Plant Physiol. (1996) 111: 949–957

Update on Photosynthetic Gene Expression

Evolution and Expression of C_4 Photosynthesis Genes

Maurice S. B. Ku[1], Yuriko Kano-Murakami, and Makoto Matsuoka*

Botany Department, Washington State University, Pullman, Washington 99164–4238 (M.S.B.K.); Fruit Tree Research Station, 2–1 Fujimoto, Tsukuba, Ibaraki 305, Japan (Y.K.-M.); and BioScience Center, Nagoya University, Chikusa-ku, Nagoya, Aichi 464–01, Japan (M.M.)

Based on the differences in the mechanism of CO_2 assimilation, land plants can be divided into three major photosynthetic types, namely C_3, C_4, and Crassulacean acid metabolism (CAM) plants. Each photosynthetic type possesses a unique set of anatomical, physiological, and biochemical features that allow them to adapt to a specific ecological niche. C_3 plants perform well in temperate climates, whereas C_4 plants thrive in high-light, arid, and warm environments. CAM plants, characterized by CO_2 uptake in the night, adapt to extreme arid conditions. Taxonomical and phylogenetic studies suggest that CAM and C_4 plants were derived from C_3 plants and the transitions occurred many times in diverse taxa during the course of evolution (Moore, 1982). A drastic decline in atmospheric CO_2 level during the late Cretaceous period (65–85 million years ago), a time of major expansion of the angiosperms, has been proposed to account for the increase of C_4 plants (Ehleringer et al., 1991).

The major function of the C_4 pathway is thought to overcome the limitation of low CO_2, which results in significant increases in photorespiration and thus reduces competitiveness. Although it is widely assumed that CAM evolved in response to selection for increased water-use efficiency, the occurrence of CAM in aquatic plants (Keeley and Morton, 1982), in which photosynthesis is often limited by low CO_2 rather than water, strongly argues that the driving force for its evolution may be low CO_2 as well. Both C_4 and CAM evolved a very similar photosynthetic biochemistry for concentrating CO_2 in the leaf, except that the CO_2-concentrating steps are spatially separated in C_4 plants but temporally separated in CAM plants. C_3 photosynthesis and CAM occur in only one cell type, the mesophyll cells, whereas C_4 photosynthesis requires the coordination of two photosynthetic cell types, the mesophyll and bundle-sheath cells. Therefore, the genetic modifications required for achieving the CO_2-concentrating mechanism are considered relatively small for CAM, as compared with those required for C_4 photosynthesis. Comparative phylogenetic studies also suggest that CAM evolved earlier than C_4 photosynthesis (Moore, 1982).

THE C_4 PATHWAY OF PHOTOSYNTHESIS

More than 90% of the land plants, including many agronomically important crop species, assimilate atmospheric CO_2 through the C_3 pathway of photosynthesis. However, photosynthesis by C_3 plants suffers from the low affinity of Rubisco toward atmospheric CO_2 and limitation by photorespiration, a process considered wasteful (reviewed by Furbank and Taylor, 1995). C_4 plants developed a biochemical mechanism to overcome the limitations of low CO_2 and photorespiration: the C_4 photosynthetic pathway serves as a "CO_2 pump" to concentrate CO_2 at the site of Rubisco and thus suppresses its oxygenase activity and the associated photorespiration (Hatch, 1987; Dai et al., 1993; Furbank and Taylor, 1995). This is achieved via the coordination of two photosynthetic cell types, namely mesophyll and bundle-sheath cells (the Kranz leaf anatomy). In the C_4 pathway (Fig. 1), atmospheric CO_2 is first hydrated to bicarbonate by CA in the cytosol of mesophyll cells and subsequently fixed into the C_4 acid oxaloacetate with the three-carbon substrate PEP through PEPC. Oxaloacetate is rapidly reduced to malate in the mesophyll chloroplasts by NADP-malate dehydrogenase or transaminated to aspartate in the cytosol by aspartate aminotransferase, depending on the C_4 acid-decarboxylating mechanism of C_4 plants (Hatch, 1987). These C_4 compounds are then transported to the inner bundle-sheath cells, where they are decarboxylated to release CO_2 by one of the three C_4 acid decarboxylation enzymes: NADP-ME, NAD-ME, or PEP carboxykinase. In the NADP-ME C_4 plants, which contain several economically important crops such as maize, sugarcane, and sorghum, malate is the predominate C_4 acid produced in the mesophyll cells. By diffusion through the intercellular plasmodesmata, malate is rapidly transported to bundle-sheath cells, where it is decarboxylated by NADP-ME in the chloroplasts, and the released CO_2 is reassimilated by Rubisco in the C_3 pathway. To complete the cycle, pyruvate produced from malate decarboxylation is shuttled back to mesophyll chloroplasts and phosphorylated to PEP via PPDK, with the input of two extra ATPs. Therefore, the C_4 pathway can be considered an additional step to the conventional C_3 pathway.

[1] Recipient of a visiting professorship from BioScience Center, Nagoya University, Japan.

* Corresponding author; e-mail j45751a@nucc.cc.nagoya-u.ac.jp; fax 81–52–789–5226.

Abbreviations: CA, carbonic anhydrase; ME, malic enzyme; PEPC, PEP carboxylase; PPDK, pyruvate, orthophosphate dikinase; *rbcS*, Rubisco small subunit.

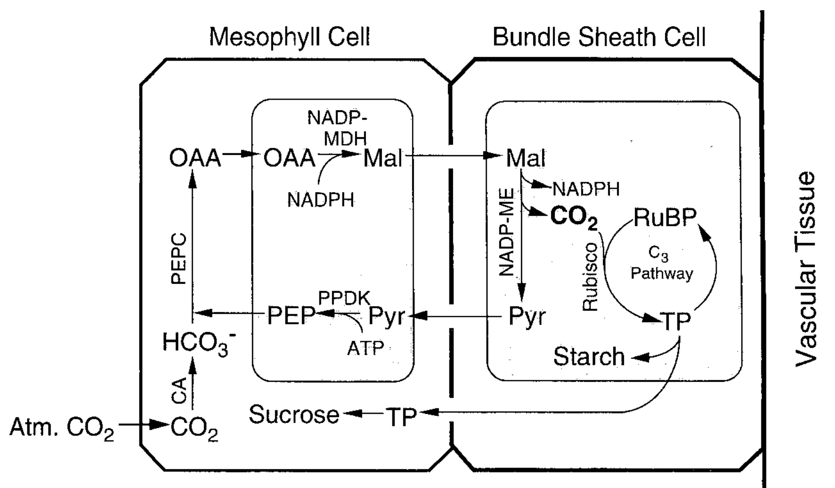

Figure 1. The C_4 pathway of photosynthesis in maize, an NADP-ME-type C_4 plant. The major function of the pathway is to concentrate CO_2 in the inner bundle-sheath cells where Rubisco is located and thus suppress its oxygenase activity and the associated photorespiration (Dai et al., 1993). The open compartments represent cytosol, and the shaded compartments represent chloroplasts. OAA, Oxaloacetate; NADP-MDH, NADP-malate dehydrogenase; Mal, malate; Pyr, pyruvate; RuBP, ribulose-1,5-bisphosphate; TP, triose-P; Atm. CO_2, atmospheric CO_2. Starch is synthesized predominantly in the bundle-sheath chloroplasts, whereas Suc is produced mainly in the cytosol of mesophyll cells. See Hatch (1987) for a variation of the pathway in other C_4 plants.

