**Summary**

C₄ photosynthesis is an adaptation that evolved to alleviate the detrimental effects of photorespiration as a result of the gradual decline in atmospheric carbon dioxide levels. In most C₄ plants, two cell types, bundle sheath and mesophyll, cooperate in carbon fixation, and, in so doing, are able to alleviate photorespiratory losses. Although much of the biochemistry is well characterized, little is known about the genetic mechanisms underlying the cell-type specificity driving C₄. However, several studies have shown that regulation acts at multiple levels, including transcriptional, post-transcriptional, post-translational and epigenetic. One example of such a regulatory mechanism is the cell-specific accumulation of major photorespiratory transcripts/proteins in bundle sheath cells, where ribulose-1,5-bisphosphate carboxylase/oxygenase is localized. Although many of the genes are expressed in the bundle sheath, some are expressed in both cell types, implicating post-transcriptional control mechanisms. Recently, ultra-high-throughput sequencing techniques and sophisticated mass spectrometry instrumentation have provided new opportunities to further our understanding of C₄ regulation. Computational pipelines are being developed to accommodate the mass of data associated with these techniques. Finally, we discuss a readily transformable C₄ grass – *Setaria viridis* – that has great potential to serve as a model for the genetic dissection of C₄ photosynthesis in the grasses.
I. Introduction

Photosynthesis is the process of converting atmospheric carbon dioxide (CO₂) into organic compounds using solar energy. It is used by autotrophic/semi-autotrophic organisms and has supported the majority of heterotrophic life as the base of the food chain for billions of years. In green plants, the photosynthetic process can be conceptually divided into two parts: the light-dependent reactions in which light-harvesting complexes and electron transport chains harvest light energy to generate ATP and NADPH; and light-independent reactions in which the Benson–Calvin cycle uses ATP and NADPH to convert CO₂ into sugars. Since the early emergence of photosynthetic organisms, evolution has made few changes to the core complexes that harvest light energy to produce ATP and NADPH (Xiong & Bauer, 2002; Vasil’ev & Bruce, 2004). By contrast, mechanisms to more efficiently deliver CO₂ to the active site of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) have evolved rapidly and independently across nearly 45 lineages of angiosperms (Sage, 2004; Osborne & Beerling, 2006; Edwards et al., 2010). For most terrestrial angiosperms, the initial carbon fixation catalyzed by RuBisCO produces two three-carbon molecules, 3-phosphoglyceric acid (3-PGA), through the fixation catalyzed by RuBisCO (Voznesenskaya et al., 2004). However, between 30 and 80 million yrs ago, atmospheric CO₂ levels began to decline to levels close to 250 ppm (Christin et al., 2008; Vicentini et al., 2008; Edwards et al., 2010). During this time, several lineages evolved utilizing new photosynthetic systems to overcome the disadvantages of photosrespiration. In plants using crassulacean acid metabolism (CAM), reductions in photosrespiration were achieved through temporal uncoupling of the light and dark reactions (Cushman & Bohnert, 1997, 1999). In C₄ plants, reductions in photosrespiration are achieved through the spatial separation of dark and light reactions, either intracellularly, as found in single-cell C₄ species, or intercellularly. In single-cell C₄, intracellular compartmentalization of enzymatic activities enables a two-step carbon fixation process where the initial carboxylation reaction occurs at one end of a cell and decarboxylation and refixation of CO₂ by RuBisCO occur at the other extreme (Voznesenskaya et al., 2004). However, the most prolific adaptation involved the spatial separation of Benson–Calvin cycle activities from the initial carbon fixation step. In angiosperms, a two-cell C₄ system has probably evolved independently in over 45 lineages, including 17 times in the grasses alone (Sage, 2004). C₄ photosynthesis is associated with a unique leaf anatomy, in which a ring of photosynthetic bundle sheath (BS) cells surrounds the vasculature and are themselves surrounded by a ring of mesophyll (ME) cells. This Kranz anatomy is also characterized by denser vein spacing, favoring the metabolic exchange of C₄ acids and sugars between BS and ME cells. C₄ plants, as the name suggests, first incorporate CO₂ into a four-carbon acid, oxaloacetic acid (OAA), through the activity of the oxygen-insensitive enzyme phosphoenolpyruvate carboxylase (PEPC) in ME cells. A more reduced four-carbon acid (malate or aspartate) then moves to adjacent BS cells. Several variations on this theme are possible (Fig. 1) and are generally distinguished on the basis of the primary decarboxylation enzyme used in the BS cell (Gutierre et al., 1974; Prendergast et al., 1987). In the grasses, the more prevalent NADP-ME subtypes use malate to transfer carbon from ME to BS cells with decarboxylation occurring in BS plastids (Sage, 2004). By contrast, NAD-ME subtypes use aspartate to transfer carbon, and the decarboxylation occurs in the BS mitochondrion. A third C₄ subtype, PEPC-carboxykinase-type (PEPCK-type), uses PEPC to catalyze the release of CO₂ from OAA to produce PEP in the BS cell cytosol. Compared with C₃ plants, C₄ and CAM plants have the distinct advantage of reduced photosrespiration; it has been reported that photosrespiration is virtually undetectable in some C₄ species (Yoshimura et al., 2004). The dual advantage of low photosrespiration and high affinity of PEPC for CO₂ ensures an efficient capture of CO₂ from the atmosphere. In arid and hot conditions, when stomatal apertures are smaller to conserve water, the advantages of C₄ are even greater (Sage, 2004).

