Identification and Expression Analysis of Phosphoenolpyruvate Carboxylase (PEPc) and PEPc Kinase Genes in C_3 Plants

by
Jonathan Stuart Sullivan

Thesis submitted for the degree of doctor of philosophy

Division of Biochemistry and Molecular Biology,
Institute of Biomedical and Life Sciences,
University of Glasgow
December, 2004

Part of the work described in this thesis has been published as:

J.T. Marsh, S. Sullivan, J. Hartwell and H.G. Nimmo. (2003) Structure and expression of phosphoenolpyruvate carboxylase kinase genes in Solanaceae. A novel gene exhibits alternative splicing. *Plant Physiology* **133**: 2021-2028

S. Sullivan, G.I. Jenkins and H.G. Nimmo (2004) Roots, cycles and leaves. Expression of the phosphoenolpyruvate carboxylase kinase gene family in soybean. *Plant Physiology* **135**: 2078-2087

Content of Thesis

List of figures and tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Abbreviations

The abbreviations used in this thesis are described in full below:

AAT, aspartate aminotransferase; ABA, abscisic acid; ADP, adenosine diphosphate; AEC, adenylate energy charge; AMP, adenosine monophosphate; ATP, adenosine triphosphate; ATPase, adenosine triphosphatase; Asp, aspartate; bp, base pair; CAB, chlorophyll a/b binding protein; CAM, crassulacean acid metabolism; CaMK, calcium/calmodulin-dependent kinase; cDNA, complementary DNA; CDPK, calcium-dependant protein kinase; CO_2, carbon dioxide; CS, citrate synthase; DAG, 1,2-diacylglycerol; DCDP, 3,3-dichloro-2-dihydroxyphosphinoylmethyl-2-propenoate; Dct-system, dicarboxylate transport system; DD, continuous dark; dH_2O, distilled water; DNA, deoxyribonucleic acid; dNTP, deoxynucleotide triphosphate; EDTA, ethylene diamine tetra-acetic acid; EGTA, ethylene glycol tetra-acetic acid; EST, expressed sequence tag; Fru-6-P, fructose-6-phosphate; gDNA, genomic DNA; Glc-1-P, glucose-1-phosphate; Glc-6-P, glucose-6-phosphate; GOGAT, glutamate synthase (glutamine 2-oxoglutarate aminotransferase); GS, glutamine synthase; ICDH, NADH-isocitrate dehydrogenase; Ins(1,4,5)P_3, inositol-1,4,5-biphosphate; Kb, kilobase; Lg, leghemoblobin; LL, continuous light; MDH, malate dehydrogenase; mRNA, messenger RNA; NIR, nitrite reductase; NR, nitrate reductase; OAA, oxaloacetate; OG, oxoglutarate; PBM, peribacteroid membrane; PBS, peribacteroid space; PCR, polymerase chain reaction; PEP, phosphoenol pyruvate; PEPc, phosphoenolpyruvate carboxylase; PFK, ATP-dependant phosphofructokinase; PFP, PPi-dependant phosphofructokinase; PGA, phosphoglycerate; Pi, inorganic phosphate; PI-PLC, phosphoinositide-specific phospholipase C; PK, pyruvate kinase; PKC, protein kinase C; PLC, phospholipase C; PPCK, phosphoenolpyruvate carboxylase kinase; PPi, inorganic pyrophosphate; RACE, rapid amplification of cDNA ends; RH, relative humidity; RNA, ribonucleic acid; RNase, ribonuclease; RT, reverse transcriptase; Rubisco, ribulose 1,5-biphosphate carboxylase / oxygenase; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; Ser, serine; TCA, tricarboxylic acid cycle; Thr, threonine; UDP, uridine diphosphate.

Throughout this thesis I will use italicised abbreviations for genes names. For example, the *Glycine max* phosphoenolpyruvate carboxylase kinase gene will be shortened to *GmPPCK*. I have also abbreviated the protein names, these will be non-italicised (GmPPCK, for example).

Acknowledgements

I would like to acknowledge my supervisors, Profs. Gareth I. Jenkins and Hugh G. Nimmo for providing me with the opportunity to study at Glasgow University and for their support and encouragement during my studentship. I am particularly grateful to Prof. Hugh Nimmo, in whose laboratory I was allowed to play. I would like to thank the members of the Bond lab, in particular Jim Jardine, James Hartwell, Justin Marsh, Veronique Fontaine, Gill Nimmo, Mathew Shenton and Zhi-Hui Chen for providing ideas, inspiration, technical knowledge, coffee, biscuits and cigarettes, i.e. all the ingredients of a happy work place. I would like to thank the members of the Plant Science group for all the help I have received during my research, the many useful discussions and trips to the pub. I am extremely grateful to the firemen of the Strathclyde Fire Brigade, whose efforts in rescuing fridges, freezers and my lab books made this thesis possible. I am also grateful to Christine Salatin for her help and support during my final year and for enabling me to remain in Glasgow during my writing-up. Finally I would like to thank my family for their continual support and encouragement, in particular my parents for their belief that one day I would finish my education and my brother for always being there.

Abstract

Phosphoenolpyruvate carboxylase (PEPc) is a cytosolic enzyme that plays a wide range of roles in different tissues of higher plants. A photosynthetic isoform of PEPc catalyses the primary fixation of CO_2 in C_4 and CAM plants. Other isoforms are thought to be responsible for different functions, including the replenishment of tricarboxylic acid cycle intermediates, pH control, and the provision of malate in guard cells, developing fruit and legume root nodules.

PEPc is an allosteric enzyme, inhibited by malate and activate by glucose 6-phosphate. Superimposed on this is control by reversible phosphorylation of a single serine residue close to the N-terminus of the protein. Phosphorylation modulates the allosteric properties of the enzyme, with the phosphorylated form being less sensitive to malate inhibition. The phosphorylation state of PEPc is controlled largely at the level of expression of PEPc kinase (*PPCK*) genes. The aim of this work was to identify and characterise *PEPC* and *PPCK* genes in C_3 plants, particularly in the leguminous plant soybean (*Glycine max*).

In legume root nodules, PEPc plays a central role in the metabolism that allows fixation of atmospheric N_2 by bacteroids. It provides the C_4 dicarboxylates that are required by the bacteroids for energy generation and also the carbon skeletons that are needed for the subsequent assimilation of ammonium into amino acids. In soybean one PEPc gene (*GmPEPC7*) is highly and relatively specifically expressed in nodules. To test the hypothesis that PEPc kinase expression would show a similar pattern, the expression of both *PPCK* and *PEPC* genes were analysed in soybean. Soybean contains at least four *PPCK* genes. The genomic and cDNA sequences of these genes were determined, and the function of the gene products demonstrated by *in vitro* expression and enzyme assays. For two of these genes, *GmPPCK2* and *GmPPCK3*, transcript abundance is highest in nodules and is markedly influenced by supply of photosynthate from the shoots. One gene, *GmPPCK4*, is under robust circadian control in leaves but not in roots. Its transcript abundance peaks in the latter stages of subjective day. This is the first report of a C_3 *PPCK* gene that is controlled in this manner. The expression patterns of five *PEPC* genes, including one encoding a bacterial-type PEPc lacking the phosphorylation site of the "plant" PEPcs, were also investigated. The *PEPC* expression patterns do not match those of any of the *PPCK* genes, arguing against the existence of specific PEPc-PEPc kinase expression

partners. Collectively, the results demonstrate that the *PEPC* and *PPCK* gene families in soybean are significantly more complex than previously understood.

Another role of PEPc is the production of organic acids in developing fruit. Previous work had identified two *PPCK* genes in tomato (*Lycopersicon esculentum*), one of which (*LePPCK2*) contained a novel second intron that exhibits alternative splicing. The correctly spliced transcript encodes a functional PEPc kinase, whereas unspliced or incorrectly spliced transcripts encode a truncated, inactive protein. The relative abundance of the transcripts varies in different tissues, both in wild-type plants and in the *greenflesh* tomato mutant. A *PPCK2* gene was identified in tobacco (*Nicotiana tabacum*), also exhibiting alternative splicing and producing various amounts of the transcript in different tissues. Similar sequences are also present in potato (*Solanum tuberosum*) and aubergine (*Solanum melongena*). The data suggest that the alternative splicing of *PPCK2* transcripts is functionally significant.

The release of the rice (*Oryza sativa*) genome sequence allowed the identification of three putative *PPCK* genes. The sequence of one of these genes (*OsPPCK2*) was unusual in containing three in frame methionine residues at the 5' end, potentially allowing the production of a larger PEPc kinase protein. The functionality of these genes was assessed, but the results were unable to provide unequivocal evidence demonstrating the phosphorylation of PEPc by the 'long' form of OsPPCK2.

Overall the data presented in this thesis reveal a novel picture of the transcriptional control of PEPc kinase in C_3 plants.

Chapter 1

Introduction

This thesis explores the roles and regulation of phosphoenolpyruvate carboxylase kinase (PPCK) in C_3 plants. The major part of this work was conducted on the leguminous plant soybean (*Glycine max*). Legumes are characterised by their ability to enter into a symbiotic association with certain soil-born bacteria. Therefore my introduction will initially give an overview of this process before moving on to consider the many roles the enzyme phosphoenolpyruvate carboxylase (PEPc) plays in higher plants. Finally I shall discuss the control of PEPc via reversible phosphorylation and the recent advances in identifying and characterising the kinase responsible.

1.1 Symbiotic nitrogen fixation

1.1.1 Introduction

Nitrogen is an essential element for plant growth and reproduction. Lack of mineral nitrogen in the soil often limits plant growth and, as a result, modern agriculture uses approximately 80 million tons of nitrogen fertiliser annually (Trevaskis *et al.*, 2002). Biological nitrogen fixation is a fundamental global process. It is the most significant natural pathway for the introduction of nitrogen into the biosphere (Serraj *et al.*, 1999). Fixation of atmospheric N_2 is restricted to organisms with prokaryotic cell structures; bacteria and cyanobacteria. Fixation occurs in the free-living state, in association with plants or in symbiosis with plants. Nitrogen fixing symbioses occur in several plant groups. Cyanobacteria form symbioses with several mosses, the water fern *Azolla*, a group of gymnosperms (the cycads) and the flowering plant *Gunnera* (Peters, 1991). The Gram-positive actinomycete *Frankia* infects nearly 200 species of flowering plants belonging to eight different families (Doyle, 1998). The best characterised symbiotic N_2-fixing system is the symbiosis between the legumes and a range of Gram-negative soil bacteria termed rhizobia. The *Frankia* and rhizobial symbioses are characterised by the formation of a novel plant organ, the nodule, on the roots (or rarely the stems) of the host plant.

Phylogenetic analysis of the angiosperms, based on the chloroplast *rbcL* sequence, suggests that all flowering plant families involved in rhizobial symbiosis or actinorhizal symbiosis (nodulation by *Frankia* spp.) belong to the same lineage, the Rosid I clade (Soltis *et al.*, 1995). However it is only a small minority of plant genera

within this large group that can be nodulated and their distribution amongst non-nodulated plants, often from the same families, suggests nodulation has originated many times independently (Doyle, 1998). It has therefore been proposed that, although nodulation may have evolved independently, the predisposition to nodulate may have a single origin (Soltis *et al.*, 1995).

1.1.2 Legume-rhizobium nodulation

It was in 1888 that Hellriegel and Wilfarth first identified rhizobia as the source of fixed nitrogen in the root nodules of legumes (Hausler *et al.*, 2001). Since then the legume-rhizobial nodulation process has been intensively studied. This is as a model system for the understanding of plant-microbe interactions, as well as for agricultural reasons; a number of important crop species are part of the legume family. These include grain legumes such as soybean, peas and lentils and fodder legumes including clover, alfalfa and vetch. A better understanding of the nodulation process will enable more efficient farming and aid the breeding of new varieties. Also the ability to engineer non-leguminous crop plants to successfully fix nitrogen has long been a goal. During symbiosis the host plant receives fixed nitrogen from the rhizobia in return for providing a source of respiratory substrates.

1.1.3 The infection process

The rhizobia-legume interaction has many features in common with plant parasite infection and development. These include the binding of microbes to the host plant, penetration into the host plant, redirection of the host plant metabolism and the morphological manifestation of the interaction, i.e. the nodule (Vance, 1983) (see Fig. 1.1). It has been suggested that nodulation can be seen as a highly regulated plant disease response. The infection process and subsequent nodule formation has been extensively researched and I will present only a brief overview of this work.

The development of nitrogen-fixing nodules is a complex multi-step process requiring the exchange and recognition of specific signals by both the host plant and the infecting rhizobia. The first stage in the process involves the migration of the bacteria towards the host plant and their attachment to the root hair surface. Flavonoid compounds present in plant-root exudate attract rhizobia toward the root surface and induce the transcription of bacterial nodulation (*nod*) genes. These genes are not expressed in the free-living bacteria, except for *nodD* that is constitutively expressed.

NodD has the ability to bind specific flavonoids in the plant-root exudate and this induces the transcription of the other *nod* genes. The products of these *nod* genes are specific lipochito-oligosacharrides called Nod factors. The specific Nod factors produced by the bacteria are associated with determining its host range. The host-specific characteristics are based on structural variations in the Nod factors produced by different rhizobial strains. In general rhizobia can only successfully infect a limited number of host plants. However, some rhizobia have a broad host range, such as *Rhizobium* spp. NGR234, which can nodulate more than 110 genera of legumes as well as the non-legume *Parasponia andersonii* (Viprey *et al.*, 1998). The basis for its broad host range is thought to be the wide variety of Nod factors it excretes (Price *et al.*, 1992).

As well as their role in specificity, Nod factors also induce several responses in the host plant. The first event, occurring in seconds after Nod factor application, is the transient influx of Ca^{2+} into the root hairs. The high Ca^{2+} concentration at the root hair tip is associated with the ability of the bacteria to cause a reorientation in the direction of growth of the root hair, causing it to curl. The growth direction of the root hair is constantly redirected towards the bacteria, which leads to it becoming entrapped. Local degradation of the cell wall allows the bacterial cells direct access to the outer surface of the plant plasma membrane. However, it is the plant itself that weakens the cell wall at the site of anticipated infection (Parniske, 2000).

A separate Ca^{2-} phenomenon is observed several minutes after Nod factor application. Ca^{2+} spiking has been observed originating from the region around the nucleus of root hairs in several legume species (Lhuissier *et al.*, 2001). The amplitude of the spikes decreases with distance from the perinuclear region and no spikes were detected at the root tip. Their role is still unclear but it is thought that they may be involved in signal transduction and gene expression.

Internalisation of the bacteria is achieved by the formation of an infection thread. After localised cell wall hydrolysis at the point of infection, the plasma membrane invaginates to form a tubular structure, the infection thread, which the bacteria then enter. The rhizobia inside the infection thread grow and divide, keeping the tubule filled with bacteria.

The processes of infection and nodule organogenesis occur simultaneously during nodule development. As the infection thread grows towards the base of the infected cell, the root cortical cells dedifferentiate and start dividing to form the

nodule primordium, from which the nodule will develop. When the infection thread reaches the primordia, rhizobia are released from the tip of the infection thread and enveloped by a membrane of plant origin called the peribacteroid membrane (PBM). Even though the rhizobia are intracellular, they are always extracytoplasmic because of the integrity of the PBM (Parniske, 2000). Within this organelle-like compartment, termed the symbiosome, the bacteria stop dividing and enlarge and differentiate into bacteroids able to fix atmospheric nitrogen.

The signals that trigger the uptake of the rhizobia from the infection thread into the PBM-enclosed compartments are unknown. There is evidence for the involvement of a type-III secretion system (TTSS). Recently, TTSSs have been shown to exist in several rhizobia (Marie *et al.*, 2001). TTSSs are multi-protein complexes that span the entire bacterial cell envelope and are conserved among animal and plant pathogenic bacteria (Pühler *et al.*, 2004). It was previously thought that TTSSs were unique to pathogenic bacteria. Mutations in two components of the TTSS of the broad-host-range *Rhizobium* spp. NGR234 strongly affect its ability to nodulate a variety of tropical legumes (Viprey *et al.*, 1998).

1.1.4 The nodule

The mature nodule is a vascularised plant organ where the fixation of atmospheric nitrogen occurs. It consists of the central tissue that contains infected and uninfected cells. In the mature nodule the enlarged infected cells may contain thousands of symbiosomes (Udvardi and Day, 1997). Fig. 1.2 shows an electron micrograph of infected and uninfected cells from a soybean root nodule. Proliferation of the PBM requires massive synthesis of lipids and proteins. An infected nodule cell has a plasma membrane area of 2800 μm^2 but a PBM area of 21500 μm^2 (Parniske, 2000). The central tissue is surrounded by the nodule parenchyma, the endodermis and outermost the cortex. Vascular bundles are located in the nodule parenchyma.

Based on their morphology nodules are classified as determinate or indeterminate. Temperate legumes (*Pisum*, *Medicago*, *Vicia* and *Trifolium*) develop club-shaped indeterminate nodules on their roots. They are elongated due to a persistent apical meristem located at their apex, which gives rise to various developmental zones within a single nodule. The youngest zone, termed the prefixation zone, is located near the apical meristem and is where infection of the cells

takes place. The central zone is where nitrogen fixation occurs. At the base of the nodule is the senescence zone where the bacteroids are degraded by the plant.

Tropical legumes (*Arachis*, *Glycine*, *Phaseolus* and *Lotus)* form spherical determinate nodules. In contrast to indeterminate nodules, they develop from a spherical meristem that stops dividing at an early stage of development and is not evident in mature nodules. There are also differences in the number of bacteroids contained within each PBM. In indeterminate nodules bacteroid division is tightly coupled with PBM division so a single bacteroid is surrounded by the PBM. In determinate nodules this coupling breaks down later in development so that several bacteroids may be enclosed (Brewin, 1991).

1.1.5 Nodule metabolism

During symbiosis the bacteroid fixes atmospheric N_2 to ammonia, which is then assimilated by the plant via the glutamine synthase (GS)/glutamate synthase (GOGAT) pathway. In exchange the plant provides the bacteroid with reduced carbon that is used as fuel for bacteroid metabolism, including nitrogen fixation. This metabolic co-operation requires several specific adaptations: protection of the oxygen labile nitrogenase enzyme complex; translocation of carbon compounds to support the high energy demands of N_2 fixation; assimilation of the product of N_2 fixation (Michiels and Vanderleyden, 1994). In order to achieve this, plants produce several proteins in nodule-specific or nodule-abundant forms. These include leghemoglobin, sucrose synthase, uricase, glutamine synthase (GS) and PEP carboxylase (PEPc).

1.1.6 Nitrogen fixation

The reduction of atmospheric nitrogen by the bacteroids is catalysed by the enzyme nitrogenase via the following reaction:

$$N_2 + 8\ H^+ + 16\ ATP + 8\ e^- \rightarrow 2\ NH_3 + H_2 + 16\ ADP + 16\ Pi$$

This is a highly energy demanding process consuming large amounts of ATP and reductant. Therefore it is unsurprising that it is a tightly regulated process at both the transcriptional and post-transcriptional level. The enzyme nitrogenase is irreversibly inactivated by oxygen, so nitrogen must be fixed under anaerobic conditions. Low O_2 concentrations are required to induce the nitrogen fixation genes. In free-living diazotrophs (nitrogen-fixing organisms) the availability of a nitrogen

source is also a major regulatory factor (Michiels and Vanderleyden, 1994), to prevent the energy-expensive reaction of N_2 fixation when an alterative source of reduced nitrogen is available. However rhizobia have evolved to provide the plant with NH_3 and the process of N_2 fixation is uncoupled from the control of nitrogen assimilation (Udvardi and Day, 1997).

In the nodule the high-energy demands of nitrogenase are met by oxidative phosphorylation. Therefore the plant must provide the bacteria with a high O_2 flux at a low O_2 concentration. Oxygen transport to the bacteroids is facilitated through the presence of an O_2-carrier protein, leghemoglobin (Bergersen and Turner, 1975). Leghemoglobin is the most abundant protein in N_2-fixing nodules and accounts for up to 25% of the plant cytoplasmic proteins (Appleby, 1984). Another adaptation to the low O_2 environment is the presence of a high- O_2-affinity terminal oxidase in the N_2-fixing bacteroid, which functions at O_2 concentrations of 5-10 nM (Vance and Heichel, 1991).

Nodules have also developed mechanisms to regulate their permeability to O_2 (Witty *et al.*, 1987; Layzell *et al.*, 1990). The free oxygen concentration within infected cells is about 10-60 nM (Kuzma *et al.*, 1993). This low free oxygen concentration results from the combination of a variable O_2 diffusion barrier in the nodule parenchyma, consisting of tightly packed cells with few intercellular air spaces, and high metabolic activity of the rhizobia (Mylona *et al.*, 1995). Expression of plant glycoproteins, which accumulate in the intercellular spaces of the nodule parenchyma, may further influence oxygen diffusion (Brewin, 1991). There are two mechanisms that control the nodule permeability to oxygen: long-term developmental adaptations such as changes in the number of cells in the nodule parenchyma, and short-term physiological mechanisms allowing rapid adjustment to changes in environmental parameters (Kuzma *et al.*, 1993).

Recently, a carbonic anhydrase isoform from *Medicago sativa*, Msca1, was shown to be localised to the nodule parenchyma and an inverse correlation was observed between the external oxygen pressure and Msca1 protein content (Gálvez *et al.*, 2000). Carbonic anhydrase catalyses the hydration of CO_2 to form bicarbonate. It is proposed that Msca1 may play a role in the modulation of the O_2 diffusion barrier.

Many experiments have shown that increasing the external partial pressure of O_2, especially in nodules in which nitrogenase has been inhibited, results in a significant increase in nodule respiration (Hartwig *et al.*, 1987; Vessey *et al.*, 1988;

Hunt *et al.*, 1989; Denison *et al.*, 1992; de Lima *et al.*, 1994). Therefore, it was concluded that legume nodule metabolism is O_2-limited. It has long been assumed that the bacteroids were the site of O_2-limited metabolism. Kuzma *et al.* (1999) investigated the site of oxygen limitation by measuring adenylate pools in soybean nodules either maintained in air or exposed to subambient O_2 levels. Using the adenylate energy charge (AEC = [ATP + 0.5 ADP] / [ATP + ADP + AMP]) value as an indicator of hypoxic metabolism, they found that in nodules in air the plant compartment had a lower value (i.e. more hypoxic) than the bacteroid compartment, which was close to that of fully aerobic organisms. When exposed to 10% O_2, the AEC value of the plant compartment remained unchanged, while that of the bacteroid was significantly decreased. This supports the view of the bacteroids being the primary site of O_2 limitation in nodules. However, cytological evidence has shown that nodule mitochondria are clustered near gas-filled intercellular spaces (Millar *et al.*, 1995), where larger free O_2 gradients are predicted to occur (Thumfort *et al.*, 1994). To resolve the apparent discrepancy between these results and the low AEC values measured for the plant compartment, Kuzma *et al.* (1999) proposed that the low AEC value is a function of the large size of, and mitochondrial localisation in, infected cells. The sinks for ATP within the plant compartment would, in many instances, occur at long distances from the mitochondria clustered near the intercellular gas-filled spaces. Therefore translocation of ATP to these sites, and the translocation of ADP and AMP back to the mitochondria may be the major factor limiting ATP production in these cells. Wei *et al.* (2004a) developed a mathematical model based on this hypothesis of large gradients in adenylate concentrations existing across the cytosol of infected nodule cells. The predictions of this model were experimentally verified (Wei *et al.*, 2004b), with the authors proposing a role for adenylates in regulating the permeability of the nodule to O_2 diffusion.

1.1.7 Nitrogen assimilation

In 1956 Burris showed that NH_4^+ was the product of atmospheric nitrogen fixation by nitrogenase and it was thought to be the sole nitrogen source secreted by the bacteroid because the levels of bacteroid ammonium assimilatory enzymes (GS/GOGAT) are insufficient for the rates of glutamate/glutamine synthesis required for nitrogen fixation by the plant (Lodwig and Poole, 2003). The NH_3 produced by the bacteroids rapidly diffuses down the concentration gradient into the peribacteroid

space (PBS). The PBS has an acidic pH due to the actions of an H^+-ATPase present in the PBM and the bacterial respiratory electron transport chain, which pumps protons out of the bacteroid into the PBS (Udvardi and Day, 1997). This promotes the formation of NH_4^+ from NH_3 and maintains the concentration gradient. Movement of NH_4^+ into the plant cytosol is probably through an ion channel on the PBM, and a putative channel has been identified in pea (Mouritzen and Rosendahl, 1997). The repression of the bacteroid ammonium uptake mechanism prevents competition with the plant for ammonium in the PBS (Lodwig and Poole, 2003).

Recent work by Lodwig *et al.* (2003) however has shown that the metabolic exchange between the plant and rhizobia is more complex than the provision of carbon by the plant in return for fixed nitrogen from the rhizobia. They demonstrated that an amino-acid cycle is essential for symbiotic nitrogen fixation by *Rhizobium* in pea nodules. They propose that, in addition to dicarboxylates, the plant provides the bacteroids with glutamate or a precursor of glutamate (Fig. 1.3). This acts as a transamination donor for the bacteroid to produce aspartate, or another amino acid such as alanine. This is then secreted to the plant in addition to NH_4^+ and is used by the plant for asparagine synthesis. This exchange of amino acids allows the bacteroid to shut down its own ammonium assimilation. In order to obtain its supply of amino acids the bacteroids must supply the plant with ammonium. However, in return the bacteroids act like plant organelles for aspartate synthesis. This dependence on each other favours the evolution of mutualism and demonstrates the level of metabolic integration in the legume-rhizobia symbiosis.

Glutamine is the primary amino compound synthesised from ammonia in the plant cytosol by the coupled activities of GS-GOGAT. These enzymes have very high activities in nodules owing to the expression of nodule-specific isoforms (Lodwig and Poole, (2003) and references therein). The assimilated nitrogen is translocated out of the nodule to other parts of the plant, but the form in which it is transported depends upon the legume species. In legumes with indeterminate nodules the amino acid asparagine is the predominant form of translocated fixed N_2. In legumes with determinate nodules glutamine is metabolised to the ureides allantoin and allantoic acid before it is translocated, via the xylem, to the shoots. Asparagine biosynthesis requires a continued supply of oxaloacetate (OAA) (Vance, 1983), which is supplied by PEPc. In contrast, the synthesis of ureides does not require the same flux of OAA (Coker and Schubert, 1981). Ureide biosynthesis involves both bacteroid infected and

neighbouring uninfected cells. Purines are synthesised in the infected cells using ammonia from the bacteroid. The purine is then oxidised to uric acid, which is translocated to an adjacent uninfected nodule cell. In the peroxisome, urate oxidase converts uric acid to allantoin, which in turn is converted to allantoic acid by allantoinase in the endoplasmic reticulum (Atkins *et al.*, 1992).

1.1.8 Carbon metabolism

Sucrose produced by photosynthesis in the plant shoots is the primary source of carbon supplied to support nodule metabolism and nitrogen fixation. In the nodule, sucrose is cleaved by sucrose synthase to UDP-glucose and fructose. A nodule-enhanced isoform of sucrose synthase has been identified in legume nodules (Thummler and Verma, 1987; Silvente *et al.*, 2003) and its activity is essential for nitrogen fixation (Gordon *et al.*, 1999; Craig *et al.*, 1999). The products of sucrose cleavage are used for cellulose or starch biosynthesis and/or metabolised to phosphoenolpyruvate (PEP) via the glycolytic pathway (Lodwig and Poole, 2003). PEP is further metabolised to malate by the actions of PEPc and malate dehydrogenase (MDH), both of which show nodule-enhanced expression in various legumes (Pathirana *et al.*, 1997; Suganuma *et al.*, 1997; Hata *et al.*, 1998; Miller *et al.*, 1998). Malate, rather than sugars, is considered to be the primary carbon source for bacteroids *in planta*. The evidence for this includes: the observation that malate, succinate and fumarate accumulate to high levels in the nodule (Vance and Heichel, 1991); nodules exposed to $^{14}CO_2$ rapidly incorporate the label into malate in the bacteroids (Vance *et al.*, 1985; Rosendahl *et al.*, 1990); a high-affinity dicarboxylate transport system (Dct-system) has been identified in rhizobia and mutants with a defective Dct-system form ineffective nodules (Michiels and Vanderleyden, 1994); symbiosomes isolated from all legumes can accumulate dicarboxylates rapidly and with high affinity (Udvardi and Day, 1997) and the presence of a dicarboxylate transporter on the PBM has been demonstrated (Udvardi *et al.*, 1988).

1.2 Roles and regulation of PEPc in plants

1.2.1 Photosynthesis

Photosynthesis is the process that allows plants to use light energy to assimilate carbon dioxide (CO_2) into organic sugars. For most plants the primary product of photosynthesis is the 3-carbon compound 3-phosphoglycerate (3-PGA) and

this is therefore termed C_3 photosynthesis. In mesophyll cell chloroplasts CO_2 is fixed by the enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) as part of the Benson-Calvin cycle. However, Rubisco also catalyses the oxygenation of ribulose 1,5-bisphosphate, which leads to the light dependent evolution of CO_2 (photorespiration). The rate of photorespiration is enhanced by increases in temperature, irradiance and the relative levels of O_2 and CO_2. Up to 40% of newly fixed carbon may be re-released as CO_2 in environments which favour photorespiration, such as the high light and temperatures of tropical and subtropical climates (Leegood, 1999). Therefore, some plants evolved mechanisms that concentrate CO_2 at the site of carboxylation.

Most C_4 photosynthetic plants have a specialised leaf anatomy known as 'Kranz' anatomy. The C_4 leaf contains two distinct chloroplast-containing cells: mesophyll and bundle sheath cells. Tightly packed bundle sheath cells are arranged in a ring around each vascular bundle and have thickened cell walls and prominent starch filled chloroplasts. The surrounding mesophyll cells are similar to those of C_3 plants; they have thin cell walls and abundant intercellular spaces. The primary products of photosynthesis are 4-carbon acids, malate or aspartate, produced by the carboxylation of phosphoenolpyruvate (PEP) by the enzyme phosphoenolpyruvate carboxylase. PEPc is localised to the cytosol of the mesophyll cells and is the site of primary CO_2 fixation. The mesophyll cell chloroplasts lack Rubisco to prevent CO_2 fixation via the Benson-Calvin cycle. The malate (or aspartate) produced in the mesophyll cell is transported to the bundle sheath cells via plasmodesmata, where it is decarboxylated. The resulting CO_2 is fixed by Rubisco via the Benson-Calvin cycle in the bundle sheath chloroplasts. This spatial separation of carbon fixation allows CO_2 to be concentrated in the chloroplasts of the bundle sheath cells, where Rubisco is localised, and therefore reduce Rubisco's oxygenase activity.

A second method for concentrating CO_2 at the site of Rubisco is found in Crassulacean acid metabolism (CAM). This form of photosynthesis was first detected in plants belonging to the Crassulaceae but has been identified in a range of angiosperms, in both monocots and dicots. Where C_4 plants have a spatial separation of CO_2 fixation, CAM plants have a temporal separation. The stomata are open during the night to allow dark CO_2 fixation by PEPc. The OAA formed is converted to malate and stored in the vacuole as malic acid. During the day the stomata close and malic acid is released from the vacuole and decarboxylated via malic enzyme or PEP

carboxykinase, depending upon the type of plant. Rubisco then fixes the CO_2 released via the Benson-Calvin cycle. Because both PEPc and Rubisco are present in the same cells their activities are regulated in time, with PEPc active during the dark but not the light period and Rubisco only active in the light. The regulation of PEPc in CAM plants is considered in section 1.2.6.

The adaptive benefit of CAM photosynthesis is the greatly increased water-use efficiency compared to C_3 or C_4 plants. The fixation of CO_2 during the night greatly reduces water loss. CAM plants typically loose 50-100 g of water for every gram of CO_2 fixed, compared to 250-300 g for C_4 plants and 400-500 g for C_3 plants (Taiz, 1998). This makes CAM plants particularly suited to dry environments such as deserts or the epiphytic environment in tropical rainforests. Many plants show facultative CAM, such as *Mesembryanthemum crystallinum*, which will shift from C_3 to CAM photosynthesis under salt or drought stress.

The distribution of C_4 and CAM photosynthesis amongst the angiosperms and the variations in the C_4 and CAM adaptations (e.g. anatomical, biochemical) in different plants strongly imply that they have both evolved independently many times. This is possible because none of the enzymes involved are unique to these plants. The dicot genus *Flaveria* contains both C_3 and C_4 species as well as species termed C_3-C_4 intermediates, as they show characteristics of both. This has been adopted as a model system to study the evolutionary steps required to change from C_3 to C_4 photosynthesis (Svensson *et al.*, 2003).

1.2.2 Structure and catalytic mechanism of PEPc

Phosphoenolpyruvate carboxylase (EC 4.1.1.31, PEPc) catalyses the ß-carboxylation of phosphoenolpyruvate (PEP) using HCO_3^- to yield oxaloacetate (OAA) and inorganic phosphate (Pi) (Andreo *et al.*, 1987). A number of divalent cations can act as cofactors for PEPc in vitro. The most effective cofactor is Mg^{2+}, but Mn^{2+} and Co^{2+} are also active (Andreo *et al.*, 1987), however Ca^{2+} is inhibitory (Andreo *et al.*, 1987). The reaction mechanism has been studied in detail. The preferred order of reactant binding is the bivalent cation, then PEP and finally HCO_3^- (Chollet *et al.*, 1996). The reaction is highly exergonic, with a $\Delta G°$ of -30 kJ mol^{-1} (Kai *et al.*, 2003), and the overall reaction is strictly irreversible (Izui *et al.*, 2004). The current reaction model shows the reaction occurring in three steps (see Fig. 1.4). In step 1, the enolate anion of pyruvate and carboxyphosphate are formed from PEP

and bicarbonate by phosphate transfer. In step 2, a conformational change positions the carbonyl carbon of the carboxyphosphate intermediate adjacent to the carbon-3 of the enolate and carboxyphosphate is cleaved into carbon dioxide and Pi inside the catalytic site. In step 3 the carbon dioxide makes an electrophilic attack on the enolate anion to form OAA and Pi is liberated.

PEPc is present in all photosynthetic organisms, including higher plants, green algae, cyanobacteria, and most nonphotosynthetic bacteria and protozoa, but is absent form animals, fungi and yeasts. Recently, the three-dimensional structures of PEPc from the bacterium *Escherichia coli* and the C_4 plant maize (*Zea mays*) have been published (Kai *et al.*, 1999; Matsumura *et al.*, 1999; Matsumura *et al.*, 2002, reviewed in Kai *et al.*, 2003; Izui *et al.*, 2004). The *E. coli* PEPc was bound to the inhibitor aspartate (Asp) and therefore the enzyme was thought to be in the inactive state (T state). The maize PEPc was bound with a sulphate anion and the crystallisation was performed in the presence of ethylene glycol, a strongly activating compatible solute, and therefore it was considered to be in the active state (R state). For *E. coli* PEPc, the structures of the complex with Mn^{2+} and Asp, and the complex with Mn^{2+}-DCDP (3,3-dichloro-2-dihydroxyphosphinoylmethyl-2-propenoate) and Asp were also determined (Matsumura *et al.*, 1999; Matsumura *et al.*, 2002). DCDP is a PEP analogue and strongly inhibits PEPc. The basic structural features are well conserved between *E. coli* and maize. PEPc consists of four identical subunits arranged as a 'dimer-of-dimers.' The maize PEPc monomer consists of an eight-stranded β barrel and 42 α helices. The α helices comprise 65% of the polypeptide, whereas the β strands comprise only 5%. Two mobile loops (Loop I and Loop II) were identified that are conserved between maize and *E. coli*. Two additional loops appear to be plant-specific.

The residues R587 (647), K773 (835), R832 (894) and N881 (968) participated directly in binding of the allosteric inhibitor Asp, where the amino acid residues are given in the *E. coli* numbering, with maize numbering in brackets. R587 (647) has been shown to be catalytically essential by site-directed mutagenesis (Yano *et al.*, 1995) and is located in the unique PEPc sequence GRGGXXGR^{587}GG (XX= TV, SI, or SV), which comprises Loop I. Comparison of the maize PEPc structure (in the R state) and the *E. coli* structure (in the T state) show that R587 (647) moves 15 Å toward the active site due to its release from direct constraint by the inhibitor

molecule Asp. This 'trapping' of R587 (647) could explain the mode of action of the allosteric inhibitor. Similarly, the essential residue H138 (177) (Terada and Izui, 1991), which is part of a short loop, is located next to the active site in the R state whereas it is 10 Å away in the T state. This suggests that R587 (647) and H138 (177) directly participate in the carboxylation reaction (Kai *et al.*, 2003).

Loop II comprises the residues 701-708 (761-768) and K702 (762), R703 (763) and R704 (764) are thought to involved in the binding of bicarbonate. Furthermore, site-directed mutagenesis studies suggest that Loop II also plays the role of a lid covering the catalytic cavity to protect catalytic intermediates from attack by surrounding water (Izui *et al.*, 2004). The roles of the additional plant-specific loops are unknown, but may be related to recognition of PEPc by the plant-specific PEPc kinase (see below and section 1.3).