The most important feature of C_4 photosynthesis is the spatial separation of the key photosynthetic enzymes and their metabolic cooperation between the two specialized cell and chloroplast types. In addition, intercellular compartmentation of nitrogen and sulfur metabolism also occurs in leaves of C_4 plants, which contributes to a more efficient utilization of light energy between the two photosynthetic cell types (Moore and Black, 1979; Gerwick et al., 1980). Because of the CO_2-concentrating mechanism, photosynthesis by C_4 plants is near saturation at atmospheric CO_2 levels and exhibits a low photosynthetic CO_2 compensation point, limited O_2 inhibition of photosynthesis, and negligible apparent photorespiration (Fig. 2). Thus, C_4 plants have a selective advantage over C_3 plants, especially under low CO_2 conditions, where carbon loss from photorespiration becomes maximal (Dai et al., 1993; Fig. 2). The operation of the C_4 pathway of photosynthesis results in several physiological advantages that make C_4 plants ideally suited for warm, high-light, and arid environments. These include high photosynthetic capacity, fast growth rate, and high nitrogen and water-use efficiencies.

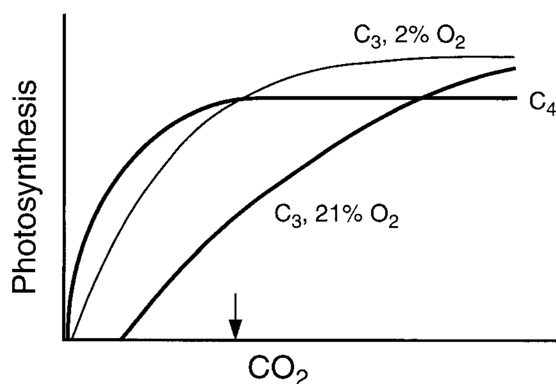

Figure 2. Photosynthetic responses to CO_2 of C_3 versus C_4 plants. Relative to C_3 plants, the major advantage of C_4 plants lies in their efficient utilization of low levels of CO_2 due to the CO_2-concentrating mechanism of the C_4 pathway of photosynthesis. C_4 plants exhibit no apparent O_2 inhibition of photosynthesis, a low CO_2 compensation point (the intercept on *y* axis), and a high carboxylation efficiency (the initial slope of CO_2-response curve), and their photosynthesis is saturated by atmospheric levels of CO_2 (indicated by arrow). On the other hand, photosynthesis by C_3 plants suffers from O_2 inhibition and photorespiration, as shown by a high CO_2 compensation point and a lower carboxylation efficiency, and is limited by atmospheric CO_2. The higher photosynthetic capacity of C_3 plants at saturating CO_2 is due to a higher Rubisco content in the leaves of C_3 plants.

EVOLUTION OF C_4-SPECIFIC GENES

The enzymes involved in C_4 photosynthesis are not unique to C_4 plants; they also operate in C_3 plants but for different metabolic processes. However, the activities of these enzymes are much lower in C_3 plants than in C_4 plants. Additionally, the C_4-specific isoforms of these enzymes usually possess different kinetic properties and, in some cases, have different inter- and intracellular compartmentation from their C_3 counterparts (Ludwig and Burnell, 1995). In other words, their function and their cellular location were modified during the course of evolution to meet the requirements of their new roles. For example, PEPC from C_4 plants has a lower affinity for its substrate PEP and a higher maximum activity and different sensitivity to its effectors, relative to its C_3 counterpart. In C_3 plants, the enzyme primarily plays an anaplerotic role in basic plant metabolism. Another example is CA. The enzyme in C_3 plants is mainly confined to the chloroplast stroma of mesophyll cells and facilitates the diffusion of CO_2 across the chloroplast stroma (Badger and Price, 1994). In contrast, in leaves of C_4 plants the enzyme is predominately located in the cytosol of mesophyll cells, mainly for the rapid conversion of CO_2 to bicarbonate, the substrate for PEPC.

The differences in kinetics, function, and localization between the photosynthetic enzymes from C_3 and C_4 plants (C_3 versus C_4 isoforms) raise many interesting questions as to how the genes for these C_4-specific isoforms (C_4-specific genes or C_4 isogenes) evolved and how their expression is regulated. The random and polyphylogenetic nature of the evolution of C_4 plants suggests that the C_4 isogenes evolved from a set of preexisting genes in ances-

tral C_3 plants. Modifications of genes that allow their protein products to have the correct enzymic properties to perform the new metabolism and with regulatory mechanisms that allow for their high levels of expression must have occurred during the course of evolution. In addition, another regulatory mechanism(s) had to be acquired for the cell- and organ-specific expression of these isogenes along with the development of Kranz leaf anatomy, which is imperative for a functional C_4 photosynthesis.

We will concentrate our discussion of the evolution of C_4-specific genes and their molecular regulatory mechanisms of expression on three key enzymes of C_4 photosynthesis: CA, PEPC, and PPDK. The genes for these enzymes have been extensively characterized. As will be shown, C_4 plants deployed different strategies to acquire the various components of C_4 photosynthesis, and different molecular mechanisms were used to create C_4-specific genes from preexisting genes in ancestral C_3 plants.

CA

CA plays a crucial role in C_4 photosynthesis by catalyzing the rapid conversion of CO_2 to HCO_3^- in the cytosol of mesophyll cells (Fig. 1). Multiple isoforms of CA located in various cellular compartments exist, and each isoform may play a different role in photosynthesis (Badger and Price, 1994). Both biochemical and molecular studies suggest that higher plants have two isoforms of CA, the cytosolic and the chloroplastic forms, and that the cytosolic form is a predominant one in C_4 plants, whereas the chloroplastic form is a major one in C_3 plants. The kinetic properties of C_4 CAs appear to be similar to those of C_3 enzymes, with a similar affinity for CO_2 and sensitivity to inhibitors (Hatch and Burnell, 1990). Distinct cDNAs encoding these isoforms have been cloned from several plant species (monocots and dicots), including C_3 and C_4 plants, and phylogenetic analysis of the deduced amino acid sequences of CAs from various organisms shows four major groupings: the prokaryotic, *Chlamydomonas*, dicot, and monocot groups (Ludwig and Burnell, 1995).