On average, crops that use C₄ photosynthesis are more productive and use less water and nitrogen than C₃ crops (Brown, 1999). This C₄ advantage, together with current increases in the world population and food prices, has spurred the development of programs to engineer C₄ traits into C₃ crops (e.g. rice) to augment global agricultural productivity (Hibberd et al., 2008; Hibberd & Covshoff, 2010; Zhu et al., 2010). To achieve this ambitious goal, a number of questions need to be addressed. How did C₄
photosynthesis evolve in the grasses? What are the signals that drive C4 cell-type differentiation? What are the mechanisms and genes that control C4 differentiation? To address these questions, new model C3/C4 plants will be needed to accelerate genetic manipulations and engineering. In this review, we focus on the known genetic controls that have contributed to our understanding of C4 photosynthesis, including the cell-specific accumulation of carbon shuttle enzymes, the photorespiratory pathway and emerging tools to understand the developmental control of C4 gene expression. We also discuss a promising C4 model, *Setaria viridis*, for furthering our understanding of C4 regulation and to assist efforts in engineering C4 photosynthesis into monocotyledonous crops.

II. Regulation of C4 differentiation

The emergence of C4 photosynthesis is a relatively recent and probably ongoing process that began c. 80 million yrs ago (Vicentini *et al.*, 2008; Christin *et al.*, 2009). Nevertheless, several distinct features in C4 species clearly differentiate them from their C3 ancestors (Sage, 2004; Edwards *et al.*, 2010). These specializations are anatomical, biochemical and ultrastructural (Hatch, 1987). To date, we understand very little about the genetic regulation of photosynthetic differentiation and almost nothing about specific regulators of C4 photosynthesis. We do know, however, that C4 photosynthesis is regulated on multiple levels: transcriptional, post-transcriptional, post-translational and epigenetic (Sheen, 1999). Below, we describe some of the best examples of regulatory controls in C4 photosynthesis.

1. Transcriptional and post-transcriptional controls

Transcriptional control is by far the most extensively studied mechanism of C4 gene regulation to date (Hibberd & Covshoff, 2010). A good example is the ME-specific localization of PEPC. In maize, a 0.6-kb segment of the PEPC promoter region is sufficient to drive β-glucuronidase (GUS) expression in developing and fully differentiated ME cells (Taniguchi *et al.*, 2000). The expression of a reporter-GUS construct is also induced by light, indicating that the expression of PEPC is suppressed under dark conditions (Kausch *et al.*, 2001). In a recent transcriptomics survey of a developing maize leaf, it was also clear that PEPC is under developmental regulation. There is a c. 300-fold increase in the abundance of ZmPEPC (GRMZM2G083841) transcripts from sink to source tissues (Li *et al.*, 2010). There are a few transcription factors implicated in PEPC regulation in both monocots and dicots. These include DOF1 (DNA Binding with One Finger 1), FhHB1 (Windhovéel *et al.*, 2001), DOF2, MNFs (Maize Nuclear Factors) and PEP-I, which have all been shown to bind to the PEPC promoter (Kanomurakami *et al.*, 1991;
Yanagisawa & Sheen, 1998). DOF1 interacts with the maize PEPC promoter and promotes PEPC expression (Yanagisawa & Sheen, 1998), which is suppressed by DOF2 (Yanagisawa, 2000). Thus, there appears to be an antagonism of action between DOF1 and DOF2. Although the importance of DOF1 for C₄ photosynthesis has been questioned in a recent report (Cavalar et al., 2007), a caveat to this report is that a single weak allele was characterized, leaving open the possibility that low levels of DOF1 accumulation are sufficient to regulate PEPC and perhaps other C₄ genes. Similar transcripational control mechanisms were discovered in C₄ Flaveria sp., where PEPC expression is also largely attributed to its promoter (Stockhaus et al., 1997).