The catalytic site is located at the C-terminal region of the β barrel, as revealed by the structure of *E. coli* complexed with Mn^{2+}-DCDP and Asp (Matsumura *et al.*, 2002). Mn^{2+} binds to the oxygen atoms of the carboxyl groups of conserved E506 (566) and D543 (603) at the C-terminal end of the β-5 and β-6 strands. DCDP interacted with Mn^{2+}, R396 (456), R699 (759) and R713 (773), which have been shown to be catalytically essential by site-directed mutagenesis (Kai *et al.*, 2003). A model of PEP binding revealed a hydrophobic pocket consisting of W248 (288), L504 (564), and M538 (598) around the methylene group of PEP and this is assumed to be the space where free CO_2 is fixed onto the enolate form of pyruvate during step 3 of the PEPc reaction.

In the maize PEPc structure the sulphate ion was surrounded by four positively charged residues: R183, R184, R231 and R372 from the adjacent subunit. This site is hypothesised to be the binding site for the allosteric activator glucose-6-phosphate (Glc-6-P). The binding of the sulphate ion resulted in the movement of several loops near the active site and the binding site of Asp. These conformational changes could relate to the allosteric activation of PEPc by Glc-6-P. Replacement of each of the arganine residues with glutamine caused desensitisation of PEPc to Glc-6-P (Izui *et al.*, 2004).

In plants, PEPc is subject to regulation via reversible phosphorylation (Chollet *et al.*, 1996; Nimmo, 2003) in addition to allosteric regulation. Alignments of plant PEPc sequences reveal the conserved motif, (E/D)(R/K)XXSIDAQ(L/M)R, around

the phosphorylatable serine residue (indicated in bold). A model for the mechanism of activation by phosphorylation suggests that the negatively charged phosphate group may compete for binding at the same site with the inhibitor Asp (Kai *et al.*, 1999). Unfortunately, the crystallised maize PEPc was truncated and lacked this region. However this model does not explain why a modified PEPc, whose N-terminal region is truncated by proteolysis shows a desensitisation to Asp (Izui *et al.*, 1986). The recent observation that a Glc-6-P insensitive mutant of PEPc was also phosphorylation insensitive (Izui *et al.*, 2004) may implicate the arginine residues R183, R184, R231 and R372 in activation by phosphorylation.

1.2.3 PEPc in higher plants

PEPc performs a wide variety of functions in higher plants. It has a very low Km for its substrate bicarbonate (in the micromolar range) and so has a general role in the carbon economy of the plant by recapturing respiratory CO_2 (Vidal and Chollet, 1997). Initially, most research was focused on the photosynthetic PEPc isoforms of C_4 and CAM plants, where it is the enzyme responsible for the primary fixation of inorganic carbon. It also plays a general housekeeping function by replenishing tricarboxylic acid cycle (TCA) intermediates withdrawn for biosynthesis of amino acids. This suggests that PEPc plays an important role coordinating carbon and nitrogen metabolism in plants. A number of specific functions have also been proposed for non-photosynthetic PEPc isoforms. These include roles in fruit ripening, seed maturation, stomatal guard cell movement and nitrogen fixation in legume nodules.

1.2.4 Transcriptional control of PEPc

Transcriptional regulation of PEPc has been shown in C_3, C_4 and CAM plants. In the leaves of the C_4 plants maize and *Sorghum*, light increases the transcript levels of the photosynthetic PEPc isoform in the mesophyll cells (Thomas *et al.*, 1990). The promoter sequence of the maize photosynthetic PEPc isoform is sufficient to confer this light dependent and cell specific expression (Kausch *et al.*, 2001). The necessary factors for this expression pattern also exist in C_3 plants, as the maize promoter also directs cell specific and light-regulated expression in rice (Matsuoka *et al.*, 1994).

Gel shift assays using nuclear extracts from mesophyll or bundle sheath cells have revealed cell specific DNA-protein interactions in maize leaves using fragments

of the photosynthetic PEPc promoter (Taniguchi *et al.*, 2000). The nuclear factors PEP_{Ib} and PEP_{Ic} were shown to be specific to mesophyll cells whereas PEP_{Ia} and PEP_{IIa} are specific to bundle sheath cells (Taniguchi *et al.*, 2000). The binding activities of PEP_{Ia}, PEP_{Ib} and PEP_{Ic} were highest in etiolated leaves and decreased during greening. In contrast PEP_{IIa} was increased during leaf greening. The binding sites and binding proteins involved in these interactions have not yet been identified. It has also been shown that Dof proteins may be involved in maize PEPc transcription (Yanagisawa and Sheen, 1998; Yanagisawa, 2000). Dof (for DNA binding with one finger) proteins are a family of zinc-finger DNA binding proteins found in plants. In maize, Dof1 and Dof2 can both bind to AAAG motifs found in the promoter of the photosynthetic PEPc (Yanagisawa and Sheen, 1998). Dof1 was shown to be a transcriptional activator whose binding is light dependent. Dof2 is proposed to act as a transcriptional repressor, as it competes with Dof1 for binding but does not activate transcription. They show different expression patterns in maize, with Dof1 constitutively expressed in leaves, stems and roots, whereas Dof2 is expressed in stems and roots. Therefore their combined action may give rise to tissue specific and light dependent gene expression patterns.

The interaction of DNA binding proteins with naked DNA does not reflect the environment *in vivo*. Transcription factors must recognise their binding sites in the chromatin context in the cell nucleus (Cavalar *et al.*, 2003). The basic unit of chromatin is the nucleosome, which consists of two molecules each of the core histones H2A, H2B, H3 and H4 with approximately 160 bp of DNA wound around the histone octamer. Recent experiments with reconstituted mononucleosomes and recombinant Dof1 and Dof2 transcription factors show they can bind to their target site on top of a nucleosome and that the position of the binding site on the nucleosome surface strongly affects binding efficiency (Cavalar *et al.*, 2003). There is also evidence of chromatin reorganisation involved in the light dependent increase in PEPc transcription in maize (Kalamajka *et al.*, 2003). Light-induced chromatin reorganisation might facilitate access of transcription factors to their binding sites.

In legume root nodules transcripts of a specific PEPc isoform accumulate during nodule development and their abundance is modulated by light (Pathirana *et al.*, 1997; Nakagawa *et al.*, 2003a). In *Glycine max* the promoter of the PEPc isoform predominantly expressed in nodules was compared to the promoter of a closely related PEPc gene which displays a general low-level expression pattern (Nakagawa

et al., 2003b). It was shown that two promoter regions are required for high expression levels in nodules. Nuclear factors binding specifically to one of the regions was demonstrated but they were not identified.

In the CAM plant *Mesembryanthemum crystallinum* the shift from C_3 to CAM photosynthesis occurs in response to salinity or drought stress. The expression of the *M. crystallinum* photosynthetic PEPc isoform (*McPpc1*) is also increased in response to salt stress (Cushman *et al.*, 1989). A DNA-binding factor (PCAT-1) binding to two AT-rich regions in the *McPpc1* promoter was identified and shown to be more abundant in extracts from salt-stressed plants (Cushman and Bohnert, 1992). Promoter deletion analysis of *McPpc1* has revealed several salt-responsive enhancer regions and a silencer region (Schaeffer *et al.*, 1995).

The above examples show the complexity of control of PEPc transcript abundance, at least in some plants. However a clear understanding of the signal(s) that modulates PEPc transcription has not been achieved. In recent years research has focused on posttranscriptional regulation of PEPc, particularly via reversible phosphorylation.

1.2.5 Posttranslational control of PEPc

PEPc is subject to allosteric regulation by various metabolites at physiologically relevant levels (reviewed in Rajagopalan *et al.*, 1994). Malate is a potent feedback inhibitor and glucose 6-phosphate (Glc-6-P) an allosteric activator. The sensitivity of PEPc to these effectors depends upon the assay pH and the phosphorylation status of a single serine residue at the N-terminus of the protein. At pH 8 the enzyme is very active and the effects of malate and Glc-6-P are significantly diminished (Schuller *et al.*, 1990). However, when assayed at suboptimal pH values that approximate those of the cytosol the enzyme is less active and more sensitive to the effectors (Chollet *et al.*, 1996). The Ki for malate of the recombinant photosynthetic PEPc from *Sorghum* is decreased 25-fold at pH 7.3 compared to pH 8.0 (Duff *et al.*, 1995). The effects of phosphorylation were also investigated using the recombinant *Sorghum* PEPc (Duff *et al.*, 1995). The Ki (L-malate) was 0.17 mM and 1.2 mM for the dephosphorylated and phosphorylated forms respectively at pH 7.3, while the Ka (Glc-6-P) was 1.3 mM and 0.28 mM respectively.

Several other metabolic effectors have been proposed to regulate PEPc activity (Rajagopalan *et al.*, 1994). Other phosphate containing compounds activate PEPc,

such as fructose 2,6-bisphosphate and dihydroxyacetone phosphate (Doncaster and Leegood, 1987). The amino acids aspartate and glutamate inhibit PEPc (O'Leary, 1982; Schuller *et al.*, 1990; Blonde and Plaxton, 2003), as do flavanoids and shikimic acid (Colombo *et al.*, 1996; Pairoba *et al.*, 1996; Moraes and Plaxton, 2000). Plants may divert PEP into the shikimic acid pathway for subsequent secondary metabolite (flavanoid) biosynthesis under conditions of excess products of glycolysis (Colombo *et al.*, 1996).

1.2.6 Roles and regulation of PEPc in CAM plants

CAM photosynthesis is characterised by a temporal separation of its two phases of CO_2 fixation. This requires that flux through PEPc only occurs during the night and not during the day. When the detached leaves of CAM plants are kept in constant environmental conditions, persistent rhythms of CO_2 exchange are observed (Wilkins, 1992). This is achieved by flux through PEPc being controlled by a circadian oscillator. It was observed that, in the constitutive CAM plant *Kalanchoë fedtschenkoi* adapted to an 8 h photoperiod, PEPc isolated from leaves during the light period was tenfold more sensitive to malate inhibition than PEPc isolated from leaves during the dark period (Nimmo *et al.*, 1984). Subsequent purification of PEPc from leaves during the day and night revealed that the night form of the enzyme is phosphorylated and the day form is dephosphorylated (Nimmo *et al.*, 1986). It is now know that this is due to the phosphorylation of a single serine residue located at the N-terminus (Chollet *et al.*, 1996). Changes in the phosphorylation status of PEPc were also observed in detached leaves of *K. fedtschenkoi* maintained in constant environmental conditions, showing that phosphorylation is controlled by a circadian rhythm as opposed to light-dark transitions (Nimmo *et al.*, 1987b). Experiments using the constitutive CAM plant *Kalanchoë daigremontiana* and the C_3-CAM intermediate *Clusia minor* allowed estimates of the *in vivo* flux through PEPc to be calculated (Borland and Griffiths, 1997). This showed a good correlation between estimated flux and the phosphorylation state of PEPc, as determined by malate sensitivity.

1.2.7 Roles and regulation of PEPc in C_4 plants

PEPc has been studied in a range of C_4 photosynthesising plants including the agriculturally important monocot crop plant maize and its close relative *Sorghum* (Chollet *et al.*, 1996; Vidal and Chollet, 1997), members of the dicot genus *Flaveria*,

as a model for C_4 evolution (reviewed in Svensson *et al.*, 2003), and the aquatic plant *Hydrilla verticillata*, which is a facultative C_4 plant that does not require Kranz anatomy to perform C_4 photosynthesis (Rao *et al.*, 2002). One of the underlying goals of such research is to introduce components of the C_4 cycle into C_3 crop plants to improve photosynthetic and water use efficiencies (Hausler *et al.*, 2002; Leegood, 2002; Miyao, 2003).

As in other plants, PEPc in C_4 plants is encoded by a small gene family (Lepiniec *et al.*, 1994; Chollet *et al.*, 1996). In *Sorghum* three PEPc genes have been identified, encoding a housekeeping isoform, a root isoform and a C_4-photosynthetic isoform (Lepiniec *et al.*, 1993). It was demonstrated that PEPc from maize leaves was phosphorylated *in vitro* when supplied with ATP (Budde and Chollet, 1986), but the kinetic consequences and physiological occurance of phosphorylation were not assessed. In contrast, *in vivo* labelling experiments with $^{32}P_i$ fed to detached maize leaves showed that PEPc was phosphorylated in illuminated leaves and dephosphorylated when extracted from leaves kept in the dark (Nimmo *et al.*, 1987a). Phosphorylation was associated with an increase in Ki for malate, 1.2 mM from illuminated leaves compared with 0.5 mM from darkened leaves (Nimmo *et al.*, 1987a).

The *in vivo* phosphorylation site was identified via Edman sequencing of *in vivo* phosphorylated PEPc from maize and *Sorghum* (Jiao *et al.*, 1991). A single N-terminal serine residue (Ser-15 in maize and Ser-8 in *Sorghum*) was the target of phosphorylation. Alignments of plant PEPc sequences show that the sequence around the phosphorylatable serine residue is highly conserved in plant PEPcs but absent from bacterial PEPcs.

1.2.8 Roles and regulation of PEPc in C_3 plants

The isoform(s) which perform the anaplerotic functions of PEPc are termed 'C_3' isoforms to differentiate them from those isoforms involved in C_4 and CAM photosynthesis. They are not exclusive to C_3 plants, but much of the work to elucidate the roles and regulation of non-photosynthetic PEPc isoforms has been carried out in C_3 plants. The results of these investigations have highlighted the role PEPc plays in connecting carbon and nitrogen metabolism.

The transcript levels and enzyme activities of key enzymes involved in nitrate and organic acid metabolism varied diurnally in the leaves of nitrogen-sufficient

tobacco (*Nicotiana tabacum*) plants (Scheible *et al.*, 2004). PEPc transcripts were maximal at the start of the light period and decreased to a minimum at the end of the light period. PEPc activity follows the changes in transcript abundance with a time lag of approximately 4 h. This is the same pattern of changes in transcripts and activities as measured for both nitrate reductase (NR) and nitrite reductase (NIR). All three enzymes were similarly affected by changes in nitrogen metabolism. The transcript abundances and enzyme activities were all very low in nitrate-limited tobacco plants, but greatly increased in transgenic tobacco plants with very low NR activity. In contrast to this pattern, the transcript levels and enzyme activities of cytosolic pyruvate kinase (PK), mitochondrial citrate synthase (CS) and NADP-isocitrate dehydrogenase (ICDH) were maximal at the end of the light period and first part of the dark period. The reciprocal changes in expression are reflected in changes in metabolite levels. During the first part of the light period, when PEPc and NR activity is maximal, malate and ammonium accumulate and citrate decreases, whereas in the later part of the light period and during the dark period, when PK, CS and ICDH activities are maximal, citrate levels increase while malate and ammonium decrease.

The documented changes in enzyme activities and metabolite levels led Scheible *et al.* (2000) to propose the following model. The assimilation of nitrate into ammonia causes alkalinisation, as the reactions catalysed by NR and NIR consume protons. PEPc and malate dehydrogenase provide the malate to act as a counter-ion to regulate pH levels. During the first part of the light period nitrate assimilation is maximal and therefore synthesis of malate is favoured. Later in the light period, nitrate assimilation rates decrease and now the production of 2-oxoglutarate is preferred to the production of malate, in order to support the metabolism of ammonia and glutamine.

The coordinated regulation of PEPc and PK has also been suggested to occur in oilseed rape (*Brassica napus*) cell culture during nitrogen assimilation (Smith *et al.*, 2000). PEPc and PK were purified from *B. napus* cells and the effects of various metabolites on enzyme activities were investigated (Moraes and Plaxton, 2000; Smith *et al.*, 2000). Similar to results obtained for other plant PEPcs, *B. napus* PEPc was activated by hexose-6-phosphates and inhibited by malate. Aspartate, isocitrate and glutamate also inhibited PEPc activity. PK is also inhibited by glutamate, but unlike PEPc it is activated by aspartate. The effect of aspartate is to effectively relieve the inhibition of PK by glutamate. They propose a model for the allosteric regulation of

PK and PEPc with the reciprocal control exerted by aspartate playing a key role. Aspartate is produced via the reactions of PEPc and aspartate aminotransferase (AAT). Increased levels of aspartate would signal sufficient nitrogen levels and protein synthesis might become more dependent upon ATP availability rather than amino acid supply. The inhibition of PEPc and synchronised activation of PK reduce the flux from PEP to aspartate and increase flux to support mitochondrial respiration.

The sequences of many C_3 PEPc isoforms have now been determined and the presence of the conserved phosphorylation site at the N-terminus suggests that they are also subject to regulation via reversible phosphorylation, as the photosynthetic PEPc isoforms are. *In vivo* phosphorylation of PEPc was demonstrated in detached wheat leaves supplied with $^{32}P_i$ (Van Quy *et al.*, 1991). Labelling of PEPc was higher in illuminated detached leaves compared to leaves kept in the dark. Enhanced PEPc labelling was also seen with leaves taken from nitrate-sufficient plants compared to nitrate-limited plants. Phosphorylation was associated with a decrease in malate inhibition, with I_{50}(malate) values of 1.45 mM from dark control leaves, 1.75 mM from illuminated low-nitrate leaves and 3.4 mM from illuminated high-nitrate leaves. Light activation of PEPc has also been reported in tobacco leaves and barley protoplasts (Li *et al.*, 1996; Smith *et al.*, 1996). However, the importance of phosphorylation in regulating C_3 PEPc isoforms has been questioned (Gupta *et al.*, 1994). In the leaves and roots of pea plants diurnal variation in the rate of dark CO_2 fixation (a measure of PEP carboxylation) was observed *in vivo* (Leport *et al.*, 1996). Rates of dark CO_2 fixation were high during the light period and low during the dark period. The diurnal modulations were not maintained under constant conditions, excluding circadian control. PEPc protein levels and activity showed no concurrent changes. This lead them to conclude that allosteric regulation of PEPc by changes in metabolite levels could account for the modulation of PEP carboxylation and that protein phosphorylation does not play a major role in regulation of PEPc in C_3 plants.

Apart from the anaplerotic roles mentioned above, PEPc has also been hypothesised to play specific roles in certain cells, tissue or organs and in response to certain environmental conditions. Some of these roles are considered below.

1.2.8 Phosphate Deficiency

Phosphorous is second only to nitrogen as the most limiting element for plant growth (Vance, 2001). Much of the phosphate in soils is unavailable for uptake as it

rapidly forms insoluble compounds with cations. This is particularly true of the acid-weathered soils of the tropical and subtropical regions, where phosphorous is considered the most limiting nutrient for the growth of leguminous crops (Raghothama, 1999; Vance and Lamb, 2001).

Plants have developed several responses for growth in phosphate-limited environments (Raghothama, 1999). These include increased root hair proliferation, proteoid root formation and association with mycorrhizal fungi, modification of metabolism to bypass phosphate-requiring steps and secretion of organic acids into the rhizosphere to enhance phosphate solubilisation. Increased activity of PEPc in response to phosphate starvation has been found in many plants. PEPc is thought to be involved in the production of malate and citrate for secretion by the roots and as part of the mechanism of an alternative glycolytic pathway (Plaxton, 1996). Pyruvate kinase requires Pi and ADP, but PEPc, malate dehydrogenase (MDH) and malic enzyme can bypass pyruvate kinase. Similarly, the PPi-dependent phosphofructokinase (PFP) can bypass the ATP-dependent phosphofructokinase (PFK).

Toyota *et al.* (2003) investigated PEPc activation by phosphorous deficiency in tobacco. Tobacco plants cultured for 8 days without phosphorous showed an increase in PEPc activity. The inclusion of 2% sucrose in the media produced a synergistic response with phosphate deficiency. Three PEPc genes have been identified in tobacco (Koizumi *et al.*, 2002) but only one (*Nsppc1*) was predominantly expressed in the 8 day old seedlings. Promoter-reporter constructs using 2kb of the *Nsppc1* promoter region showed increased expression under phosphate starvation and a much greater induction when sucrose was included. This suggests that increased *Nsppc1* transcription accounts for the observed increases in PEPc activity. Analysis of the *Nsppc1* promoter by deletion studies and gain-of-function experiments suggest that the promoter sequence between –539 and –442 is necessary and sufficient for the phosphate-deficiency response.

White lupin (*Lupinus albus*) is a leguminous plant that has an extreme tolerance for low phosphate levels. It develops proteoid roots, which are clusters of rootlets covered in a mat of root hairs and are specialised in the synthesis and secretion of organic acids (Raghothama, 1999). A PEPc gene was cloned from white lupin (*LaPEPC1*) that shows increased expression levels in proteoid roots under phosphate deficiency (Uhde-Stone *et al.*, 2003). *LaPEPC1* transcripts were higher in

mature proteoid roots than in recently emerged proteoid roots, which correlates with the timing of organic acid exudation. A macroarray analysis of 1250 ESTs from phosphate-deficient proteoid roots found that many genes in the glycolytic pathway were induced by phosphate-deficiency (Uhde-Stone *et al.*, 2003). These included MDH and PFP, suggesting a role for the phosphate-conserving alternative metabolic pathways in phosphate-deficient proteoid roots of white lupin.

1.2.10 Guard cells

Stomata are composed of pairs of specialised epidermal cells called guard cells, which surround the stomatal pore. Changes in guard cell turgor pressure regulate the aperture of the stomatal pore. Environmental factors such as light intensity and quality, temperature, relative humidity and intercellular CO_2 concentrations are sensed by guard cells (Taiz, 1998). Stomatal opening is initiated by the activation of a proton pump, H^+-ATPase, which hyperpolarizes the plasma membrane. This provides the driving force for passive uptake of solutes, primarily potassium ions (K^+), which increases the osmotic potential of the guard cells. This leads to osmotic water uptake, the guard cells swell and due to their cell wall structure the aperture of the stomatal pore increases.

The changes in the K^+ concentration are electrically balanced by chloride and malate anions. The malate is synthesised in the cytosol by the actions of PEPc and malate dehydrogenase. Thus PEPc in guard cells is a key enzyme for malate accumulation during stomatal opening (Cotelle *et al.*, 1999). The guard cells of open stomata in *Vicia faba* contain approximately 5-fold as much malate as those of closed stomata (Outlaw and Lowry, 1977). A cDNA encoding PEPc has been isolated from the guard cells of *Solanum tuberosum* (Müller-Röber *et al.*, 1998). This cDNA hybridised to two different sized mRNAs that were derived from the same gene. The smaller transcripts were exclusively localised to the guard cells, while the larger transcripts were found in all tissues. However, a promoter-GUS reporter construct of this gene showed GUS staining only in the guard cells.

Fusicoccin is a fungal toxin which induces stomatal opening. It activates the H^+-ATPases of the guard cells and stimulates the accumulation of malate (Du *et al.*, 1997). Abscisic acid (ABA), an antagonist of fusicoccin, causes stomatal closure and induces malate efflux (Du *et al.*, 1997). Du *et al.* (1997) examined the changes in phosphorylation status of guard cell PEPc in *Vicia faba* during fusicoccin-induced

stomatal opening and ABA-induced stomatal closure. PEPc was more phosphorylated when stomata were open and stomatal closure reduced PEPc phosphorylation. Further work by Outlaw *et al.* (2002), investigating the signalling pathway that leads to PEPc phosphorylation in fusicoccin-treated guard cells, showed that activation of the H^+-ATPase was essential, but not sufficient, to cause guard cell PEPc phosphorylation. They found a correlation between malate accumulation and PEPc phosphorylation but not necessarily a correlation between stomatal opening and PEPc phosphorylation.

1.2.11 Fruit ripening

The ripening process of fleshy fruits makes the fruit attractive and palatable to seed-dispersing organisms. Before ripening the accumulation of high levels of organic acids, such as malate and citrate, makes the fruit unpalatable. During fruit ripening the organic acid levels decline and soluble carbohydrates accumulate. Therefore, fruit development involves large changes in the composition of sugars and organic acids. Other processes that occur during fruit ripening include: the enzymatic breakdown of cell walls, starch hydrolysis, alterations in pigment biosynthesis and accumulation, and increased levels of flavour and aromatic volatiles (Giovannoni, 2001).

Due to its importance as a crop species, tomato (*Lycopersicon esculentum*) has long been the primary model for climacteric fruit ripening. In tomato, the accumulation of organic acids commences at the end of the cell division stage of fruit development and the levels peak at the end of cellular expansion, reaching up to 75 mM (Guillet *et al.*, 2002). PEPc may play an important role in the synthesis of these organic acids.

PEPc activity has been demonstrated in various fruits, including: banana (Law and Plaxton, 1995), avocado (Notton and Blanke, 1993), peach (Moing *et al.*, 2000), grape (Or *et al.*, 2000), cherimoya (Munoz *et al.*, 2001), as well as tomato (Guillet *et al.*, 2002). Recently, Guillet *et al.* (2002) reported the isolation of two PEPc cDNA sequences from tomato, termed *LePpc1* and *LePpc2*. The expression patterns of both genes were determined and *LePpc1* transcripts were detected at various levels in all parts of the plant, indicating a possible housekeeping function. However, *LePpc2* transcripts accumulated strongly and specifically in the fruit tissues (gel and pericarp) but not other tissues. Transcript levels of *LePpc2* were maximal at 8 days post anthesis (DPA) and remained high until 20 DPA, which correlates with the cell expansion phase of fruit development and maximal growth rates. PEPc protein and

activity levels closely followed *LePpc2* transcript levels. *In situ* localisation of *LePpc2* in 8 DPA fruits showed an association between cell expansion and *LePpc2* transcripts. *LePpc2* transcripts were localised to the cells that show the greatest levels of expansion. This led the authors to suggest a possible role for LePpc2 in the synthesis of organic acids to provide the turgor pressure necessary for cell expansion.

1.2.12 Seed development and germination

The role of PEPc in seeds has been examined in both monocotyledonous and dicotyledonous plants. In developing and germinating wheat grains two PEPc polypeptides were detected with molecular masses of 103 and 108 kDa (Gonzalez *et al.*, 1998). The 103 kDa form was the most abundant and was present throughout grain development and in the scutellum and aleurone layer of germinating grains. The 108 kDa form progressively declines from 15 DPA until it is barely detectable by 31 DPA. However, it was also present in the scutellum and aleurone layer of germinating grains. Both isoforms were found in tissues with high metabolic activity. PEPc may play an anaplerotic role by re-fixing CO_2 released through respiration and supplying carbon skeletons for amino acid biosynthesis. During grain development the starch endosperm accumulates malate, which may represent a further role for PEPc. PEPc was also detected at high amounts in the vascular tissue of developing grains.

Two PEPc cDNA clones have been isolated from a *Vicia faba* seed library, termed *VfPEPCase1* and *VfPEPCase2* (Golombek *et al.*, 1999). They show different expression patterns in various *Vicia faba* organs but are both expressed in cotyledons. The peak of expression occurs during the early to mid cotyledon stage (around 22 to 28 days after pollination (DAP)), just before storage protein accumulation, and then declines. However, PEPc activity during cotyledon development does not follow transcript abundance as it was shown to increase from 15 to 40 DAP. Similarly, PEPc from 36 DAP cotyledons was less sensitive to inhibition by malate than PEPc from 16 DAP cotyledons, suggesting that the increase in activity may be due to an increase in PEPc phosphorylation state. The increase in PEPc activity occurs at the time of storage protein synthesis and therefore may indicate a role for PEPc in shifting the carbon flux to amino acid synthesis.

The transcripts of *VfPEPCase1*, but not *VfPEPCase2*, were modulated by nitrogen starvation. Under nitrogen starvation *VfPEPCase1* mRNA levels decreased, but only in the presence of glucose and fructose. Nitrogen starvation treatments in the

presence of sucrose did not affect *VfPEPCase1* transcript levels. PEPc activity levels followed this same pattern. It is suggested that sucrose overrides the effect of nitrogen starvation as sucrose is thought to induce storage activity in sink tissues.

PEPc has also been investigated in the oil accumulating seeds of castor bean (*Ricinus communis*) (Blonde and Plaxton, 2003). Two immunologically distinct PEPc isoforms were detected, PEPC1 and PEPC2. While PEPC1 was a typical homotetramer, PEPC2 was heteromeric, consisting of the same subunit found in PEPC1 together with an immunologically unrelated subunit. The composition of PEPC2 is considered further in the section below on algal PEPc. Protein levels of PEPC1 increased during castoroil seed maturation, similar to the trend of storage protein accumulation. PEPC1 is proposed to fulfil an anaplerotic role by providing carbon skeletons to support storage protein synthesis. The developmental profile for PEPC2 was different to PEPC1; it peaked earlier during seed maturation and then rapidly declined so as not to be detectable at the maturation phase, when PEPC1 protein levels are maximal. This mimics the accumulation of triglycerides (storage lipids) in castor oilseeds. Triglyceride accumulation requires synthesis of long chain fatty acids, which occurs in specialised plastids called leucoplasts. Malate is the carbon source for fatty acid synthesis and therefore PEPC2 may support leucoplast fatty acid synthesis. The metabolite regulation of PEPC1 and 2 support these hypotheses. PEPC1 is similar to other plant PEPcs; it is activated by hexose mono-phosphates (Glc-6-P, Fru-6-P) and inhibited by malate, aspartate and glutamate. The inhibition via amino acids may act as feedback control. In contrast, PEPC2 is much less affected by all the metabolites tested and also shows a broader pH range. The reduced inhibition by malate may allow PEPC2 to function in the presence of significant malate levels.

1.2.13 Algal PEPc

In contrast to the attention given to PEPc in vascular plants, relatively little work has been concentrated on the enzyme in unicellular green algae. Investigations into algal PEPc have been restricted to two members of the Chlorophyceae, *Chlamydomonas reinhardtii* and *Selenastrum minutum* (Rivoal *et al.*, 1996; Rivoal *et al.*, 1998; Rivoal *et al.*, 2001; Rivoal *et al.*, 2002). *Selenastrum minutum* has been used as a model to study the interactions between carbon and nitrogen metabolism in plants (Huppe and Turpin, 1994). Resupply of NH_4^+ to nitrogen starved cells leads to

a large increase in the flux of CO_2 through PEPc (Schuller *et al.*, 1990), which allows the rapid synthesis of amino acids via GS-GOGAT (Turpin *et al.*, 1990).

Two distinct classes of PEPc have been purified from *C. reinhardtii* and *S. minutum* (Rivoal *et al.*, 1996; Rivoal *et al.*, 1998; Rivoal *et al.*, 2001). Class 1 PEPc (PEPC1) is a typical homotetramer composed of 100 kDa or 102 kDa (p102) subunits in *C. reinhardtii* and *S. minutum* respectively. Class 2 PEPcs consist of large protein complexes containing the PEPc1 monomer and other distinct PEPc catalytic polypeptides, and represent the bulk of PEPc activity extractable from the algae. In *C. reinhardtii* a single isoform of the Class 2 PEPc was isolated (PEPc2) (Rivoal *et al.*, 1998), whereas in *S. minutum* there are three isoforms (PEPc2, PEPc3 and PEPc4) of the Class 2 PEPcs (Rivoal *et al.*, 1996).

Initial reports of the purification of PEPc2, 3 and 4 from *S. minutum* revealed native molecular weights of 984, 1186 and 1590 kDa respectively (Rivoal *et al.*, 1996). The isoforms contained the p102 polypeptide of PEPc1 and various amounts of immunologically unrelated polypeptides of 73, 70, 65 and 61 kDa. However, later purification using more protease inhibitors showed that the Class 2 PEPcs contained a 130 kDa polypeptide (p130) as well as 73 and 65 kDa polypeptides (Rivoal *et al.*, 2001). As the size of the Class 2 PEPcs increased, p130 became more abundant and p102 less abundant. During purification p130 and p102 coeluted and efforts to separate them failed. Even though they could not be assayed individually the data suggested that both subunits had PEPc activity and that p102 has a higher K_m for PEP (1.23 mM) than p130 (0.13 mM). Experiments investigating the allosteric regulation of PEPc1 showed that it was activated by glutamine and inhibited by glutamate, 2-oxoglutarate, aspartate and malate (Rivoal *et al.*, 1996). These properties suggest that PEPc1 is the isoform responsible for the rapid increase in PEPc activity upon nitrogen supply to nitrogen-starved cells. Rivoal *et al.* (2001) proposed that the p130 PEPc plays a housekeeping role when nitrogen is limited. The co-expression of both p102 and p130 may allow the algae to react quickly to nutrient changes in the environment.

Rivoal *et al.* (2002) showed that in *S. minutum* a Ca^{2+}-independent protein kinase with molecular weight of 40 kDa co-purifies with PEPc1 but not PEPc2. However, the p130 subunit of PEPc2 was phosphorylated by a protein kinase present in clarified cell extract that modulated the structural organisation and activity of PEPc2. Dephosphorylation of PEPc2 by protein phosphatase 2A was associated with an increase in the native molecular weight of PEPc2 and an increase in activity. N-

terminal sequencing of the 100 kDa PEPc polypeptide of PEPc2 from *C. reinhardtii* and of p130 from *S. minutum* showed no sequence similarity to the conserved phosphorylatable region of plant PEPcs (Rivoal *et al.*, 1998; Rivoal *et al.*, 2001). Recently the cDNA sequences of two PEPcs from *C. reinhardtii* were deposited in Genbank (Accession numbers AY517644 and AY517643). *C. reinhardtii PPC1* encodes a protein of 974 amino acids in length with a predicted mass of 109 kDa, while *C. reinhardtii PPC2* contains 1221 amino acids with a predicted mass of 131 kDa. The sequences confirm the lack of similarity with plant PEPc sequences at the N-terminal end and are discussed further in chapter 6.

The first evidence that plants may also contain two distinct types of PEPc came from experiments on developing castor oilseeds (*Ricinus communis*) (Blonde and Plaxton, 2003). PEPc was isolated from the endosperm of castor oilseeds as two distinct isoforms; PEPc1 was a typical 410 kDa homotetramer composed of 107 kDa (p107) subunits whereas PEPc2 existed as a 681 kDa hetero-octamer comprising the same p107 polypeptide found in PEPc1 and a 64 kDa protein (p64) which was structurally and immunologically unrelated. PEPc2 from castor bean exhibited similar kinetic properties to the Class 2 PEPc isoforms of algae, namely, enhanced thermal stability relative to PEPc1 and much lower sensitivity to allosteric regulation and pH. Q-TOF MS/MS of the p64 polypeptide revealed it to be most closely related to two putative PEPcs from the rice and Arabidopsis genome sequence databases. Both these sequences contained the conserved regions required for PEPc activity but possessed N-terminal regions lacking the conserved phosphorylation motif found in all previously reported plant PEPc sequences (Chollet *et al.*, 1996). This was confirmed when the expression of these novel PEPc isoforms from rice and Arabidopsis, termed *Osppc-b* and *Atppc4* respectively, was reported (Sanchez and Cejudo, 2003). A phylogenetic analysis showed that both *Atppc4* and *Osppc-b* are more closely related to PEPc sequences from γ-proteobacteria, such as *E. coli*, than to PEPc sequences from plants, including other PEPc isoforms from Arabidopsis and rice. These results will be discussed further in chapter 6.

1.2.14 PEPc in legume nodules

Several functions have been proposed for PEPc in root nodules (Deroche and Carrayol, 1988). These include re-fixation of respiratory CO_2, provision of malate for

metabolism by the bacteroids in support of nitrogenase activity (King *et al.*, 1986), synthesis of carbon skeletons for the assimilation of ammonium into amino acids (Coker and Schubert, 1981), provision of oxaloacetate for asparagine biosynthesis (particularly in amide-transporting legumes) (Sawhney *et al.*, 1987), production of malate for export in the xylem to counterbalance cation transport and regulate pH (Israel and Jackson, 1982) and production of malate which may be a factor in the control of the variable O_2 diffusion barrier (Pathirana *et al.*, 1997). Enhanced levels of PEPc activity, relative to non-nodulated roots, have been measured in the nodules of many legume species, as well as the non-legume *Alnus glutinosa* (Deroche and Carrayol, 1988). Activity of bacterial PEPc has not been detected in bacteroids (Miller *et al.*, 1987).

PEPc accounts for approximately 0.5 to 2.0% of total soluble protein in alfalfa (*Medicago sativa*) and soybean root nodules (Miller *et al.*, 1987; Zhang *et al.*, 1995). The presence of a nodule-enhanced PEPc isoform in legume root nodules is well documented (Pathirana *et al.*, 1997; Suganuma *et al.*, 1997; Hata *et al.*, 1998; Nakagawa *et al.*, 2003a). During nodule development increases in PEPc transcript, protein and activity levels are associated with the onset of nitrogenase activity and nitrogen fixation (Vance and Gantt, 1992; Pathirana *et al.*, 1997; Dolgikh *et al.*, 1998). *In situ* localisation and immunogold localisation studies together with the production of transgenic plants expressing promoter-reporter constructs have revealed the cellular localisation of the nodule-enhanced PEPc mRNA and encoded protein in alfalfa nodules (Robinson *et al.*, 1996; Pathirana *et al.*, 1997). PEPc was present in the meristematic zone, infection zone and the nitrogen-fixing zone of the nodule. PEPc was present in both infected and uninfected cells within the nitrogen fixing zone. PEPc was also detected in the nodule parenchyma and vascular tissue, particularly pericycle cells. These cells are located inside the vascular bundles and are characterised by protrusions of the cell wall, which are thought to facilitate the transport of assimilated nitrogen into nodule xylem. This may reflect the role of PEPc in pH control or, as these cells contain many mitochondria and are thought to be highly metabolically active, it may reflect PEPc's role in re-fixing respired CO_2. Very similar results to those obtained in alfalfa have been reported for *in situ* hybridisation studies in nodules from pea and soybean (Hata *et al.*, 1998; Fedorova *et al.*, 1999).