Within each higher plant group, the cytosolic isoform is invariably more closely related to the chloroplastic isoform from the same species than to the cytosolic form from another species (Ludwig and Burnell, 1995). These results suggest that the genes for the two isoforms must have evolved from a common ancestral gene long before the divergence of monocots and dicots (approximately 200 million years ago) and that the rates of divergence for both the cytosolic and chloroplastic forms from the same species remained relatively constant. If only a differential expression of the two isoforms (i.e. high level of expression of the cytosolic isoform in mesophyll cells) is required for achieving this component of C_4 photosynthesis (provision of substrate for PEP carboxylation) during its evolution, then this mechanism represents a simple genetic alteration. This is consistent with the notion that a major adaptation in photosynthetically related CAs may have occurred in the regulation of their quantity rather than their kinetic property (Badger and Price, 1994). In this regard, it would be interesting to compare the promoter regions of CA genes for the cytosolic and chloroplastic isoforms between closely related C_3 and C_4 plants to identify the possible molecular mechanism(s) that mediates the differential expression of the two isoforms.

PEPC

Molecular studies show that PEPCs of higher plants are encoded by a multigene family PEPC. The genes from the C_4 plant maize have been extensively studied, and at least three isoforms of PEPC have been recognized in the species: C_4- and C_3-specific forms in the leaves and a root-specific form. An earlier study by Grula and Hudspeth (1987) suggested that the PEPC gene family in maize consists of at least five genes. These gene members can be classified into three distinct groups or subfamilies, based on differences in genomic structure (Schaffner and Sheen, 1992). One has been termed a C_4-specific gene because of its high level of expression in green leaves (light dependency) and low level of expression in etiolated leaves and roots (organ specificity), whereas three others are designated as non-C_4 or root-specific PEPC genes because of their preferential expression in roots (Grula and Hudspeth, 1987). The C_4-specific gene was shown by restriction fragment length polymorphism analysis to be located on chromosome 9, whereas the presumed root-specific genes were found in duplicate on chromosomes 4 and 5. A root-specific PEPC gene and a C_3-type leaf PEPC gene were subsequently isolated from maize and characterized by Kawamura et al. (1990, 1992). The C_3 leaf gene has been located on chromosome 7. Relative to the C_4-specific gene, the C_3 gene is constitutively expressed in both green and etiolated leaves (light independence) at a much lower level. It is also expressed at low levels in the roots, as compared with that in leaves. Comparisons of the two leaf-specific PEPC genes of maize show that they share a high homology (71%), but the C_4 gene evolved under strong G/C pressure (Kawamura et al., 1992). It is speculated that the higher G+C preference of the C_4 gene (62%), relative to that of the C_3 gene (51%), may reflect a functional G/C pressure due to constraints exerted at the transcriptional level. Genes with high G+C preference are known to have a strong expression in response to endogenous and exogenous stimuli (Quigley et al., 1989). Many photosynthetic genes from grasses that are expressed at high levels tend to have a high G+C content (Matsuoka et al., 1988).

In contrast to maize, the sorghum PEPC gene family is composed of only three members: one is specific for C_4 photosynthesis and two others are non-C_4 types (Lepiniec et al., 1993). One of the two non-C_4-type genes is designated as a root gene, which also shows a high homology (76%) in the 3′ noncoding region to the maize root PEPC gene (Kawamura et al., 1990). However, cross-hybridization and phylogenetic analyses indicated that the sorghum C_4-specific gene is more closely related to the root gene than to the other non-C_4 type gene. Whether the second non-C_4 type gene is a leaf-specific C_3 PEPC gene in sorghum remains to be established. Similar to maize, the sorghum C_4 gene also had evolved under high G/C pressure (61% for the coding region, 83% for

the third position), relative to the non-C_4 genes (51–52% for the coding region, 55–56% for the third position) (Lepiniec et al., 1993).

Plants in the dicot *Flaveria* also have been used to gain insights into the molecular changes responsible for the transition from C_3 to C_4 photosynthesis. This small genus is characterized by having a wide range of C_3, C_4, C_4-like, and C_3–C_4 intermediate species (Ku et al., 1990), and it is argued that a comparative analysis of C_3- versus C_4-specific isogenes for a particular enzyme in closely related C_3 and C_4 species would allow detection of sequence differences that are relevant for the functioning of the gene in C_4 photosynthesis (Hermans and Westhoff, 1992). It was demonstrated that PEPCs in several *Flaveria* species examined are encoded by multigene families and both the C_4 species *Flaveria trinervia* and the C_3 species *Flaveria pringlei* have a similar organization of the gene families (Hermans and Westhoff, 1990, 1992; Poetsch et al., 1991). In contrast to maize and sorghum, in which the C_4 PEPC is encoded by a single-copy gene, C_4 PEPC in the C_4 dicot *F. trinervia* appears to be encoded by a separate subfamily of closely related genes.

Several interesting points or suggestions concerning the evolutionary relationship of PEPCs can be made from parsimonious phylogenetic and rooted phenetic analyses of the enzyme from various organisms, including bacteria and plants (Hermans and Westhoff, 1992; Kawamura et al., 1992; Lepiniec et al., 1993; Toh et al., 1994). The comparisons show two distinct groupings of PEPCs, the prokaryotic and eukaryotic branches, which suggests that all of the PEPC genes evolved from a common ancestral gene. Since the Ser residue at position 15 involved in PEPC light/dark modulation through phosphorylation/dephosphorylation is found only in plant enzymes (Schaffner and Sheen, 1992; Lepiniec et al., 1993), this alteration in enzymic property must have occurred during the divergence of the two distinct groups. This property is highly conserved during the evolution of plants and is present not only in C_3- and C_4- but also in root-specific PEPCs.