It has been shown that a G-to-A substitution and an insertion of the tetrarnucleotide CACT in the 41-bp ME expression module in the distal region of the Flaveria PEPC promoter (from −1981 to −1940 bp relative to the transcription start site) contribute to the ME-specific expression pattern by suppressing BS expression of PEPC (Gowik et al., 2004; Akyildiz et al., 2007). A proximal region (up to −570 bp) probably interacts with the distal region to regulate the expression of PEPC and is primarily responsible for inducing the high-level expression of PEPC in leaves (Engelmann et al., 2008). The homeodomain proteins FbHB1, 3 and 4 from Flaveria have also been shown to interact with the PEPC promoter in a yeast one-hybrid assay (Windhovel et al., 2001). However, loss-of-function alleles of these genes do not exist and the deletion of a putative FhHB binding site in the 5'-untranslated region (5'-UTR) of the PEPC gene has no significant effect on ME-specific expression of the PEPC gene (Engelmann et al., 2008), making it difficult to assess the biological significance of this interaction.

It is also probable that PEPC expression is subject to additional regulatory controls. For instance, in Thalassiosira pseudonana, a putative single-cell-type C₄ marine diatom, the PEPC transcript is inducible by low CO₂ adaptation, but the mechanism of induction is unknown (McGinn & Morel, 2008). In summary, despite extensive characterizations of the PEPC promoter region and DNA binding elements, the mechanisms that direct ME-specific expression of PEPC are not well understood in any C₄ species.

The expression of NADP-ME is also regulated at the transcriptional level. In maize, NADP-ME expression is BS specific and increases along the leaf developmental gradient (Sheen & Bogorad, 1987; Langdale et al., 1988). However, it is not clear whether transcript accumulation in maize is regulated through transcriptional or post-transcriptional controls. In Flaveria bidentis, a C₄ eudicot, FbME1 transcripts accumulate preferentially in BS cells, are light inducible and require the interaction of 5'- and 3'-noncoding sequences (Marshall et al., 1996, 1997). A mechanism of suppression of NADP-ME transcript accumulation in ME cells is also likely, as the exposure of dark-grown plants to light decreased transcript levels in ME cells (Lai et al., 2002). Interestingly, cis-elements that affect BS-specific expression of ME are different for F. bidentis and a close C₄ relative, F. trinervia. In F. bidentis, the endogenous NADP-ME gene may require elements within the coding region at the 5'- and 3'-ends in addition to specific sequences in the promoter region, whereas, in F. trinervia, only sequences from the 5'-end are required for strong BS-specific expression (Marshall et al., 1997; Ali & Taylor, 2001; Lai et al., 2002). These data suggest that cis-determinates regulating ME1 expression may be rapidly evolving. Indeed, a recent comparative analysis of C₄ ME genes revealed that several amino acid changes are conserved in independently evolved C₄ grasses of the NADP-ME subtype (Wang et al., 2009), suggesting strong positive selection leading to sequence convergence (Christin et al., 2009).