The increase in PEPc transcript levels during nodule development and the corresponding increases in protein and activity levels suggest regulation at the

transcriptional level. Pathirana *et al.* (1997) performed promoter deletion studies in transgenic alfalfa plants. They showed that a promoter-GUS construct containing 1.3 kb of sequence upstream of the transcription start site of the nodule-enhanced PEPc conferred nodule-enhanced reporter expression and gave results very similar to those obtained with *in situ* hybridisation studies. When a reduced promoter consisting of 0.7 kb of upstream sequence was used, GUS activity was greatly reduced in nodules. A further 98 bp deletion of this construct almost abolished GUS activity in all tissues except root and nodule vascular tissue, where very high levels of activity were measured. These results suggest that the region between –0.7 and –1.3 kb contain elements required for high levels of expression in the nodules and the region from –0.7 to –0.6 kb contains a negative element controlling expression in vascular tissue.

The promoter of the nodule PEPc isoform has also been investigated in soybean. In soybean the nodule-enhanced PEPc is termed *GmPEPC7* and shows high sequence similarity with that of another soybean PEPc isoform, *GmPEPC15*, which has been shown to be expressed at a low level in all tissues (Hata *et al.*, 1998). Nakagawa *et al.* (2003b) showed that a promoter-GUS construct consisting of 466 bp of sequence upstream from the transcription start site of *GmPEPc7* was sufficient to direct expression to the nodules, with no GUS activity detected in the roots. Deletion of 66 bp from this promoter resulted in loss of GUS activity. This critical region of the *GmPEPc7* promoter (–400 to –466) was termed the switch region (SR). Gel-shift assays found no nodule-specific proteins interacting with the SR of *GmPEPC7* but DNA-binding proteins were detected when another region, termed the amplifier region (AR), consisting of –400 to –290 bp, was used. In hybrid promoter experiments, the SR region from *GmPEPc7* was incorporated into the corresponding promoter region of *GmPEPc15*. This lead to weak but significantly increased levels of activity observed in nodules. When both the SR and AR were incorporated nodule activity was comparable to that of the *GmPEPc7* construct.

Much of the early work on nodule PEPc was concerned with post-translational control and the role of metabolites. Schuller *et al.* (1990) examined the effect of a range of metabolites on the activity of PEPc partially purified from soybean nodules. Similar to the results obtained with PEPc isolated from other plant tissues, soybean nodule PEPc is activated by sugar phosphates (Glc-6-P, Fru-6-P, Glc-1-P) and inhibited by malate, citrate, 2-oxoglutarate, aspartate and glutamate when assayed at pH 7. Raising the pH to 8 abolishes the effects of the activators and decreases or

abolishes the effects of the inhibitors. The exceptions to this were 2-phosphoglycerate (2-PGA), 3-phosphoglycerate (3-PGA) and inorganic phosphate (Pi). Both 2-PGA and 3-PGA had no effect on PEPc activity at pH 7 but were inhibitors at pH 8. Similarly Pi, which was activating at pH 7, became inhibitory at pH 8. Based on comparisons between the estimated concentrations of metabolites in the nodule and the I_{50} values obtained, they suggested malate and aspartate are likely to be the most significant inhibitors *in vivo* and that pH fluctuations could play a significant role in modulating their effects. There is evidence that the sugar phosphate activators could also modulate the effects of pH and malate inhibition (Ocana *et al.*, 1996; Woo and Xu, 1996). Also using partially purified PEPc from soybean root nodules, Woo and Xu (1996) showed that Glc-6-P broadens the pH optimum of PEPc, stimulates PEPc activity at physiological PEP concentrations by increasing the affinity for PEP and protects the enzyme from malate inhibition. Ocana *et al.* (1996), working with partially purified PEPc from *Vicia faba*, also showed reversal of malate inhibition by Glc-6-P.

Schuller and Werner, (1993) demonstrated that soybean nodule PEPc can be phosphorylated *in vitro* by an endogenous protein kinase present in nodule crude extracts. When phosphorylated, PEPc was less sensitive to malate inhibition with I_{50} (malate) values of 0.34 and 1.24 mM before and after phosphorylation respectively. Zhang *et al.* (1995) were able to show *in vivo* phosphorylation of PEPc in detached intact soybean nodules. They also investigated the apparent phosphorylation status of PEPc, as indicated by its malate sensitivity, in response to treatments that perturb the flow of photosynthate from the shoots. They showed that stem girdling reduced the phosphorylation status of PEPc, as did extended periods of darkness (26 h) but not shorter periods of darkness (14 h). Subsequent illumination for 3 h reversed the effect of extended darkness and this reversal was prevented by concomitant stem girdling. Complete removal of all shoot material (detopping) also increased the sensitivity of PEPc to malate (Wadham *et al.*, 1996). Measurements of metabolite levels in nodules from control and detopped plants show a decrease in malate and an increase in PEP levels, which is consistent with reduced PEPc activity (Wadham *et al.*, 1996). These results suggest that the phosphorylation status of soybean nodule PEPc is modulated by photosynthate transported from the shoots. Zhang and Chollet, (1997) partially purified PEPc kinase from soybean root nodules and demonstrated that the effects of

photosynthate supply from the shoots on PEPc phosphorylation status were mediated through parallel changes in PEPc kinase activity.

Treatments that affect the carbon and nitrogen metabolism of root nodules, such as stem girdling, detopping or replacement of N_2 in the rhizosphere with argon ($Ar:O_2$ treatment), inhibit the activity of nitrogenase in the bacteroids (Hartwig *et al.*, 1987; Vessey *et al.*, 1988; Hunt *et al.*, 1989; Denison *et al.*, 1992; de Lima *et al.*, 1994). Curioni *et al.* (1999) measured the concentrations of intermediates of the glycolytic cycle in alfalfa nodules after detopping or $Ar:O_2$ treatment. Their results suggested that both treatments lead to inhibition of glycolysis at the steps catalysed by phosphofructokinase, pyruvate kinase and PEPc. This inhibition occurred after the decrease in nitrogenase activity, thought to be due to a decrease in the permeability of the nodule to oxygen, and was associated with an increase in the ratio of glutamate to glutamine in the nodule. They proposed that the decrease in nitrogenase activity, and therefore reduced levels of ammonium, causes the GS/GOGAT cycle to become nitrogen limited. This results in an increase in glutamate levels, which inhibit PEPc and PK activity. This in turn leads to the accumulation of the substrate PEP, which is a potent inhibitor of PFK. Such a mechanism could represent a method to coordinate the supply of respiratory substrates to the bacteroids, and the supply of 2-oxoglutarate for N_2 assimilation, to the supply of ammonium from the bacteroids. Wei *et al.* (2004b) proposed that the reduced flux through glycolysis might be due to changes in adenylate levels. They put forward a model where large cytosolic adenylate concentration gradients exist across the infected cells in nodules (Wei *et al.*, 2004a). This is due to the mitochondria being clustered at the cell periphery near the intercellular gas-filled spaces, while the main site of ATP utilisation is the cytosol surrounding the symbiosomes. Wei *et al.* (2004b) measured adenylate levels in soybean nodules before and after $Ar:O_2$ treatment. Using a nonaqueous fractionation technique they were able to separate subcellular compartments of the central infected zone of the nodule, which contained 94% of the total adenylate pool of the whole nodule, and calculate adenylate levels for the bacteroids, plant cytosol and mitochondria. After $Ar:O_2$ treatment the levels of ATP in the cytosol increased, while those of ADP and AMP decreased. This is consistent with a decrease in nitrogenase activity and therefore a reduction in ATP demand for ammonium assimilation by glutamine synthetase. The increase in the cytosolic ATP:ADP ratio could inhibit certain glycolytic enzymes and reduce carbon-flux through the pathway. Cytosolic PK

from several different plants and tissues is inhibited by ATP (Ambasht and Kayastha, 2002), while PFK from chickpea (*Cicer arietinum*) nodules has a lower affinity for its substrate in the presence of ATP (Lee and Copeland, 1996). They also suggest that adenylates may be the signal molecules controlling nodule oxygen permeability through activation of ion pumps or ion channels at the plasma membrane. A potential role for ATP acting as a signal molecule, similar to its role as a neurotransmitter in animals, is envisioned, with an extracellular ATP pool reflecting intracellular adenylate levels.

1.3 Regulation of PEPc by phosphorylation

1.3.1 Control of PEPc phosphorylation status

With the early work from the Nimmo group demonstrating the circadian control of PEPc phosphorylation in the CAM plant *Kalanchoë fedtschenkoi* (Nimmo *et al.*, 1984; Nimmo *et al.*, 1986; Nimmo *et al.*, 1987b) and the subsequent observation of PEPc phosphorylation in leaves of the C_4 plant maize during illumination (Budde and Chollet, 1986; Nimmo *et al.*, 1987a), research focused on identifying the factors responsible for control of PEPc phosphorylation. The state of PEPc phosphorylation is dependent upon the activities of the kinase and phosphatase. Therefore the identification of these proteins and the measurements of their activities under conditions known to affect the phosphorylation status of PEPc was required.

1.3.2 PEPc phosphatase

In yeast and mammalian cells four main Ser/Thr protein phosphatases have been identified, PP1, PP2A, PP2B and PP2C. They are classified on the basis of substrate specificity, pharmacological properties and divalent cation requirements. Homologues of PP1, PP2A and PP2C have been identified in plants (Hrabak *et al.*, 2003) and Carter *et al.* (1990) showed that PP2A purified from rabbit skeletal muscle could dephosphorylate PEPc from *K. fedtschenkoi*, which lead to an increase in the sensitivity of PEPc to malate inhibition. Partial purification of a PP2A catalytic subunit from *K. fedtschenkoi* could also dephosphorylate PEPc (Carter *et al.*, 1990) and its activity showed no variation during a 24 h period (Carter *et al.*, 1991). McNaughton *et al.* (1991) investigated PP2A activity in maize leaf extracts using glycogen phosphorylase-a as a substrate. No change in activity was observed between extracts from illuminated or darkened leaves. These results suggest that, at least in C_4

and CAM plants, the main regulator of PEPcs phosphorylation status is the kinase and not the phosphatase. Recently Dong *et al.* (2001) purified and characterised PP2A from illuminated maize leaves. Similar to other eukaryotic PP2As it consisted of a heterotrimer. The enzyme was not specific for PEPc as it dephosphorylated mammalian phosphorylase-a and casein *in vitro*. Some difference in PP2A activity was observed between extracts from illuminated or darkened leaves. The activity from the light sample was about 1.5- to 2-fold higher and was correlated with increased amounts of PP2A subunits present in the sample. However it is thought unlikely that these modest changes would account for the large differences in PEPc phosphorylation status observed *in vivo*.

1.3.3 PEPc kinase

Carter *et al.* (1991) partially purified a kinase from the leaves of *K. fedtschenkoi* which could phosphorylate *K. fedtschenkoi* PEPc. The activity of this kinase was unaffected by calcium. Measurements of the kinase activity over a diurnal cycle in leaf extracts revealed high activity in the middle of the night period, when PEPc is phosphorylated, and low or no activity during the light period, when PEPc is dephosphorylated. These results indicate that the activity of PEPc kinase is responsible for the change in PEPc phosphorylation status previously observed by Nimmo *et al.* (1987b). Carter *et al.* (1991) also showed that the appearance of PEPc kinase during the night was prevented by feeding leaves RNA or protein synthesis inhibitors, indicating *de novo* protein synthesis is required for the appearance of PEPc kinase activity. Hartwell *et al.* (1996) developed a novel assay; total RNA isolated from plant leaves was used in *in vitro* translation reactions and the products of translation were assayed for PEPc kinase activity. They showed that PEPc kinase mRNA levels were 20-fold higher during the night than in the day in *K. fedtschenkoi* leaves, and in leaves kept in constant conditions the kinase mRNA exhibited circadian oscillations which paralleled the phosphorylation state of PEPc. Inhibitor studies also showed that the appearance of the kinase mRNA and activity were blocked by inhibitors of protein and RNA synthesis.

The identity of the PEPc phosphorylating enzyme in C_4 plants was initially confusing as both calcium-dependent and independent kinases were isolated. A Ca^{2+}-dependent PEPc kinase, with a molecular mass of approximately 50 kDa, was isolated from maize (Ogawa *et al.*, 1992; Ogawa *et al.*, 1998). It was inhibited by specific

inhibitors of Ca^{2+}-calmodulin dependent kinases, a calmodulin antagonist (W7) and in the presence of EGTA, which chelates Ca^{2+}. Although it phosphorylated PEPc this did not cause significant changes in PEPc's sensitivity to malate. Wang and Chollet, (1993b) also partially purified a PEPc kinase from illuminated maize leaves; however, it was a Ca^{2+}-independent kinase of approximately 30 kDa. This kinase phosphorylated *Sorghum* PEPc specifically on the Ser-8 residue and this lead to a decrease in sensitivity to malate inhibition. Using an "in gel" kinase assay technique, Li and Chollet, (1993) further investigated the number and nature of PEPc kinases in maize leaves. Protein extracts from illuminated leaves were separated by SDS-PAGE followed by *in situ* renaturation of the proteins in the gel. PEPc kinase activity was then assayed by incubating gel segments with ^{32}P-ATP and dephosphorylated maize PEPc. They detected three polypeptides which could phosphorylate PEPc *in vitro*, one Ca^{2+}-dependent kinase (57 kDa) and two Ca^{2+}-independent kinases (37 and 30 kDa). The 30 kDa kinase appeared to correspond to the kinase previously identified by (Wang and Chollet, 1993b). Protein samples from leaves placed in darkness or treated with inhibitors of photosynthesis (methyl viologen) or cytosolic protein synthesis (cycloheximide) all showed a large reduction in the activities of the 37 and 30 kDa Ca^{2+}-independent PEPc kinases (Li and Chollet, 1993). However, none of these treatments had similar inhibitory effects on the 57 kDa Ca^{2+}-dependent PEPc kinase. These results suggest that the physiologically relevant PEPc kinase, which undergoes light activation in the leaves of C_4 plants, is Ca^{2+}-independent and is either one or both of the 37 and 30 kDa polypeptides.

As was shown for the nocturnal increase in PEPc kinase activity in CAM plants, Hartwell *et al.* (1996) demonstrated that the light induced increase in PEPc kinase in maize leaves was accompanied by a 7-fold increase in kinase mRNA levels. They also showed that RNA synthesis inhibitors blocked this increase in transcript levels and the phosphorylation of PEPc (Hartwell *et al.*, 1999b). However, while treatment with protein synthesis inhibitors blocked the light-induced increase in PEPc phosphorylation, they did not prevent the increase in mRNA levels (Hartwell *et al.*, 1999b). In fact the mRNA levels were increased by protein synthesis inhibitor treatments. These results suggests that the transcriptional requirement for the light-induced increase in PEPc kinase mRNA in maize leaves is that of PEPc kinase itself and not of another protein which affects the translatability of the kinase mRNA. The authors suggested that the increase in mRNA levels seen after treatment with protein

synthesis inhibitors is due to the perturbation of a metabolic signalling system which influences the transcription of PEPc kinase gene(s).

Results from C_3 plants revealed PEPc kinase to have similar properties in C_3, C_4 and CAM plants. Wang and Chollet, (1993a) partially purified PEPc kinase from illuminated leaves of tobacco (*Nicotiana tabacum*). They demonstrated that this kinase was Ca^{2+}-independent. Using "in gel" kinase assays, Li *et al.* (1996) resolved two Ca^{2+}-independent PEPc kinase polypeptides with molecular weights of 37 and 30 kDa. They showed that kinase activity was not present in the leaves during the dark but was induced by exposure to light. Cycloheximide treatment prevented the light activation of PEPc kinase activity, as did the use of an inhibitor of photosynthesis (DCMU). They also found that the light-induced kinase activity was blocked when glutamine synthase activity was inhibited by methionine sulphoxide treatment, and this could be partially alleviated by feeding glutamine to the leaves. Light activation of PEPc kinase activity in C_3 leaves was also investigated using barley. Hartwell *et al.* (1996) observed an approximate 2-fold increase in PEPc kinase translatable mRNA extracted from illuminated barley leaves compared to darkened leaves. Using barley leaf mesophyll protoplasts, Smith *et al.* (1996) showed PEPc kinase activity was present in the dark but increased after exposure to light. Cycloheximide treatment prevented PEPc kinase light induction but DCMU treatment did not. However, approximately 50% of the kinase activity was unaffected by prolonged cycloheximide treatment. The authors suggest that barley contains two different PEPc kinase activities, one which has a slow turnover and is present in the dark and a second that is light-induced and rapidly turned over.

1.3.4 Cloning and identification of PEPc kinase genes

Purification of PEPc kinase from plant tissues has proved difficult due to its very low abundance, less than 1 in 10^5 of soluble leaf protein in *K. fedtschenkoi* (Hartwell *et al.*, 1999a). Saze *et al.* (2001) recently achieved high purification of maize PEPc kinase after a 1.4 million-fold enrichment, which required 2.6 kg of leaf tissue and resulted in the recovery of less than 1μg of enzyme. This has prevented direct biochemical attempts to isolate the gene by obtaining partial amino acid sequence data. The cloning of a PEPc kinase cDNA was therefore a high priority. An attempt to isolate PEPc kinase using a yeast two-hybrid screen with a C_4 PEPc cDNA

as bait was unsuccessful (Alvarez *et al.*, 2003). The first PEPc kinase gene (*PPCK*) was cloned from *K. fedtschenkoi* using a novel approach (Hartwell *et al.*, 1999a). A cDNA library was constructed from RNA isolated from mature *K. fedtschenkoi* leaves collected in the middle of the dark period. The cDNA library could then be screened by transcribing linearised plasmid, translating and then assaying for PEPc kinase activity. By successively subdividing the library and selecting pools with the most PEPc kinase activity, a single clone encoding a 30 kDa Ca^{2+}-independent PEPc kinase was isolated. Northern analysis showed the transcript abundance of *PPCK* increased during leaf development, which parallels the induction of CAM photosynthesis and increasing levels of PEPc kinase activity seen as the leaf matures. The *PPCK* transcript levels varied over a diurnal cycle and this variation was maintained in detached leaves under constant environmental conditions. The correlation seen between levels of *PPCK* transcripts, translatable mRNA and PEPc kinase activity strongly suggest that the cloned *PPCK* gene is responsible for controlling the phosphorylation state of the CAM photosynthetic PEPc.

Based on comparisons with the *K. fedtschenkoi PPCK* (*KfPPCK*) sequence two *PPCK* genes were identified in *Arabidopsis thaliana* (*AtPPCK1* and *AtPPCK2*) (Hartwell *et al.*, 1999a; Fontaine *et al.*, 2002). They both encode Ca^{2+}-independent PEPc kinase proteins of 31 kDa. Expression patterns of the two genes was investigated by reverse transcriptase PCR (RT-PCR), as a signal was not obtained with Northern blotting experiments indicating their low expression levels (Fontaine *et al.*, 2002). *AtPPCK1* transcripts were highest in rosette leaves, clearly detectable, but lower, in flowers and roots, and barely detectable in siliques, cauline leaves and stems. In contrast, *AtPPCK2* was most highly expressed in roots and flowers, moderately expressed in cauline leaves and expressed at a low level in rosette leaves and stems. Their differing expression patterns suggest that *AtPPCK1* and *AtPPCK2* perform different functional roles within the plant.

A second *PPCK* clone was isolated from a CAM photosynthetic plant, the facultative CAM plant *Mesembryanthemum crystallinum* (Taybi *et al.*, 2000). *M. crystallinum* can be induced to perform CAM photosynthesis by salinity stress. As with *K. fedtschenkoi*, *M. crystallinum* leaves performing CAM have high PEPc kinase activity during the night and low activity during the day (Li and Chollet, 1994). Taybi *et al.* (2000) used this expression pattern as the basis for a protein-kinase-targeted differential display RT-PCR cloning strategy. A PCR product showing salt and night

up-regulated expression was subjected to RACE PCR (rapid amplification of cDNA ends) to obtain a full length cDNA clone. The resulting clone, termed *McPPCK1*, encodes a 32 kDa Ca^{2+}-independent PEPc kinase polypeptide. *McPPCK1* transcripts showed circadian oscillation in abundance, with expression peaking during the second half of a 12 h dark period. Southern blot experiments revealed the presence of another sequence with high homology to *McPPCK1* in the *M. crystallinum* genome, suggesting that, as in *A. thaliana* (Fontaine *et al.*, 2002), PEPc kinase is encoded by a small gene family.

To further characterise McPPCK1, a recombinant form of the kinase was produced in *E. coli* and purified (Ermolova *et al.*, 2003). Phosphorylation experiments using recombinant McPPCK1 in the presence of wild-type or phosphorylation site mutants of *M. crystallinum*, sorghum and maize PEPc demonstrated that McPPCK1 only phosphorylates PEPc on the single invariant serine residue located near the N-terminus of PEPc. Phosphorylation of *M. crystallinum* PEPc extracted from leaves of salt-stressed plants collected during the day resulted in a significant decrease in PEPc's sensitivity to malate inhibition. When the experiment was repeated with PEPc collected from leaves during the night there was no further reduction in malate sensitivity. These results demonstrate the role of McPPCK1 in controlling PEPc activity and therefore nocturnal fixation of CO_2 in *M. crystallinum*.

Tsuchida *et al.* (2001) reported the first cloning of a *PPCK* gene from a C_4 plant, the dicotyledonous *Flaveria trinervia*. Using degenerate primers based on conserved regions of previously reported *PPCK* genes (Hartwell *et al.*, 1999a; Taybi *et al.*, 2000), a 520 bp fragment was obtained from *F. trinervia* cDNA. This fragment was then used to screen a *F. trinervia* illuminated leaf cDNA library from which a full length *PPCK* sequence was obtained (*FtPPCK1*). This clone was expressed in *E. coli* and the recombinant protein was able to phosphorylate wild-type maize PEPc but not a phosphorylation site mutant maize PEPc. Analysis of *FtPPCK1* transcript levels by Northern blot showed highest expression in leaves during the day but little expression in leaves during the night. Expression levels in roots and stems were also very low. This suggests that *FtPPCK1* encodes the C_4 photosynthetic *PPCK*. Southern blot analysis indicates the presence of other sequences with high homology to *FtPPCK1*.

1.3.5 Structure of PEPc kinase genes

The sequences of the *PPCK* genes revealed that it is the smallest protein kinase known (Hartwell *et al.*, 1999a). PPCK comprises a protein kinase catalytic domain with minimal N- and C-terminal extensions. Its sequence shows close similarity to members of the CaMK (Ca^{2+}/calmodulin-dependent kinase) group of protein kinases (Hanks and Hunter, 1995), particularly the higher plant family of calcium-dependent protein kinases (CDPKs). However, PPCK lacks the autoinhibitory region and Ca^{2+}-binding EF hands of CDPKs, explaining its Ca^{2+}-independent nature. Many protein kinases are activated by phosphorylation of residues located in the 'activation loop' (Johnson *et al.*, 1996). PPCK does not contain these phosphorylatable residues, indicating it is constitutively active. This agrees with the high activity of PPCK generated in *in vitro* transcription/translation reactions (Hartwell *et al.*, 1999a).

The genomic sequences of *KfPPCK*, *AtPPCK1* and *AtPPCK2* (Hartwell *et al.*, 1999a; Fontaine *et al.*, 2002) revealed the presence of a single intron present very close to the 3' end of the coding region, the presence and position of which is conserved in all *PPCK* sequences identified to date (Nimmo, 2003). This intron interrupts the codon for a conserved Arg residue prior to the His.X.aromatic.hydrophobic consensus sequence that is taken to mark the end of the catalytic domain of protein kinases. In most *PPCK* sequences the intron contains an in frame stop codon, so a translated genomic protein would lack this consensus sequence. Fontaine *et al.* (2002) assayed the transcription/translation product of the genomic clone of *AtPPCK2* and showed that it did not possess kinase activity.

The results of Hartwell *et al.* (1999a), Taybi *et al.* (2000), Fontaine *et al.* (2002) and Tschida *et al.* (2001), together with the analysis of sequence databases presented in Nimmo (2003) all suggest that PPCK is encoded by a small gene family in plant species. All of the available sequences to date have indicated that PPCK is a Ca^{2+}-independent kinase of approximately 30 kDa. The identity of the 37 kDa polypeptide revealed in studies using 'in-gel' kinase assays remains unclear.

1.3.6 Regulation of PEPc kinase

Most progress in characterising the signalling pathway controlling the transcript abundance of *PPCK* has been achieved in C_4 photosynthetic plants. Much of this work has made use of isolated mesophyll cells and mesophyll cell protoplasts

from the C_4 grass *Digitaria sanguinalis*. One advantage is that isolated mesophyll cells of *D. sanguinalis* contain functional plasmodesmata facilitating introduction of certain compounds into the cells (Duff *et al.*, 1996). Treatment of *D. sanguinalis* protoplasts with light and a weak base, such as NH_4Cl or methylamine, resulted in a decrease in PEPc's sensitivity to malate due to an increase in PEPc kinase activity (Giglioli-Guivarc'h *et al.*, 1996; Duff *et al.*, 1996). The effect of weak base treatment could be mimicked by application of exogenous 3-phosphoglycerate (3-PGA) and these treatments were shown to cause an increase in cytosolic pH using pH sensitive fluorescent dyes (Giglioli-Guivarc'h *et al.*, 1996). Treatment with EGTA and a Ca^{2+} ionophore (A23187), to deplete intracellular Ca^{2+} levels, prevented PEPc phosphorylation *in situ* (Duff *et al.*, 1996) and demonstrated a role for endogenous Ca^{2+}. Furthermore, experiments using various Ca^{2+}-channel blockers suggested the release of Ca^{2+} from vacuolar stores (Giglioli-Guivarc'h *et al.*, 1996). This suggests that a Ca^{2+}-dependent step acts upstream of the Ca^{2+}-independent PEPc kinase. The light and weak base induced increase in PEPc kinase activity could also be blocked by photosynthesis inhibitors (DCMU and gramicidin), the calcium/calmodulin antagonists (W5, W7 and compound 40/80), and the cytosolic protein synthesis inhibitor cycloheximide.

Coursol *et al.* (2000) extended this work by demonstrating the involvement of Ca^{2+}-dependent phosphoinositide-specific phospholipase C (PI-PLC) in the light dependent pathway leading to increased PEPc kinase activity. In animal cells, PI-PLC catalyses the hydrolysis of membrane bound phosphatidylinositol-4,5-bisphosphate to give inositol-1,4,5-triphosphate ($Ins(1,4,5)P_3$) and 1,2-diacylglycerol (DAG). $Ins(1,4,5)P_3$ causes the release of Ca^{2+} from internal stores, while DAG leads to the activation of protein kinase C (PKC). A similar pathway is believed to exist in plants. Coursol *et al.* (2000) measured a transient increase in *in situ* levels of $Ins(1,4,5)P_3$ in light and weak base treated *D. sanguinalis* protoplasts. A PI-PLC inhibitor, U-73122, blocked the increase in $Ins(1,4,5)P_3$ and the increase in PEPc kinase activity, suggesting that PI-PLC signalling is a required step in the activation of PEPc kinase. Based on these results a model for the light transduction pathway leading to phosphorylation of PEPc in mesophyll cells of C_4 plants has been proposed (Fig. 1.5). In this model it is envisioned that 3-PGA, which is generated in the bundle sheath cell chloroplasts during photosynthesis, diffuses into the mesophyll cell and acts as a signal indicating the rate of photosynthesis occurring in the adjacent bundle sheath

cell. 3-PGA is transported into the mesophyll chloroplast in a protonated form, which causes alkalinisation of the cytosol. This is followed by activation of PI-PLC and a transient increase in Ins(1,4,5)P_3 and DAG. The increased levels of Ins(1,4,5)P_3 cause tonoplast calcium channels to open and an efflux of Ca^{2+} into the cytosol. The increased cytosolic Ca^{2+} concentration, together with DAG, activates a PKC-like enzyme. This then leads to *de novo* protein synthesis of PEPc kinase, possibly through the phosphorylation of a transcription factor that increases *PPCK* transcription. Recently, Osuna *et al.* (2004) have identified a Ca^{2+}-dependent protein kinase from *D. sanguinalis* mesophyll cell protoplasts which has some characteristics of mammalian PKCs and the authors suggest that it may represent the Ca^{2+}-activated component of the C_4 PEPc signalling pathway. However, the link between alkalinisation of the cytosol and activation of PI-PLC is unclear as the enzyme has an acidic pH optimum (Coursol *et al.*, 2000).

In CAM plants there is evidence that at least some of the components identified in the C_4 signalling pathway are also involved in the nocturnal induction of PEPc kinase in *M. crystallinum* leaves (Bakrim *et al.*, 2001). In detached leaves the nocturnal increase in PEPc kinase activity could be prevented by treatment with various inhibitors. Cycloheximide, the PI-PLC inhibitor U-73122, the tonoplast Ca^{2+}-channel blocker TMB-8 and the Ca^{2+}-dependent protein kinase antagonist W7 were all effective in preventing PEPc kinase induction. These results suggest that, as in the C_4 model, the second messenger Ins(1,4,5)P_3 is involved in Ca^{2+} release from the vacuole leading to activation of a Ca^{2+}-dependent protein kinase. However, whether cytosolic alkalinisation, as suggested by Bakrim *et al.* (2001), due to the accumulation of malic acid in the vacuole during the night could trigger the phosphorylation cascade in CAM plants, in a similar way as proposed for C_4 plants, is unclear. Paterson and Nimmo, (2000) investigated the effects of pH on PEPc kinase induction in *K. fedtschenkoi*. Leaf discs were treated with the weak base NH_4Cl but this did not affect induction of the kinase.

Other work has shown the potential importance of metabolite regulation in control of *PPCK* in CAM plants. PEPc kinase mRNA and activity levels were measured in leaves of *Kalanchoë daigremontiana* subjected to treatments that affect the levels of cytosolic malate (Borland *et al.*, 1999). Leaves were enclosed in an atmosphere of nitrogen during the night, which prevented the nocturnal accumulation of malate. When the leaves were released into normal air at the start of the

photoperiod an increase in the rate of CO_2 fixation and the duration of CO_2 fixation was observed in comparison to control leaves. This was due to the levels of PEPc mRNA and activity, and PEPc phosphorylation state, remaining elevated for longer than in control leaves. Another treatment was to expose the leaves to a temperature increase in the middle of the night, which lead to a reduction in PEPc kinase mRNA and activity levels. This effect of elevated temperature is thought to be due to an increased permeability of the tonoplast to malate and therefore efflux of malate into the cytosol. In leaves which had been prevented from accumulating malate, the effect of an increase in temperature was an increase in kinase mRNA. These results suggest that malate could be a feedback inhibitor of *PPCK* expression. This lead to a proposed model of circadian control of *PPCK* expression being a secondary response to fluctuations in metabolism, shown in Fig 1.6 (Nimmo, 2000). In this model the primary output from the circadian oscillator is control of malate transport across the tonoplast. This would result in circadian changes in metabolite levels, which in turn could regulate *PPCK* expression. A possible direct connection between the oscillator and *PPCK* expression and malate feedback to the oscillator is also proposed.

Recently, a further level of control of PEPc kinase activity has been suggested. Nimmo *et al.* (2001) detected the presence of a protein that reversibly inhibits PEPc kinase activity. The presence of this inhibitor has been detected in the leaves of *K. fedtschenkoi*, maize and *Hordeum vulgare* (barley), thereby demonstrating its presence in C_3, C_4 and CAM plants. It was demonstrated not to be a protease, ATPase or protein phosphatase. The inhibitor appeared to be specific for PEPc kinase as it had no effect on the activities of casein kinase I, histone kinase or cyclic AMP-dependent protein kinase. In *K. fedtschenkoi* leaves the amount of inhibitor varied over a diurnal cycle, increasing at the end of the dark period and decreasing towards the end of the photoperiod. This is the opposite pattern to that of PEPc kinase activity. The authors suggest that the function of the inhibitor may be to inhibit the low levels of kinase activity present during the light period.

Much less is known about the signalling pathway controlling *PPCK* expression in C_3 plants. The results of Van Quy *et al.* (1991) and Duff and Chollet, (1995) demonstrated that illumination and supply of nitrate to detached nitrate-deficient wheat leaves led to an increase in PPCK activity and the phosphorylation status of PEPc. Li *et al.* (1996) also reported that PEPc kinase is reversibly light activated *in vivo* in tobacco leaves and this light activation is dependent upon the

nitrogen status of the leaf. The level of translatable mRNA in barley leaves and the transcript levels of both *AtPPCK1* (in rosette leaves) and *AtPPCK2* (in flowers) were increased in response to illumination (Hartwell *et al.*, 1996; Fontaine *et al.*, 2002). In legume root nodules it appears that photosynthate supply from the shoots is involved in the regulation of PEPc kinase activity (Zhang and Chollet, 1997). In germinating barley seeds the phosphorylation of PEPc increased during imbibition (Osuna *et al.*, 1999). However, PEPc kinase activity was present in the dry seeds and did not increase following imbibition. Furthermore, the increase in phosphorylation was not affected by cycloheximide, W7or TMB-8 treatment, which have all been used to prevent upregulation of PEPc kinase in other systems. These results suggest that the PEPc kinase present in barley seeds is not subject to rapid protein turnover.

1.4 Objectives

In C_3 plants, PEPc has been proposed to play a variety of roles, as well as fulfilling a more general anaplerotic function. As was shown for C_4 and CAM photosynthetic PEPc isoforms, the activity of PEPc in C_3 plants is modulated by reversible phosphorylation by a specific protein kinase, PEPc kinase. However, at the start of this project only a single *PPCK* gene had been cloned from a C_3 plant (*Arabidopsis thaliana*). In the root nodules of leguminous plants PEPc plays an important metabolic role and a *PEPC* gene with nodule-enhanced expression has been cloned. Further work partially purified PEPc kinase from soybean root nodules and demonstrated that treatments which perturb the flow of photosynthate from the shoots affected the activity of PEPc kinase in the nodules. The major aim of this project was to clone *PEPC* and *PPCK* genes and characterise their expression patterns in the leguminous plant soybean. During the course of the project Xu *et al.* (2003) published the cloning of two *PPCK* genes from soybean, one of which showed nodule-enhanced expression. They proposed the existence of *PEPC-PPCK* 'expression-partners' in plants, exemplified by the expression of a nodule-enhanced *PEPC* and *PPCK* gene. However, the results presented in this thesis show that the situation is appreciably more complex.

Further work arose from the identification of two *PPCK* genes in tomato, one of which was a novel *PPCK* gene subject to alternative splicing (Marsh, 2002). To extend this work the expression patterns of these genes was assessed in a tomato mutant and the closely related solanaceous species tobacco.

Finally database searches of the rice genome allowed identification of three putative *PPCK* genes. The sequence of one of these genes was unusual in containing three in frame methionine residues at the 5' end, potentially allowing the production of a larger PEPc kinase protein. In collaboration with Dr Hiroshi Fukayama (National Institute of Agrobiological Resources, Japan) the functionality of these genes was assessed.

Figure 1.1 Cartoon showing the different stages in the infection process of a leguminous plant.

Taken from

http://homepage.smc.edu/ishida_Carolyn/Week01/EnviromentalMicro.htm.

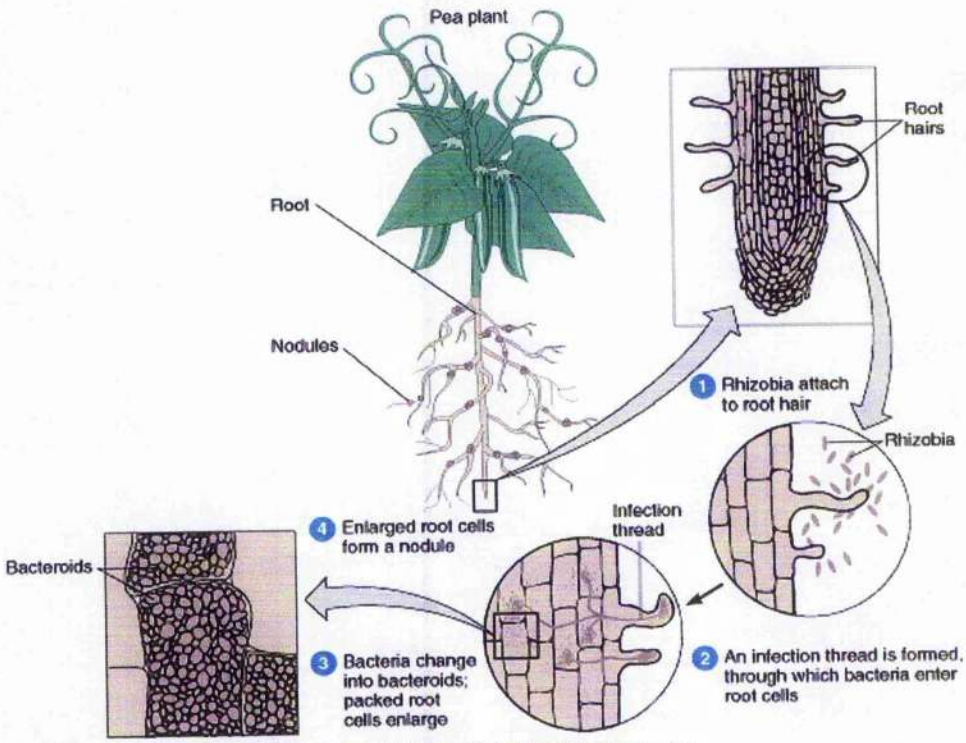


Figure 1.2 Electron micrograph showing cells in the infection zone of a soybean root nodule.