Among the 14 plant PEPCs characterized, several distinct subgroups can be recognized: (a) the monocot C_4 subgroup, (b) the monocot non-C_4 (leaf and root) subgroup, (c) the dicot non-C_4 subgroup, and (d) the dicot *Flaveria* subgroup (Toh et al., 1994). The evolutionary origin of monocot C_4-PEPCs is clearly different from those of other PEPC subgroups. The maize C_4 PEPC gene is more closely related to the C_4 gene from another C_4 monocot plant, sorghum, than to its own non-C_4-type PEPC genes or the C_4 PEPC gene from the dicot plant *F. trinervia*. This suggests that the monocot and dicot C_4 plants evolved independently after their divergence. It also implies that a common prototype gene might have given rise to the C_4 genes in both maize and sorghum, and it is suggested that the divergence of this prototype gene from other plant PEPC genes might have preceded the monocot-dicot divergence (Toh et al., 1994). It is interesting that PEPCs from all dicot species examined, irrespective of their roles in carbon metabolism (C_3, C_4, or CAM), are more closely related to each other and to the non-C_4-type PEPCs from C_4 monocot plants. It is possible that after the monocot-dicot divergence C_4 PEPCs in dicot plants evolved at a much slower rate, as compared with C_4 PEPCs in monocot plants. Alternatively, the C_4 dicot plant *F. trinervia* could have evolved more recently than the C_4 monocot plants, maize and sorghum.

Comparisons of both PEPC cDNAs and genomes from the C_4 species *F. trinervia* with those from the C_3 species *F. pringlei* showed that the C_3 and C_4 PEPCs from the genus are more highly related to each other than to other plant PEPCs. However, the degree of heterogeneity in the PEPC gene families within one species is greater than that between the two photosynthetic species (Hermans and Westhoff, 1990, 1992). Both the C_3 and C_4 species contain pairs of closely related genes, with similar expression patterns with regard to organ specificity. In other words, the C_3 *Flaveria* sp. possesses PEPC genes that are closely related to the C_4-specific genes in the C_4 *Flaveria* sp. Thus, these homologous PEPC genes from the C_3 and C_4 species might have evolved from a common ancestral gene, and the *F. pringlei* gene may well retain the ancient features of the C_4 PEPC genes. Examination of this relationship between more closely related C_3 and C_4 *Flaveria* sp. will be interesting.

These results also raise many intriguing questions as to how the C_4-specific PEPC genes might have evolved. The *Flaveria* data support the hypothesis that C_4-prototype PEPC genes had already been created in the C_3 species. It is tempting to postulate whether the C_4-prototype PEPCs in C_3 plants already possessed the correct kinetic property to adapt to C_4 photosynthesis, or was it perhaps acquired later? In any case, further mutations in the promoter regions to confer a high level of expression and cell and organ specificity must have occurred for the eventual rise of the functional C_4 isogenes. Whether this mode of evolution is universal for all C_4 PEPCs remains to be seen. The homology comparisons also reveal that only a few C_4-specific amino acid positions are found in the proteins (Hermans and Westhoff, 1992). If these changes are genuinely related to the changes in kinetic properties of C_4-specific isoforms, then the changes are rather limited.

PPDK

The genes for PPDK have been cloned and characterized from a diverse range of organisms, and the structure, regulation, and evolution of these genes have been recently reviewed (Matsuoka, 1995). The quaternary structures of PPDK differ considerably among organisms: the bacterial enzyme is a homodimer, whereas the plant enzyme is a homotetramer. Also, the enzyme functions in the synthesis of ATP from PEP in the lower organisms, whereas it catalyzes the phosphorylation of pyruvate by ATP to form PEP in higher plants. In spite of these differences, the primary structures for the enzyme are quite similar among various organisms, except that the plant enzymes contain transit peptide sequences and the protozoal and bacterial enzymes do not. The high degree of homology in primary sequence among the PPDKs suggests that an ancestral PPDK gene evolved before the divergence of prokaryotes and eu-

karyotes (Sheen, 1991; Matsuoka, 1995). Furthermore, phylogenetic tree analysis of amino acid sequences of the enzymes from various organisms shows that the C_4 PPDK gene of maize is more closely related to the C_3 PPDK gene from another gramineous plant, rice (C_3), than to the C_4 PPDK gene of a C_4 dicot plant from *Flaveria* (Matsuoka, 1995). The amino acid sequence of the transit peptide of maize PPDK is also more homologous to that of rice PPDK, as compared with that of the C_4 PPDK from *Flaveria*. These results suggest that both the C_3- and C_4-specific genes evolved from a common ancestral gene, which existed before the divergence of C_3 and C_4 plants in gramineous species. As in the case with PEPC, this is consistent with the notion that C_4 monocots and C_4 dicots evolved independently from their respective C_3 ancestors, with the maize C_4 genes evolving from ancestral genes of a C_3 gramineous species, and the *Flaveria* C_4 genes evolving from ancestral genes of a C_3 dicot species.

Molecular studies showed that the C_4-specific PPDK gene is transcribed from two different initiation sites under the control of two promoters, producing two mRNAs of different sizes (Glackin and Grula, 1990; Sheen, 1991; Matsuoka, 1995; Rosche and Westhoff, 1995). The larger mRNA is transcribed under the control of the first promoter and contains the chloroplast transit sequences, and its product is targeted to the chloroplast. The smaller one is transcribed under the control of the second promoter and does not contain the transit sequences, and its product remains in the cytosol. It is most interesting that this unusual dual promoter system occurs not only in C_4 but also in C_3 plants, including both monocots and dicots (Imaizumi et al., 1992; Matsuoka, 1995; Rosche and Westhoff, 1995). The occurrence of this unique transcription system in both C_3 and C_4 plants suggests that the genetic alterations required for the evolution of C_4 PPDK genes from their ancestral genes in C_3 plants again may be relatively small. In contrast to the C_4 chloroplastic isoform, which is expressed at high levels predominantly in the mesophyll chloroplasts of C_4 leaves, the cytosolic isoform is expressed at extremely low levels in the leaves of C_4 plants (Sheen, 1991). In C_3 plants, the chloroplastic isoform is also predominately expressed in green leaves and is induced by light in the same manner as the maize PPDK but at a much lower level than that in C_4 plants (Hata and Matsuoka, 1987; Matsuoka and Yamamoto, 1989). This suggests that the light regulatory machinery of chloroplastic PPDK genes had already been acquired during the evolution of angiosperms. In the C_4 dicot *F. trinervia*, the larger transcript is usually found in leaves and stems and the smaller one is usually found in roots (Rosche and Westhoff, 1995). The steady-state level of the larger transcript accumulates in the light and decreases in the dark, whereas that of the smaller transcript in the stem is only induced by darkness and can be reversed by light. This interesting observation suggests that light may exert a differential effect on the expression of the two transcripts.