Pyruvate orthophosphate dikinase (PPDK) is another important C₄ enzyme that is required for the regeneration of PEP in ME cells and is regulated at the transcriptional level (Hibberd & Covshoff, 2010). Its expression is driven by two distinct promoters. In maize, the longer transcript encodes a transit peptide that targets the enzyme to the chloroplast; the shorter transcript lacks the transit peptide and is localized to the cytosol (Glackin & Grula, 1990; Sheen, 1991). The same gene structure is conserved in a broad range of plant species, including both C₃ and C₄ types, such as Oryza sativa, Brachypodium distachyon, Sorghum bicolor, Setaria viridis and Setaria italica. Additional PPDK genes have been identified in both maize and C₄ Flaveria sp. that code for the cytosolic version of PPDK, suggesting that the cytosolic version is necessary in both C₃ and C₄ plants (Parsley & Hibberd, 2006; Hennen-Bierwagen et al., 2009). Although PPDK is detectable in BS cells of maize, its expression is highly enriched in ME cells and is also under developmental regulation (Li et al., 2010). In maize, a nuclear binding factor, PPD-1, was found to interact with the PPDK promoter. It was shown that the region from −301 to −296 bp from the transcription start site is crucial for PPD-1 binding, and deletion of the cis element drastically reduces ME-specific expression in a transient microbombardment experiment (Matsuoka & Numazawa, 1991). However, the gene encoding PPD-1 is still unknown and, as discussed above for PEPC and NADP-ME, no transcription factors have been identified in any species that confer BS or ME cell-specific expression.

One of the first transcription factors identified through genetic analysis as potentially playing a role in C₄ photosynthesis was the maize Golden2(G2)/Bsd1 gene (Langdale & Kidner, 1994). In maize, a loss-of-function allele of g2 results in retarded chloroplast development in light-grown BS cells and in both BS and ME cells in dark-grown tissues (Langdale & Kidner, 1994). In the g2 mutant, transcripts of major C₄ carbon shuttle genes are expressed at lower levels compared with the wild-type, and the major C₄
enzymes do not accumulate in BS cells (Langdale & Kidner, 1994). To further elucidate the function of G2 and the related paralog GLK1, homologs were identified and characterized in Arabidopsis thaliana. A double mutant Atglk1 Atglk2 (Golden2-like) remained pale green throughout development and showed reduced thylakoid stacking (Fitter et al., 2002). A characterization of transcriptional targets of AtGLK1 and AtGLK2 suggested that there are at least 100 targets, most of which encode components of the photosynthetic apparatus, such as photosystem I and II, as well as tetrapyrrole biosynthesis (Waters et al., 2009). Furthermore, this A. thaliana study revealed that AtGLK1 and AtGLK2 are functionally redundant. However, differential expression of each paralog may provide a mechanism to fine tune gene expression. In maize, the Golden2 transcript was enriched in BS cells, and GLK1 was enriched in ME cells (Hall et al., 1998; Rossini et al., 2001). A recent transcriptomics study in maize revealed similar findings with G2 (GRMZM2G026833) and Glk1 (GRMZM2G087804) expressed predominantly in BS and ME cells, respectively, in leaf tip tissues (Li et al., 2010). Complete loss-of-function mutants of both Golden2 and Glk1 will hopefully lead to the resolution of G2/GLK1 function in contributing to C₄ photosynthetic development.

2. Post-translational regulation

Modification or degradation of proteins is also implicated in the control of C₄ cell-type differentiation (Majeran & van Wijk, 2009). One example is the reversible phosphorylation of PEPC, the primary carboxylation enzyme that provides oxalacetate to the C₄ cycle. The regulation of PEPC has been extensively characterized in plants that utilize both C₄ and CAM pathways (Chollet et al., 1996; Nimmo, 2003). The phosphorylation of PEPC occurs at a conserved N-terminal serine, which limits feedback inhibition by malate and increases sensitivity to activation by glucose-6-phosphate (Echevarria et al., 1994; Duff et al., 1995). In maize, the phosphorylation of leaf PEPC occurs prior to dawn and decreased phosphorylation begins well before dark, suggesting potential circadian and/or metabolic control mechanisms in addition to light regulation (Ueno et al., 2000). Interestingly, PEPC kinase (PEPCK, GRMZM2G178074), the enzyme that phosphorylates PEPC, also appears to be regulated at the protein level through redox control (Saze et al., 2001). Rapid degradation of PEPCK has been observed in ME cells in response to light and probably involves the ubiquitination of PEPCK (Agetsuma et al., 2005). It is worth noting that PEPCK mRNA is preferentially expressed in ME cells, suggesting cell-specific transcriptional control as well (Li et al., 2010).