The large infected cells (ic) containing large numbers of symbiosomes can be seen on either side of an uninfected cell (uc) and an air space (as). Taken from Udvardi and Day, (1997)

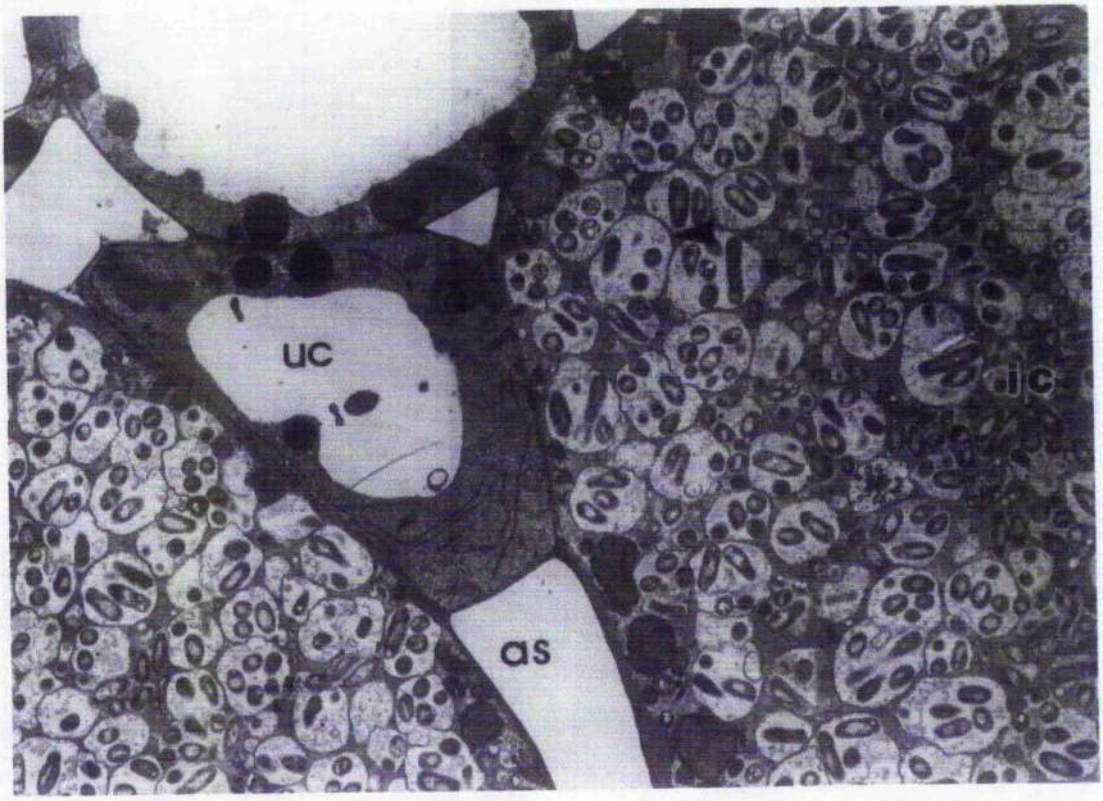

uc
ic
as

Figure 1.3 Nutrient exchange between a nitrogen fixing bacteroid and the host plant cell, including the proposed amino-acid cycle.

Transport systems of the peribacteroid membrane that have been kinetically but not genetically characterised are shown in blue, while those that are hypothetical are shown in yellow. PHB, polyhydroxybutyrate; OAA, oxaloacetate; TCA, tricarboxylic acid cycle. Taken from Lodwig *et al.* (2003).

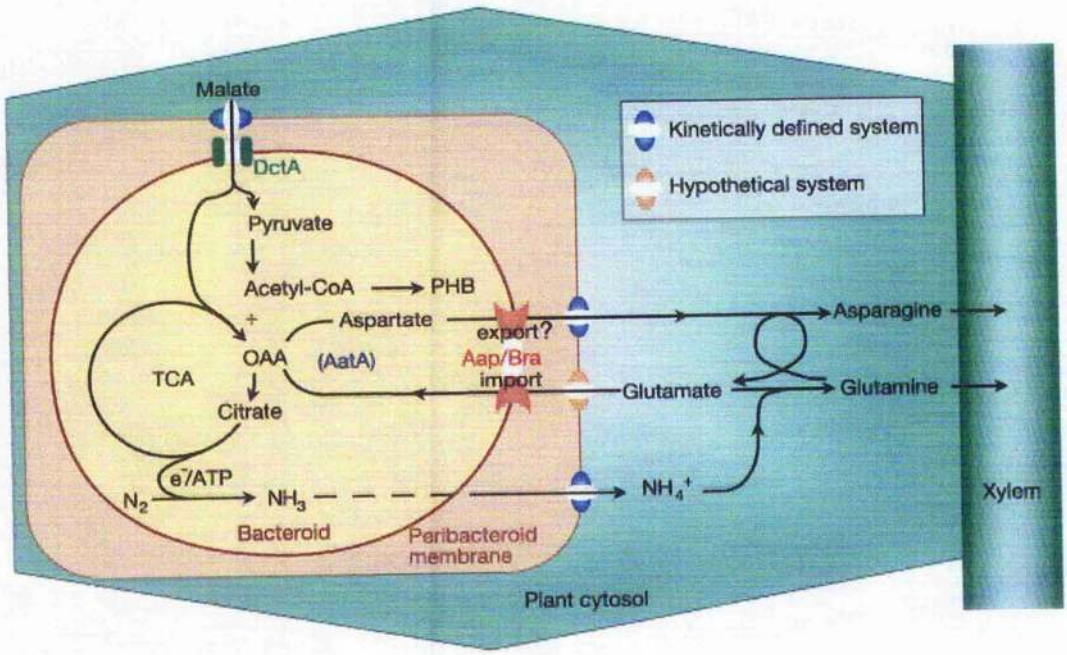
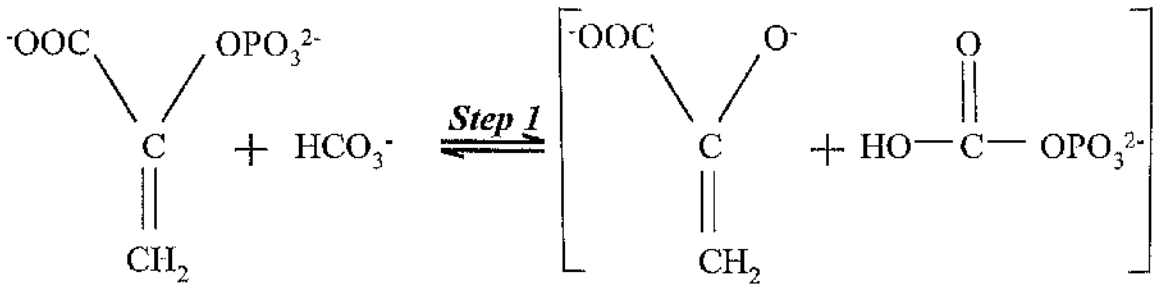
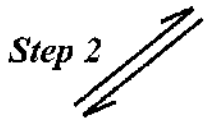
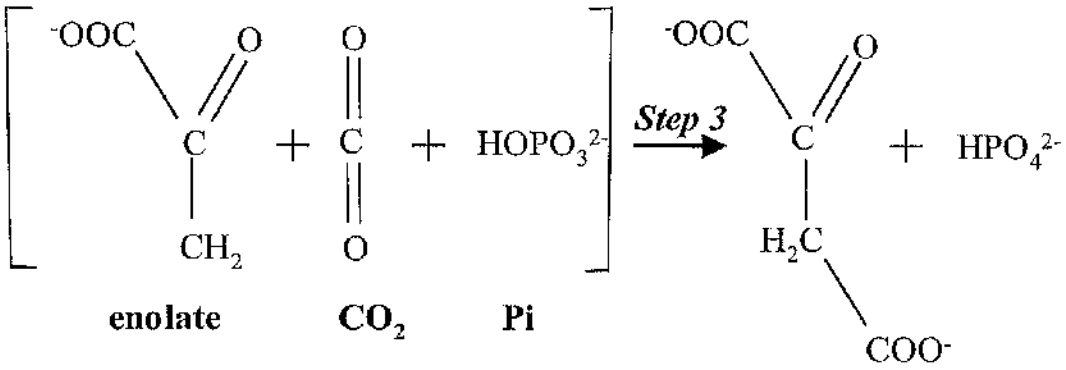

Malate
DctA
Pyruvate
Acetyl-CoA
PHB
Aspartate
OAA
(AatA)
TCA
Citrate
e⁻/ATP
N2
NH3
Bacteroid
Peribacteroid membrane
export?
Aap/Bra
import
Glutamate
NH4+
Asparagine
Glutamine
Xylem
Plant cytosol
Kinetically defined system
Hypothetical system

Figure 1.4 Proposed reaction mechanism of phosphoenolpyruvate carboxylase. Step 1: the enolate anion of pyruvate and carboxyphosphate are formed from PEP and bicarbonate by phosphate transfer. Step 2: a conformational change positions the carbonyl carbon of the carboxyphosphate intermediate adjacent to the carbon-3 of the enolate and carboxyphosphate is cleaved into carbon dioxide and Pi inside the catalytic site. Step 3: the carbon dioxide makes an electrophilic attack on the enolate anion to form oxaloacetate (OAA) and Pi is liberated. Adapted from Izui *et al.* (2004).



PEP **bicarbonate** **enolate** **carboxyphosphate**





enolate **CO_2** **Pi**

oxaloacetate **Pi**

Figure 1.5 Model of the signal transduction cascade in the C_4 leaf resulting in increased PPCK activity in response to light.

An increase in 3-phosphoglyceric acid (3-PGA) ensures the intercellular coupling via pHc changes in the mesophyll cell cytosol. This is followed by activation of a mesophyll cell phospholipase C (PI-PLC) and transient production of the second messenger inositol-1,4,5-triphosphate ($Ins(1,4,5)P_3$) and diacylglycerol (DAG). $Ins(1,4,5)P_3$ causes tonoplast calcium channels to open and efflux of calcium into the cytosol while DAG and calcium activate a PKC-like protein. The rapid synthesis of the calcium-independent PEPc kinase leads to phosphorylation of C_4 PEPc. Question marks indicate the steps that remain to be elucidated. AOA, oxaloacetate; PiP2, phosphatidylinositol-4,5-bisphosphate; TP, triose-phosphate. Taken from Jeanneau *et al.* (2002).

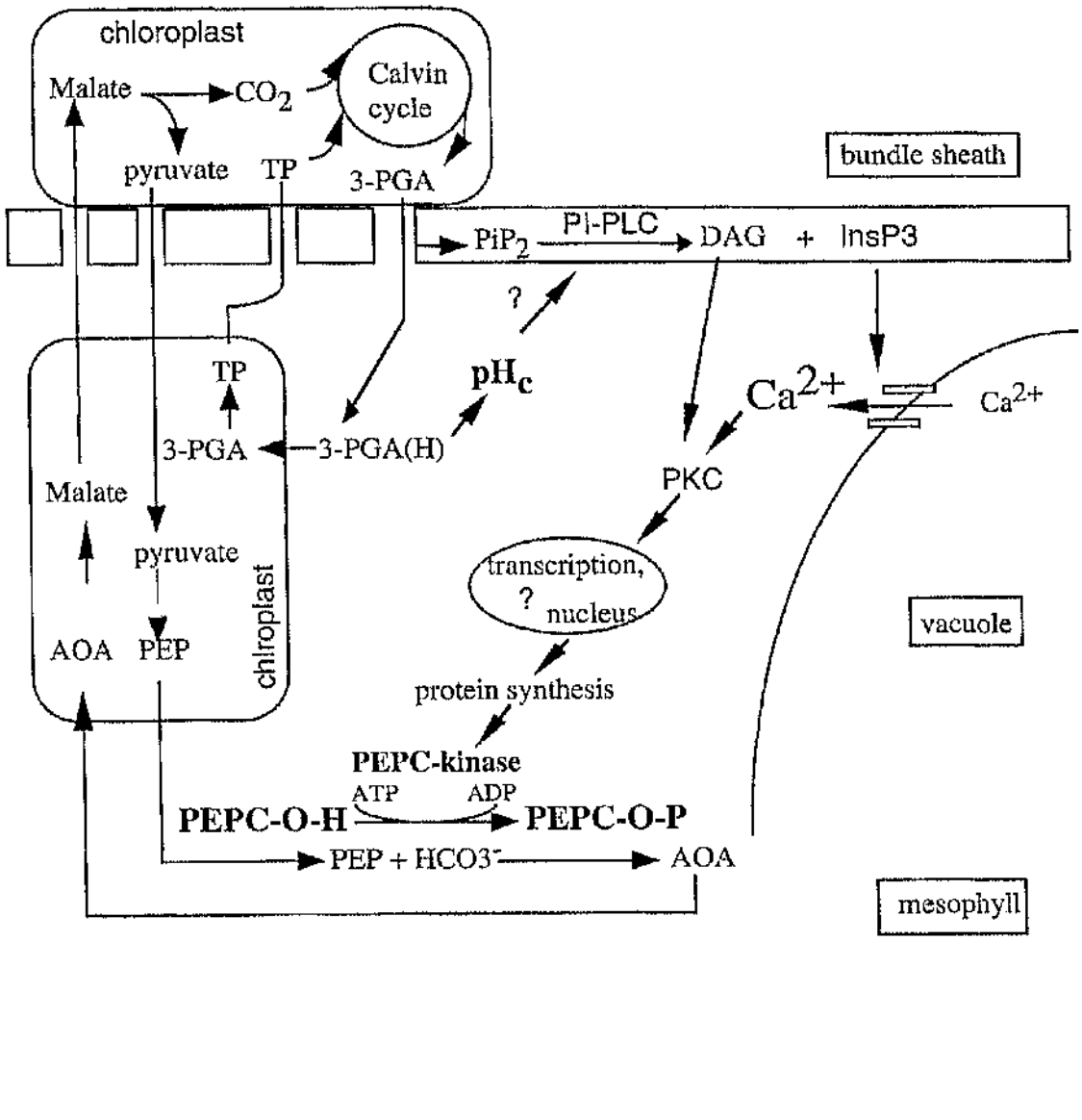
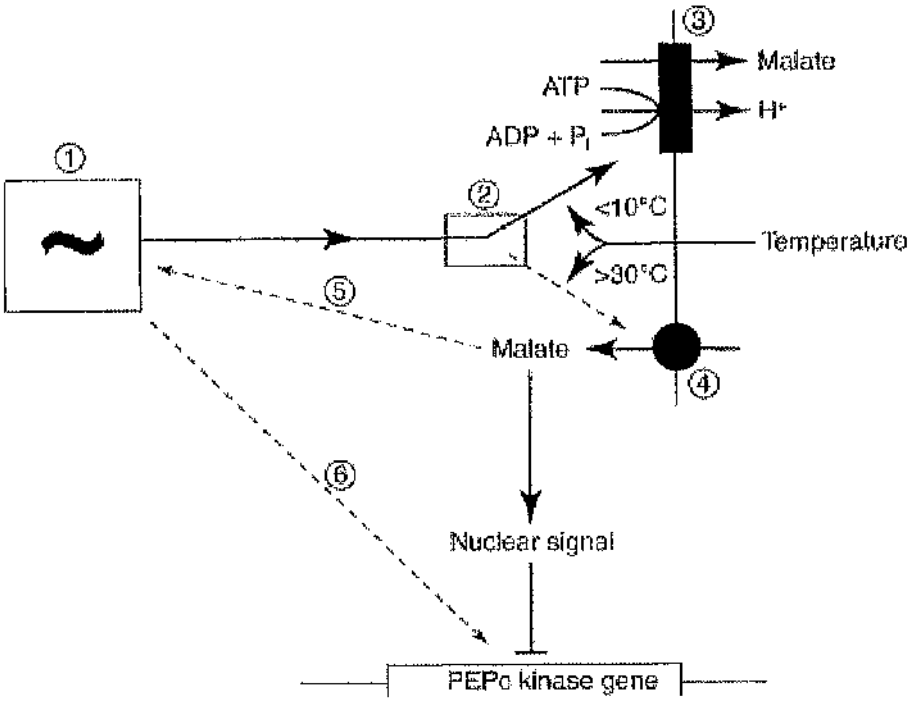

chloroplast
Malate
CO_2
Calvin cycle
pyruvate
TP
3-PGA
bundle sheath
PiP_2
PI-PLC
DAG + InsP3
?
pH_c
Ca^{2+}
Ca^{2+}
TP
3-PGA
3-PGA(H)
PKC
Malate
pyruvate
transcription, ? nucleus
vacuole
AOA
PEP
chloroplast
protein synthesis
PEPC-kinase
ATP
ADP
PEPC-O-H
PEPC-O-P
PEP + HCO3⁻
AOA
mesophyll

Figure 1.6 Hypothetical model for the circadian control of PEPc by metabolites. The primary oscillator (1) drives a switch (2) that affects the tonoplast, controlling the rates of malate uptake energised by an H^+-ATPase (3) and malate release (4). Temperature extremes affect the switch operation as shown. The nuclear signal inhibits the expression of PEPc kinase. In addition, there might be feedback from cytosolic malate to the oscillator (5) and some direct connection between the oscillator and the expression of PEPc kinase (6). Broken arrows represent connections that might occur. Taken from Nimmo (2000).


③
Malate
ATP
H+
ADP + Pi
①
②
<10°C
Temperature
>30°C
⑤
Malate
④
⑥
Nuclear signal
PEPc kinase gene

Chapter 2

Materials and Methods

2.1 Materials

All chemicals were from Sigma-Aldrich Ltd (Poole, Dorset, U.K.), Fisher Scientific UK (Southampton) or VWR International Ltd. (Poole, Dorset, U.K.), unless otherwise stated.

All restriction enzymes were from New England Biolabs Ltd. (Hitchin, Hertfordshire, U.K.) or Promega UK (Southampton).

AMV Reverse transcriptase, DNA size markers, dNTPs and RNasin RNase inhibitor were from Promega UK (Southampton).

[γ-^{32}P] ATP (triethylammonium salt, 3000 Ci mmol^{-1}), rabbit reticulocyte lysate *in vitro* translation kit and Redivue [^{35}S]-methionine (1000 Ci mmol^{-1}) were from Amersham Biosciences UK Ltd. (Bucks., U.K.).

DNA sequencing and primer synthesis were carried out by MWG-Biotech, (Ebersberg, Germany).

DNA-free kit and mMessage mMachine T3 and T7 *in vitro* transcription kits were from Ambion (Huntingdon, UK.).

GeneRacer RACE kit and TOPO TA cloning kit were from Invitrogen Ltd. (Paisley, U.K.).

Genome Walker kit was from BD Biosciences Clontech (Cowley, Oxford, U.K.).

Hydroxyapatite (Bio-gel HTP) and 37.5:1 acrylamide:bis-acrylamide solution were from Bio-Rad Laboratories (England) Ltd., Bramley, Kent, U.K.

Okadaic acid was from Moana Bioproducts, Honolulu, U.S.A.

PUREgene DNA isolation kit was from Gentra Systems (Minneapolis, MN).

QIA-quick gel extraction kit, QIA-prep spin plasmid miniprep kit and RNeasy Plant RNA extraction kit were from Qiagen Ltd. (Crawley, West Sussex, U.K.).

ReddyMix PCR master mix was from ABgene (Epsom, UK).

2.2 Plant and bacterial material

2.2.1 *Glycine max*

Nodulated soybean plants (*Glycine max*, cv. Williams 82) were grown in a perlite/sand mix in a controlled environment chamber (Microclima 1750, Snijders

Scientific) at 22°C/18°C day/night temperature and 60% relative humidity (RH), with a 16 h photoperiod and a light intensity of 400 μmol $m^{-2}s^{-1}$ at the top of the leaf canopy. Plants were inoculated with *Bradyrhizobium japonicum* (strain 3402) after germination. Nodulated plants were watered daily with a nitrogen-free nutrient solution (Ryle *et al.*, 1978). Plant material was collected from 5 week old plants, except for flowers, which were collected from 7-8 week old plants. For the experiments conducted in constant conditions, non-nodulated soybean plants were grown in a perlite/compost mix at 22°C, 60% RH with a 12 h photoperiod.

2.2.2 *Lycopersicon esculentum*

Tomato plants (*Lycopersicon esculentum* cv Alicante and the *gf* mutant) were grown from seed on bedding compost (William Sinclair Horticulture Ltd., Lincoln, U.K.) in a greenhouse under a 16 h photoperiod supplemented with mercury-vapour lamps (300 μmol m^{-2} s^{-1}). Plants were re-potted at 4 weeks and then watered every 2 days. Tomato fruit were allowed to ripen on the vine. Stages of ripening were assigned according to the colour classification by the U.S Department of Agriculture (Standards for Grade of Fresh Tomatoes). Quartered segments were used to isolate RNA. Young leaves were harvested as the primary leaf 14 days after sowing, while mature leaves were harvested after 6 weeks.

2.2.3 *Nicotiana tabacum*

Tobacco plants (*Nicotiana tabacum* cv Petite Havana) were grown in bedding compost (William Sinclair Horticulture Ltd., Lincoln, U.K.) in a greenhouse under a 16 h photoperiod supplemented with mercury-vapour lamps (300 μmol m^{-2} s^{-1}). Plants were watered every 2 days. Plant material was collected from 6 week old plants, except for flowers, which were collected from 8-9 week old plants.

2.2.4 *Bradyrhizobium japonicum*

Bradyrhizobium japonicum (strain 3402) was grown on yeast mannitol agar (Vincent, 1970) slopes in universal bottles at 28°C. Once grown, cultures were stored at 4°C. New agar slopes were inoculated every 4-6 months. Congo Red plates were

used to check the purity of the rhizobium stocks. Colonies of rhizobium absorb very little of the Congo Red dye, whereas many other bacteria strongly absorb it (Hahn, 1966; Vincent, 1970). For inoculation of soybean seeds, 100 mls of yeast mannitol media was inoculated with *B. japonicum* and grown in a shaking incubator at 28°C, 200 rpm, for 7-10 days. The culture was divided into two 50 ml centrifuge tubes and spun for 5 min at 3000 rpm in a Sorvall Legend RT bench top centrifuge using a swing-out rotor (Number 75006441). The supernatant was decanted and the cells resuspended in dH_2O and watered onto the soybean seeds.

2.3 Nucleic acid isolation

2.3.1 Isolation of genomic DNA from plant tissue

Plant leaf material was collected and immediately frozen under liquid nitrogen and stored at -70°C. Genomic DNA was isolated using either the PUREgene DNA isolation kit (Gentra Systems, Minneapolis, MN) or the DNeasy DNA isolation kit (Qiagen, U.K.) according to the manufacturer's protocol. The quantity and purity of the DNA was determined spectrophotometrically according to the method described by Sambrook *et al.* (1989). To check the integrity of genomic DNA a sample was resolved by electrophoresis on a 0.8 % agarose gel and visualised with ethidium bromide.

2.3.2 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from 10 ml overnight cultures of *E.coli* containing the plasmid of interest using QIA-prep spin plasmid miniprep kit (Qiagen, U.K.) according to the manufacturer's protocol.

2.3.3 RNA isolation from plant tissue

Plant material was collected and immediately frozen under liquid nitrogen and stored at -70°C. Total RNA was isolated using the RNeasy plant RNA isolation kit (Qiagen, U.K.) according to the manufacturer's protocol. To ensure the elimination of any contaminating genomic DNA the RNA samples were treated with the DNA-free kit (Ambion, U.K.) according to the manufacturer's protocol. The quantity and purity of the RNA was determined spectrophotometrically according to the method described by Sambrook *et al.* (1989). The integrity of the RNA was determined by

detection of ribosomal RNA bands after denaturing agarose gel electrophoresis and ethidium bromide staining.

2.4 Polymerase chain reaction techniques

2.4.1 Primer design

Primers were designed using the online Primer3 software (Rozen and Skaletsky, 2000).

Table 1. Primers used in RT-PCR reactions and cloning.

RT-PCR primers	
GmPPCK-F	GAGTACGATGAGAAGGTTGATGTGT
GmPPCK1-R	ATCCGAAACTGAAACTCTTTAAAC
GmPPCK2-R	CACTACTACTACAGCATGTAGCTAC
GmPPCK3-R	AAAATCTACAGAAGCGTGCTCTGTTCC
GmPPCK4-F	CAAGATCGTTCAGCTTCTCTCC
GmPPCK4-R	AAAAACACGAGTCGGAAACCT
GmUbiquitin-F	GTACCCTCGCCGACTACAAC
GmUbiputin-R	GGGTCCTTCCATCCTCTAGC
GmCAB-F	TTGGTGTTGGAAGGGTTAGC
GmCAB-R	TCTATCCAGGTGGCAGCTTT
GmNod26-F	AACATGATAACATTTCCTGGGATAG
GmNod26-R	GTGCCAACAAATATATCCATATTCC
GmpLHY-F	ATACAGCTCCATCGAACACAAGT
GmpLHY-R	CTGTAGTCACTGATGCAGGACAG
GmPEPC7-F	ATGGTTGCAGCACTGAAAGTTA
GmPEPC7-R	TTCCTTGTTATTTCTCCCCTCA
GmPEPC4-F	TCGAAGAAGGATCAAGTTACTGAAGA
GmPEPC4-R	CCATGAAACAATCACAACACACA
GmPEPC1-F	GTTTGTTTGAATACTGAAAGTG
GmPEPC1-R	GGAATTCTAGAGCAGACTACCAAA
GmPEPC16-F	TGTCTTCTCTTTGGTTCCACTTTT
GmPEPC16-R	GACTTTCAGCCTAATGACAACAAA
GmPEPC17-F	TACAAGCCTTCCACAATTCACATA
GmPEPC17-R	GGAAAAACTTAATTATTTTGCATC
TC178069-F	CCTGTGTTGCCTTCCTATACAAGT
TC178069-R	AGAATCCAGTTAAAATCCAAGCTG
LePPCK1-F	TCAAATTTGCGAAGAAATCG
LePPCK1-R	CCTTCTCCTTCTCTCTTCCACA
LePPCK2-F	GAGCTATTCGCCGTCAAGTC
LePPCK2-R	AAATCAGCCAATTTCAGTTCG
LeUbiquitin-F	GACAACGTCAAGGCTAAGATCC
LeUbiquitin-R	TTACCAGCGAAAATCAACCTCT
Degenerate nested PCR primers	
GmPEPC17 degenerate-F	ATGACGGCAACNACNGAYGAYATCGC
GmPEPC17-R1	AAGGGTGAACATATGTGAATTGTG
GmPEPC17-R2	GAAGCAAGGTCAACCTGTATTCTT
Primers for cloning	
GmPPCK1-F	ATGTATAGAGAAGCAGCGGTGAAGA
GmPPCK1-R	ATCCGAAACTGAAACTCTTTAAAC
GmPPCK2-F	GGAGTCGGAATTGAATAGAATGTGGT
GmPPCK2-R	CACTACTACTACAGCATGTAGCTAC
GmPPCK3-F	GCTAAGAAAGAAAAGAGAAGAAACGAA

GmPPCK3-R	AAAATCTACAGAAGCGTGCTCTGTTCC
GmPPCK4 full-length F	ATCACAGTGGAAACCAGTGAAGT
GmPPCK4 full-length R	TACAGAGAGATCAACCACCTCCA
GmPEPC4-F	TGGTTCATATTGGTTGGTTCAG
GmPEPC4-R	CCATGAAACAATCACAACACACA
GmPEPC17-F	TTTCTCCACCTTTCTTGCTTTC
GmPEPC17-R	GGAAAAACTTAATTATTTTGCATC
StPPCK2-F	GAGTCATTCGCCGTCAAGTC
StPPCK2-R	AAATCAGCCAATTTCAGCTCG

2.4.2 Polymerase chain reaction

Polymerase chain reaction (PCR) experiments were performed using a PTC-220 DNA Engine DYAD Peltier thermo cycler (MJ Research). Unless otherwise stated the reaction mixture (usually 25 µl) contained DNA or cDNA template, 1x ReddyMix PCR master mix with 1.5 mM $MgCl_2$ (Abgene, Epsom, UK) and 0.5 µM of each primer. Amplification was performed using a suitable number of cycles consisting of an initial denaturation step of 2 min at 94°C, then each cycle would contain a denaturing step of 94°C for 20 s, an annealing step for 30 s usually at a temperature 5°C below the predicted primer melting temperature, and an extension step at 72°C for 1 min per kb of product length. A final extension step at 72°C for 5 min was then performed before the reactions were maintained at 4°C. The reaction products were then resolved by electrophoresis on an agarose gel and stained with ethidium bromide.

2.4.3 Genome Walking

To obtain upstream and downstream genomic sequence information from an area of known sequence the Universal GenomeWalker kit (BD Biosciences, U.K.) was used.

2.4.4 RACE-PCR

5' and 3' RACE-PCR was carried out using the GeneRacer RACE kit (Invitrogen, U.K.).

2.4.5 RT-PCR

Total RNA samples (2.5 µg) were incubated with 0.25 µM oligo dT for 10 min at 70°C and then cooled to 4°C. Reverse transcription was carried out in a

reaction mixture (25 µL) containing 1x AMV reverse transcriptase buffer, 1 mM dNTPs, 1 U μl^{-1} RNase inhibitor and 0.4 U μL^{-1} AMV reverse transcriptase. The reaction was performed at 48°C for 45 min. The enzyme was then heat-inactivated at 95°C for 5 min and the samples cooled to 4°C before they were used for PCR.

PCR reactions were performed with 2.0 µl of cDNA in a reaction mixture (25 µl) containing 0.5 µM of each primer and 1x ReddyMix PCR master mix with 1.5 mM $MgCl_2$ (Abgene, Epsom, UK). The reaction products were then resolved by electrophoresis on an agarose gel and stained with ethidium bromide. RT-PCR experiments were optimised by performing PCR reactions over a range of amplification cycles in order to ensure that the reactions were within the linear range of amplification. A suitable number of cycles was chosen which allowed detection of transcripts while still revealing any variation in expression.

2.5 Determination of PEPc kinase activity

2.5.1 *In vitro* transcription

Full-length cDNA or genomic PEPc kinase clones were used as templates for *in vitro* transcription from the T3 or T7 priming site using the mMessage mMachine kit (Ambion, Huntingdon, UK). The plasmid DNA carrying the clone was first linearised by endonuclease digestion in the multiple cloning site after the 3' end of the insert. The digested DNA fragment was then resolved by electrophoresis on an agarose gel stained with ethidium bromide and the band excised with a sterile razor blade. The DNA was purified from the gel slice using the QIA-quick gel extraction kit (Qiagen, U.K.). The quantity and purity of the RNA was determined spectrophotometrically and the size of the RNA transcripts determined by denaturing agarose gel electrophoresis and ethidium bromide staining.

2.5.2 *In vitro* translation

A rabbit reticulocyte lysate system (Amersham, U.K.) was used for *in vitro* translation of the RNA from the *in vitro* trascription reactions according to the manufacturer's protocols and Hartwell *et al.* (1996), using Redivue [^{35}S]-methionine as the labelled amino acid. Incubations were for 45 min at 30 °C.

2.5.3 PEPc kinase activity in *in vitro* translation products

The activity of PPCK in the *in vitro* translated products was assayed (Hartwell *et al.*, 1996) by incubating 5 µl of translation products in a total volume of 25 µl with 50 mM Tris/HCl, pH 7.8 containing 3 mM $MgCl_2$, 1 mM benzamidine hydrochloride, antipain (10 µg ml^{-1}), leupeptin (10 µg ml^{-1}), 5 nM okadaic acid, 0.03 units of purified dephosphorylated *K. fedtschenkoi* PEPc and 10 µCi [γ-^{32}P]ATP. After 30 min incubation at 30°C, the reaction was halted by adding 10 µl of polyclonal rabbit anti-*K. fedtschenkoi* PEPc antibody (Nimmo *et al.*, 1986) and incubating, upon ice, for 1 h. The immunoprecipitated PEPc was pelleted by centrifugation at 13000 rpm in a microfuge for 10 min. The supernatant was discarded and the pellet was washed in 200 µl of 1.3 M NaCl, 2 mM EDTA (pH 7.0). The PEPc was re-pelleted by centrifugation at 13000 rpm in a microfuge for 10 min, and re-suspended in 30 µl of 5x SDS sample buffer, followed by 4 min incubation at 100°C. ^{32}P incorporation into PEPc was measured by SDS-PAGE and autoradiography or phosphorimaging.

2.6 DNA cloning, transformation and culture of *Eschericia coli*

2.6.1 DNA cloning and transformation of competent *E. coli*

PCR amplified DNA inserts were cloned into pCR4-TOPO vectors and transformed into TOP10 chemically competent *E. coli* cells using a TOPO TA Cloning Kit (Invitrogen Life Sciences), according to the manufacturer's protocols.

2.6.2 Culture of *E. coli* strains possessing plasmids containing cloned cDNA/gDNA inserts

As outlined in (Sambrook *et al.*, 1989), transformed cells (10 µl and 90 µl) were plated out onto LB broth plates (1% (w/v) Bactotryptone, 0.5% (w/v) Bactoyeast extract, 1% (w/v) NaCl, 1.5% (w/v) Bactoagar, pH 7.5) containing 100 µg/ml ampicillin or 50 µg/ml kanamycin (filter sterilised and added after autoclaved media had cooled to ~ 45°C), and then incubated overnight at 37°C.

Single colonies were "picked" from the selection plates by using a sterile loop. The colony was used to inoculate 10 ml of LB broth (1% (w/v) Bactotryptone, 0.5%

(w/v) Bactoyeast extract, 1% (w/v) NaCL, pH 7.5) containing 100 µg/ml ampicillin or 50 µg/ml kanamycin. The culture was allowed to grow overnight in a shaking incubator (at 37°C and 200 rpm). Glycerol stocks were made by mixing 700 µl of the overnight culture with 300 µl of sterile glycerol. The mixture was vortexed and flash-frozen in liquid nitrogen before storage at –70°C. This method was adapted from Sambrook *et al.* (1989).

2.6.3 DNA Sequencing

DNA vector inserts were sequenced by MWG-Biotech using universal primers after plasmid DNA extraction and purification.

2.7 Gel electrophoresis techniques

2.7.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins

Proteins were separated by discontinuous SDS-PAGE according to the method originally outlined by Laemmli (1970). All SDS-PAGE gels used in this work contained an 8% or 12.5% polyacrylamide separating gel and a 3 % stacking gel. Samples were denatured by adding one fifth volume of 5 x concentration sample buffer (125 mM Tris-HCl, pH 6.8, 2.5% (w/v) SDS, 25% (v/v) glycerol, 0.05% Bromophenol Blue and 2.5% (v/v) 2-mercaptoethanol to achieve a final concentration of 1 x and boiling the samples in a water bath or heating block for 4 min. Samples were then loaded onto the gel and electrophoresed at 60-70 mA until the tracking dye had left the bottom of the separating gel.

2.7.2 SDS-PAGE gel staining

The stacking gel was removed and discarded and the separating gel was stained for 30 min. at 37°C in 0.1 % Coomassie Brilliant Blue G250, 50% (v/v) methanol, 10% (v/v) glacial acetic acid and destained in several changes of 10% (v/v) methanol, 10% (v/v) glacial acetic acid at 37°C.

2.7.3 Drying, autoradiography and phosphor imaging of SDS-PAGE gels

SDS-PAGE gels were dried onto Whatmann 3MM chromatography paper using a Scie Plas Gel Drier (model GD4535) connected to Vacuubrand MZ C2

vacuum pump. Dried radioactive gels and blots were exposed onto X-ray film (Fuji RX) using two intensifying screens at -70°C for 1-3 days. X-ray film was developed using a Kodak X-OMAT Processor (model ME-3). Radioactive gels and blots were also routinely phosphor imaged using a Fuji FLA-5000 phosphor imager (Fuji Photo Film Co. Ltd., Japan) by exposing the gel to a pre-blanked imaging plate for 1-24 h in a cassette at room temperature. Exposed plates were developed automatically by inserting them in the Fuji FLA-5000 phosphor imager and the images were captured onto an Apple Power Mac G4 computer running Image Reader and Image Gauge software (Fuji Photo Film Co. Ltd., Japan).

2.7.4 Agarose gel electrophoresis of RNA and DNA

Samples of purified DNA and RNA were routinely checked for integrity and molecular weight distribution using horizontal agarose gel electrophoresis. 0.8-2.0 % (w/v) agarose gels were prepared and run in 0.5 x TBE (45 mM Tris-borate, pH 8.0, 1 mM EDTA) buffer containing 0.25 µg/ml ethidium bromide. Samples of DNA or RNA were mixed with 6 x Blue/Orange loading buffer (Promga, U.K.) to facilitate loading into the gel, and electrophoresed at 5 V/cm for 1-2 h depending on the dimensions of the gel. For DNA samples, 1 Kb or 100 bp ladder (Promega,U.K.) was run alongside the samples to allow the subsequent estimation of the molecular weight of the sample DNA. Ethidium stained agarose gels were visualised under ultraviolet light using a Bio-Rad Gel-Doc 2000 connected to an Apple Power Mac G4 computer running Quantity One software (Bio-Rad Laboratories, CA)

2.7.5 Denaturing agarose gel electrophoresis of RNA

RNA samples were routinely checked for integrity and molecular weight distribution by electrophoresis through denaturing formaldehyde/MOPS agarose gels. This method ensures accurate separation of RNA according to size as the denaturing conditions remove secondary structure that can cause RNA to run at aberrant molecular weights in non-denaturing agarose gels.

