimony analysis.

**Plant material and growth conditions**

Arabidopsis thaliana ecotype Columbia was used in all experiments. Plants were grown in a Sanyo Versatile Environmental Test Chamber (Gunma, Japan) with a daily cycle of 16 h light (100 μmol m⁻² sec⁻¹) at 22°C and 8 h dark at 18°C (long-day conditions).

**Extraction of DNA**

DNA was isolated from plant tissue using a rapid DNA extraction method. Gridding sticks were prepared from disposable pipette tips by twisting molten tip ends into 1.5 ml Eppendorf tubes. Tissue was homogenized to a paste in an Eppendorf using a grinding stick, then resuspended immediately in 100 mM Tris–HCl pH 8.0, 50 mM EDTA, 500 mM NaCl, 10 mM β-mercaptoethanol and 1.4% SDS. After incubation at 65°C for 10 min, samples were adjusted to 1 m KOAc and after 5 min the debris was pelleted by centrifugation. DNA was precipitated from the supernate using isopropanol and NaOAc.

**DNA gel-blot analysis**

DNA gel-blot analysis was performed as described by (Langdale et al., 1991). Hybridizations were performed in 0.45 M NaCl at 65°C.

**Extraction of RNA and gel-blot analysis**

RNA was isolated, electrophoresed, blotted and hybridized as described by Langdale et al. (1988). Denstometry of autoradiographs and analysis of fluorescence in ethidium bromide-stained gels were performed using the BioRad Fluor-S Multimager and BioRad (Hercules, CA, USA) MultiAnalyst software. Gene-specific probes used for RNA gel-blot analysis were as follows: AtGLK1, 224 bp fragment corresponding to positions 217–441 of Accession number AY026772; AtGLK2, 261 bp fragment corresponding to positions 1–261 of Accession number AY026773; TUBULIN, 1: 1: 1: 1 mix of TUBULIN 1–4 (a gift from Dr M. Knight, Department of Plant Sciences, University of Oxford); LHCP1, EST AP2L15fo3R; LHCP6, EST 23C1T7; SRP43, 71A11T7; SRP54, 847 bp fragment corresponding to positions 797–1844 of Accession number AF092168; CAO, EST 103D24T7; CHLH, EST AP2L15fo3R Hybridizations were performed in 0.45 M NaCl at 65°C.

**Identification of insertional mutants**

To identify lines with dSpm insertions in or near to the GLK genes, the Sainsbury Laboratory Arabidopsis Transposant (Tissier et al., 1999) lines were screened. DNA was obtained from 48 pools of 50 lines and the inverse polymerase chain reaction (iPCR; Ochman et al., 1988), using dSpm primers, was used to generate products containing DNA sequences flanking each insertion. iPCR products were spotted onto nylon filters by Dr S. Rutherford (Department of Plant Sciences, University of Oxford) and filters were hybridized with AtGLK1 and AtGLK2 cDNA sequences. In each case, a subpool of 50 lines was identified. Seed corresponding to the subpool (kindly donated by Profesor Jonathan Jones, Sainsbury Laboratory, John Innes Centre, Norwich, UK) were germinated on peat plugs, and those containing a dSpm element were selected by spraying phosphi-

notrocin (PPT) at 100 mg l\(^{-1}\) after 7 days. DNA was extracted from PPT-resistant plants that were pooled into 16 groups of six. PCR using GLK gene-specific primers and dSpm-specific primers was then used to identify pools of plants containing an insertion line. PCR conditions were: 35 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 2 min. DNA amplification was achieved for AtGLK1 using primers Smp5 (5’-CGGGATCCGACACTTCTTAATT-AACTGACACTC-3’) and 2 bgp2 (5’-AAGTGGTTTACTGATCCG-ATTGTTCTT-3’). For AtGLK2, positive amplification was achieved using primers Smp1 (5’-CTTAATTCAAGAAGTTGCGTTTTT-GG-3’) and 5.2 (5’-CTCTCTGACGGTTGACTCGG-3’) or Smp8 (5’-GGTTTCTGCGACACTTCTTACC3’) and ara4 (5’TCCGATGTGAC-CTATATTTCC-3’). Having identified a pool of six plants that contained an insertion line, the PCR was repeated using DNA from individual plants. In each case, copy number of the insertion allele was determined by restriction digestion (HindIII for AtGLK1 and EcoRV for AtGLK2) and DNA gel-blot analysis. Both lines were homozygous for the dSpm insertion and the mutant alleles were named Atglk1.1 and Atglk2.1, respectively.