PPDK is another carbon shuttle enzyme that is also tightly regulated by post-translational phosphorylation. The regulation of PPDK by phosphorylation differs from that of PEPC in a few important ways. First, PPDK activity decreases with phosphorylation, whereas PEPC activity increases. Second, PPDK regulatory protein (PPDK-RP) is the sole enzyme that catalyzes both the phosphorylation and dephosphorylation of PPDK, whereas PEPC is phosphorylated by one enzyme (PEPCk) and is dephosphorylated by a heteromeric protein phosphatase 2A complex (Budde et al., 1985; Dong et al., 2001). PPDK-RP, a very low abundance (< 0.04% of soluble maize leaf protein) ME-localized protein, mediates the light-induced phosphorylation of a threonine residue on PPDK (Thr-456 in maize) and is also capable of catalyzing the dephosphorylation of PPDK at night (Ashton et al., 1984; Budde et al., 1985; Burnell & Hatch, 1985). Third, the regulatory mechanism of PPDK-RP is closely modulated by light and presumably more directly by the availability of ADP, the substrate for PPDK-RP. Under dark conditions, stromal ADP increases, leading to increased PPDK-RP kinase activity and phosphorylation of PPDK, whereas, in the light, stromal ADP decreases, leading to increased PPDK-RP phosphatase activity and dephosphorylation of PPDK, and activation of the enzyme (Chastain & Chollet, 2003). As a result PPDK activity is closely correlated with dark–light cycles. Unlike PEPCk, there is no evidence that PPDK-RP is post-translationally regulated (Burnell & Chastain, 2006). Unexpectedly, the transcript encoding PPDK-RP (GRMZM2G004880) is enriched in BS cells, whereas the PPDK transcript is primarily localized to ME (Li et al., 2010). However, both enzymes accumulate predominantly in ME cells, suggesting an uncoupling of transcriptional and translational or post-translational controls regulating PPDK-RP accumulation (Majeran et al., 2008).

In addition to phosphorylation, PPDK abundance may also be controlled through protein turnover. A study performed in Miscanthus giganteus showed that cold treatment leads to decreased accumulation of PPDK protein, whereas quantitative reverse transcription-polymerase chain reaction (RT-PCR) suggested the PPDK transcript was not affected (Naidu et al., 2003). It is interesting to speculate that ubiquitination of PPDK at low temperatures and ubiquitination of PEPCk in the light may play roles in maintaining the cell specificity of these enzymes. Although this is clearly a speculation, further studies into protein turnover may reveal new mechanisms of cell-specific control.

Although studies focusing on the post-translational control of C₄ differentiation are still limited, it is safe to assume that such regulatory mechanisms will apply beyond the phosphorylation of PEPC and PPDK. For instance, post-translational protein turnover is likely to be involved in the reduction of photosystem II complexes in mature BS cells (Woo et al., 1970; Schuster et al., 1985). For example, in maize, photosystem II activity and granal stacking are clearly observed in immature BS cells, which gradually disappear through development (Woo et al., 1970; Schuster...
et al., 1985). Further research combining transcriptomics and proteomics datasets may help us to identify more C₄ control mechanisms at the post-translational level.

3. Epigenetic control

To date, the link to epigenetic control of C₄ gene expression has been predominantly associated with the expression of PEPC. In maize, ME-specific accumulation of PEPC transcript has long been attributed to specific promoter sequences (Taniyuki et al., 2000; Kausch et al., 2001). Interestingly, the methylation state of a restriction site in the PEPC promoter has been correlated with cell-specific gene expression (Langdale et al., 1991). However, it is clear that methylation at this site cannot be the only regulatory mechanism, because a promoter-GUS fusion lacking this methylation site is still capable of initiating ME-specific accumulation of GUS activities in both developing and fully differentiated ME cells (Taniyuki et al., 2000; Kausch et al., 2001; Hibberd & Covshoff, 2010). Chromatin modification also seems to be important for PEPC and malic enzyme expression in maize, with methylation and histone acetylation being implicated in tissue-specific and light-regulated control of gene expression (Danker et al., 2008; Offermann et al., 2008). With the advent of next-generation sequencing technology, it is likely that genome-wide mapping of methylation patterns (e.g. bisulfite-seq) will provide new insight into the potential regulation of C₄ gene expression (Feng et al., 2010; Laird, 2010).