The C_4 dicot *F. trinervia* contains only one PPDK gene (Rosche and Westhoff, 1995), whereas maize has three PPDK genes, with the second and the third ones encoding cytosolic isoforms (Sheen, 1991). These differences are intriguing because in rice, a C_3 plant, at least two PPDK genes have been identified (N. Imaizumi, K. Ishihara, M. Samejima, S. Kaneko, M. Matsuoka, unpublished data). In addition, the *Flaveria* C_4 PPDK gene contains an extra intron and an intervening sequence in the 5′ untranslated leader region, as compared with that of maize, and it is proposed that the *Flaveria* gene may be more similar to the primordial PPDK gene than the maize gene (Rosche and Westhoff, 1995). The rice PPDK gene also contains an extra intron in its 5′ noncoding region (N. Imaizumi, K. Ishihara, M. Samejima, S. Kaneko, M. Matsuoka, unpublished data). In any event, the plant PPDKs appear to be of eubacterial origin (Pokalyco et al., 1990) and are highly conserved during evolution (Rosche and Westhoff, 1995; Matsuoka, 1995). This is in strong contrast to plant PEPCs, which show low homology to the bacterial enzyme and are encoded by a multigene family, presumably through duplications during evolution.

From these comparative studies, a possible sequence may be proposed for the evolution of plant PPDKs. It is conceivable that an ancestral gene encoding the dimeric, cytosolic enzyme in the primitive, lower organisms gave rise to the plant genes for the cytosolic housekeeping enzyme. The dual-promoter system of plant PPDK genes must have developed during the early stage of plant evolution. Because in maize the gene encoding the C_4 chloroplast PPDK is highly related to one cytosolic PPDK gene and overlaps another cytosolic PPDK gene, it is proposed that this could have occurred through duplication of the gene for the cytosolic isoform, followed by genomic rearrangement to attach the transit peptide sequences to the gene (Sheen, 1991). This would allow an additional isoform to occur in the plant chloroplasts. In the meantime, the quaternary structure of the enzyme was modified from dimer to tetramer, which may have endowed the enzyme with a higher stability and specificity and with more regulatory features. Subsequent evolution of C_4 photosynthesis might have been accomplished by the acquisition of changes in the primary structure and in the upstream regulatory regions, which enable the enhanced expression of a C_4-specific PPDK to occur predominantly in the mesophyll chloroplasts. At present, it is not known whether C_4 chloroplastic PPDK has the same enzymic property as its C_3 counterpart.

REGULATION OF EXPRESSION OF C_4-SPECIFIC GENES

The key features of the expression pattern of C_4-specific genes are high level of expression, induction by light, and organ and cell specificity. Although the exact regulatory elements that are responsible for these differential expression patterns of C_4-specific genes cannot be pinpointed presently, some recent comparative studies have begun to show that the molecular modifications essential for C_4 photosynthesis may not be large.

Rate of Expression

The differential expression of C_4-specific genes, such as the PEPC and PPDK genes, is regulated mainly at the

transcriptional level in a light-dependent manner (Hudspeth et al., 1986; Sheen and Bogorad, 1987). However, very few data are available to decipher the exact regulatory mechanisms that confer the high rate of expression of C_4-specific genes, relative to their C_3 counterparts. Common promoter sequences, such as the TATA box, a CAAT-like sequence, and GC-rich elements, are found in maize, sorghum, and *Flaveria* C_4-type PEPC and PPDK genes (Hudspeth and Grula, 1989; Matsuoka and Minami, 1989; Matsuoka, 1990; Rosche and Westhoff, 1995). In general, interactions of these common sequences and other upstream regulatory elements have been shown to mediate the differential expression of C_4 genes (Schaffner and Sheen, 1991, 1992; Sheen, 1991).

Using a novel, homologous transient expression assay, Sheen (1991) showed that two separate, *cis*-acting elements are important for the leaf-specific expression of the maize C_4 PPDK gene. A distal element between −347 and −109 is essential for its expression in a light-mediated manner, whereas the proximal element −108 to −52 is responsible for its light-independent expression. Consistent with this observation, Matsuoka and Numazawa (1991), using microprojectile bombardment with maize leaves to study the promoter activity of the maize C_4 PPDK gene in vivo, found that the sequence between −308 and −289 is essential for the high-level expression in mesophyll cells. A DNA-binding protein interacting specifically with the sequence of −301 to −296 has been demonstrated and proposed as a positive factor. In the case of maize C_4 PEPC gene, it has been shown that the acquisition of a new promoter is at least partially responsible for its specific expression patterns (Schaffner and Sheen, 1992). It was deduced, by comparing the sequences in the 5′ untranslated regions (up to −616) of three different maize PEPC genes and their promoter activities in leaf, stem, and root protoplasts, that the addition of a TATTT box and upstream elements are important for the strength of the promoter and the tissue specificity of the C_4 isogene. The most proximal of the four 14-bp C-rich direct repeats may be a very important sequence element in the regulation of PEPC expression (Yanagisawa and Izui, 1990; Kano-Murakami et al., 1991; Schaffner and Sheen, 1992).

To identify the molecular differences between C_4-specific genes in C_4 plants and their homologous genes in C_3 plants, comparisons of the genomic structures (up to −550 from the translation site) of a C_4-specific PEPC gene from the C_4 species *F. trinervia* and its closely related C_3-specific PEPC gene from the C_3 species *F. pringlei* have been performed (Hermans and Westhoff, 1992). It was suggested that a few minor alterations in the promoter regions may be responsible for the high rate of expression of the C_4 isogene, since these altered elements are missing in the promoter of the C_3 isogene. These include (a) an upstream box (GTGTTAATGATG), which in combination with another box is thought to confer a high level of expression; (b) TATA boxes and their immediate sequence environment, which is important for the assembly of the transcription complex; and (c) a scaffold attachment region, which is often associated with strongly expressed genes. If these factors indeed contributed to the high rate of expression of C_4 PEPC genes, then the genetic modifications are relatively small because the 5′ flanking regions of the two genes are essentially homologous. However, using a transgenic approach, Stockhaus et al. (1994) clearly showed that the C_4-specific PEPC isogene of *F. trinervia* contains a novel leaf-specific enhancer-like element located between positions −2118 and −500 of its promoter region. The shorter promoter of this isogene, encompassed position −505 to +66, gave weak activities in the transgenic tobacco leaves. It is interesting that similar trends in promoter strength were observed between a short (−596 to +94) and a long (−2489 to +94) promoter of the C_3 isogene from *F. pringlei*, but the overall strengths of the C_3 promoters were much lower than those of the corresponding C_4 promoters. More detailed analysis of the promoters in the upper regions of the two isogenes is needed to dissect the key elements specific to the C_4 isogene.

One of the interesting questions in C_4 evolution is whether all C_4-specific genes, such as C_4 PEPC and PPDK genes, share a common regulatory element to mediate high-level expression. It appears that this is not the case, since the promoters of the C_4-specific PEPC and PPDK genes from *F. trinervia* or from maize are totally dissimilar (Sheen, 1991; Hermans and Westhoff, 1992; Schaffner and Sheen, 1992; Rosche and Westhoff, 1995). Thus, the regulatory elements that are responsible for the high rate of expression of C_4 isogenes may be different for the various C_4 genes. Certainly, they were not generated by simple duplication.