**Preparation of total leaf protein and gel-blot analysis**

Plant tissue (1 g) was harvested, flash frozen in liquid N\(_2\) and homogenized using a pipette grinder in 1.5 ml Eppendorf tubes. Laemmli SDS-PAGE buffer (1 ml): 4% SDS, 20% glycerol, 10% mercaptoethanol, 0.004% bromophenol blue (BPB), 0.125 M Tris–HCl pH 6.8) was added and equal volumes of samples were then electrophoresed on 12% SDS-polyacrylamide gels. Proteins were blotted onto 0.2 \(\mu\)m nitrocellulose membrane (Schleicher and Schuell, Dassel, Germany) using a Bio-Rad Mini-PROTEAN 3 apparatus according to the manufacturer’s instructions. Immunoreactions were carried out using horseradish peroxidase-conjugated secondary antibody and 4-chloro-1-naphthol. Immune complexes were visualized using autoradiography (AgFA Scientia EM film 23 D 56 (Leverkusen, Germany). Protein bands were quantified by densitometry using a Molecular Dynamics (CARY) System. 

**Chlorophyll assays**

0.5 cm\(^2\) leaf sections were sampled under green safe light using a bore with an 8 mm diameter. Discs were flash frozen in liquid N\(_2\) and ground in an Eppendorf using a grinding tip, then suspended in 1 ml 80% acetone. Following centrifugation to remove debris, the OD of the supernate was measured at 645 and 663 nm. Chlorophyll concentration was calculated using the formula:

\[
\text{Chlorophyll concentration (µg ml\(^{-1}\)) = } \frac{\text{OD}_{645} \times 20.2 + \text{OD}_{663} \times 8.0}{5}\]

(Arnon, 1949). This value was then divided by 0.5 to obtain the total chlorophyll concentration per cm\(^2\) tissue.

**ANOVA**

The experimental design of the chlorophyll assay resulted in an unbalanced data set, with each genotype represented by different numbers of plants. A general linear model (GLM) was applied using the model 'Chlorophyll concentration' = 'LeafType' + 'Copies GLK1' + 'Copies GLK2' + 'Copies GLK1' 'Copies GLK2'. Reference to ANOVA is included at the same time. Analysis was carried out using MINITAB software (Coventry, UK).

**Transmission electron microscopy**

TEM analysis was carried out essentially according to Langdale and Kidner (1994), with the following exceptions. Tissue samples were cut under and vacuum infiltrated with Trumps fixative (3% formaldehyde, 3% gluteraldehyde, 0.025 M phosphate buffer pH 7.2) for 2 h at room temperature. After sectioning, samples were mounted on Butvar B98 slots (Agar Aids, Essex, UK) and stained using a 2168 Ultrastainer Carlsberg System (Leica, Heerbrugg, Switzerland) in Ultrastain1 for 2 h and Ultrastain2 for 10 min according to the manufacturer’s instructions. Sections were photographed using AgFA Scientia EM film 23 D 56 (Leverkusen, Germany).

**Acknowledgements**

We thank Maria Rossini (University of Milan, Italy) and members of the laboratory for discussions stimulating throughout the course of this work, and Cledwyn Merriman for technical assistance. This work was supported by grants from the BBSRC to J.A.L.; M.J.C. was the recipient of a BBSRC postgraduate studentship.

**References**