III. The photorespiratory pathway: a case study for transcriptional and post-transcriptional controls

The evolution of C₄ photosynthesis has resulted in the partitioning of many biochemical activities between BS and ME of C₄ plants. In maize, over 20% of the transcriptome is differentially expressed between BS and ME cells (Sawers et al., 2007; Li et al., 2010) and surveys of the maize plastid proteome confirm a high degree of specialization in each cell type (Majeran et al., 2008, 2010; Majeran & van Wijk, 2009; Friso et al., 2010). One pathway that is ubiquitous in all plants and was probably the target of C₄ evolution (Sage, 2004) is the photorespiratory pathway. Photorespiration is detrimental to plant growth, resulting in the production of 2PG, a process that diverts energy and carbon from photosynthesis. In C₄ plants, it has been estimated that CO₂ levels in BS cells may approach 10–100 times the levels reached in ME cells of C₃ plants, effectively eliminating the oxygenase activity of RuBisCO (Furbank & Hatch, 1987). Nevertheless, in maize, a functional photorespiratory cycle is required for photosynthetic development (Zelitch et al., 2009), indicating an essential role for photorespiration in both C₃ and C₄ plants. Here, we discuss the cell specificity of transcripts/enzymes that are involved in the C₂ cycle and suggest that a detailed study of the BS cell-specific accumulation of these enzymes may prove fruitful in elucidating novel mechanisms that drive differential gene expression in C₄ plants.

Photorespiration involves a series of enzyme-catalyzed reactions that results in lower net fixation of carbon and the addition of NADPH and ATP (Bauwe et al., 2010). As illustrated in Fig. 2, a primary product of oxygenation of RuBP, 2PG, is hydrolyzed to glycolate through a specific chloroplast-localized 2PG-phosphatase (PGP). In A. thaliana, a knockout of PGP is seedling lethal under ambient conditions, but can be rescued under high CO₂ (Somerville & Ogren, 1979; Schwarte & Bauwe, 2007). Maize B73 also contains two copies of PGP genes (GRMZM2G018441 and GRMZM2G417843) based on the B73 5a.59 build of the maize genome (Schnable et al., 2009). In maize, the plastid isoform of PGP (GRMZM2G018441) is expressed at high levels in BS cells of photosynthetically active cells (Li et al., 2010). Once 2PG is converted to glycolate, it is transported to the peroxisome. This is achieved first by an unknown glycoglycerate transporter(s) which moves glycolate from the chloroplast. Glycolate is then imported into the peroxisome through porin-like channels on the peroxisomal outer membrane (Reumann & Weber, 2006). Once in the peroxisome, glycolate is oxidized to glyoxylate by glycolate oxidase (GOX). A molecule of O₂ is required for this irreversible reaction, which also produces hydrogen peroxide as a byproduct. In C₃ species, GOX is indispensable for the photorespiratory cycle. For instance, a reduction in GOX activity in rice led to a reduced photosynthetic rate, together with other typical photosynthesis-deficient phenotypes (Xu et al., 2009). There has been much debate regarding the role of the photorespiratory cycle in C₄ plants, as photorespiration is undetectable in many C₄ species (Yoshimura et al., 2004). However, Zelitch et al. (2009) showed that a maize GOX (GRMZM2G129246) mutant is lethal under normal growth conditions in air, but can be rescued with elevated CO₂. They also showed that glycolate accumulated in the GOX mutants, indicating that, despite the partitioning of RuBisCO to BS, some photorespiration is probably occurring in C₄ plants. This study revealed the essential role of the photorespiratory pathway in maize and strongly suggests that this pathway is operational in additional C₄ plants (Zelitch et al., 2009). Reducing the hydrogen peroxide produced by GOX is also an important part of photorespiration, which is normally catalyzed by catalase (CAT). There are three isoforms of CAT in Arabidopsis, and CAT2, which is expressed at highest levels in the leaf, is required for the photorespiratory cycle (Queval et al., 2007). There are also three putative CAT genes in the maize genome (GRMZM2G079348, GRMZM2G088212 and GRMZM2G090568) that are probably orthologs of the
A. thaliana genes. Among the maize CAT paralogs, GRMZM2G090568 is the most similar to Arabidopsis CAT2 and is highly enriched in BS cells (Fig. 2), suggesting that it may be involved in the maize photorespiratory cycle. As shown in Fig. 2, many transcripts encoding photorespiratory genes accumulate in a cell-specific manner. These include transcripts for the glutamate:glyoxylate aminotransferase (GGT), which catalyzes the conversion of glyoxylate to glycine in the peroxisome, glycine decarboxylase (GDC) and serine hydroxymethyltransferase (SHMT), which jointly convert glycine to serine in the mitochondrion, and peroxisomal NADH-dependent hydroxypyruvate reductase (pHPR), which converts hydroxypyruvate into glyceraldehyde. One exception to this pattern is the accumulation of transcripts for serine:glyoxylate aminotransferase (SGT) in ME cells. SGT converts serine to hydroxypyruvate and thus could be expected to accumulate in peroxisomes of BS cells. However, serine:glyoxylate aminotransferase proteins are usually multifunctional, and thus SGT may be required for alanine and aspartate metabolism when alanine is used as the amino-group donor (Truszkiewicz & Paszkowski, 2005). This non-photorespiratory activity may demand the expression of SGT in both BS and ME cells. It is also possible that SGT is being controlled post-transcriptionally. A systemic study of peroxisomal proteins would help to resolve some of these discrepancies.