Gels (0.8-1.5 % (w/v) agarose) were prepared in 1 x MOPS buffer, pH 8.0 (10 x MOPS buffer: 200 mM MOPS, pH 8.0, 50 mM sodium acetate and 10 mM EDTA) and 10 % (v/v) formaldehyde. The agarose was dissolved in water by microwaving. After cooling to approximately 60°C the formaldehyde and 10 x MOPS were added in a fume hood and mixed thoroughly. The gel was poured and left to set in the fume

hood. Gels were run in 1 x MOPS buffer, pH 7.0 (10 x stock as above but with the pH adjusted to 7.0).

RNA samples were diluted to a final volume of 8 µl with dH_2O and combined with 6 µl of sample buffer (72 µg/ml ethidium bromide, 2 x MOPS, pH 8.0, 6 % (v/v) formaldehyde and 70 % (v/v) formamide) and heated to 65°C for 2.5 min. 1-2 µg of RNA size markers (Promega, UK) were treated in the same manner, and run alongside the samples to allow the later estimation of the molecular weight of RNA bands of interest. Samples were then snap cooled on ice for 5 min and loaded onto the gel.

2.8 Sequence alignments and phylogentic analysis

2.8.1 Nucleotide alignments

Pairwise nucleotide alignments were performed using LALIGN at http://www.ch.embnet.org/. The global alignment method was chosen, with all other parameters set to default values (opening gap penalty –14 and extending gap penalty –4). The multiple nucleotide alignment for Fig. 5.5 was produced using the GeneJockey II programme with all parameters set to default values (fixed gap penalty of 10, floating gap penalty of 10 and weighted DNA transitions). Translation maps of nucleotide sequences were produced using the GeneJockey II programme.

2.8.2 Amino acid alignments

Amino acid alignments were performed using the GeneJockey II programme with all parameters set to default values (fixed gap penalty of 10, floating gap penalty of 10 and PAM 250 weight matrix).

2.8.3 Phylogenetic analysis

ClustalX Multiple Sequence Alignment Program (version 1.64b) (Thompson *et al.*, 1997) was used for phylogenetic analysis. Deduced amino acid sequences were aligned using the default parameters (gap opening 10, gap extension 0.05, BLOSUM protein weight matrix selected, residue-specific penalties on, hydrophobic penalties on, gap separation distance 8 and end gap separation off) and 1000 bootstrap replicates were performed on the Neighbour-Joining tree. Trees were viewed using TreeView (Page, 1996).

Chapter 3

Identification and expression of *PEPC* and *PPCK* genes in *Glycine max*

3.1 Introduction

PEPc has been proposed to play a variety of roles in the root nodules of leguminous plants. Its major function is thought to be the provision of malate as a respiratory substrate to support bacteroid metabolism and the high-energy demands of nitrogen fixation (King *et al.*, 1986). The high levels of PEPc protein present in alfalfa and soybean root nodules, 0.5-2.0 % of the total soluble protein, indicate the importance of this enzyme in the nodule (Miller *et al.*, 1987; Zhang *et al.*, 1995) and PEPc genes showing nodule-enhanced or nodule-specific expression patterns have been reported from several legume species (Pathirana *et al.*, 1997; Suganuma *et al.*, 1997; Hata *et al.*, 1998; Nakagawa *et al.*, 2003a). With the work from CAM and C_4 plants demonstrating the importance of reversible phosphorylation in modulating PEPc activity, research focused on characterising PEPc kinase from root nodules.

Zhang and Chollet, (1997) partially purified two PEPc kinase polypeptides from soybean root nodules with molecular masses of 32 and 37 kDa. However at the beginning of this project no *PPCK* genes had been isolated from legume nodules. The recent cloning of *PPCK* genes from *K. fedtschenkoi* and *A. thaliana* offered the possibility of identifying *PPCK* genes from other species based on sequence similarity. Soybean was a good candidate for such a project; it is a model plant for the study of symbiotic nitrogen fixation and much of the early work investigating PEPc in nodules used soybean, it is an important C_3 crop species and a large collection of expressed sequence tag (EST) sequences were available. A similar method of sequence similarity was also employed to identify additional PEPc genes from soybean. Previous work on legume PEPc was focused on the identification and characterisation of the nodule-enhanced isoform. By investigating the expression patterns of previously identified but little characterised PEPc isoforms together with newly identified isoforms it was hoped a fuller picture of *PEPC* and *PPCK* genes in C_3 plants could be achieved.

During the course of this project Nakagawa *et al.* (2003a) and Xu *et al.* (2003) reported the cloning of *PPCK* genes from *Lotus japonicus* and soybean respectively. *L. japonicus* is an unusual legume because it forms determinate nodules, like soybean, but also transports the products of nitrogen fixation as amides rather than ureides, like

indeterminate legumes. Nakagawa *et al.* (2003a) cloned two *PEPC* genes and one *PPCK* gene from *L. japonicus* and expression of all three genes was investigated. Xu *et al.* (2003) isolated and characterised two *PPCK* clones from a soybean nodule cDNA library. Comparisons between these recent results and those presented in this chapter will be presented in the discussion.

3.2 Results

3.2.1 Identification of *PPCK* genes in *Glycine max*

Potential PEPc kinase genes were identified by searching soybean (EST) databases with known plant PEPc kinase amino acid sequences. Putative PEPc kinase clones (Genbank accession numbers AI736847, AW0990717, and AW756453) were sequenced completely and found to have different 3'untranslated regions (UTRs). The corresponding genes are termed *GmPPCK1*, *GmPPCK2* and *GmPPCK3*. Gene-specific primers were designed and 5' RACE was used to obtain full-length sequences of *GmPPCK1*, *GmPPCK2* and *GmPPCK3*. *GmPPCK4* was identified from EST CA783260. Table 2 gives the accession numbers of all the soybean ESTs encoding PEPc kinase and the libraries from which they were cloned. A virtual expression analysis reveals that none of the sequences were isolated from nodule libraries. This may be because only one nodule cDNA library has been used as part of the soybean EST project.

Adapter ligation PCR was used to obtain 5' and 3' sequences from genomic DNA. Full-length PCR products for all of the genes were obtained from genomic DNA and cDNA, and the sequences have been deposited in GenBank under accession numbers AY144180 to AY144185, AY568713 and AY568714. Fig. 3.1 shows the alignment of the deduced amino acid sequences. These sequences are typical of plant PEPc kinase genes. The encoded proteins comprise a Ser/Thr kinase catalytic domain similar to the catalytic region of plant calcium-dependent protein kinases but lacking the N-terminal extensions, the auto-inhibitory region and the Ca^{2+}-binding EF hands of that family. The genomic sequences revealed a single intron (81-107 bp long) close to the 3' end of the coding sequence, the position of which is conserved in all plant PEPc kinase sequences identified to date (Nimmo, 2003). Fig. 3.2 shows the translation map of the genomic sequence of GmPPCK2. The *GmPPCK3* sequence is unusual in not having an in frame stop codon in the intron, unlike almost all other *PPCK* genes. Hence unprocessed *GmPPCK3* transcripts would encode a protein

identical to GmPPCK3 over most of its length but with one amino acid alteration (the codon interrupted by the intron) and an insertion of 30 residues.

Pairwise alignments of the soybean PPCK sequences show that GmPPCK2 and GmPPCK3 are the most similar. At the nucleotide level, *GmPPCK2* and *GmPPCK3* are 94% identical in the coding region and 81% identical in the 3'UTR. *GmPPCK1* is 71% and 70% identical to *GmPPCK2* and *GmPPCK3* respectively in the coding region, and 42% and 40% identical in the 3'UTR. *GmPPCK4* is more divergent being only 66% and 67% identical to *GmPPCK2* and *GmPPCK3* respectively in the coding region.

To test the function of the gene products, full-length cDNA sequences for each *GmPPCK* and the full length *GmPPCK3* genomic DNA sequence were cloned behind a T7 promoter. The plasmids were linearised, transcribed and translated, and the translation products were assayed for PEPc kinase activity. Fig. 3.3 shows the phosphor images of the ^{35}S-labelled translation products separated by SDS-PAGE and of the ^{32}P incorporated into purified *K. fedtschenkoi* PEPc in PEPc kinase assays. All of the cDNA sequences gave rise to translation products of the expected size (31-32 kDa), while the genomic *GmPPCK3* sequence encoded a larger protein (34 kDa) owing to the translation of the intron sequence. However, while the cDNA translation products were able to phosphorylate PEPc, confirming their identification as functional PEPc kinases, the product from the genomic *GmPPCK3* sequence showed no kinase activity.

3.2.2 Expression of *PPCK* genes in soybean

Although *PPCK* transcripts in soybean nodules are sufficiently abundant to be detected by Northern analysis (Xu *et al.*, 2003), their expression in other organs was expected to be very low, as is the case in other C_3 plants such as Arabidopsis (Fontaine *et al.*, 2002). Therefore the expression of the four soybean *PPCK* genes was studied using RT-PCR (Fig 3.5). Various tissues were collected from 5 week old nodulated plants (Fig. 3.4), except flowers which were collected from 7 week old nodulated plants. As a control for the equal loading of RNA into each cDNA reaction, primers were designed to amplify a soybean ubiquitin gene (Genbank accession D16248). The use of ubiquitin as a constitutive control in legumes has been previously reported (Horvath *et al.*, 1993; Moran *et al.*, 2002). Controls for tissue contamination were also used. Nodulin-26 (*Nod26*) encodes an aquaporin channel

protein, which is found in the symbiosome membrane (Dean *et al.*, 1999). Primers were designed based on the Genbank sequence X04782. Chlorophyll a/b binding protein (*CAB*) is also a membrane protein, found in the thylakoid membrane in chloroplasts. Primers were designed based on the Genbank sequence U01964.

Of the four *GmPPCK* genes, *GmPPCK1* displays the most tissue specific expression pattern. Its expression is highest in stem and root tissue, with low levels of expression in all other tissues examined. *GmPPCK2* and *GmPPCK3* have very similar expression patterns. They are both preferentially expressed in nodule tissue, with lower expression levels in all other tissues. However slight differences are observed; *GmPPCK3* is more highly expressed in flowers and stems than *GmPPCK2*. *GmPPCK4* is also expressed in all tissues examined, with highest transcript levels in nodules, roots and flowers. As expected *CAB* is expressed only in green tissues, while *Nod26* is expressed only in nodules.

The timing of expression of the *PPCK* genes during nodule development was examined. Nodule and root tissue was collected at one week intervals starting one week after inoculation with *B. japonicum*. For weeks 1 and 2 nodule and root tissue were not separated due to the absence of nodules at 1 week post-inoculation and the small size of nodules at 2 weeks post-inoculation. For weeks 3-5 nodule and root tissue was separated. Cotyledons from one week old seedlings were also collected. The results of the RT-PCR analysis are shown in Fig. 3.6. Transcripts of *GmPPCK1* are detectable only during weeks 1 and 2, presumably due to the presence of root tissue. Transcript levels of both *GmPPCK2* and *GmPPCK3* increase during nodule development. The faint bands observed at week 1 are probably due to their low levels of expression observed in root tissue (Fig. 3.5). However by week 2, transcripts of both genes are clearly detectable. This suggests that *GmPPCK2* and *3* expression is not increased by *B. japonicum* infection but by nodule emergence. The expression of Nodulin-26 shows that bacteroids were present from week 2 onwards. *GmPPCK4* expression also increased during development. The increase in expression shown from weeks 1 to 3 probably reflect the higher levels of *GmPPCK4* expression in nodule tissue compared to root tissue (Fig. 3.5). The expression in roots from weeks 3 to 5 and in cotyledons is also shown in Fig. 3.6. The transcript levels of all four kinases remain stable in roots over this period. *GmPPCK4* appears to be the main PEPc kinase isoform expressed in cotyledons.

3.2.3 Identification and expression analysis of soybean *PEPC* genes

The presence of a nodule-enhanced PEPc isoform in legume root nodules is well documented (Pathirana *et al.*, 1997; Suganuma *et al.*, 1997; Hata *et al.*, 1998). The nodule-enhanced gene in soybean is *GmPEPC7*. Three other PEPc genes have also been identified in soybean, *GmPEPC1* (Vazquez-Tello *et al.*, 1993), also termed *GmPEPC15* (Hata *et al.*, 1998), *GmPEPC4* (Hata *et al.*, 1998) and *GmPEPC16* (Sugimoto *et al.*, 1992). Results from Northern blotting experiments showed that *GmPEPC1* and *GmPEPC16* were expressed at low levels in the tissues examined (Sugimoto *et al.*, 1992; Hata *et al.*, 1998). However, only a small amount of sequence data for *GmPEPC4* was available and its expression pattern was not examined (Hata *et al.*, 1998). Based on the partial sequence and using 5' RACE a full-length cDNA sequence of *GmPEPC4* was obtained (Genbank accession AY563044).

Querying of the TIGR soybean EST database (see http://www.tigr.org/) revealed a further three Tentative Consensus sequences with high sequence similarity to plant PEPc genes, TC192656, TC178069 and TC202374. Recently it was shown that *Arabidopsis thaliana* contains a PEPc gene termed *Atppc4* encoding a putative protein that resembles bacterial PEPcs much more than higher plant PEPcs; for example it lacks the N-terminal phosphorylation site present in other plant PEPc genes (Sanchez and Cejudo, 2003). One of the soybean Tentative Consensus sequences, TC192656, closely resembles *Atppc4*. Using a degenerate 5' primer based on the sequence of *Atppc4* and a 3' primer based on TC192656, an RT-PCR product was isolated, termed *GmPEPC17*, and a full-length sequence was obtained by 5' RACE (Genbank accession AY563043). Fig. 3.7 shows an alignment of the deduced amino acid sequences of the full length soybean *PEPC* genes and Fig 3.8 shows the deduced amino acid sequence of *GmPEPC17* compared with the Arabidopsis homologue *Atppc4*, *Atppc1*, which is expressed in all organs (Sanchez and Cejudo, 2003), and *GmPEPC7*, which encodes the nodule-enhanced PEPc. As expected, *GmPEPC17* is much more similar to *Atppc4* than to the other soybean PEPc sequences, and lacks the N-terminal phosphorylation site.

Primers were designed to the 3'UTR regions of *GmPEPC1*, *GmPEPC7*, *GmPEPC16*, *GmPEPC17* and *TC178069* for RT-PCR analysis. For *GmPEPC4* the forward primer was in the 5' portion of the coding region and the reverse primer bound to the 3'UTR. The expression of TC202374 was not investigated due to the small amount of 3'UTR sequence available and the high degree of similarity to

TC178069 in the coding region, preventing the design of gene-specific primers. The tissue expression pattern of the soybean *PEPC* genes is presented in Fig 3.9. The data confirm the preferential expression of *GmPEPC7* in nodules, and the broad but low-level expression pattern of *GmPEPC16* and *GmPEPC1*, although *GmPEPC1* is not detectably expressed in leaves. TC178069 showed a broad expression pattern while both *GmPEPC4* and *GmPEPC17* were expressed mainly in above ground organs.

3.3 Discussion

The results in this chapter demonstrate the presence of four *PPCK* genes in soybean, the transcribed cDNA products of which are all able to phosphorylate purified *K. fedtschenkoi* PEPc in an *in vitro* kinase assay. All previously identified genomic *PPCK* sequences contained a stop codon in the intron. Fontaine *et al.* (1999) demonstrated that the translated product of the genomic *AtPPCK2* sequence was non-functional. This proved that the conserved His.X.aromatic.hydrophobic sequence is required for kinase activity. The conserved presence of a stop codon in the intron could suggest its function was to prevent unspliced transcripts from producing a protein with undesirable properties. The lack of an in-frame stop codon in the intron of *GmPPCK3* allowed us to test the effect of the extra amino acids encoded by the intron on kinase activity. The genomic *GmPPCK3* translation products had no detectable kinase activity *in vitro*, suggesting that unspliced transcripts would not be functional *in vivo*.

The results of the RT-PCR analysis revealed that two *PPCK* genes showed nodule-enhanced expression (Fig. 3.5). This was an unexpected result, as our working hypothesis was that *PPCK* expression would mirror *PEPC* expression in having a single nodule-enhanced isoform. The recent work of Xu *et al.* (2003) supports such a hypothesis. Xu *et al.* (2003) reported the sequence and expression pattern for two soybean *PPCK* genes. These are a nodule-enhanced gene termed *NE-PpcK*, identical to *GmPPCK2* in this work, and a more broadly expressed gene termed *GmPpcK*, identical to our *GmPPCK1*. However, the results presented in this chapter reveal a more complex picture with both *GmPPCK2* and *GmPPCK3* showing nodule enhanced expression patterns. *GmPPCK2* and *3* are very similar at the sequence level and their expression patterns are also similar, but not identical. However, it is not possible to ascribe different functions to the two genes based on expression data. Xu *et al.* (2003) suggested that the expression of *NE-PpcK* and *GmPEPC7* (the soybean

nodule-enhanced PEPc isoform) argued for the existence of *PPCK-PEPC* 'expression-partners.' In that regard, it is noteworthy that *GmPPCK2* and *GmPPCK3* are expressed in immature and mature leaves as well as in nodules, whereas *GmPEPC7* is not (Fig. 3.9). Hence, if the target of the GmPPCK2 and 3 in nodules is GmPEPC7, their targets must be different in leaf compared to nodule tissue.

Nakagawa *et al.* (2003a) recently reported the cloning of two PEPc genes and a PPCK gene from a root nodule cDNA library of *L. japonicus*. One of the PEPc genes, *LjPEPC1*, was preferentially expressed in nodules, while the other, *LjPEPC2*, was mainly expressed in shoots and roots at low levels. Northern blot analysis of the expression of the isolated PPCK gene, *LjPEPC-PK*, showed transcript levels highest in nodules but also present in shoots and roots. Nakagawa *et al.* (2003a) also investigated the *in situ* localisation of *LjPEPC1* and *LjPEPC-PK* transcripts in both mature and rapidly growing immature nodules. In mature nodules *LjPEPC1* and *LjPEPC-PK* transcripts were present in all cell types with the strongest signal in the vascular bundles of infected cells. However, in immature nodules *LjPEPC1* transcripts were located at the base of the nodules, while *LjPEPC-PK* transcripts were present in the central and upper portions of the infected zone. Xu *et al.* (2003) also investigated the *in situ* localisation of *NE-PpcK* in mature soybean nodules. *NE-PpcK* transcripts were present in both infected and uninfected cells in the central zone of the nodule but not the vascular bundles or the cortex. Clearly a more detailed analysis of the expression of *GmPPCK2* and *3* in root nodules is required, and may shed some light on the possible roles of these genes. Further analysis of other legumes such as *Lotus* and *Medicago* to reveal whether they, like soybean, also contain two nodule-enhanced PPCK genes would indicate if two nodule-enhanced PPCK genes is a requirement for nodule functioning. However, the selective disruption of the genes individually and together, through strategies such as anti-sense or insertional mutagenesis, would allow an evaluation of the importance of each gene and an assessment of the degree of functional redundancy.

The expression of *GmPPCK1* is greatest in stems and roots. This pattern is somewhat different from the data reported by Xu *et al.* (2003) in which the highest expression of *GmPPCK1* was observed in seeds, flowers and leaves. Their presented expression data was obtained by Northern blot experiments, however they reported that similar results were also obtained with RT-PCR. Therefore the RT-PCR experiment was repeated using the primers reported by Xu *et al.* (2003) and the results

are shown in Fig. 3.10. These results are very similar to those in Fig. 3.5. Therefore the differences between our data and those of Xu *et al.* (2003) are not due to differences in experimental technique but from the samples examined. There is no obvious explanation for the disparity between the sets of results but the differences may be due to growth conditions and/or a cultivar difference.

The analysis of PEPc genes in soybean revealed the existence of a "bacterial-type" PEPc gene. Sanchez and Cejudo (2003) identified these novel PEPc isoforms, termed *Atppc4* and *Osppc-b*, in the genomes of Arabidopsis and rice. Phylogenetic analysis showed that both *Atppc4* and *Osppc-b* were more similar to PEPc sequences from γ-proteobacteria, such as *E. coli*, than to PEPc sequences from plants, including other PEPc isoforms from Arabidopsis and rice. These bacterial-type PEPc isoforms lack the conserved Ser close to the N-terminus that is the target of PEPc kinase (Sanchez and Cejudo, 2003) and therefore are probably not regulated by phosphorylation. *Atppc4* is expressed exclusively in siliques, flowers and roots (Sanchez and Cejudo, 2003), whereas *GmPEPC17* shows a much more uniform pattern of expression (Fig. 3.9). This expression pattern offers no obvious clues to the function of the bacterial-type PEPc. However, EST database searches have revealed the presence of several other potential homologues of *Atppc4*, *Osppc-b* and *GmPEPC17* in both monocotyledonous and dicotyledonous species (not shown) leading to the conclusion that bacterial-type PEPc genes are a feature of many plant species.

Table 2 Accession numbers of soybean *PPCK* ESTs.

Genbank Accession Number	cDNA Library
GmPPCK1	
AI736847	Apical shoots of 9-10 day old etiolated seedlings
AW100476	Immature flowers of field grown plants
AW620843	Root tips of 7 day old seedlings.
BE659798	Reracked set of clones from immature cotyledons, immature seed coats, flowers and young seed pods. Contains intron.
BI784996	Roots of 8 day old Bragg seedlings, mock-infected 72 hours prior to harvesting. Contains intron.
BG507686	Stem tissue of 1 month old greenhouse grown plants.
CA852232	Root tissue from 6 and 8 day old seedlings.
BG790358	Epicotyls of 2 week old seedlings treated with 100 ppm solution of auxin 24 hours prior to harvesting. Contains intron.
GmPPCK2	
AW0990717	Apical shoots of 9-10 day old etiolated seedlings.
BU927033	Somatic embryos (2-9 months old) grown on MSD20.
GmPPCK3	
AW278795	Immature seed coats of greenhouse grown plants. 52bp deletion in coding region.
AW755870	Somatic embryos (2-9 months old) grown on MSD20.
AW756033	Somatic embryos (2-9 months old) grown on MSD20.
AW756453	Somatic embryos (2-9 months old) grown on MSD20.
BE822598	Reracked set of clones from immature cotyledons, immature seed coats, flowers and young seed pods. AW278795 also on clone.
BQ134073	Whole seedling of 1 week old greenhouse grown plants.
BU547915	Reracked set of clones from germinating cotyledons, immature seed coats, tissue culture derived somatic embryos, germinating seedlings, shoot tips, or leaves exposed to various stresses and young leaves exposed to bacterial and fungal pathogens. AW756453 also on clone.
CA819910	Whole seedling of 1 week old greenhouse grown plants.
GmPPCK4	
CA783260	Somatic embryos (2-9 months old) grown on MSD20.

Figure 3.1 Sequence analysis of the soybean *PPCK* genes

Alignment of deduced amino acid sequences of the soybean *PPCK* genes. Dots and vertical lines denote identical and similar residues respectively.

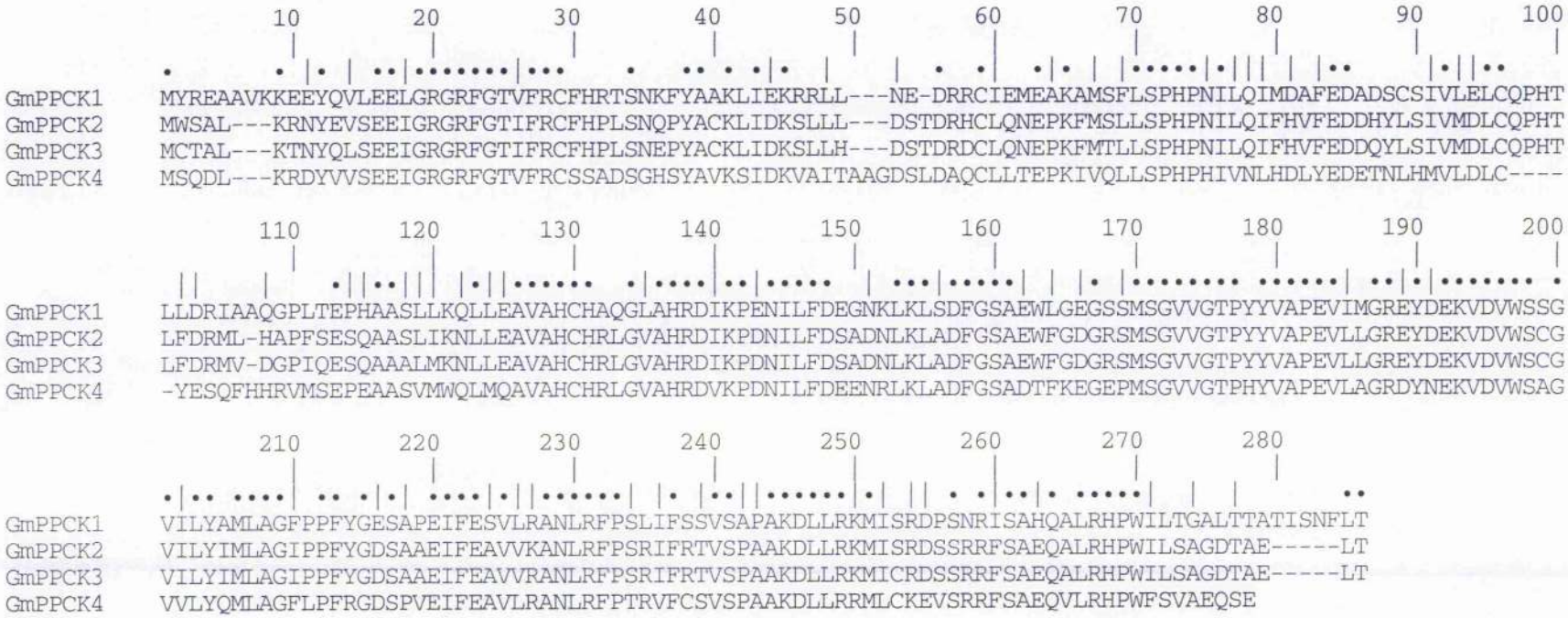

10 20 30 40 50 60 70 80 90 100
GmPPCK1 MYREAAVKKEEYQVLEELGRGRFGTVFRCFHRTSNKFYAAKLIEKRRLL---NE-DRRCIEMEAKAMSFLSPHPNILQIMDAFEDADSCSIVLELCQPHT
GmPPCK2 MWSAL---KRNYEVSEEIGRGRFGTIFRCFHPLSNQPYACKLIDKSLLL---DSTDRHCLQNEPKFMSLLSPHPNILQIFHVFEDDHYLSIVMDLCQPHT
GmPPCK3 MCTAL---KTNYQLSEEIGRGRFGTIFRCFHPLSNEPYACKLIDKSLLH---DSTDRDCLQNEPKFMTLLSPHPNILQIFHVFEDDQYLSIVMDLCQPHT
GmPPCK4 MSQDL---KRDYVVSEEIGRGRFGTVFRCSSADSGHSYAVKSIDKVAITAAGDSLDAQCLLTEPKIVQLLSPHPHIVNLHDLYEDETNLHMVLDLC----
110 120 130 140 150 160 170 180 190 200
GmPPCK1 LLDRIAAQGPLTEPHAASLLKQLLEAVAHCHAQGLAHRDIKPENILFDEGNKLKLSDFGSAEWLGEGSSMSGVVGTPYYVAPEVIMGREYDEKVDVWSSG
GmPPCK2 LFDRML-HAPFSESQAASLIKNLLEAVAHCHRLGVAHRDIKPDNILFDSADNLKLADFGSAEWFGDGRSMSGVVGTPYYVAPEVLLGREYDEKVDVWSCG
GmPPCK3 LFDRMV-DGPIQESQAAALMKNLLEAVAHCHRLGVAHRDIKPDNILFDSADNLKLADFGSAEWFGDGRSMSGVVGTPYYVAPEVLLGREYDEKVDVWSCG
GmPPCK4 -YESQFHHRVMSEPEAASVMWQLMQAVAHCHRLGVAHRDVKPDNILFDEENRLKLADFGSADTFKEGEPMSGVVGTPHYVAPEVLAGRDYNEKVDVWSAG
210 220 230 240 250 260 270 280
GmPPCK1 VILYAMLAGFPPFYGESAPEIFESVLRANLRFPSLIFSSVSAPAKDLLRKMISRDPSNRISAHQALRHPWILTGALTTATISNFLT
GmPPCK2 VILYIMLAGIPPFYGDSAAEIFEAVVKANLRFPSRIFRTVSPAAKDLLRKMISRDSSRRFSAEQALRHPWILSAGDTAE-----LT
GmPPCK3 VILYIMLAGIPPFYGDSAAEIFEAVVRANLRFPSRIFRTVSPAAKDLLRKMICRDSSRRFSAEQALRHPWILSAGDTAE-----LT
GmPPCK4 VVLYQMLAGFLPFRGDSPVEIFEAVLRANLRFPTRVFCSVSPAAKDLLRRMLCKEVSRRFSAEQVLRHPWFSVAEQSE

Figure 3.2 Translation map of the genomic sequence of *GmPPCK2*

DNA sequence in black denotes exon sequence, DNA sequence in red denotes intron sequence.

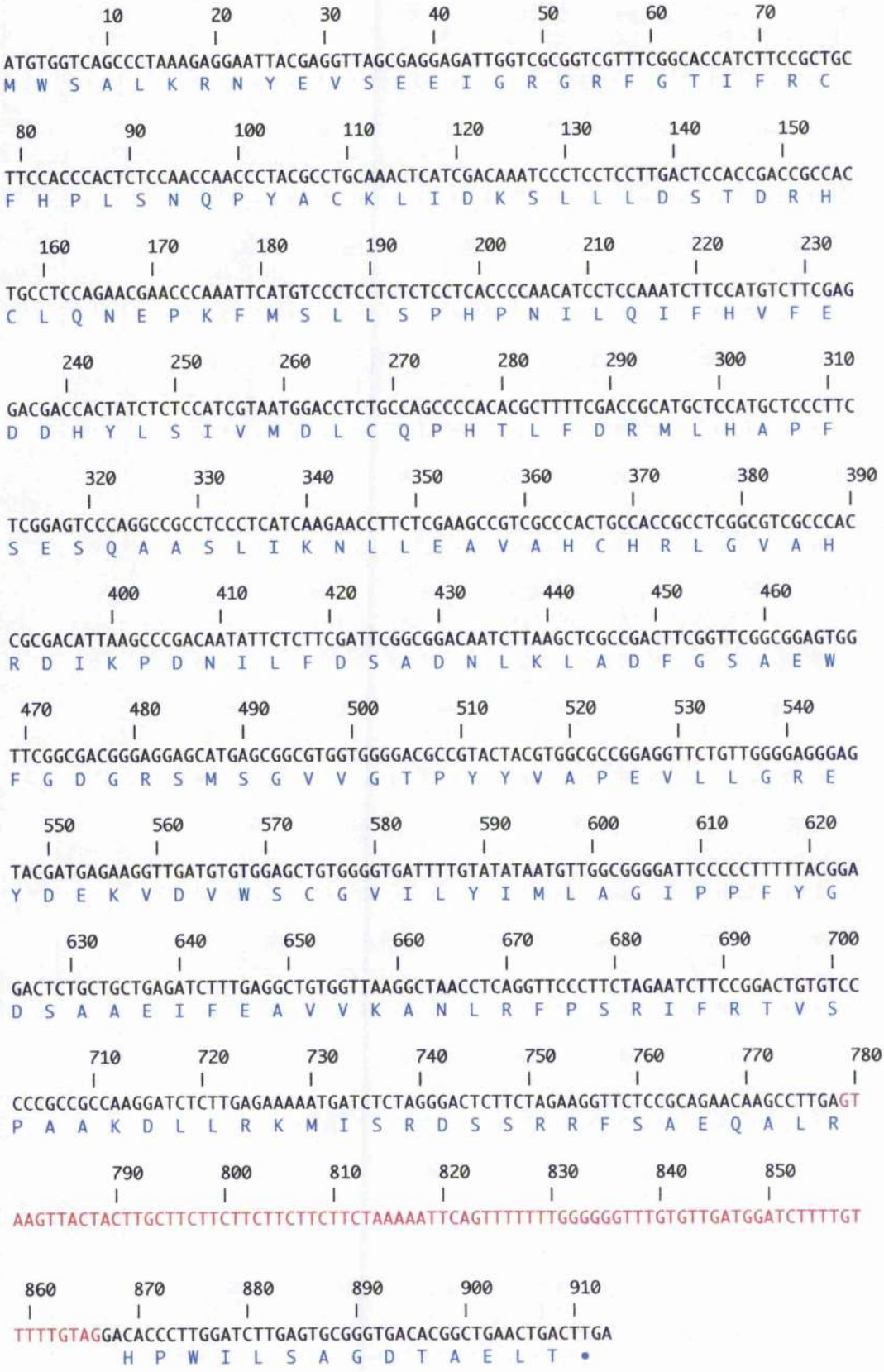

10 20 30 40 50 60 70
ATGTGGTCAGCCCTAAAGAGGAATTACGAGGTTAGCGAGGAGATTGGTCGCGGTCGTTTCGGCACCATCTTCCGCTGC
M W S A L K R N Y E V S E E I G R G R F G T I F R C
80 90 100 110 120 130 140 150
TTCCACCCACTCTCCAACCAACCCTACGCCTGCAAACTCATCGACAAATCCCTCCTCCTTGACTCCACCGACCGCCAC
F H P L S N Q P Y A C K L I D K S L L L D S T D R H
160 170 180 190 200 210 220 230
TGCCTCCAGAACGAACCCAAATTCATGTCCCTCCTCTCTCCTCACCCCAACATCCTCCAAATCTTCCATGTCTTCGAG
C L Q N E P K F M S L L S P H P N I L Q I F H V F E
240 250 260 270 280 290 300 310
GACGACCACTATCTCTCCATCGTAATGGACCTCTGCCAGCCCCACACGCTTTTCGACCGCATGCTCCATGCTCCCTTC
D D H Y L S I V M D L C Q P H T L F D R M L H A P F
320 330 340 350 360 370 380 390
TCGGAGTCCCAGGCCGCCTCCCTCATCAAGAACCTTCTCGAAGCCGTCGCCCACTGCCACCGCCTCGGCGTCGCCCAC
S E S Q A A S L I K N L L E A V A H C H R L G V A H
400 410 420 430 440 450 460
CGCGACATTAAGCCCGACAATATTCTCTTCGATTCGGCGGACAATCTTAAGCTCGCCGACTTCGGTTCGGCGGAGTGG
R D I K P D N I L F D S A D N L K L A D F G S A E W
470 480 490 500 510 520 530 540
TTCGGCGACGGGAGGAGCATGAGCGGCGTGGTGGGGACGCCGTACTACGTGGCGCCGGAGGTTCTGTTGGGGAGGGAG
F G D G R S M S G V V G T P Y Y V A P E V L L G R E
550 560 570 580 590 600 610 620
TACGATGAGAAGGTTGATGTGTGGAGCTGTGGGGTGATTTTGTATATAATGTTGGCGGGGATTCCCCCTTTTTACGGA
Y D E K V D V W S C G V I L Y I M L A G I P P F Y G
630 640 650 660 670 680 690 700
GACTCTGCTGCTGAGATCTTTGAGGCTGTGGTTAAGGCTAACCTCAGGTTCCCTTCTAGAATCTTCCGGACTGTGTCC
D S A A E I F E A V V K A N L R F P S R I F R T V S
710 720 730 740 750 760 770 780
CCCGCCGCCAAGGATCTCTTGAGAAAAATGATCTCTAGGGACTCTTCTAGAAGGTTCTCCGCAGAACAAGCCTTGAGT
P A A K D L L R K M I S R D S S R R F S A E Q A L R
790 800 810 820 830 840 850
AAGTTACTACTTGCTTCTTCTTCTTCTTCTAAAAATTCAGTTTTTTTGGGGGGTTTGTGTTGATGGATCTTTTGT
860 870 880 890 900 910
TTTTGTAGGACACCCTTGGATCTTGAGTGCGGGTGACACGGCTGAACTGACTTGA
H P W I L S A G D T A E L T •

Figure 3.3 Functional analysis of full-length *GmPPCK* clones

(A) Phosphor images of [^{35}S] Met-labelled products from *in vitro* translation of RNA samples separated on a 12.5% SDS-PAGE gel.

(B) Phosphor images of immunoprecipitated, ^{32}P-labelled PEPc from assays of the PEPc kinase activity of translation products separated on an 8% SDS-PAGE gel. Lanes are: 1, *GmPPCK1* cDNA; 2, *GmPPCK2* cDNA; 3, *GmPPCK3* genomic clone; 4, *GmPPCK3* cDNA; 5, no RNA control and 6, *GmPPCK4* cDNA. The arrows indicate the two subunit types of *K. fedtshenkoi* PEPc (Nimmo *et al.*, 1986).