Light Induction

There appears to be some mechanistic differences in light regulation of expression among C_4 genes. The question of concern is whether light induction of C_4 gene expression is coupled to leaf development (indirect effect) or is a result of an immediate activation of transcription (direct effect). Using mesophyll protoplasts isolated from well-differentiated etiolated, greening, and green leaves of maize for transient expression of promoter activity, Schaffner and Sheen (1992) showed that the light induction of the maize C_4 PEPC gene is coupled to leaf development. The light-dependent induction of maize C_4 PEPC gene expression during leaf greening may be controlled by a long-term, light-dependent developmental program rather than by an immediate light activation of transcription. In contrast, there is a clear, short-term, immediate light induction for maize C_4 PPDK gene (Sheen, 1991), chlorophyll *a*/*b*-binding protein gene, and *rbcS* (Schaffner and Sheen, 1991, 1992). However, an indirect effect of light, albeit much smaller in magnitude, was observed with the expressions of the maize PPDK and *rbcS* genes, which is attributed to light-triggered chloroplast development. This is not too surprising, since both proteins are synthesized in the cytosol and subsequently targeted to the chloroplasts, and development of chloroplasts may influence their transcriptional activities. On the other hand, PEPC, which is not subject to light induction, is a cytosolic enzyme. Consistent with this, no light activation of the promoters for the two

cytosolic PPDK genes from maize was observed. It is thus proposed that at least two signal transduction pathways are responsible for light-induced expression of C_4 genes in maize (Schaffner and Sheen, 1992).

Using a series of deleted promoters coupled with transient expression assay, Sheen (1991) located the *cis*-acting element mediating the light induction of maize C_4 PPDK gene between −347 and −109. For the maize *rbcS* gene, multiple *cis*-regulatory elements located between −229 and −58 were found to be important for its light-dependent expression (Schaffner and Sheen, 1991). However, no consensus sequences can be found between the light-responsive regulatory *cis*-element of maize C_4 PPDK gene and known *cis*-elements of photosynthetic genes (Sheen, 1991). These results indicate the complexity of the light regulatory mechanisms of photosynthetic genes.

Cell-Specific Expression

Currently, very little is known about the molecular mechanisms underlying the cell-specific expression patterns of C_4 genes. To account for the high levels of expression of C_4 isogenes in a cell-specific manner in leaves of C_4 plants, two regulatory models have been proposed by Matsuoka and colleagues (Matsuoka et al., 1993; Matsuoka, 1995). In the first model, which is shown in Figure 3, it is assumed that the C_3- and C_4-specific genes differ in *cis*-acting elements in their promoter regions and that the machineries that regulate the cell-specific expression (e.g. cell-specific *trans*-acting elements) do not differ between the two photosynthetic types. In contrast, in the second model it is assumed that there are no differences in the *cis*-acting elements between the C_3- and C_4-specific genes. Rather, differences in *trans*-acting elements would account for the differential expression of C_4 genes in C_3 versus C_4 plants. Using transgenic rice (C_3) plants transformed with chimeric genes fusing the 5′ flanking region of the maize PPDK and PEPC genes to the coding region of GUS gene, Matsuoka et al. (1993, 1994) demonstrated that the introduced genes are expressed at high levels almost exclusively in the mesophyll cells of rice leaves, with no or very little activity in other cells. Also, the expression of the transgenes was induced by light, in the same manner as in maize. These results are consistent with the first model, which showed that C_3 plants such as rice possess the regulatory machinery sufficient for the light-dependent, cell-specific expression of the maize C_4 genes. It also suggests that rice, a C_3 plant, has the necessary *trans*-acting factors required for the expression of C_4-specific genes. This is further supported by the demonstration that a nuclear protein(s) with similar DNA-binding specificity to that in maize is present in rice (Matsuoka et al., 1993, 1994).

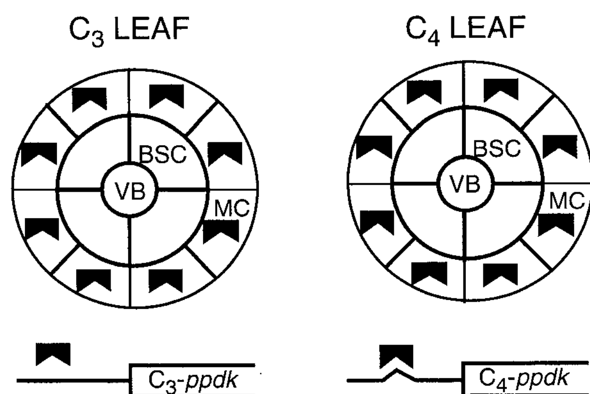

Figure 3. A hypothetical model that accounts for the cell-specific expression of C_4 genes, such as *ppdk* and *pepc*, in mesophyll cells of a C_4 leaf and in mesophyll cells of a C_3 leaf. The model predicts that C_3- and C_4-specific genes differ in *cis*-acting elements in their promoter regions and that both photosynthetic types possess the necessary machineries (e.g. cell-specific *trans*-acting elements, indicated by the solid symbols in mesophyll cells) for the high rate of expression of the C_4 genes. Specific *cis*-regulatory elements allow C_4 genes to be expressed at high levels. VB, Vascular bundle; MC, mesophyll cell; BSC, bundle-sheath cell.

The notion that C_3 plants possess the necessary *trans*-regulatory elements that can interact with the *cis*-regulatory elements of C_4 isogenes for cell-specific expression is further supported by another transformation experiment (Stockhaus et al., 1994). The host tobacco plants were transformed with various promoter-GUS fusions. The promoter of the C_4 PEPC gene from *F. trinervia* (C_4) induced high levels of expression of the reporter gene specifically in the palisade mesophyll cells of transgenic tobacco leaves but low levels of expression in stem and root tissues. On the other hand, the promoter of a closely related C_3 PEPC gene from *F. pringlei* (C_3) gave rise to low levels of expression in leaves and elevated levels in stems and roots. Thus, the promoter of the C_4 gene acts as a strong organ- and cell-specific promoter in the C_3 plant tobacco. The implication of these promoter studies with transgenic plants is that during the transition from C_3 to C_4 photosynthesis the major molecular events may be the acquisition of *cis*-regulatory elements that increase the strength of the promoter and in addition confer organ- and cell-specific patterns of expression. Whether these events were accomplished simultaneously or separately remains to be elucidated.