A second potential exception to the localization of transcripts with photorespiratory activity to BS is glycerate kinase (GLYK). This enzyme catalyzes the last step of the C2 cycle by converting glycerate to 3-PGA in the chloroplast. In maize, there appear to be two genes that encode GLYK (GRMZM2G018786 and GRMZM2G054663) with the former being expressed at a much higher level (Li et al., 2010). Both ZmGLYK transcripts accumulate to high levels in ME and BS cells, but appear to be enriched in ME cells. Interestingly, a recent report showed that the maize GLYK differs from the C3 version of GLYK by the existence of
of a short carboxy-terminal extension, and is subject to redox control. A disulfide bond is formed by the oxidation of two cysteine residues which inhibits GLYK activity; this inhibition is reversed by cleavage of the disulfide bond (Bartsch et al., 2010). This report strongly indicates that GLYK is regulated by the redox state. Photosystem I and II activities reside in the ME cell plastids (Oswald et al., 1990) resulting in a redox environment that is comparable with that in chloroplasts of C3 plants, but potentially less favorable for GLYK activity. Moreover, a maize proteomic study showed that the NAD(P)H-dehydrogenase complex is enriched in BS cells (Majeran et al., 2008), consistent with the hypothesis that BS cells may have a more reduced environment that favors GLYK activity, thus resulting in enhanced GLYK activity in BS relative to ME cells.

Because energy and carbon are lost during photorespiration, it has been assumed that reduced photorespiration would lead to increased photosynthetic efficiency and enhanced biomass production. However, this hypothesis has never been tested experimentally until recently, when the Escherichia coli glycolate catabolic pathway was engineered into A. thaliana chloroplasts (Kebeish et al., 2007). As illustrated in Fig. 2, this photorespiratory bypass converts glycolate to glyceraldehyde in three steps all within the chloroplast. The transgenic plants expressing the transgenes grew faster and accumulated greater shoot and root biomass, presumably as the result of a shortened photorespiratory cycle (Kebeish et al., 2007). This experiment clearly demonstrates the feasibility of improving the photorespiratory cycle. Given the essential role of GOX in maize, a similar shunt may have application in C4 plants to reduce the carbon loss associated with photorespiration. An additional application of a photorespiratory bypass should be considered in efforts to engineer C4 rice (http://irri.org/c4rice), as the localization of high levels of RuBisCO to BS in rice may necessitate the maintenance of a photorespiratory cycle as well. In summary, the photorespiratory cycle represents a useful target for examining the multiple levels of control that have evolved in C4 photosynthesis.

IV. Future perspectives of C4 research – an introduction to a C4 model

Photosynthesis is a fundamental process that has been the subject of intense study for decades, yet not much progress has been made in understanding the major regulatory networks that underlie the C4 differentiation process. However, a better understanding of these networks is likely to come through recent technological innovations. These include the advent of next-generation sequencing technologies and sophisticated mass spectrometry instrumentation that have dramatically increased our capability to study C4 photosynthesis at the transcriptomic and proteomic levels (Brautigam et al., 2008; Majeran et al., 2008; Li et al., 2010). Indeed, in the past 2 yr, the cell-specific accumulation of thousands of transcripts and proteins has been defined in developing BS and ME cells in maize (Li et al., 2010; Majeran et al., 2010) and in Cleome (Brautigam et al., 2011). In a recent study, c. 80% of the transcripts that have been annotated from the maize genome were detected using Illumina-based sequencing of leaf transcript pools (Schnable et al., 2009). In addition, a developmental series of gene expression was conducted which should allow a deep exploration of the developmental dynamics of C4 cell-type differentiation (Li et al., 2010). Comparative proteomics studies of maize ME and BS cells have also revealed the partitioning of many activities between these cell types (Brautigam et al., 2008; Majeran et al., 2008, 2010; Friso et al., 2010). Yet, despite these large inventories of transcripts and proteins, the systems analysis of these datasets has only just begun.