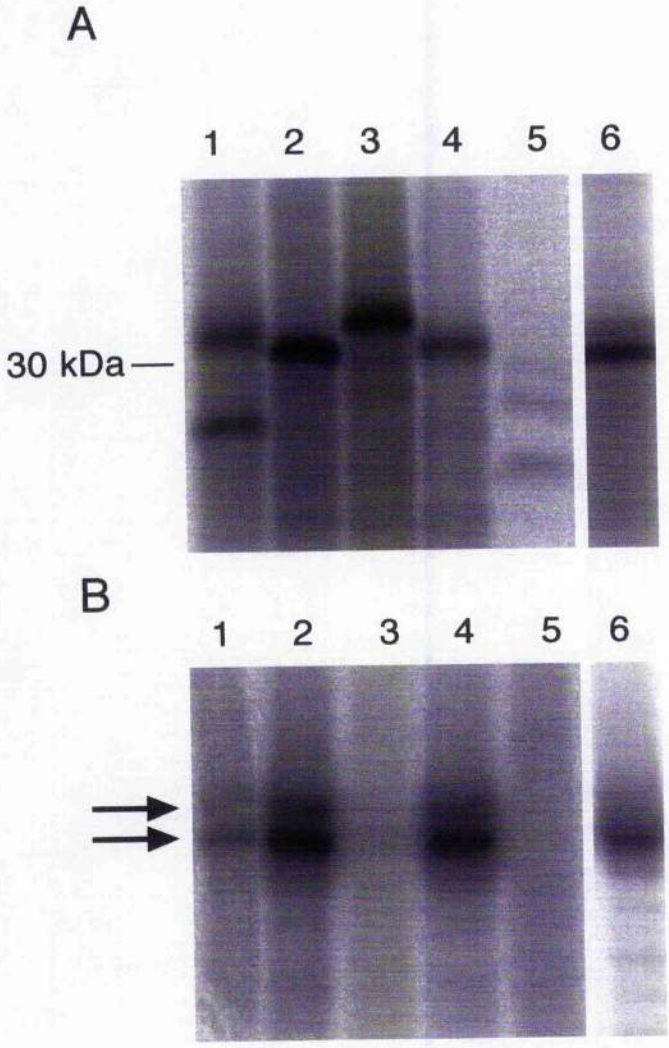

A
1 2 3 4 5 6
30 kDa
B
1 2 3 4 5 6

Figure 3.4 Picture of a 5 week old nodulated soybean plant

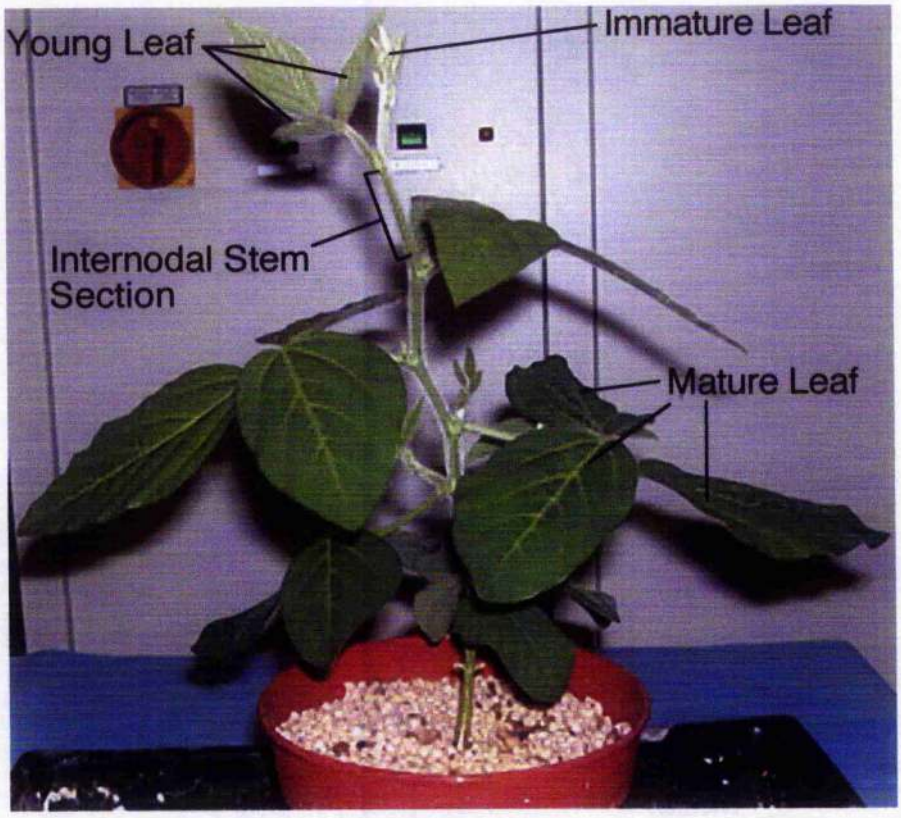

Young Leaf
Immature Leaf
Internodal Stem Section
Mature Leaf

Figure 3.5 Expression patterns of *GmPPCK* genes

This shows the products obtained from RT-PCR analysis of *GmPPCK* genes from nodulated soybean plants. Ubiquitin was used as a control for equal loading. 27 cycles of amplification were used for all GmPPCK genes and 20 cycles for Ubiquitin, CAB and Nodulin 26. All tissues were collected from a single plant. The presented results are representative of at least three biological replicates (indiviual plants) and three RT-PCR experiments on the same RNA samples.

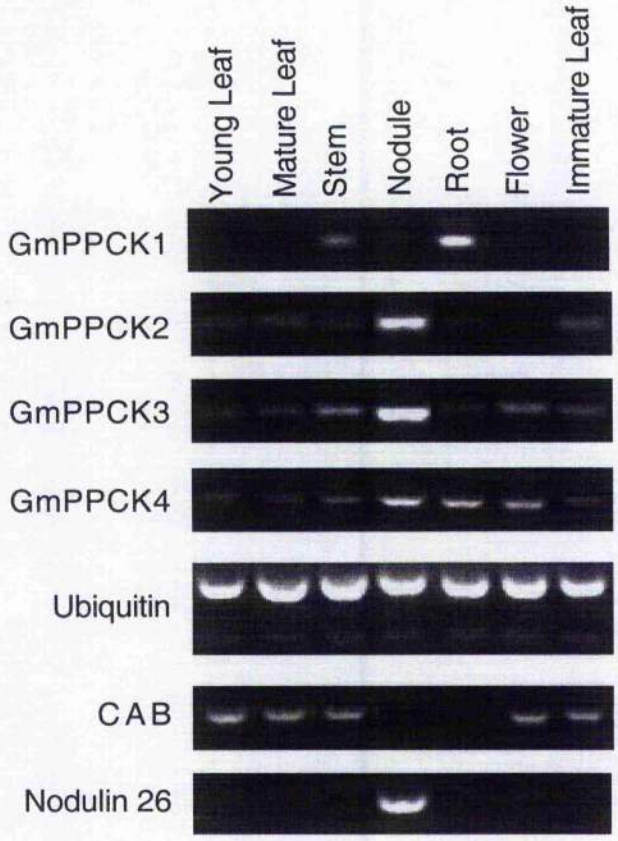

Young Leaf
Mature Leaf
Stem
Nodule
Root
Flower
Immature Leaf
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
Ubiquitin
CAB
Nodulin 26

Figure 3.6 Expression of *GmPPCK* genes during development

This shows the products obtained from RT-PCR analysis of *GmPPCK* genes from nodulated soybean plants. Ubiquitin was used as a control for equal loading. 28 cycles of amplification were used for GmPPCK1, 27 cycles for GmPPCK2, 3 and 4, and 20 cycles for Ubiquitin and Nodulin 26. For each sample tissue was collected from three plants, which were mixed prior to RNA extraction.

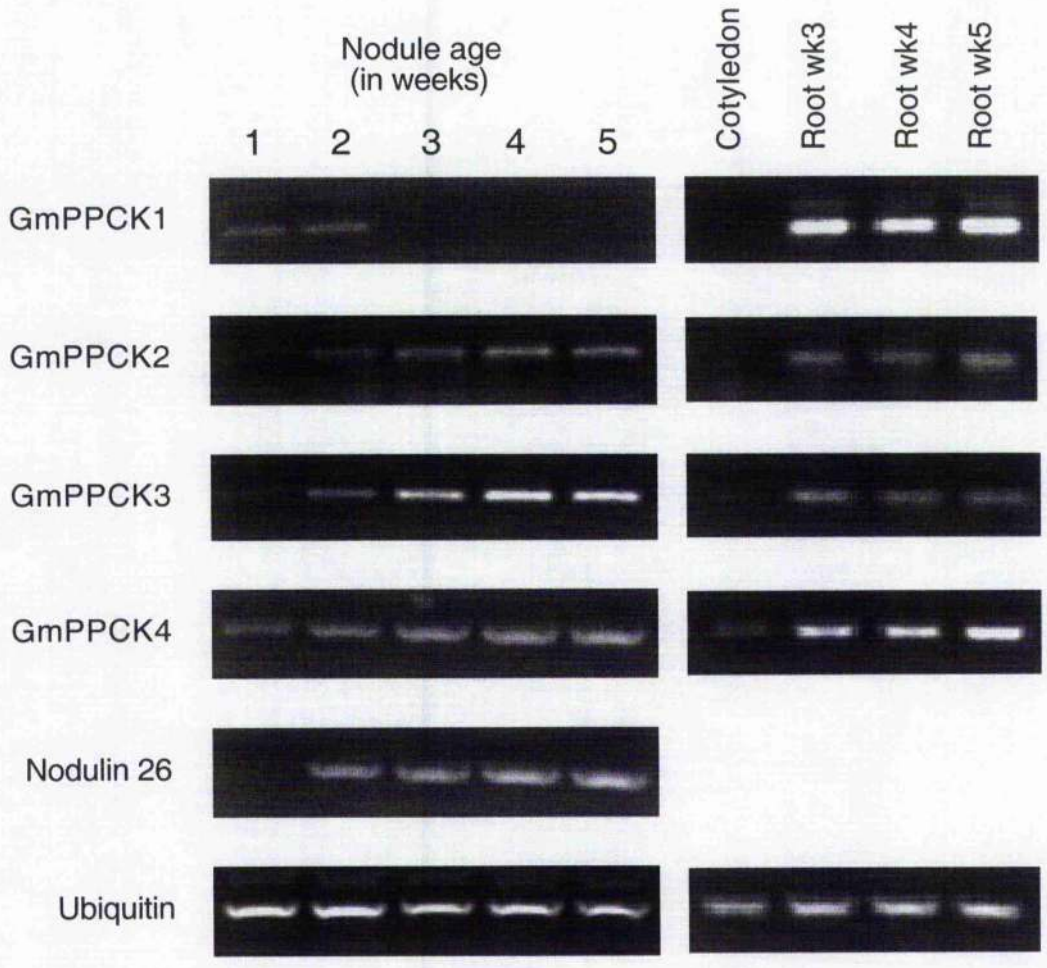
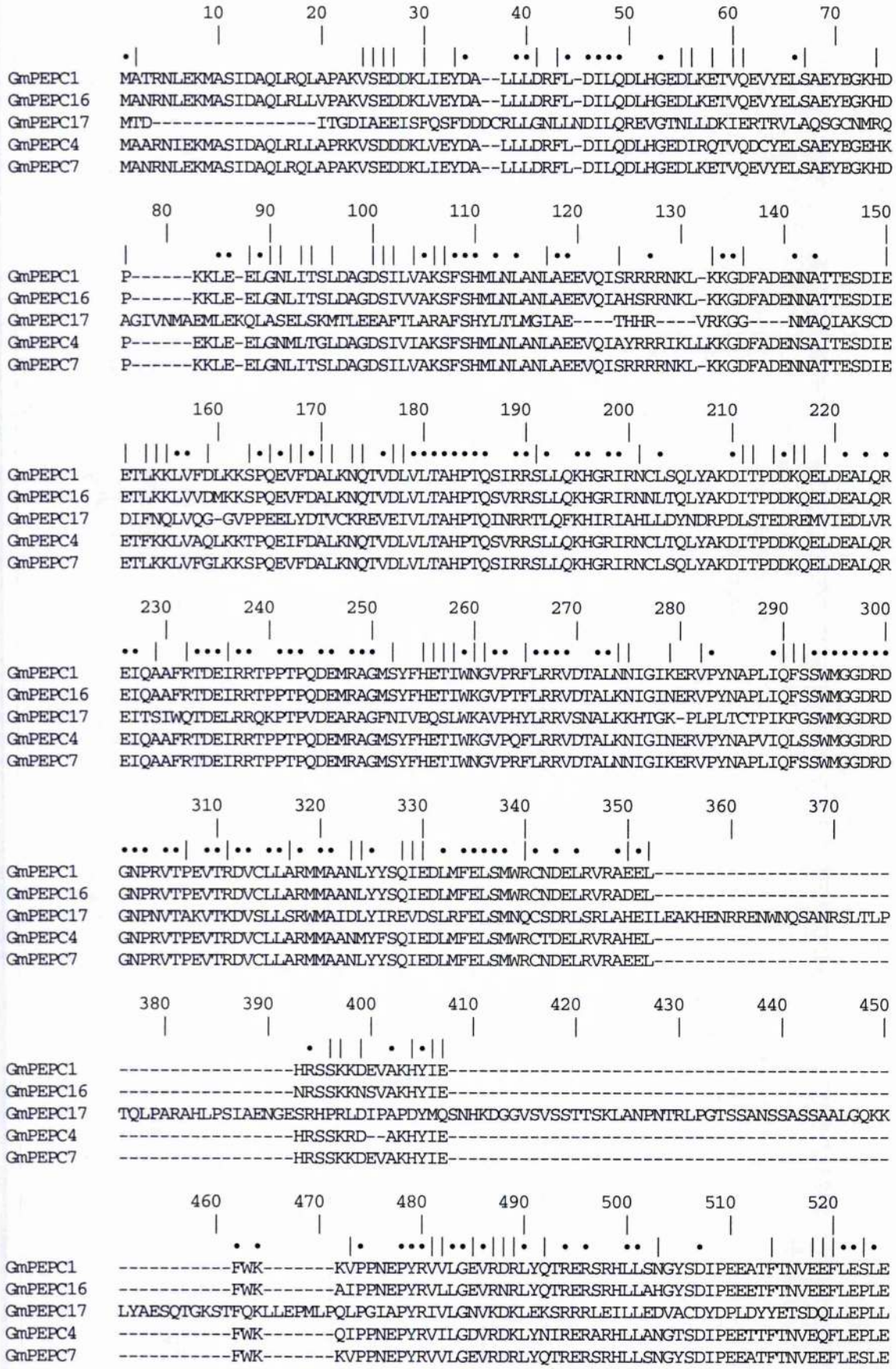

Nodule age
(in weeks)
1
2
3
4
5
Cotyledon
Root wk3
Root wk4
Root wk5
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
Nodulin 26
Ubiquitin

Figure 3.7 Sequence analysis of soybean *PEPC* genes

Alignment of deduced amino acid sequences of the soybean *PEPC* genes. Dots and vertical lines denote identical and similar residues respectively.


GmPEPC1
GmPEPC16
GmPEPC17
GmPEPC4
GmPEPC7

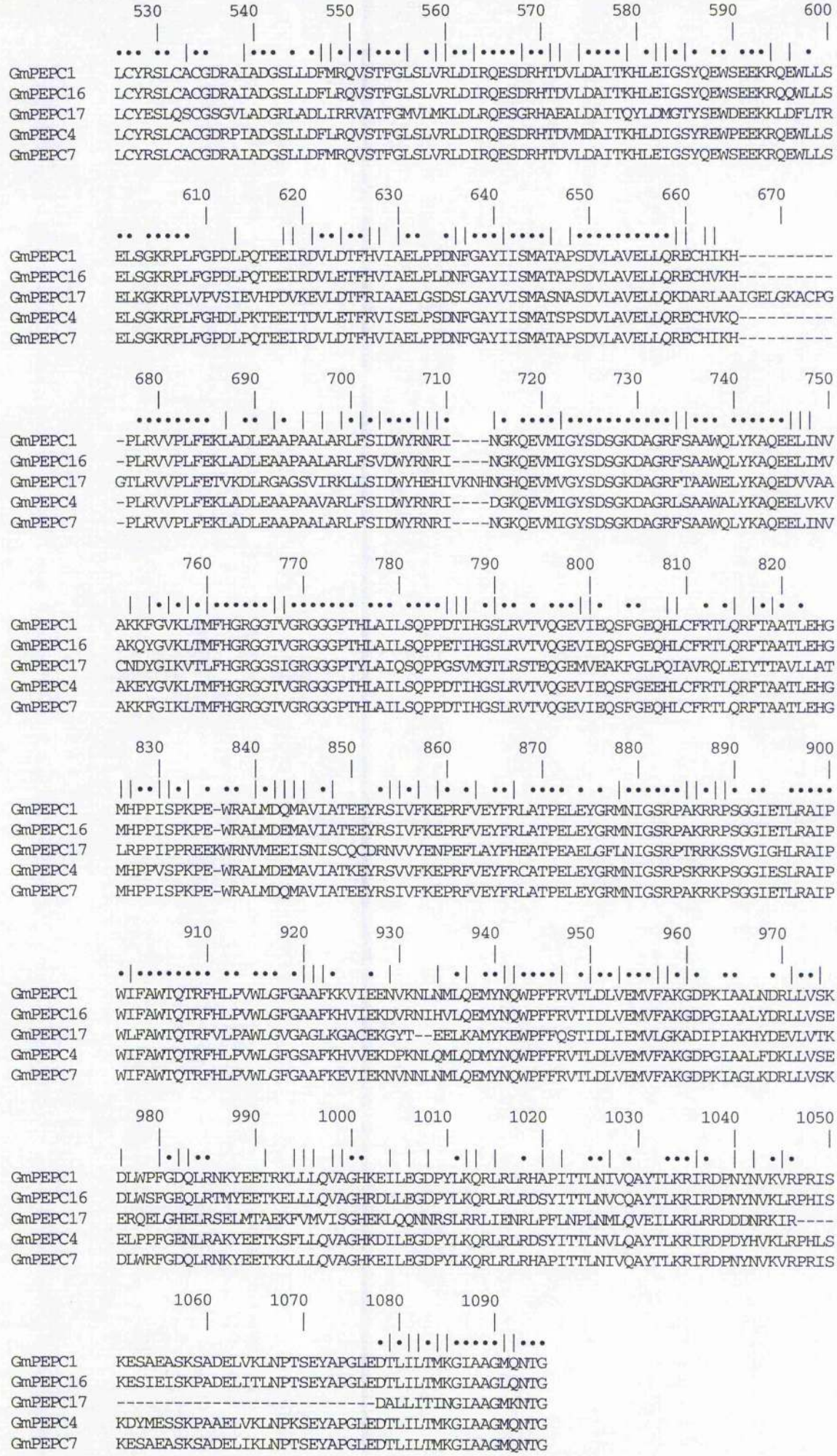
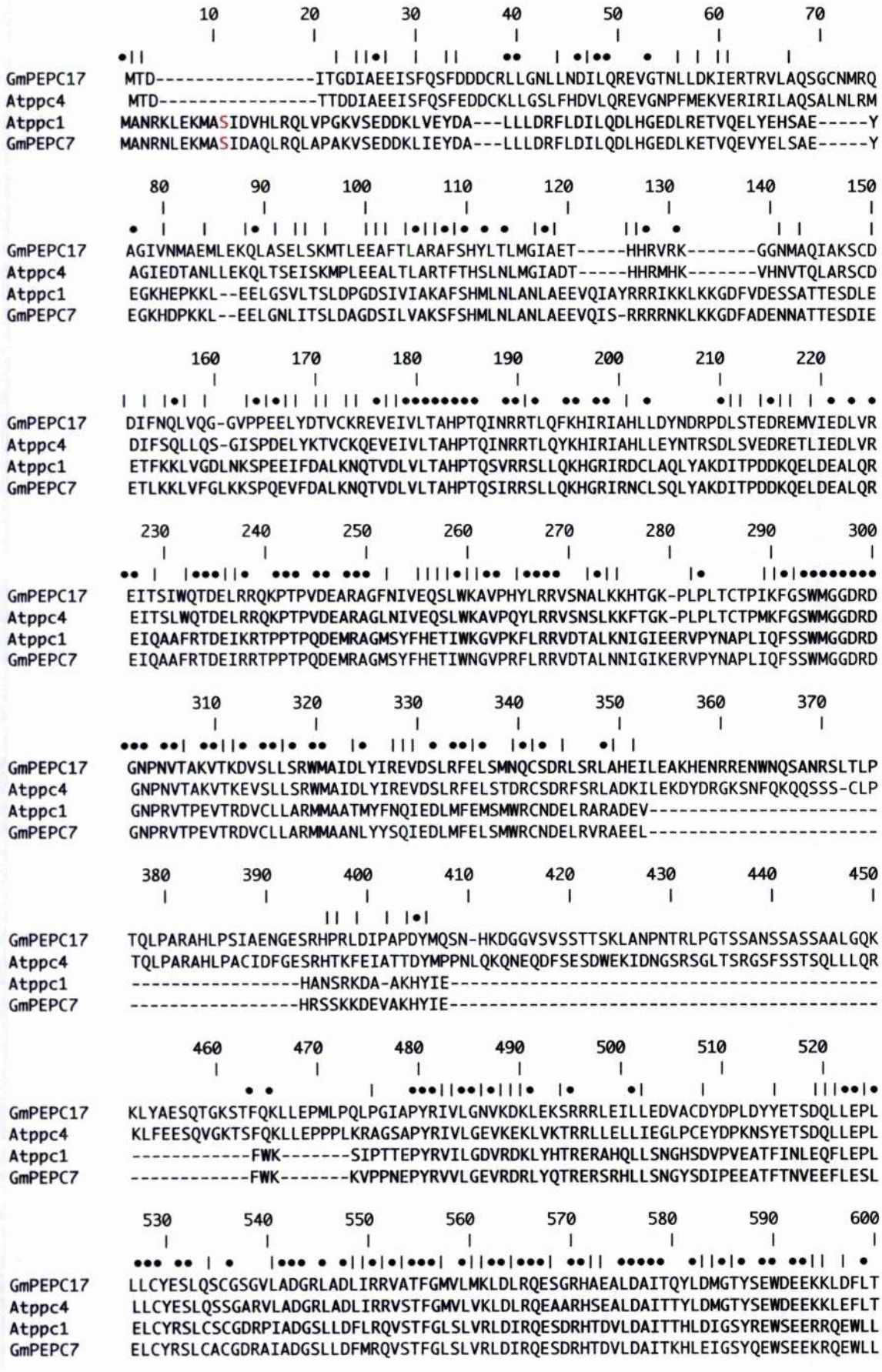

530 540 550 560 570 580 590 600
GmPEPC1 LCYRSLCACGDRAIADGSLLDFMRQVSTFGLSLVRLDIRQESDRHTDVLDAITKHLEIGSYQEWSEEKRQEWLLS
GmPEPC16 LCYRSLCACGDRAIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITKHLEIGSYQEWSEEKRQQWLLS
GmPEPC17 LCYESLQSCGSGVLADGRLADLIRRVATFGMVLMKLDLRQESGRHAEALDAITQYLDMGTYSEWDEEKKLDFLTR
GmPEPC4 LCYRSLCACGDRPIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVMDAITKHLDIGSYREWPEEKRQEWLLS
GmPEPC7 LCYRSLCACGDRAIADGSLLDFMRQVSTFGLSLVRLDIRQESDRHTDVLDAITKHLEIGSYQEWSEEKRQEWLLS
610 620 630 640 650 660 670
GmPEPC1 ELSGKRPLFGPDLPQTEEIRDVLDTFHVIAELPPDNFGAYIISMATAPSDVLAVELLQRECHIKH----------
GmPEPC16 ELSGKRPLFGPDLPQTEEIRDVLETFHVIAELPLDNFGAYIISMATAPSDVLAVELLQRECHVKH----------
GmPEPC17 ELKGKRPLVPVSIEVHPDVKEVLDTFRIAAELGSDSLGAYVISMASNASDVLAVELLQKDARLAAIGELGKACPG
GmPEPC4 ELSGKRPLFGHDLPKTEEITDVLETFRVISELPSDNFGAYIISMATSPSDVLAVELLQRECHVKQ----------
GmPEPC7 ELSGKRPLFGPDLPQTEEIRDVLDTFHVIAELPPDNFGAYIISMATAPSDVLAVELLQRECHIKH----------
680 690 700 710 720 730 740 750
GmPEPC1 -PLRVVPLFEKLADLEAAPAALARLFSIDWYRNRI----NGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELINV
GmPEPC16 -PLRVVPLFEKLADLEAAPAALARLFSVDWYRNRI----NGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELIMV
GmPEPC17 GTLRVVPLFETVKDLRGAGSVIRKLLSIDWYHEHIVKNHNGHQEVMVGYSDSGKDAGRFTAAWELYKAQEDVVAA
GmPEPC4 -PLRVVPLFEKLADLEAAPAAVARLFSIDWYRNRI----DGKQEVMIGYSDSGKDAGRLSAAWALYKAQEELVKV
GmPEPC7 -PLRVVPLFEKLADLEAAPAALARLFSIDWYRNRI----NGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELINV
760 770 780 790 800 810 820
GmPEPC1 AKKFGVKLYMFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIEQSFGEQHLCFRTLQRFTAATLEHG
GmPEPC16 AKQYGVKLYMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVIEQSFGEQHLCFRTLQRFTAATLEHG
GmPEPC17 CNDYGIKVTLFHGRGGSIGRGGGPTYLAIQSQPPGSVMGTLRSTEQGEMVEAKFGLPQIAVRQLEIYTTAVLLAT
GmPEPC4 AKEYGVKLYMFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIEQSFGEEHLCFRTLQRFTAATLEHG
GmPEPC7 AKKFGIKLYMFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIEQSFGEQHLCFRTLQRFTAATLEHG
830 840 850 860 870 880 890 900
GmPEPC1 MHPPISPKPE-WRALMDQMAVIATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPAKRRPSGGIETLRAIP
GmPEPC16 MHPPISPKPE-WRALMDEMAVIATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPAKRRPSGGIETLRAIP
GmPEPC17 LRPPIPPREEKWRNVMEEISNISCQCDRNVVYENPEFLAYFHEATPEAELGFLNIGSRPTRRKSSVGIGHLRAIP
GmPEPC4 MHPPVSPKPE-WRALMDEMAVIATKEYRSVVFKEPRFVEYFRCATPELEYGRMNIGSRPSKRKPSGGIESLRAIP
GmPEPC7 MHPPISPKPE-WRALMDQMAVIATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPAKRKPSGGIETLRAIP
910 920 930 940 950 960 970
GmPEPC1 WIFAWTQTRFHLPVWLGFGAAFKKVIEENVKNLNMLQEMYNQWPFFRVTLDLVEMVFAKGDPKIAALNDRLLVSK
GmPEPC16 WIFAWTQTRFHLPVWLGFGAAFKHVIEKDVRNIHVLQEMYNQWPFFRVTIDLVEMVFAKGDPGIAALYDRLLVSE
GmPEPC17 WLFAWTQTRFVLPAWLGVGAGLKGACEKGYT--EELKAMYKEWPFFQSTIDLIEMVLGKADIPIAKHYDEVLVTK
GmPEPC4 WIFAWTQTRFHLPVWLGFGSAFKHVVEKDPKNLQMLQDMYNQWPFFRVTLDLVEMVFAKGDPGIAALFDKLLVSE
GmPEPC7 WIFAWTQTRFHLPVWLGFGAAFKEVIEKNVNNLNMLQEMYNQWPFFRVTLDLVEMVFAKGDPKIAGLKDRLLVSK
980 990 1000 1010 1020 1030 1040 1050
GmPEPC1 DLWPFGDQLRNKYEETRKLLLQVAGHKEILEGDPYLKQRLRLRHAPITTLNIVQAYTLKRIRDPNYNVKVRPRIS
GmPEPC16 DLWSFGEQLRTMYEETKELLLQVAGHRDLLEGDPYLKQRLRLRDSYITTLNVCQAYTLKRIRDPNYNVKLRPHIS
GmPEPC17 ERQELGHELRSELMTAEKFVMVISGHEKLQQNNRSLRRLIENRLPFLNPLNMLQVEILKRLRRDDDNRKIR----
GmPEPC4 ELPPFGENLRAKYEETKSFLLQVAGHKDILEGDPYLKQRLRLRDSYITTLNVLQAYTLKRIRDPDYHVKLRPHLS
GmPEPC7 DLWRFGDQLRNKYEETKKLLLQVAGHKEILEGDPYLKQRLRLRHAPITTLNIVQAYTLKRIRDPNYNVKVRPRIS
1060 1070 1080 1090
GmPEPC1 KESAEASKSADELVKLNPTSEYAPGLEDTLILTMKGIAAGMQNTG
GmPEPC16 KESIEISKPADELITLNPTSEYAPGLEDTLILTMKGIAAGLQNTG
GmPEPC17 ----------DALLITINGIAAGMKNTG
GmPEPC4 KDYMESSKPAAELVKLNPKSEYAPGLEDTLILTMKGIAAGMQNTG
GmPEPC7 KESAEASKSADELIKLNPTSEYAPGLEDTLILTMKGIAAGMQNTG

Figure 3.8 Sequence analysis of 'bacterial-type' *PEPC* genes

Alignment of deduced amino acid sequences of soybean and Arabidopsis PEPcs. *Atppc1* (AJ532901) and *Atppc4* (AJ532903) are from Sanchez and Cejudo (2003). *GmPEPC17* and *Atppc4* both encode bacterial-type PEPcs. The serine residue in red is the target for phosphorylation by PEPc kinase. Dots and vertical lines denote identical and similar residues respectively.


GmPEPC17 MTD----------ITGDIAEEISFQSFDDDCRLLGNLLNDILQREVGTNLLDKIERTRVLAQSGCNMRQ
Atppc4 MTD----------TTDDIAEEISFQSFEDDCKLLGSLFHDVLQREVGNPFMEKVERIRILAQSALNLRM
Atppc1 MANRKLEKMASIDVHLRQLVPGKVSEDDKLVEYDA---LLLDRFLDILQDLHGEDLRETVQELYEHSAE-----Y
GmPEPC7 MANRNLEKMASIDAQLRQLAPAKVSEDDKLIEYDA---LLLDRFLDILQDLHGEDLKETVQEVYELSAE-----Y
GmPEPC17 AGIVNMAEMLEKQLASELSKMTLEEAFTLARAFSHYLTLMGIAET-----HHRVRK-------GGNMAQIAKSCD
Atppc4 AGIEDTANLLEKQLTSEISKMPLEEALTLARTFTHSLNLMGIADT-----HHRMHK-------VHNVTQLARSCD
Atppc1 EGKHEPKKL--EELGSVLTSLDPGDSIVIAKAFSHMLNLANLAEEVQIAYRRRIKKLKKGDFVDESSATTESDLE
GmPEPC7 EGKHDPKKL--EELGNLITSLDAGDSILVAKSFSHMLNLANLAEEVQIS-RRRRNKLKKGDFADENNATTESDIE
GmPEPC17 DIFNQLVQG-GVPPEELYDTVCKREVEIVLTAHPTQINRRTLQFKHIRIAHLLDYNDRPDLSTEDREMVIEDLVR
Atppc4 DIFSQLLQS-GISPDELYKTVCKQEVEIVLTAHPTQINRRTLQYKHIRIAHLLEYNTRSDLSVEDRETLIEDLVR
Atppc1 ETFKKLVGDLNKSPEEIFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRDCLAQLYAKDITPDDKQELDEALQR
GmPEPC7 ETLKKLVFGLKKSPQEVFDALKNQTVDLVLTAHPTQSIRRSLLQKHGRIRNCLSQLYAKDITPDDKQELDEALQR
GmPEPC17 EITSIWQTDELRRQKPTPVDEARAGFNIVEQSLWKAVPHYLRRVSNALKKHTGK-PLPLTCTPIKFGSWMGGDRD
Atppc4 EITSLWQTDELRRQKPTPVDEARAGLNIVEQSLWKAVPQYLRRVSNSLKKFTGK-PLPLTCTPMKFGSWMGGDRD
Atppc1 EIQAAFRTDEIKRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGIEERVPYNAPLIQFSSWMGGDRD
GmPEPC7 EIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWNGVPRFLRRVDTALNNIGIKERVPYNAPLIQFSSWMGGDRD
GmPEPC17 GNPNVTAKVTKDVSLLSRWMAIDLYIREVDSLRFELSMNQCSDRLSRLAHEILEAKHENRRENWNQSANRSLTLP
Atppc4 GNPNVTAKVTKEVSLLSRWMAIDLYIREVDSLRFELSTDRCSDRFSRLADKILEKDYDRGKSNFQKQQSSS-CLP
Atppc1 GNPRVTPEVTRDVCLLARMMAATMYFNQIEDLMFEMSMWRCNDELRARADEV----------
GmPEPC7 GNPRVTPEVTRDVCLLARMMAANLYYSQIEDLMFELSMWRCNDELRVRAEEL----------
GmPEPC17 TQLPARAHLPSIAENGESRHPRLDIPAPDYMQSN-HKDGGVSVSSTTSKLANPNTRLPGTSSANSSASSAALGQK
Atppc4 TQLPARAHLPACIDFGESRHTKFEIATTDYMPPNLQKQNEQDFSESDWEKIDNGSRSGLTSRGSFSSTSQLLLQR
Atppc1 ----------HANSRKDA-AKHYIE----------
GmPEPC7 ----------HRSSKKDEVAKHYIE----------
GmPEPC17 KLYAESQTGKSTFQKLLEPMLPQLPGIAPYRIVLGNVKDKLEKSRRRLEILLEDVACDYDPLDYYETSDQLLEPL
Atppc4 KLFEESQVGKTSFQKLLEPPPLKRAGSAPYRIVLGEVKEKLVKTRRLLELLIEGLPCEYDPKNSYETSDQLLEPL
Atppc1 ----------FWK-------SIPTTEPYRVILGDVRDKLYHTRERAHQLLSNGHSDVPVEATFINLEQFLEPL
GmPEPC7 ----------FWK-------KVPPNEPYRVVLGEVRDRLYQTRERSRHLLSNGYSDIPEEATFTNVEEFLESL
GmPEPC17 LLCYESLQSCGSGVLADGRLADLIRRVATFGMVLMKLDLRQESGRHAEALDAITQYLDMGTYSEWDEEKKLDFLT
Atppc4 LLCYESLQSSGARVLADGRLADLIRRVSTFGMVLVKLDLRQEAARHSEALDAITTYLDMGTYSEWDEEKKLEFLT
Atppc1 ELCYRSLCSCGDRPIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITTHLDIGSYREWSEERRQEWLL
GmPEPC7 ELCYRSLCACGDRAIADGSLLDFMRQVSTFGLSLVRLDIRQESDRHTDVLDAITKHLEIGSYQEWSEEKRQEWLL