The low levels of expression of PEPC and PPDK in C_3 plants and the cytosolic PPDK in C_4 plants have been attributed to an absence of upstream elements and weak activity of the TATA promoter element (Sheen, 1991; Schaffner and Sheen, 1992). Thus, gain of important *cis*-acting element(s) to confer high levels of expression and cell specificity may provide a simple and efficient means to develop the biochemical components of C_4 photosynthesis (Schaffner and Sheen, 1992). However, other C_4 genes, such as *rbcS* and NADP-ME genes that are differentially expressed in the bundle-sheath cells, maybe regulated by a different mechanism. A posttranscriptional process plays a significant role in bundle-sheath cell accumulation of *rbcS* transcript (Schaffner and Sheen, 1991; Bansal et al., 1992; Nelson and Langdale, 1992; Viret et al., 1994). Run-off transcription assays with nuclei isolated from maize leaves show that *rbcS* is transcribed in both mesophyll and bundle-sheath cells (Schaffner and Sheen, 1991). Consistent with this observation, the maize *rbcS* promoter is transcriptionally active in mesophyll protoplasts when introduced

by electropolation (Schaffner and Sheen, 1991) and in both cell types when introduced by microbombardment of leaves (Bansal et al., 1992; Viret et al., 1994). The activity of the maize *rbcS* promoter is stimulated by light in bundle-sheath cells but not mesophyll cells, and its 3′ flanking sequences appear to suppress transcription in mesophyll cells (Viret et al., 1994).

FUTURE DIRECTIONS

C_4 photosynthesis requires the coordination of adaptation at the biochemical, anatomical, and ultrastructural levels (Hatch, 1987). During the past three decades, the biochemistry of the C_4 pathway has been clearly elucidated, and recent molecular studies of the structures of C_4-specific genes and their regulatory mechanisms are beginning to shed some light on the molecular events that took place during the evolution of C_4 plants. However, more questions have been raised than answered. More comparative studies of the structure-function relations of C_4 genes and their counterparts in closely related C_3 and C_4 plants will be needed. In this regard, the variation in expression of Kranz leaf anatomy and C_4 biochemical characteristics in *Flaveria* is rather unique (Ku et al., 1990) and may provide the opportunity for elucidating the molecular events that occurred during the transition from a C_3 to a C_4 plant. However, very little is known about the biochemistry and differentiation of the Kranz anatomy, which is essential for a functional C_4 pathway of photosynthesis (i.e. the CO_2-concentrating mechanism). A combination of genetic and molecular approaches may prove fruitful (Langdale and Kidner, 1994). Further investigation is also needed to address the developmental signals and molecular mechanisms that initiate and maintain the C_4 patterns of gene expression (Nelson and Langdale, 1992). Finally, the need to have three different mechanisms of C_4 biochemistry for the same purpose of concentrating CO_2 still remains an unresolved mystery in C_4 photosynthesis. It is interesting that a recent ecological survey of the relative abundance of three subtypes of C_4 grasses and rainfall in Australia suggests that NAD-ME C_4 plants perform well under severe drought conditions, whereas NADP-ME C_4 plants do well in moist environments, with PEP carboxykinase C_4 plants, a minor group of the three, relatively unaffected by moisture (Henderson et al., 1994). The biochemical and molecular bases for this ecophysiological differentiation of C_4 photosynthesis await further investigation.

Received February 29, 1996; accepted April 2, 1996.
Copyright Clearance Center: 0032–0889/96/111/0949/09.

LITERATURE CITED

Badger RM, Price GD (1994) The role of carbonic anhydrase in photosynthesis. Annu Rev Plant Physiol Plant Mol Biol **45:** 369–392

Bansal KC, Viret J-F, Haley J, Khan BM, Schantz R, Bogorad L (1992) Transient expression from cab-m1 and rbcS-m3 promoter sequences is different in mesophyll and bundle sheath cells in maize leaves. Proc Natl Acad Sci **89:** 3654–3658

Dai Z, Ku MSB, Edwards GE (1993) C_4 photosynthesis. The CO_2-concentrating mechanism and photorespiration. Plant Physiol **103:** 83–90

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW (1991) Climate change and the evolution of C_4 photosynthesis. Trends Ecol Evol **6:** 95–99

Furbank RT, Taylor WC (1995) Regulation of photosynthesis in C_3 and C_4 plants: a molecular approach. Plant Cell **7:** 797–807

Gerwick BC, Ku MSB, Black CC (1980) Initiation of sulfate activation: a variation in C_4 photosynthesis plants. Science **209:** 513–515

Glackin CA, Grula JW (1990) Organ-specific transcripts of different size and abundance derive from the same pyruvate, orthophosphate dikinase gene in maize. Proc Natl Acad Sci USA **87:** 3004–3008

Grula JW, Hudspeth RL (1987) The phosphoenolpyruvate carboxylase gene family of maize. *In* JL Key, L McIntosh, eds, Plant Gene Systems and Their Biology. Liss, New York, pp 207–216

Hata S, Matsuoka M (1987) Immunological studies on pyruvate, orthophosphate dikinase in C_3 plants. Plant Cell Physiol **28:** 635–641

Hatch MD (1987) C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim Biophys Acta **895:** 81–106

Hatch MD, Burnell JN (1990) Carbonic anhydrase activity in leaves and its role in the first step in C_4 photosynthesis. Plant Physiol **93:** 380–383

Henderson S, Hattersley P, von Caemmerer S, Osmond CB (1994) Are C_4 pathway plants threatened by global climatic change? *In* E-D Schulze, MM Caldwell, eds, Ecophysiology of Photosynthesis, Ecological Studies, Vol 100. Springer-Verlag, New York, pp 529–552

Hermans J, Westhoff P (1990) Analysis of expression and evolutionary relationships of phosphoenolpyruvate carboxylase genes in *Flaveria trinervia* (C_4) and *F. pringlei* (C_3). Mol Gen Genet **224:** 459–468

Hermans J, Westhoff P (1992) Homologous genes for the C_4 isoform of phosphoenolpyruvate carboxylase in a C_3 and a C_4 *Flaveria* species. Mol Gen Genet **234:** 275–284

Hudspeth RL, Glackin CA, Bonner J, Grula JW (1986) Genomic and cDNA clones for maize phosphoenolpyruvate carboxylase and pyruvate, orthophosphate dikinase: expression of different gene-family members in leaves and roots. Proc Natl Acad Sci USA **83:** 2884–2888

Hudspeth RL, Grula JW (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C_4 photosynthesis. Plant Mol Biol **12:** 579–589