One current limitation to a systems biology approach is the lack of computational tools needed to integrate diverse datasets, such as proteins, metabolites and transcripts (Wang et al., 2010). Tools such as MetGenMAP (Joung et al., 2009), the Reactome (Croft et al., 2011) and the Tomato Functional Genomics Database (Fei et al., 2011) are offering some avenues to explore and integrate these datasets, but many challenges remain. A few of the most significant are the problems of data integration, gene/protein/metabolite annotation and data mining. It will become increasingly important to adopt standards for RNA-seq (e.g. SAM/BAM) (Li et al., 2009) and proteomics analysis (e.g. MIAPE) (Taylor, 2006). Furthermore, data repositories will soon become overwhelmed with the massive output capacity of the latest sequencing techniques. For instance, the latest HiSeq2000® machine from Illumina is capable of producing 150–200 gigabytes of sequence data per run (http://www.Illumina.com). If we also consider the storage space needed for post-processing, intermediate files and final output, the amount can easily run into the terabyte range. A centralized datacenter with remote access would be one option for researchers to process high-throughput data. Once such center is TACC (Texas Advanced Computing Center) that can provide up to 10 petabytes of storage for NSF-funded projects. The other option is cloud-based storage. There are several commercially available options that will provide storage and bioinformatics analysis options (e.g. Amazon, Microsoft Azure, DNAnexus, GenomeQuest). The other limitation is the availability of standard data analysis pipelines. Individual groups have developed computational tools for analyzing RNA-seq data, including BWA, Tophat and Cufflinks, which map read data and calculate expression levels of individual transcripts (Langmead et al., 2009; Li & Durbin, 2009; Trapnell et al., 2009, 2010). Although these toolsets perform well, compatibility issues remain for comparing datasets from different species and sequencing methods. The iPlant Collaborative (http://www.iplantcollaborative.org/)
is currently building a public NextGen data analysis framework that will enable scientists to process their data remotely and help to standardize the process. These bottlenecks in data storage, processing and integration will need to be overcome if we are going to truly exploit the wealth of information that is now being generated for a systems analysis of C₄ photosynthesis.

One of the greatest limitations to understanding C₄ is the inadequacy of the current model genetic systems. Most of our understanding of C₄ photosynthesis has come from studies of organisms with large genomes, long generation cycles and low transformation efficiencies (e.g. sorghum, maize, Cleome, Flaveria). To accelerate the pace of discovery, we have recently proposed using Setaria viridis (Fig. 3a) as a model C₄ plant (Brutnell et al., 2010). Setaria viridis is a wild relative of the foxtail millet (Setaria italica) and uses the NADP-ME subtype of C₄ photosynthesis. Setaria viridis has a small stature that is comparable with Arabidopsis. Moreover, under short-day growth conditions, its lifecycle can be completed in < 6 wk. The most important aspects of Setaria that make it a promising model system are the availability of a complete genome sequence and a protocol for high-efficiency stable and transient transformation (Fig. 3b,c) (Brutnell et al., 2010). This system will be ideal for conducting high-throughput genetic screens, examining cis-regulatory elements and overexpressing or ‘knocking-out’ genes that are involved in C₄ photosynthesis. By combining the power of informatics-based approaches with this emerging model system, it should be possible to move forward quickly in the elucidation of the molecular mechanisms that underlie C₄ photosynthesis and to develop a platform for engineering C₄ traits into some of the world’s most important crop plants.

Acknowledgements

We would like to thank Pinghua Li and Hugues Barbier for helpful comments on the manuscript. This work was supported through an NSF grant (IOS-0701736 award to T.P.B.) and funding through the iPlant Collaborative (L.W. and T.P.B.).

References