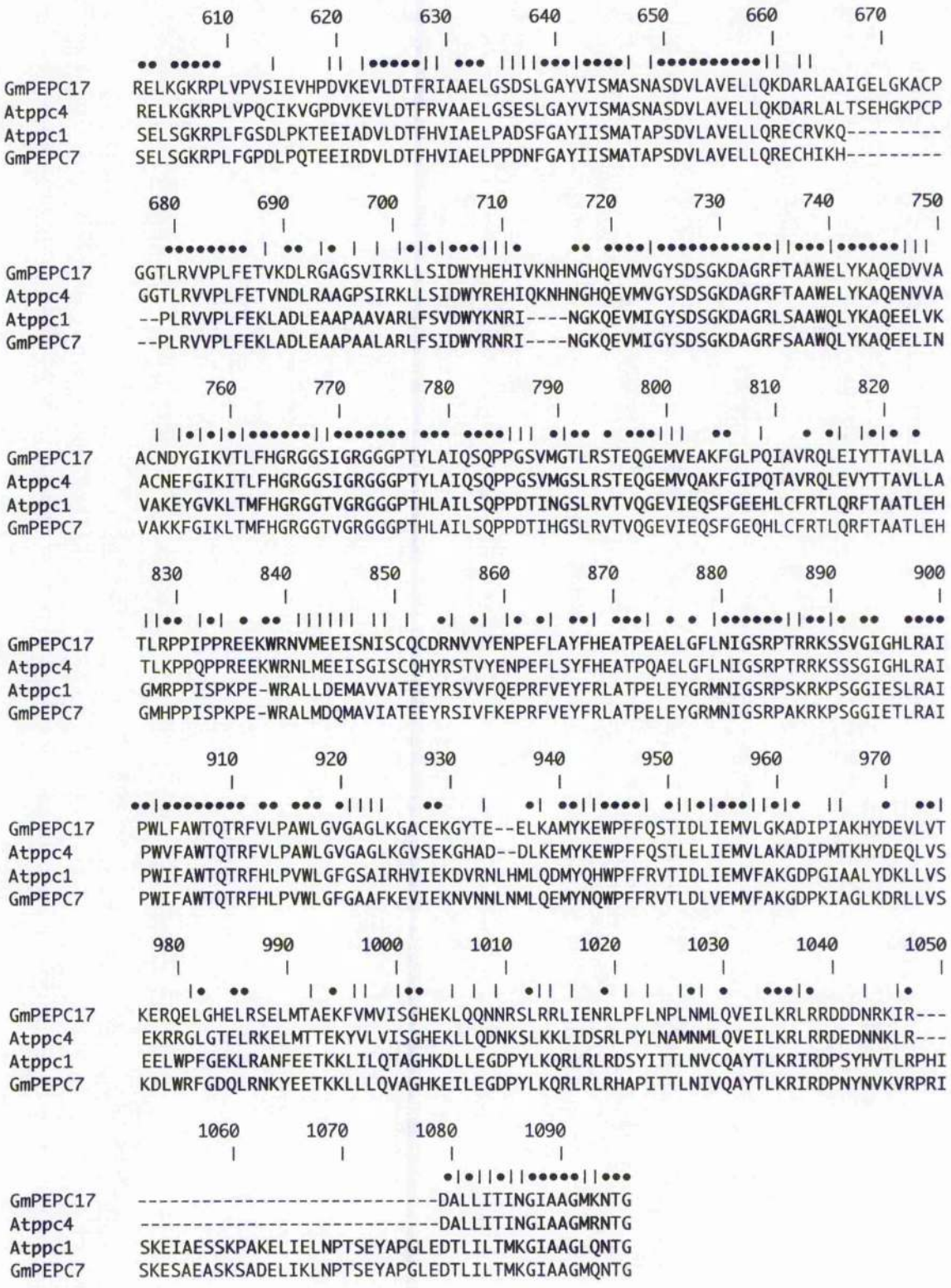

610 620 630 640 650 660 670
GmPEPC17 RELKGKRPLVPVSIEVHPDVKEVLDTFRIAAELGSDSLGAYVISMASNASDVLAVELLQKDARLAAIGELGKACP
Atppc4 RELKGKRPLVPQCIKVGPDVKEVLDTFRVAAELGSESLGAYVISMASNASDVLAVELLQKDARLALTSEHGKPCP
Atppc1 SELSGKRPLFGSDLPKTEEIADVLDTFHVIAELPADSFGAYIISMATAPSDVLAVELLQRECRVKQ---------
GmPEPC7 SELSGKRPLFGPDLPQTEEIRDVLDTFHVIAELPPDNFGAYIISMATAPSDVLAVELLQRECHIKH---------
680 690 700 710 720 730 740 750
GmPEPC17 GGTLRVVPLFETVKDLRGAGSVIRKLLSIDWYHEHIVKNHNGHQEVMVGYSDSGKDAGRFTAAWELYKAQEDVVA
Atppc4 GGTLRVVPLFETVNDLRAAGPSIRKLLSIDWYREHIQKNHNGHQEVMVGYSDSGKDAGRFTAAWELYKAQENVVA
Atppc1 --PLRVVPLFEKLADLEAAPAAVARLFSVDWYKNRI----NGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELVK
GmPEPC7 --PLRVVPLFEKLADLEAAPAALARLFSIDWYRNRI----NGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELIN
760 770 780 790 800 810 820
GmPEPC17 ACNDYGIKVTLFHGRGGSIGRGGGPTYLAIQSQPPGSVMGTLRSTEQGEMVEAKFGLPQIAVRQLEIYTTAVLLA
Atppc4 ACNEFGIKITLFHGRGGSIGRGGGPTYLAIQSQPPGSVMGSLRSTEQGEMVQAKFGIPQTAVRQLEVYTTAVLLA
Atppc1 VAKEYGVKLTMFHGRGGTVGRGGGPTHLAILSQPPDTINGSLRVTVQGEVIEQSFGEEHLCFRTLQRFTAATLEH
GmPEPC7 VAKKFGIKLTMFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIEQSFGEQHLCFRTLQRFTAATLEH
830 840 850 860 870 880 890 900
GmPEPC17 TLRPPIPPREEKWRNVMEEISNISCQCDRNVVYENPEFLAYFHEATPEAELGFLNIGSRPTRRKSSVGIGHLRAI
Atppc4 TLKPPQPPREEKWRNLMEEISGISCQHYRSTVYENPEFLSYFHEATPQAELGFLNIGSRPTRRKSSSGIGHLRAI
Atppc1 GMRPPISPKPE-WRALLDEMAVVATEEYRSVVFQEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRAI
GmPEPC7 GMHPPISPKPE-WRALMDQMAVIATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPAKRKPSGGIETLRAI
910 920 930 940 950 960 970
GmPEPC17 PWLFAWTQTRFVLPAWLGVGAGLKGACEKGYTE--ELKAMYKEWPFFQSTIDLIEMVLGKADIPIAKHYDEVLVT
Atppc4 PWVFAWTQTRFVLPAWLGVGAGLKGVSEKGHAD--DLKEMYKEWPFFQSTLELIEMVLAKADIPMTKHYDEQLVS
Atppc1 PWIFAWTQTRFHLPVWLGFGSAIRHVIEKDVRNLHMLQDMYQHWPFFRVTIDLIEMVFAKGDPGIAALYDKLLVS
GmPEPC7 PWIFAWTQTRFHLPVWLGFGAAFKEVIEKNVNNLNMLQEMYNQWPFFRVTLDLVEMVFAKGDPKIAGLKDRLLVS
980 990 1000 1010 1020 1030 1040 1050
GmPEPC17 KERQELGHELRSELMTAEKFVMVISGHEKLQQNNRSLRRLIENRLPFLNPLNMLQVEILKRLRRDDDNRKIR---
Atppc4 EKRRGLGTELRKELMTTEKYVLVISGHEKLLQDNKSLKKLIDSRLPYLNAMNMLQVEILKRLRRDEDNNKLR---
Atppc1 EELWPFGEKLRANFEETKKLILQTAGHKDLLEGDPYLKQRLRLRDSYITTLNVCQAYTLKRIRDPSYHVTLRPHI
GmPEPC7 KDLWRFGDQLRNKYEETKKLLLQVAGHKEILEGDPYLKQRLRLRHAPITTLNIVQAYTLKRIRDPNYNVKVRPRI
1060 1070 1080 1090
GmPEPC17 ----------DALLITINGIAAGMKNTG
Atppc4 ----------DALLITINGIAAGMRNTG
Atppc1 SKEIAESSKPAKELIELNPTSEYAPGLEDTLILTMKGIAAGLQNTG
GmPEPC7 SKESAEASKSADELIKLNPTSEYAPGLEDTLILTMKGIAAGMQNTG

Figure 3.9 Expression patterns of *GmPEPC* genes

This shows the products obtained from RT-PCR analysis of *GmPEPC* genes from nodulated soybean plants. Ubiquitin was used as a control for equal loading. 27 cycles of amplification were used for GmPEPc genes and 20 cycles for Ubiquitin. All tissues were collected from a single plant. The presented results are representative of at least three biological replicates (individual plants) and three RT-PCR experiments on the same RNA samples.

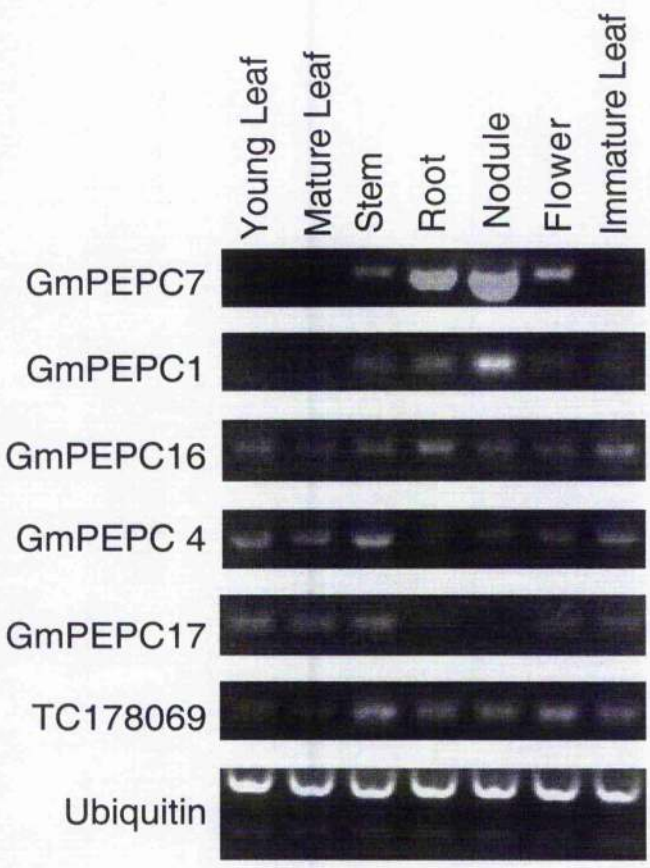

Young Leaf
Mature Leaf
Stem
Root
Nodule
Flower
Immature Leaf
GmPEPC7
GmPEPC1
GmPEPC16
GmPEPC 4
GmPEPC17
TC178069
Ubiquitin

Figure 3.10 Expression pattern of *GmPPCK1* using different primer pairs

This shows the products obtained from RT-PCR analysis of *GmPPCK1* from nodulated soybean plants using either the primers give in Table 1 or those published by Xu *et al.* (2003). Ubiquitin was used as a control for equal loading. The same RNA samples were used as in Fig. 3.5. 27 cycles of amplification were used for both sets of GmPPCK1 primers and 20 cycles for Ubiquitin.

Young Leaf
Mature Leaf
Stem
Nodule
Root
Flower
Immature Leaf
GmPPCK1
(primers given
in table 1)
GmPPCK1
(Xu et al., 2003
primers)
Ubiquitin

Chapter 4

Expression of GmPPCK genes under constant conditions

4.1 Introduction

The work of Hartwell *et al.* (1999a) on *K. fedtschenkoi* showed that the abundance of *PPCK* transcripts in leaves increased with leaf age, following the development of CAM photosynthesis. In mature leaves a good correlation between the level of *PPCK* transcripts and the level of translatable PEPc kinase mRNA was observed, with expression peaking in the middle of the night. PEPc kinase activity and PEPc phosphorylation state, as judged by malate sensitivity, tracked the circadian rhythm in *PPCK* transcript abundance, with a small time lag. These results strongly suggested that the control of transcript abundance is an important factor governing PEPc kinase activity. Subsequent work by Taybi *et al.* (2000) and Tsuchida *et al.* (2001) was consistent with this hypothesis. The deduced amino acid sequence of PEPc kinase itself suggested that it was essentially unregulated when expressed (Hartwell *et al.*, 1999a).

Recently an inhibitor of PEPc kinase has been identified in the leaves of *K. fedtschenkoi*, barley and maize, demonstrating the presence of other factors involved in PEPc kinase regulation (Nimmo *et al.*, 2001). The work of Osuna *et al.* (1999) demonstrates that PEPc kinase may not be regulated at the level of transcript abundance in all systems. Working with barley seeds, they showed that PEPc becomes phosphorylated following imbibition. However, PEPc kinase activity was detected in dry seeds and no increase was observed following imbibition.

With the results of the previous chapter demonstrating the presence of two *PPCK* genes showing nodule-enhanced expression and transcripts of a third *PPCK* gene clearly detectable in soybean nodules, it was important to assess the effects on *PPCK* transcript abundance of treatments that had previously been shown to reduce nodule PEPc kinase activity. Treatments such as shoot removal (detopping), stem girdling and extended darkness, which all prevent or reduce the flow of photosynthate from the leaves, have been shown to reduce PEPc activity, PEPc phosphorylation status and PEPc kinase activity (Zhang *et al.*, 1995; Wadham *et al.*, 1996; Zhang and Chollet, 1997). The *PPCK* gene or genes regulated by these treatments would therefore possess the predicted properties of the nodule-enhanced PEPc kinase.

4.2 Results

4.2.1 Effect of darkness and detopping on *PPCK* expression

In order to examine the effects of light, soybean plants were subjected to 2, 12 or 24 h of darkness or 24 h darkness followed by 3 h of illumination. In other experiments, plants were detopped and tissue collected after 2, 12 and 24 h. Fig. 4.1 shows the results of the RT-PCR experiments from these plants. The transcript abundance of both *GmPPCK2* and *GmPPCK3* shows a reduction in nodules as a result of reduced photosynthate supply from the shoots. The effect of detopping is much more dramatic than that of darkness, with transcripts barely detectable after 12 h. The effects of 24 h darkness are reversible by 3 h of illumination. In contrast, the transcript abundance of *GmPPCK4* shows little change in response to the treatments. The expression of the nodule-enhanced PEPc gene, *GmPEPC7*, is greatly reduced after 24 h of darkness and detopping. This is in agreement with previously reported findings Nakagawa *et al.* (2003a, 2003b), both for soybean and *L. japonicus* nodule-enhanced PEPc.

Leaf tissue was also collected from the control and dark treated plants to examine the effects of darkness on PEPc kinase expression in a photosynthetic organ. Figs. 4.2 and 4.3 show the results obtained using mature and young leaf tissue respectively. The most notable feature is the change in expression levels of *GmPPCK4* in control and dark-treated plants. In mature leaves *GmPPCK4* transcripts are readily detectable at 2 and 24h but no transcripts are detected at 12h, both for control plants in a light/dark cycle and treated plants maintained in constant darkness. A similar pattern of expression is observed in young leaves, though less pronounced. This result is in stark contrast to the expression of *GmPPCK4* in nodules from these plants (Fig. 4.1). None of the other *GmPPCKs* show a similar pattern. *GmPPCK2* and *3* both have lower transcript levels in the dark-treated plants compared to the control plants. *GmPPCK1* showed no pattern of expression in mature leaves and was not detectable in the young leaf samples.

4.2.2 Expression of soybean PPCK genes under constant conditions

The expression pattern of *GmPPCK4* in leaves of dark-treated plants (Figs. 4.2 and 4.3) raised the possibility that its transcript abundance could be under circadian control. However the observed expression pattern could also be due to day/night

temperature changes experienced by the plants, to which even those treated with 24 h of constant darkness were subjected. In order to test the hypothesis of circadian regulation, 2-3 week old non-nodulated plants were transferred to continuous light (LL) and 22°C at the end of a normal dark period and leaf samples were collected every 3h. Fig. 4.4 shows the results of RT-PCR analysis of the four *GmPPCK* genes. The cyclical pattern of *GmPPCK4* expression is evident, with transcript levels peaking in the later half of the day. Both *GmPPCK2* and *3* transcript levels are higher during the first light period than they are during the preceding dark period, consistent with the previous results (Figs. 4.2 and 4.3), suggesting light-enhanced expression in leaves. The expression of *GmPPCK1* also varied during the time-course, with transcript abundance appearing maximal around dawn on the first day and midday of the second day. This experiment showed that the changes in *GmPPCK4* transcript abundance were not related to environmental factors (i.e. temperature) and were present in the leaves of non-nodulated seedlings as well as 5 week old nodulated plants.

More extensive time-course experiments were performed to further investigate the circadian control of *GmPPCK* genes. Two to three week old non-nodulated plants were transferred to continuous light (LL) and 22°C at the end of a normal dark period or to continuous dark (DD) and 22°C at the end of a normal light period. Leaf and root samples were taken over the following 60 h and the expression of all four *PPCK* genes was assessed by RT-PCR. In leaves, although there may be some cycling of *GmPPCK1* transcript levels in LL (Fig. 4.5), this is not seen in DD (Fig. 4.6). Again *GmPPCK2* and *3* transcript abundance was higher in light that the preceding or subsequent dark period, but levels did not cycle in constant conditions. In contrast *GmPPCK4* was clearly under robust circadian control in leaves. In LL, transcript abundance was highest in the second half of subjective day but was low in subjective night (Fig. 4.5). *GmPPCK4* transcript abundance also cycled in DD (Fig. 4.6), though the amplitude of the second peak in DD was lower than that of the first peak. In roots, however, *GmPPCK4* transcript abundance was not under clear circadian control in either LL (Fig. 4.7) or DD (Fig. 4.8). In order to confirm the presence of the circadian oscillator in these tissues primers were designed to amplify a partial length soybean homologue of the Arabidopsis clock gene *LHY*, which is part of the central machinery of the clock (Alabadi *et al.*, 2001), using tentative consensus sequence TC189020 (see http://www.tigr.org/). The soybean sequence (hereafter termed *GmpLHY* for *Glycine*

max putative *LHY*) has been deposited in GenBank (AY568715). The putative soybean LHY is 80% identical to a LHY orthologue isolated from *Phaseolus vulgaris* (Kaldis *et al.*, 2003) and 49% identical to the Arabidopsis LHY at the amino acid level. Fig. 4.9 shows an amino acid alignment of these sequences and the positions of the primers used to amplify *GmpLHY*. As expected from the behaviour of Arabidopsis *LHY* (Schaffer *et al.*, 1998), *GmpLHY* was expressed in a circadian fashion in both LL and DD in leaves and roots, with transcript abundance maximal early in subjective day (Figs. 4.5 – 4.8). As with *GmPPCK4*, the amplitude of the second peak in *GmpLHY* transcript abundance in DD was lower than that of the first peak.

4.3 Discussion

The results from the experiments investigating the effects of darkness and of detopping show that the transcript abundance of both *GmPPCK2* and *GmPPCK3* are modulated by these treatments, while *GmPPCK4* is largely unaffected (Fig. 4.1). The effect of extended darkness on the expression levels of *GmPPCK2* and *3* is reversible by a short period of illumination. These results are very similar to previous reports measuring the effects of extended darkness and stem girdling on PEPc kinase activity and I_{50} (malate) values (Zhang *et al.*, 1995; Zhang and Chollet, 1997). This would suggest that GmPPCK2 and GmPPCK3 represent the majority of the measured kinase activity in soybean nodules and are responsible for controlling the phosphorylation status of PEPc. However, it is clearly important to assess how much of the nodule PPCK activity is contributed by these three genes (*GmPPCK2-4*) under different conditions.

The transcript levels of the nodule-enhanced PEPc, *GmPEPC7*, were very similar to those of the nodule-enhanced kinases in response to the darkness and detopping treatments (Fig. 4.1). However, in contrast to the results from PEPc kinase, previous work has demonstrated that PEPc protein and activity levels are not affected by such treatments (Zhang *et al.*, 1995). Therefore the nodule GmPEPC7 protein is not rapidly turned over and control of nodule PEPc is largely at the post-translational level.

The decline in PEPc kinase activity and transcript abundance following treatments that perturb the flow of photosynthate from the leaves has been interpreted as control of PEPc kinase activity by a constituent of photosynthate, such as sucrose or a metabolite of sucrose (Zhang and Chollet, 1997; Xu *et al.*, 2003; Nakagawa *et al.*,

2003a). Wadham *et al.* (1996) measured metabolite levels in soybean nodules 6 h after detopping. They observed a 50% decline in sucrose and malate levels and a 100% increase in PEP levels, in comparison to control plants. No difference was observed in total amount of hexoses, suggesting they play no role in *PPCK* regulation, but only total nodule metabolite pools were measured and at only a single time-point. Detopping and stem girdling treatments have been shown to lead to a decline in nodule O_2 permeability and cause inhibition of nitrogenase activity (Hartwig *et al.*, 1987; Vessey *et al.*, 1988; Denison *et al.*, 1992; de Lima *et al.*, 1994). Therefore the interpretation of these experiments is complicated by the gross effect they have on nodule metabolism.

The investigation into the effects of darkness on the transcript levels in soybean plants lead to a surprising discovery concerning the regulation of *GmPPCK4*. It is the first *PPCK* gene shown to be under circadian control other than those in CAM species. Previously, a report from Ueno *et al.* (2000) suggested that there might be a circadian component to the control of the phosphorylation state of PEPc in mature maize leaves. Using phosphorylation state-specific antibodies to the C_4 maize PEPc, they observed that the phosphorylation and dephosphorylation of PEPc commenced before dawn and before sunset, respectively. However no further characterisation of this phenomenon has yet been published. Fontaine *et al.* (2002) demonstrated that in Arabidopsis *AtPPCK1* and *AtPPCK2* are light regulated, but no circadian rhythm in transcript abundance was detected in constant light for either gene.

Another notable feature of *GmPPCK4* expression is that, under constant conditions, the transcript abundance cycles in leaves but not roots. To our knowledge, this is the first example of a plant gene where circadian control of expression is organ-specific. Many studies of circadian rhythms in plants have involved the use of whole seedlings, and organ specificity has rarely been addressed. However circadian-controlled expression of Arabidopsis PHYTOCHROME A has been observed in both leaves and roots, and similar observations have been made with CHALCONE SYNTHASE in aerial tissue and roots (Hall *et al.*, 2001; Thain *et al.*, 2002). Many plant genes show robust rhythms of transcript abundance in LL that dampen rapidly in DD, as is the case for *GmPPCK4* in leaves (Fig. 4.6). However this is not due to the damping of the oscillator itself. As judged by the behaviour of *GmpLHY*, the circadian clock operates in both roots and leaves of soybean, in DD and LL. In tobacco, such behaviour is thought to involve the *ZGT* gene, which has been shown to link the

circadian oscillator to *LHCB* expression (Xu and Johnson, 2001). *LHCB* is a member of the light-harvesting protein-pigment complex (LHC) gene family; its transcript abundance is under circadian regulation in LD and LL, but this rhythm rapidly dampens in DD. In transgenic tobacco plants constitutively expressing ZGT, the circadian pattern of *LHCB* expression is maintained in DD (Xu and Johnson, 2001). One explanation of our data concerning *GmPPCK4* is that machinery such as *ZGT* exists to connect the expression of *GmPPCK4* to the clock in soybean leaves but not in roots. It is interesting to note that in tobacco *ZGT* transcripts have two alternate forms that are differentially expressed in different tissues (Xu and Johnson, 2001).

Harmer *et al.* (2000) analysed the circadian control of transcript levels of over 8000 genes in Arabidopsis using microarrays. Approximately 6% (453) of the genes on the chip were classified as being circadian. Through promoter analysis they identified a nine-nucleotide 'evening element', AAAATATCT, that plays an important role in conferring rhythmic gene expression towards the end of the day in Arabidopsis. Since *GmPPCK4* shows this pattern of expression we analysed the limited sequence information available for all four *GmPPCK* promoters that was obtained from the genome walker experiments (see Chapter 2). The *GmPPCK4* promoter contains the sequence ATAATATCT, which differs from the 'evening element' by just one base and includes the 'core' seven-nucleotide sequence AATATCT noted by Xu and Johnson, (2001). It also contains the sequence CACGTG, one version of the 'G box core' light-regulation motif (Chattopadhyay *et al.*, 1998). None of the other *GmPPCK* promoters contain such a close match to the nine-nucleotide 'evening element', and a 'G-box core', and it is possible that these features contribute to the control of *GmPPCK4* expression.

Harmer *et al.* (2000) took samples every 4 h, allowing the circadian genes to be placed into one of six groups depending upon peak expression time. The largest group consisted of more than 25% of the total number of genes and their expression peaked 8 h into the 12 h photoperiod. *GmPPCK4* would also fall into this group of late afternoon expressed genes. Harmer *et al.* (2000) noted that when the genes were also grouped according to function, the transcript abundance of several genes whose products are involved in the consumption, translocation or storage of sugars peaked in this late afternoon period. They suggested that the circadian clock might play an important role in allocating sugar carbon to different pathways. This is consistent with

the circadian control of *GmPPCK4* in leaves and the potential role of the gene product in directing carbon flow towards biosynthesis.

Figure 4.1 Expression of *GmPPCK* genes and *GmPEPC7* in root nodules in response to darkness or shoot detopping

This shows the products obtained from RT-PCR analysis of nodulated soybean plants. Control plants were grown under normal light/dark conditions. All experiments were started 6 h into the light period. For dark treatments plants were transferred to an identical environmental chamber under the same growth conditions but without lighting. For detopping all above-ground organs were removed from below the point at which the cotyledons were attached. Ubiquitin was used as a control for equal loading. 27 cycles of amplification were used for all GmPPCK genes and 20 cycles for GmPEPc7 and Ubiquitin. For each time point tissue was collected from a single plant.

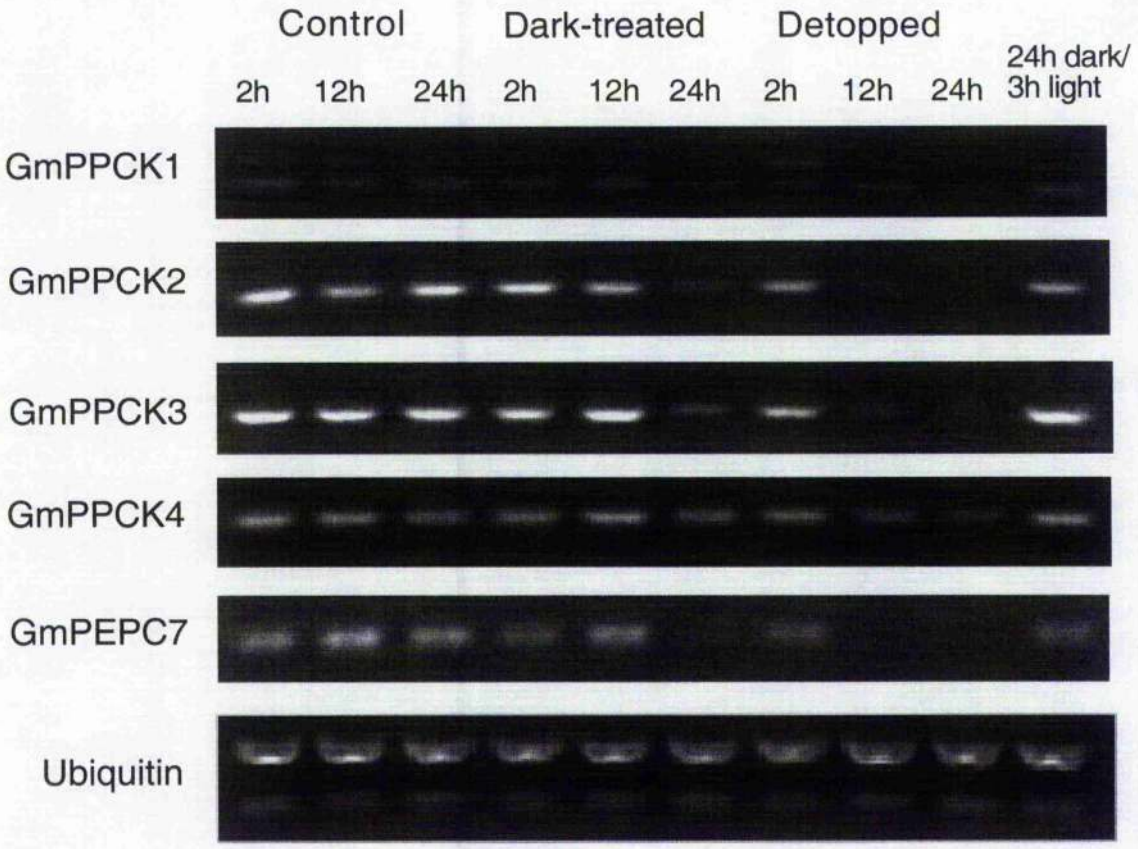

Control
Dark-treated
Detopped
2h
12h
24h
2h
12h
24h
2h
12h
24h
24h dark/
3h light
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
GmPEPC7
Ubiquitin

Figure 4.2 Expression of *GmPPCK* genes in mature leaves in response to darkness

This shows the products obtained from RT-PCR analysis of nodulated soybean plants. Experimental conditions are given in Fig 4.1. Ubiquitin was used as a control for equal loading. 27 cycles of amplification were used for all GmPPCK genes and 25 cycles for Ubiquitin. For each time point tissue was collected from a single plant.

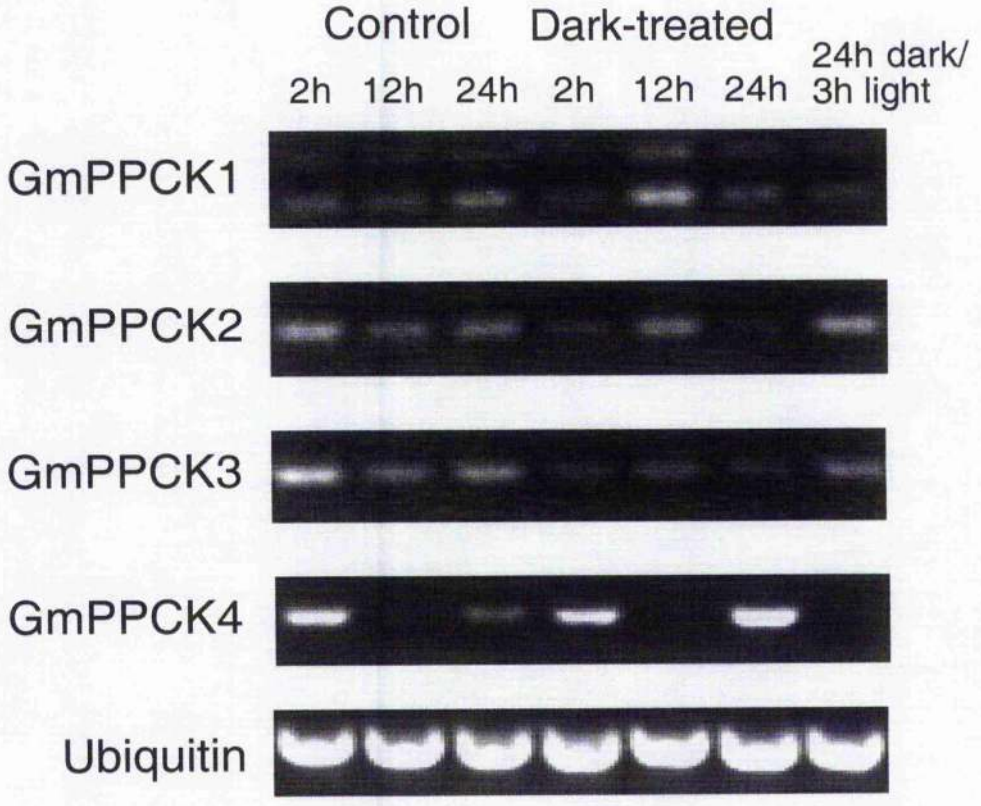

Control
Dark-treated
2h
12h
24h
2h
12h
24h
24h dark/
3h light
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
Ubiquitin

Figure 4.3 Expression of *GmPPCK* genes in young leaves in response to darkness
This shows the products obtained from RT-PCR analysis of nodulated soybean plants. Experimental conditions are given in Fig 4.1. Ubiquitin was used as a control for equal loading. 27 cycles of amplification were used for all GmPPCK genes and 20 cycles for Ubiquitin. For each time point tissue was collected from a single plant.

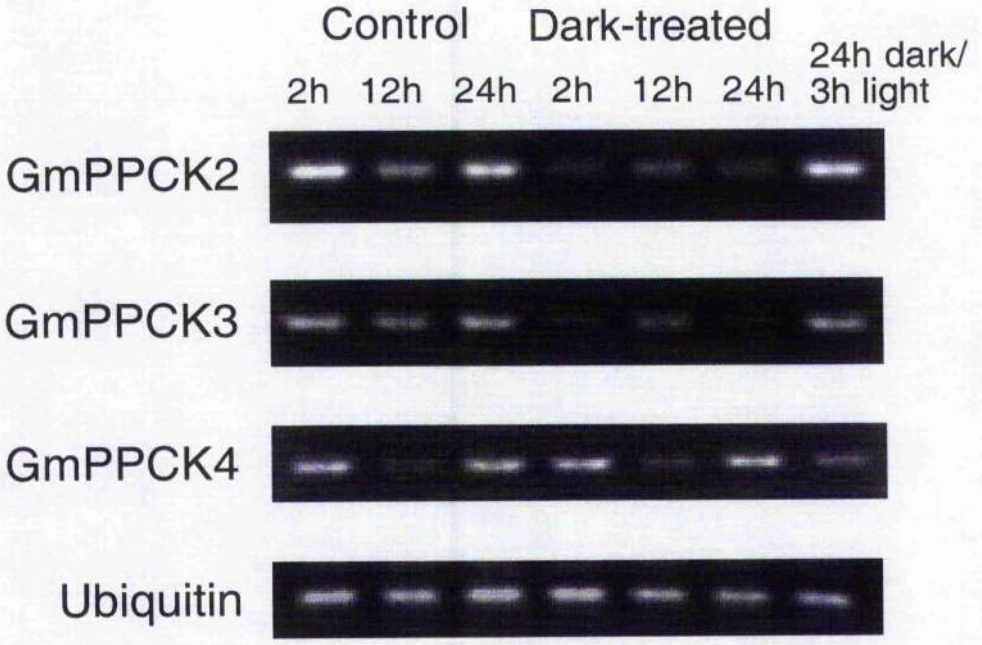

Control
Dark-treated
2h
12h
24h
2h
12h
24h
24h dark/
3h light
GmPPCK2
GmPPCK3
GmPPCK4
Ubiquitin

Figure 4.4 Pilot experiment to study the expression of *GmPPCK* genes in constant light

This shows the products obtained from RT-PCR analysis of leaves from non-nodulated soybean plants. The time points indicate hours in constant light. The black box indicates the preceding dark period, white boxes indicate subjective day and the shaded boxes indicate subjective night. 27 cycles of amplification were used for all GmPPCK genes. For each time point tissue was collected from a single plant.

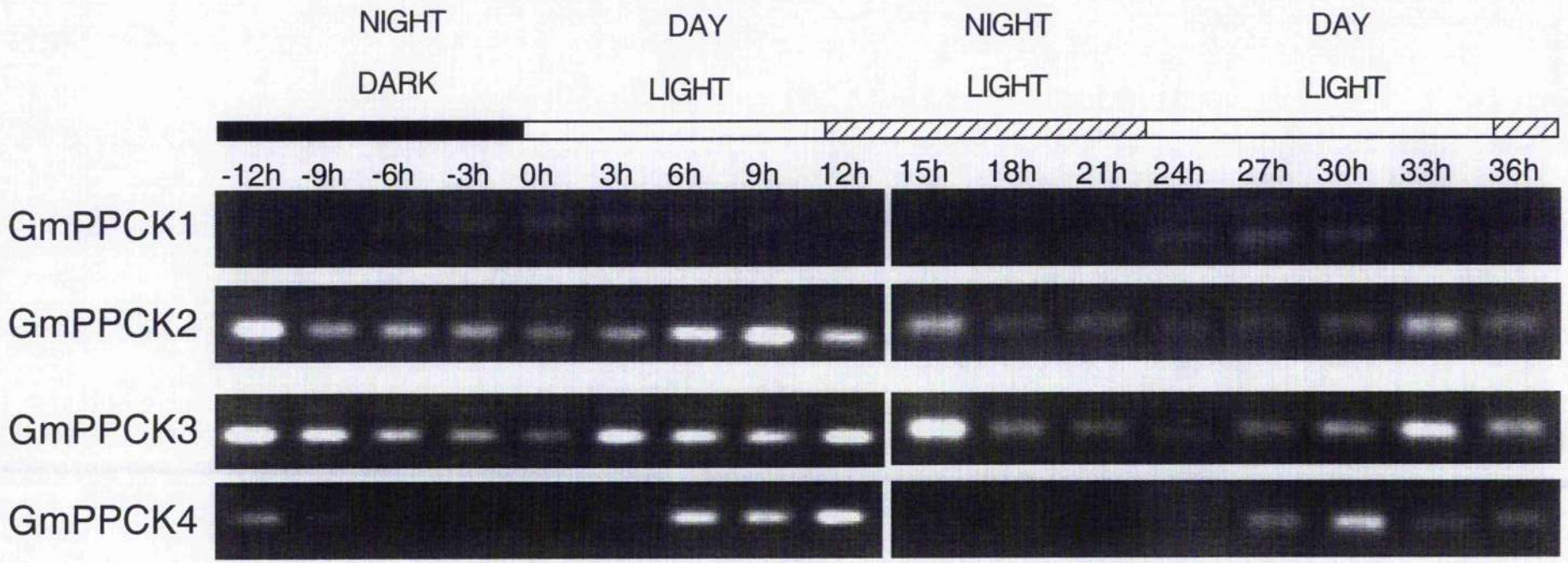

NIGHT
DARK
DAY
LIGHT
NIGHT
LIGHT
DAY
LIGHT
-12h -9h -6h -3h 0h 3h 6h 9h 12h 15h 18h 21h 24h 27h 30h 33h 36h
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4

Figure 4.5 Expression of *GmPPCK* genes in leaves in constant light

This shows the products obtained from RT-PCR analysis of leaves from non-nodulated soybean plants. The time points indicate hours in constant light. The black box indicates the preceding dark period, white boxes indicate subjective day and the shaded boxes indicate subjective night. *GmpLHY* was used as a circadian control and ubiquitin as a control for equal loading. 27 cycles of amplification were used for all GmPPCK genes, 25 cycles for GmpLHY and 20 cycles for Ubiquitin. For each time point tissue was collected from three plants and mixed prior to RNA extraction.