Imaizumi N, Ishihara K, Samejima M, Kaneko S, Matsuoka M (1992) Structure and expression of pyruvate, orthophosphate dikinase gene in C_3 plant. *In* N Murata, ed, Research in Photosynthesis. Kluwer Academic, Dordrecht, The Netherlands, pp 875–878

Kano-Murakami Y, Suzuki I, Sugiyama T, Matsuoka M (1991) Sequence-specific interactions of a maize factor with a GC-rich repeat in the phosphoenolpyruvate carboxylase gene. Mol Gen Genet **225:** 203–205

Kawamura T, Shigesada K, Toh H, Okumura S, Yanagisawa S, Izui K (1992) Molecular evolution of phosphoenolpyruvate carboxylase for C_4 photosynthesis in maize: comparison of its cDNA sequence with a newly isolated cDNA encoding an isozyme involved in the anapleirotic function. J Biochem **112:** 147–154

Kawamura T, Shigesada K, Yanagisawa S, Izui K (1990) Phosphoenolpyruvate carboxylase prevalent in maize roots: isolation of a cDNA clone and its use for analyses of the gene and gene expression. J Biochem **107:** 165–168

Keeley JE, Morton BA (1982) Distribution of diurnal acid metabolism in submerged aquatic plants outside the genus *Isoetes*. Photosynthetica **16:** 546–553

Ku MSB, Wu J, Dai Z, Scott RS, Edwards GE (1990) Photosynthetic and photorespiratory characteristics of *Flaveria* species. Plant Physiol **96:** 518–528

Langdale JA, Kidner CA (1994) Bundle sheath defective, a mutation that disrupts cellular differentiation in maize leaves. Development **120:** 673–681

Lepiniec L, Keryer E, Philippe H, Gadal P, Cretin C (1993) Sorghum phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution. Plant Mol Biol **21:** 487–502
Ludwig M, Burnell JN (1995) Molecular comparison of carbonic anhydrase from *Flaveria* demonstrating different photosynthetic pathways. Plant Mol Biol **29:** 353–365
Matsuoka M (1990) Structure, genetic mapping, and expression of the gene for pyruvate, orthophosphate dikinase from maize. J Biol Chem **265:** 16772–16777
Matsuoka M (1995) The gene for pyruvate, orthophosphate dikinase in C_4 plants: structure, regulation and evolution. Plant Cell Physiol **36:** 1–7
Matsuoka M, Kano-Murakami Y, Tanaka Y, Ozeki Y, Yamamoto N (1988) Classification and nucleotide sequence of cDNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase from rice. Plant Cell Physiol **29:** 1015–1022
Matsuoka M, Kyozuka J, Shimamoto K, Kano-Murakami Y (1994) The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific, light-regulated expression in a C_3 plant (rice). Plant J **6:** 311–319
Matsuoka M, Minami E (1989) Complete structure of the gene for phosphoenolpyruvate carboxylase from maize. Eur J Biochem **181:** 593–598
Matsuoka M, Numazawa T (1991) *Cis*-acting elements in the pyruvate, orthophosphate dikinase gene in maize. Mol Gen Genet **228:** 143–152
Matsuoka M, Tada Y, Fujimura T, Kano-Murakami Y (1993) Tissue-specific light-regulated expression directed by the promoter of a C_4 gene, maize pyruvate, orthophosphate dikinase, in a C_3 plant, rice. Proc Natl Acad Sci USA **90:** 9586–9590
Matsuoka M, Yamamoto N (1989) Induction of mRNAs for phosphoenolpyruvate carboxylase and pyruvate, orthophosphate dikinase in leaves of a C_3 plant exposed to light. Plant Cell Physiol **30:** 479–486
Moore P (1982) Evolution of photosynthetic pathways in flowering plants. Nature **295:** 647–648
Moore R, Black CC (1979) Nitrogen assimilation pathways in leaf mesophyll and bundle sheath cells of C_4 photosynthesis plants formulated from comparative studies with *Digitaria sanguinalis* (L.) Scop. Plant Physiol **64:** 309–313
Nelson T, Langdale JA (1992) Development and genetics of C_4 photosynthesis. Annu Rev Plant Physiol Mol Biol **43:** 25–47
Poetsch W, Hermans J, Westhoff P (1991) Multiple cDNAs of phosphoenolpyruvate carboxylase in the C_4 dicot *Flaveria trinervia*. FEBS Lett **292:** 133–136
Pokalyco DJ, Carroll LJ, Martin BM, Babbitt PC, Dunaway-Mariano D (1990) Analysis of sequence homologies in plant and bacterial pyruvate phosphate dikinase, enzyme I of the bacterial phosphoenolpyruvate: sugar phosphotransferase system and other PEP-utilizing enzymes. Identification of potential catalytic and regulatory motifs. Biochemistry **29:** 10757–10765
Quigley F, Brinkmann H, Martin WF, Cerff R (1989) Strong functional GC pressure in a light-regulated maize gene encoding subunit GAPA of chloroplast glyceraldehyde-3-phosphate dehydrogenase: implications for the evolution of GAPA pseudogenes. J Mol Evol **29:** 412–421
Rosche E, Westhoff P (1995) Genomic structure and expression of the pyruvate, orthophosphate dikinase gene of the dicotyledonous C_4 plant *Flaveria trinervia* (Asteraceae). Plant Mol Biol **29:** 663–678
Schaffner AR, Sheen J (1991) Maize *rbcS* promoter activity depends on sequence elements not found in dicot *rbcS* promoters. Plant Cell **3:** 997–1012
Schaffner AR, Sheen J (1992) Maize C_4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes. Plant J **2:** 221–232
Sheen J (1991) Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell **3:** 225–245
Sheen J-Y, Bogorad L (1987) Differential expression of C_4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. J Biol Chem **262:** 11726–11730
Stockhaus J, Poetsch W, Steinmuller K, Westhoff P (1994) Evolution of the C_4 phosphoenolpyruvate carboxylase promoter of the C_4 dicot *Flaveria trinervia* and expression analysis in the C_3 plant tobacco. Mol Gen Genet **245:** 286–293
Toh H, Kawamura T, Izui K (1994) Molecular evolution of phosphoenolpyruvate carboxylase. Plant Cell Environ **17:** 31–43
Viret J-F, Mabrouk Y, Bogorad L (1994) Transcriptional photoregulation of cell-type-preferred expression of maize *rbcS*-m3: 3′ and 5′ sequences are involved. Proc Natl Acad Sci USA **91:** 8577–8581
Yanagisawa S, Izui K (1990) Multiple interactions between tissue-specific nuclear proteins and the promoter of the phosphoenolpyruvate carboxylase gene for C_4 photosynthesis in *Zea mays*. Mol Gen Genet **224:** 325–332