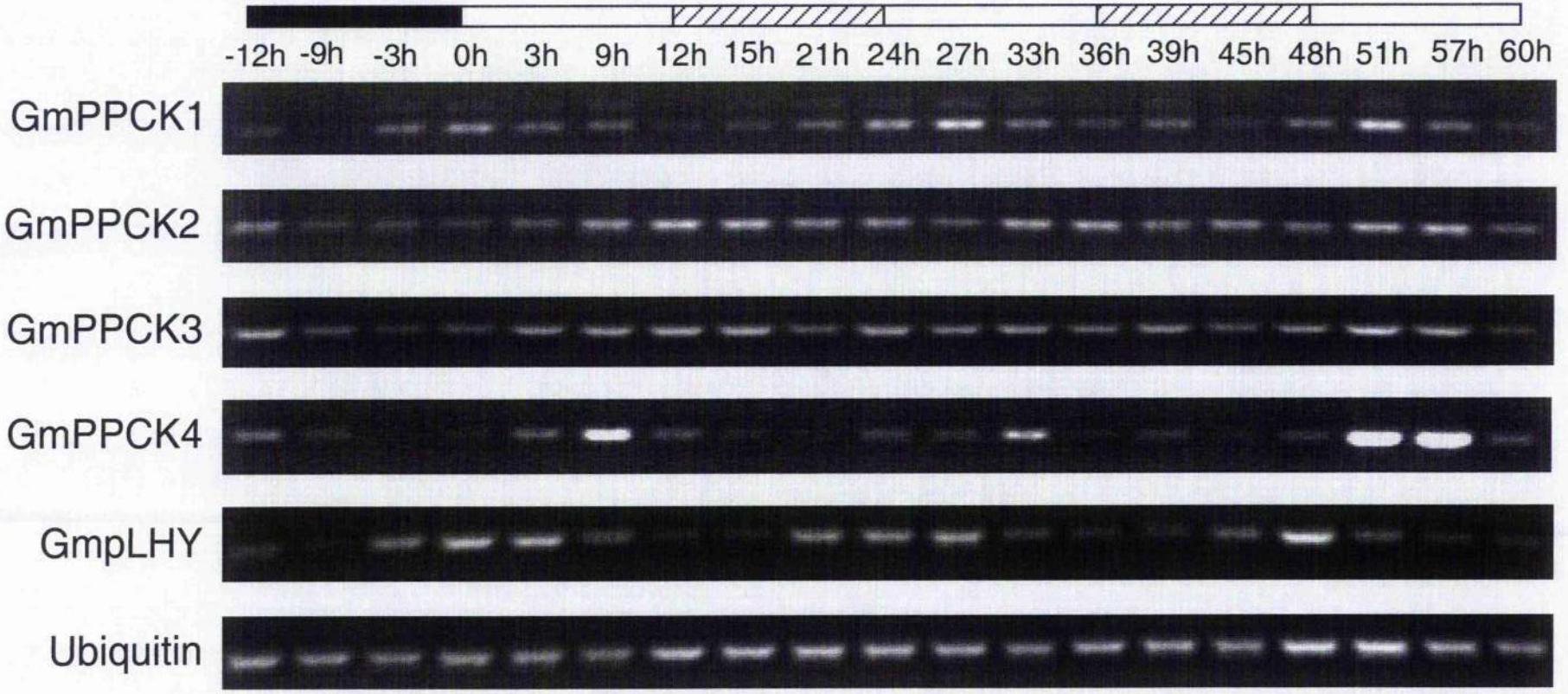

-12h -9h -3h 0h 3h 9h 12h 15h 21h 24h 27h 33h 36h 39h 45h 48h 51h 57h 60h
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
GmpLHY
Ubiquitin

Figure 4.6 Expression of *GmPPCK* genes in leaves in constant darkness

This shows the products obtained from RT-PCR analysis of leaves from non-nodulated soybean plants. The time points indicate hours in constant darkness. The white box indicates the preceding light period, black boxes indicate subjective night and the shaded boxes indicate subjective day. *GmpLHY* was used as a circadian control and ubiquitin as a control for equal loading. 27 cycles of amplification were used for all GmPPCK genes, 25 cycles for GmpLHY and 20 cycles for Ubiquitin. For each time point tissue was collected from three plants and mixed prior to RNA extraction.

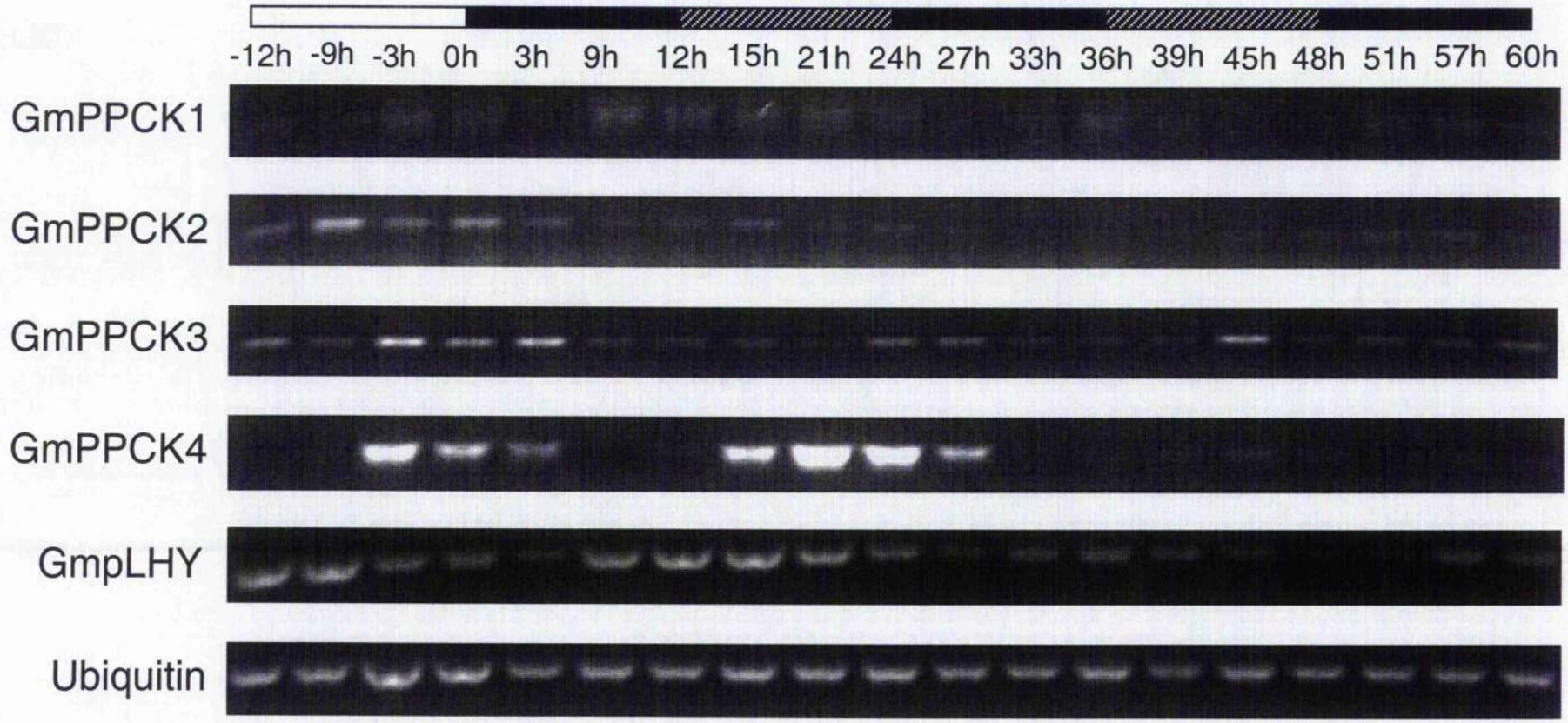

-12h -9h -3h 0h 3h 9h 12h 15h 21h 24h 27h 33h 36h 39h 45h 48h 51h 57h 60h
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
GmpLHY
Ubiquitin

Figure 4.7 Expression of *GmPPCK* genes in roots in constant light

This shows the products obtained from RT-PCR analysis of roots from non-nodulated soybean plants. The time points indicate hours in constant light. The black box indicates the preceding dark period, white boxes indicate subjective day and the shaded boxes indicate subjective night. *GmpLHY* was used as a circadian control and ubiquitin as a control for equal loading. 27 cycles of amplification were used for GmPPCK1 and GmPPCK4, 28 cycles for GmPPCK2 and GmPPCK3, 25 cycles for GmpLHY and 20 cycles for Ubiquitin. For each time point tissue was collected from three plants and mixed prior to RNA extraction.

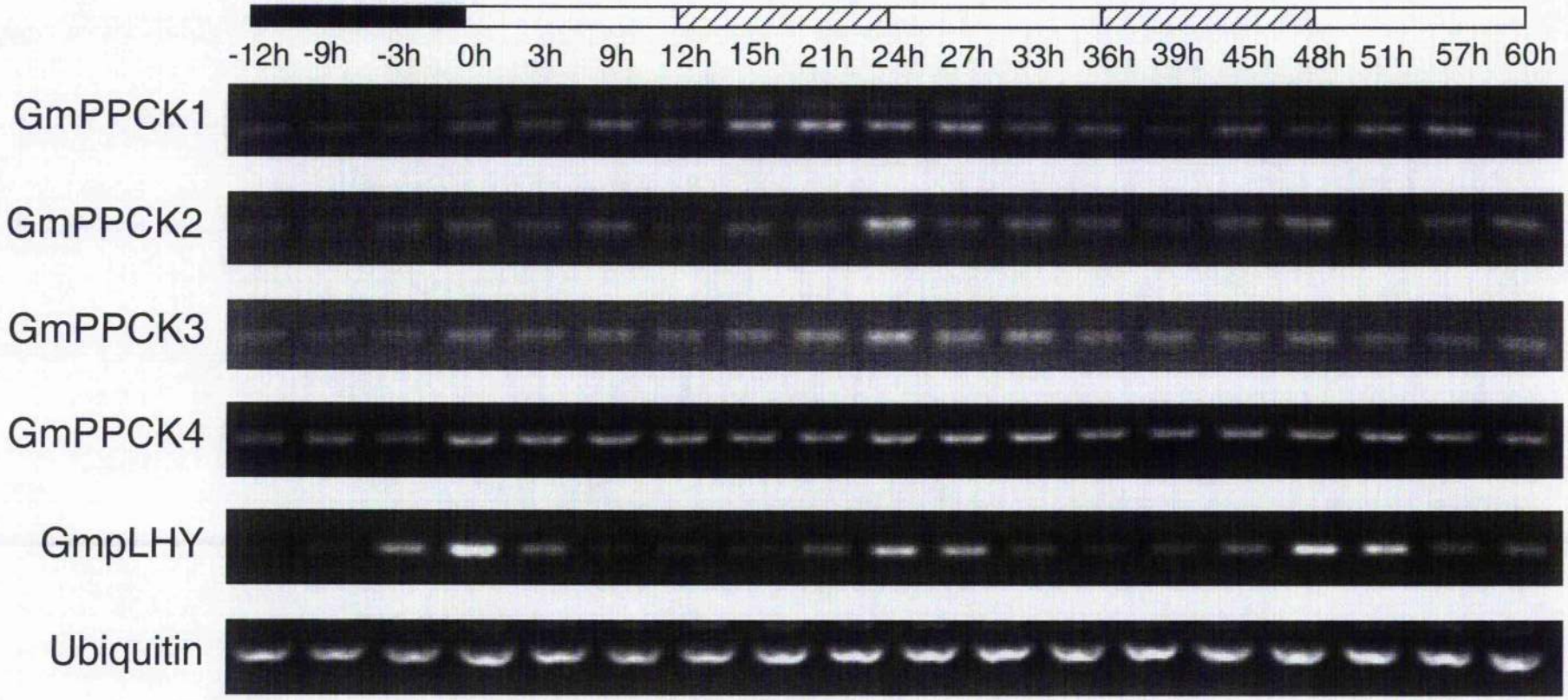

-12h -9h -3h 0h 3h 9h 12h 15h 21h 24h 27h 33h 36h 39h 45h 48h 51h 57h 60h
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
GmpLHY
Ubiquitin

Figure 4.8 Expression of *GmPPCK4* in roots in constant darkness

This shows the products obtained from RT-PCR analysis of roots from non-nodulated soybean plants. The time points indicate hours in constant darkness. The white box indicates the preceding light period, black boxes indicate subjective night and the shaded boxes indicate subjective day. *GmpLHY* was used as a circadian control and ubiquitin as a control for equal loading. 27 cycles of amplification were used for GmPPCK1 and GmPPCK4, 28 cycles for GmPPCK2 and GmPPCK3, 25 cycles for GmpLHY and 20 cycles for Ubiquitin. For each time point tissue was collected from three plants and mixed prior to RNA extraction.

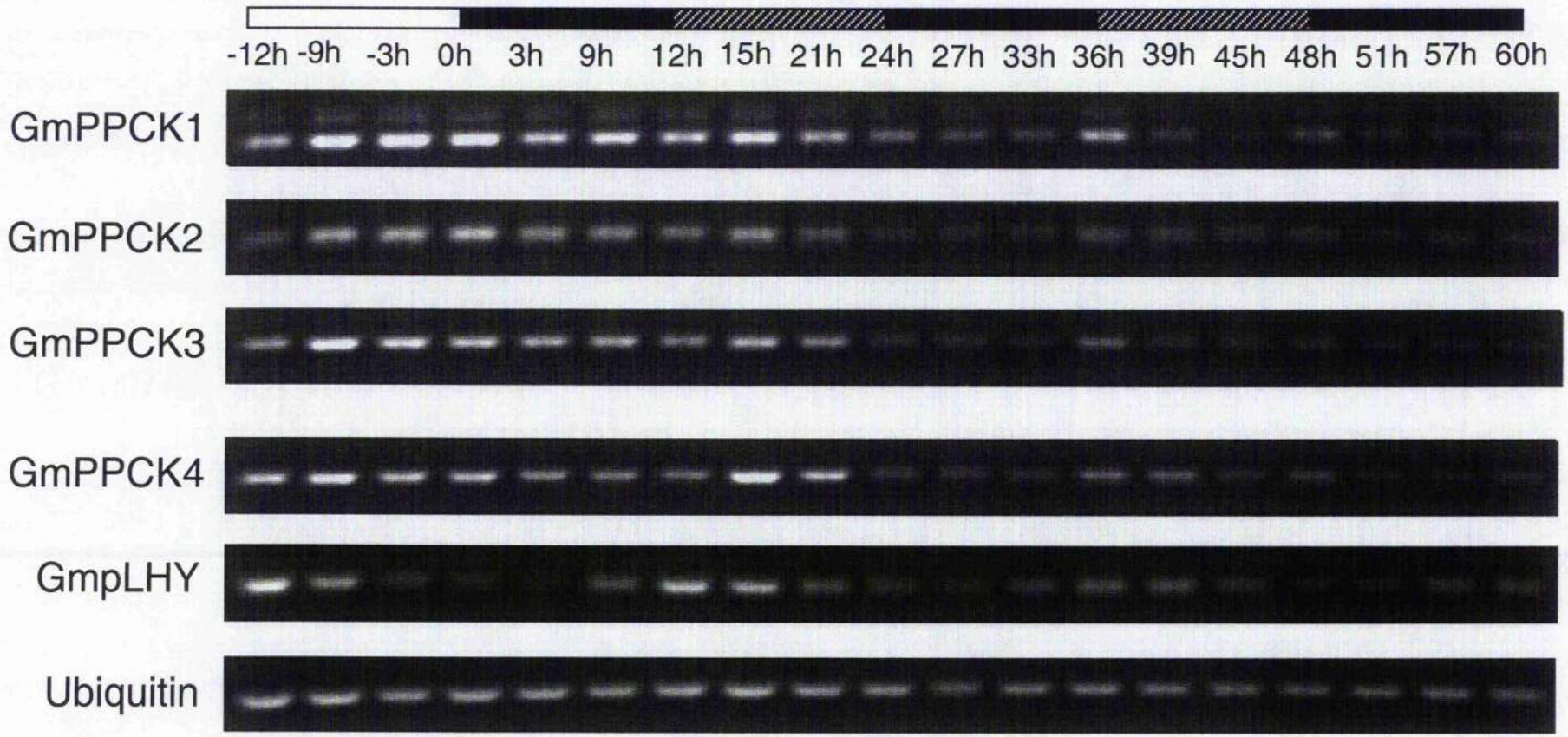
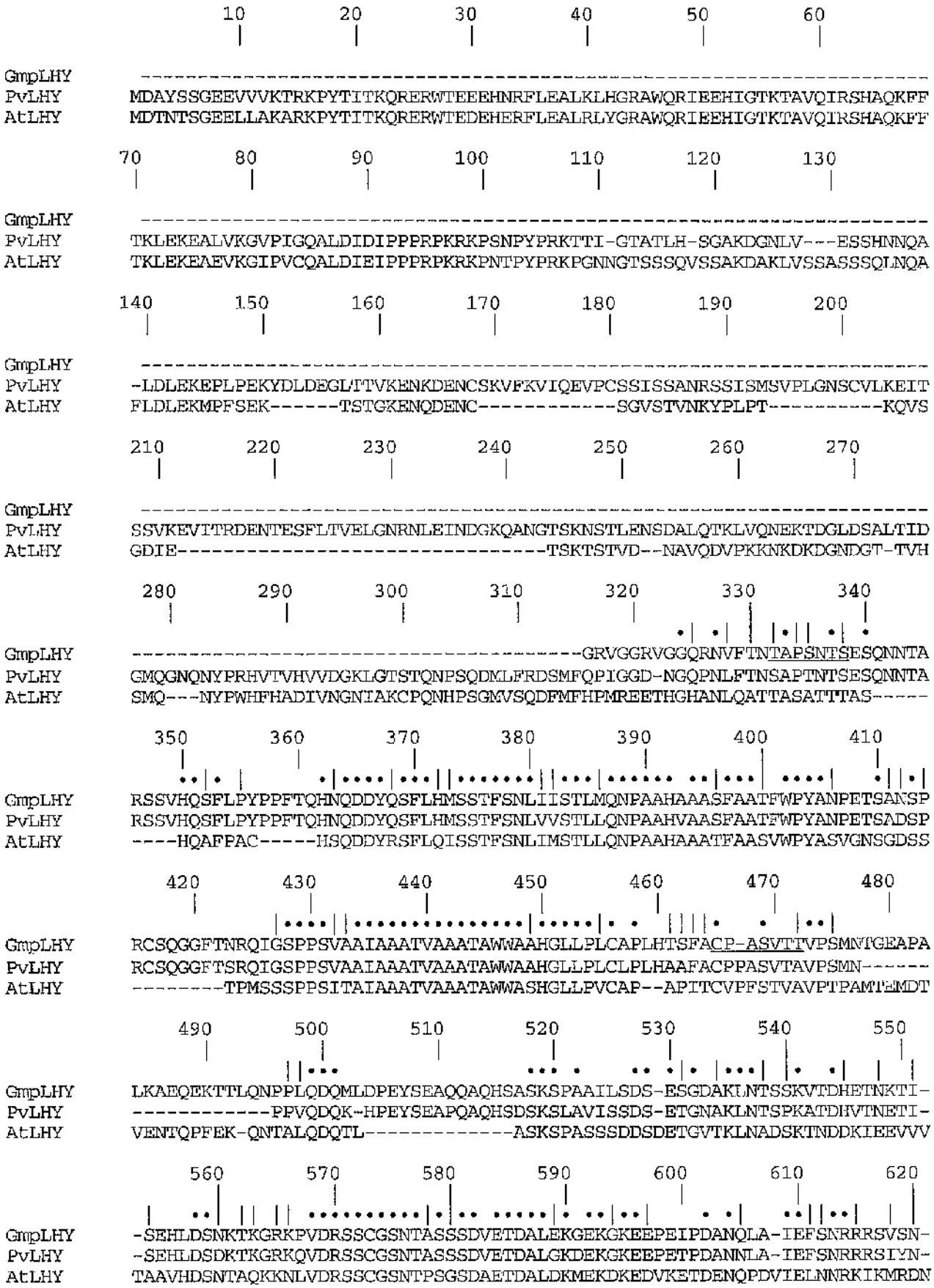
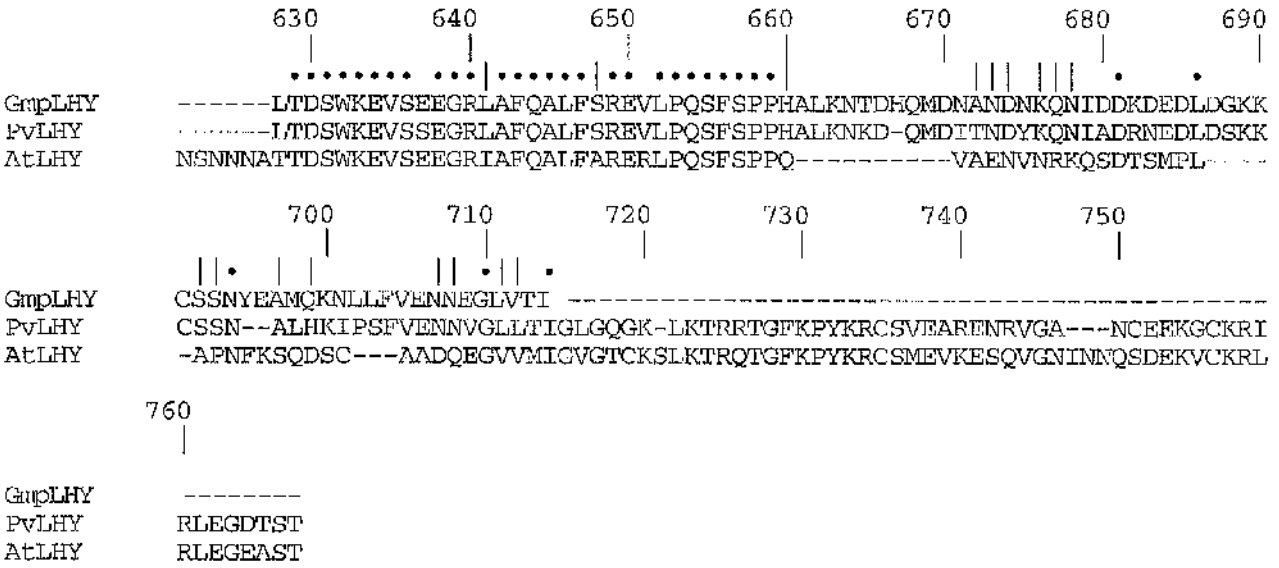

-12h -9h -3h 0h 3h 9h 12h 15h 21h 24h 27h 33h 36h 39h 45h 48h 51h 57h 60h
GmPPCK1
GmPPCK2
GmPPCK3
GmPPCK4
GmpLHY
Ubiquitin

Figure 4.9 Sequence analysis of *LHY* genes

Alignment of deduced amino acid sequences of Arabidopsis, common bean and soybean *LHY* genes. AtLHY (CAA07004) is from Schaffer *et al.* (1998) and PvLHY (AJ420902) from Kaldis *et al.* (2003). Underlined residues show the positions of the primer binding sites used to amplify GmpLHY. Dots and vertical lines denote identical and similar residues respectively.


GmpLHY
PvLHY
AtLHY
MDAYSSGEEVVVKTRKPYTITKQRERWTEEEHNRFLEALKLHGRAWQRIEEHIGTKTAVQIRSHAQKFF
MDTNTSGEELLAKARKPYTITKQRERWTEDEHERFLEALRLYGRAWQRIEEHIGTKTAVQIRSHAQKFF
TKLEKEALVKGVPIGQALDIDIPPPRPKRKPSNPYPRKTTI-GTATLH-SGAKDGNLV---ESSHNNQA
TKLEKEAEVKGIPVCQALDIEIPPPRPKRKPNTPYPRKPGNNGTSSSQVSSAKDAKLVSSASSSQLNQA
-LDLEKEPLPEKYDLDEGLTTVKENKDENCSKVFKVIQEVPCSSISSANRSSISMSVPLGNSCVLKEIT
FLDLEKMPFSEK------TSTGKENQDENC----------SGVSTVNKYPLPT----------KQVS
SSVKEVITRDENTESFLTVELGNRNLEINDGKQANGTSKNSTLENSDALQTKLVQNEKTDGLDSALTID
GDIE----------TSKTSTVD--NAVQDVPKKNKDKDGNDGT-TVH
GRVGGRVGGQRNVFTNTAPSNTSESQNNTA
GMQGNQNYPRHVTVHVVDGKLGTSTQNPSQDMLFRDSMFQPIGGD-NGQPNLFTNSAPTNTSESQNNTA
SMQ---NYPWHFHADIVNGNIAKCPQNHPSGMVSQDFMFHPMREETHGHANLQATTASATTTAS-----
RSSVHQSFLPYPPFTQHNQDDYQSFLHMSSTFSNLIISTLMQNPAAHAAASFAATFWPYANPETSANSP
RSSVHQSFLPYPPFTQHNQDDYQSFLHMSSTFSNLVVSTLLQNPAAHVAASFAATFWPYANPETSADSP
----HQAFPAC-----HSQDDYRSFLQISSTFSNLIMSTLLQNPAAHAAATFAASVWPYASVGNSGDSS
RCSQGGFTNRQIGSPPSVAAIAAATVAAATAWWAAHGLLPLCAPLHTSFACP-ASVTTVPSMNTGEAPA
RCSQGGFTSRQIGSPPSVAAIAAATVAAATAWWAAHGLLPLCLPLHAAFACPPASVTAVPSMN------
--------TPMSSSPPSITAIAAATVAAATAWWASHGLLPVCAP--APITCVPFSTVAVPTPAMTEMDT
LKAEQEKTTLQNPPLQDQMLDPEYSEAQQAQHSASKSPAAILSDS-ESGDAKLNTSSKVTDHETNKTI-
----------PPVQDQK-HPEYSEAPQAQHSDSKSLAVISSDS-ETGNAKLNTSPKATDHVTNETI-
VENTQPFEK-QNTALQDQTL----------ASKSPASSSDDSDETGVTKLNADSKTNDDKIEEVVV
-SEHLDSNKTKGRKPVDRSSCGSNTASSSDVETDALEKGEKGKEEPEIPDANQLA-IEFSNRRRSVSN-
-SEHLDSDKTKGRKQVDRSSCGSNTASSSDVETDALGKDEKGKEEPETPDANNLA-IEFSNRRRSIYN-
TAAVHDSNTAQKKNLVDRSSCGSNTPSGSDAETDALDKMEKDKEDVKETDENQPDVIELNNRKIKMRDN


630 640 650 660 670 680 690
GmpLHY ------LTDSWKEVSEEGRLAFQALFSREVLPQSFSPPHALKNTDHQMDNANDNKQNIDDKDEDLDGKK
PvLHY ------LTDSWKEVSSEGRLAFQALFSREVLPQSFSPPHALKNKD-QMDITNDYKQNIADRNEDLDSKK
AtLHY NSNNNATTDSWKEVSEEGRIAFQALFARERLPQSFSPPQ----------VAENVNRKQSDTSMPL
700 710 720 730 740 750
GmpLHY CSSNYEAMQKNLLFVENNEGLVTI ----------
PvLHY CSSN--ALHKIPSFVENNVGLLTIGLGQGK-LKTRRTGFKPYKRCSVEARENRVGA---NCEFKGCKRI
AtLHY -APNFKSQDSC---AADQEGVVMICVGTCKSLKTRQTGFKPYKRCSMEVKESQVGNINNQSDEKVCKRL
760
GmpLHY --------
PvLHY RLEGDTST
AtLHY RLEGEAST

Chapter 5

Analysis of *PPCK* genes in other C_3 plants

5.1 Introduction

Although the main focus of the work presented in this thesis is the analysis of PEPc kinase in soybean, unusual features of *PPCK* genes identified in other C_3 plants led to investigations of *PPCK* genes in tomato (*Lycopersicon esculentum*) and rice (*Oryza sativa*).

One of the many roles of PEPc in C_3 plants is the production of organic acids in developing fruits. In the model fruit tomato, malic and citric acid levels increase at the end of cell division phase of fruit development, peak at the end of the cell expansion phase, and decline during ripening (Guillet *et al.*, 2002). As in guard cells, organic acids provide osmotic pressure to drive cellular expansion (Cotelle *et al.*, 1999). PEPc has also been proposed to play a role during the later stages of fruit development (Chollet *et al.*, 1996) through the re-fixation of respiratory CO_2, which is produced in high amounts in ripening fruits. Two PEPc isoforms have been identified in tomato (Guillet *et al.*, 2002). One of the isoforms (*LePpc1*) is expressed in a range of tissues, while *LePpc2* is expressed mainly in expanding fruit tissue (Guillet *et al.*, 2002). The expression of *LePpc2* during tomato fruit development is tightly correlated with PEPc protein levels, PEPc activity and the accumulation of organic acids. This raised the question of the involvement of PEPc kinase in fruit development.

Work in our lab led to the identification of two PEPc kinase isoforms from tomato (Marsh, 2002). The genomic sequence of *LePPCK1* showed it was structurally similar to previously reported *PPCK* genes (Hartwell *et al.*, 1999a; Taybi *et al.*, 2000; Tsuchida *et al.*, 2001; Fontaine *et al.*, 2002). The coding sequence comprised two exons, with one 89 bp intron at the 3' end of the coding sequence. The sequence of the other isoform, *LePPCK2*, showed a surprising feature, it contains two introns. One intron, as in other *PPCK* genes, is located at the 3' of the coding sequence. The second intron is located close to the middle of the coding region and possesses two alternative 5' splice sites, therefore allowing alternative splicing. The region between these two splice sites contains an in-frame stop codon. Fig. 5.1 shows the organisation of *LePPCK2*. Splicing can therefore give rise to three different transcripts: transcript 1 encodes a functional PEPc kinase; transcript 2 is incorrectly spliced and contains a

premature stop codon; transcript 3 is an unspliced transcript with a premature stop codon. Transcripts 2 and 3 both encode a truncated, non-functional PEPc kinase.

Work by Marsh (2002) showed that *LePPCK1* transcripts were present in all tissues examined, with lowest expression levels early on in fruit development (expansion phase 1) and highest in mature leaves (Fig. 5.2). *LePPCK2* expression levels increased during fruit ripening, being highest in ripe red fruit (Fig. 5.2). Interestingly the relative abundance of the three transcripts varied in different tissues. This was also evident when *LePPCK2* transcript abundance in different fruit compartments was examined (Fig. 5.3). These results led us to examine the relative abundance of *LePPCK2* transcripts in a tomato mutant and to identify homologous genes in other solanaceous species.

The production of a draft rice genome sequence presented us with the opportunity to identify the number of PEPc kinase genes in a C_3 monocot species. There has been much research into the transferral of the beneficial traits of C_4 photosynthesis into C_3 crop plants (reviewed in Hausler *et al.*, 2002; Leegood, 2002; Miyao, 2003). To this end, transgenic rice plants expressing the maize photosynthetic PEPc isoform have been produced (Ku *et al.*, 1999). Levels of PEPc expression two to three times greater than those found in maize leaves were achieved, but rates of photosynthesis were not significantly different to untransformed rice plants under atmospheric conditions. Fukayama *et al.* (2002) demonstrated that the maize PEPc expressed in the leaves of transgenic rice plants underwent phosphorylation. However, in contrast to the light induced increase in PEPc phosphorylation observed in maize leaves, the maize PEPc in rice became phosphorylated during the night and remained dephosphorylated in the day. Therefore, knowledge of the control of PPCK activity in rice is required if the desired pattern of PEPc activity is to be achieved.

5.2 Results

5.2.1 Analysis of *PPCK* genes in the Solanaceae

With previous results demonstrating variation of *PPCK* transcript abundance in different tissues of wild-type tomato (Figs. 5.2 and 5.3) further experiments were conducted to assess the expression of *LePPCK1* and *LePPCK2* in the tissues of the *green flesh* (*gf*) tomato mutant. In this mutant the fruit retain substantial amounts of chlorophyll during ripening, while other ripening changes, such as ethylene synthesis and lycopene accumulation are normal (Akhtar *et al.*, 1999). The results of these

experiments are shown in Fig. 5.4. *LePPCK1* is highly expressed in leaves, roots and flowers. It is present in the skin and pericarp tissue of ripe fruit but hard to detect in locule tissue. *LePPCK2* shows differences in the relative levels of transcript types depending upon the tissue. This is evident in leaves where predominately transcript 3 is produced, compared to roots where little difference between the levels of transcript types is observed. This is in contrast to the expression pattern in the wild-type Alicante tomato cultivar, where leaves expressed mainly transcripts 2 and 3 and roots express mainly transcript 1. Expression studies of *LePPCK2* in the different compartments of ripe *gf* fruit show that all three transcripts are present in skin and pericarp tissue, like in Alicante, but transcript 1 and 2 are present in the locule and seeds whereas in Alicante only transcript 1 was detected in these tissues.

The presence of sequences homologous to *LePPCK2* in other members of the Solanaceae was also investigated. Analysis of EST databases revealed sequences very similar to *LePPCK2* were present in potato (*Solanum tuberosum*). Primers were designed based on the EST sequences so they would flank the putative additional intron site. Three products of the expected mobilities were obtained from potato leaf cDNA and a single product from genomic DNA. The genomic product was cloned and sequenced (555 bp, Genbank accession no. AY293738) and found to match the EST sequences almost exactly. Using the primers designed for RT-PCR of *LePPCK2*, which span the unusual intron in *LePPCK2*, products were also amplified from the genomic DNA of aubergine (*Solanum melongena*) and tobacco (*Nicotiana tabacum*). These partial sequences (553 bp from aubergine, Genbank accession no. AY236482, and 538 bp from tobacco, Genbank accession no. AY347261) are very similar to the *LePPCK2* and potato sequences. Fig. 5.5 shows an alignment of the intron sequences, which comprise 233, 234, 232 and 217 bp in tomato, potato, aubergine and tobacco, respectively. The nucleotide sequence of the intron is highly conserved between these four members of the Solanaceae. In fact the intron and exon sequences are equally similar. The tomato and aubergine sequences are 94% identical over 321 bases of exon sequence (not shown) and 92% identical over 233 intron bases (see Fig. 5.5). The two donor splice sites and the acceptor site are conserved. All four sequences contain at least one in frame stop codon 70 – 80 bp into the intron (bold in Fig. 5.5); aubergine has two closely adjacent stop codons. The tobacco intron has a deletion of 18 bp relative to the other introns just 3' to the second ("incorrect") donor splice site. Using the same primers as for *LePPCK2*, products were obtained from mature leaf,

root and flower cDNA from tobacco (Fig. 5.6). The three transcript types are present in these tissues, but as with tomato, the relative abundance of the transcript types differs between organs.

5.2.2 *PPCK* genes in rice: identification and demonstration of function

Analysis of the rice genome revealed the presence of three gene sequences very similar to other plant PEPc kinases. These putative rice *PPCK* genes were termed *OsPPCK1*, *OsPPCK2* and *OsPPCK3* respectively. Analysis of the genomic sequences shows that *OsPPCK1* and *OsPPCK3* have similar gene structures to other plant *PPCK* genes. They consist of two exons with an intron at the 3' end. The deduced amino acid sequences consist of 289 and 285 residues for *OsPPCK1* and *OsPPCK3* respectively. The *OsPPCK2* sequence shows two interesting features (Fig. 5.7). While it also consists of two exons with an intron at the 3' end, the intron does not contain an in frame stop codon, as is usual for plant *PPCK* genes. The only other identified plant *PPCK* sequence that shares this trait is *GmPPCK3*. Therefore unspliced *OsPPCK2* transcripts, like *GmPPCK3*, could encode a larger PPCK protein. The other unusual feature of the *OsPPCK2* sequence is the presence of three in frame methionine codons at the 5' end of the gene. Amino acid alignments of the three rice *PPCK* sequences with other plant *PPCK* sequences (Fig. 5.8) show that the third methionine residue of *OsPPCK2* aligns with the start methionine from the other plant *PPCK* genes. Translation initiation from the first methionine would encode a larger protein of 344 amino acids with a putative molecular mass of 36.6 kDa. This is a similar mass to the 37 kDa PEPc kinase that has been observed in several plant species, but no *PPCK* gene that would encode a protein of this size has yet been identified. Attempts to obtain a full length PCR product of *OsPPCK2* from rice leaf cDNA were unsuccessful. However, we were provided with full-length cDNA clones of *OsPPCK1*, *2* and *3*, a genomic clone of *OsPPCK2* and a cDNA clone of *OsPPCK2* from the first methionine by Dr Hiroshi Fukayama (National Institute of Agrobiological Resources, Japan), in order to test their ability to encode a functional PEPc kinase. Fig. 5.9 shows the *in vitro* transcription products of the five clones separated on a formaldehyde/MOPS gel. All the clones produce transcripts of the expected sizes. The transcription products were then translated and the translation products assayed for their ability to phosphorylate purified *Kalanchoë fedtschenkoi* PEPc. Fig. 5.10A shows the translation products and Fig. 5.10B shows the results of

the kinase assays. The translation products from *OsPPCK1* and *OsPPCK3* cDNA clones were both able to phosphorylate PEPc. For *OsPPCK2*, both cDNA clones made active PEPc kinase, while the *OSPPCK2* genomic clone did not show any kinase activity. This is the same result as was seen with *GmPPCK3* genomic clone and suggests that the presence of the translated intron sequence abolishes kinase activity. The 'long' form of *OsPPCK2*, containing all three methionine codons, produces three protein products. This is presumably due to the initiation of translation from each of the three methionines. The use of KCl instead of KAc in the *in vitro* translation reaction buffer, as recommended in the manufacturer's protocol to favour translation from the first methionine codon, did not prevent this from occurring (Fig. 5.10A, lane 7). Therefore it is not possible to determine whether the kinase activity from this clone is solely due to the protein translated from the 'normal' third methionine, and whether translation from either of the other upstream methionine codons will encode an active PEPc kinase.

5.3 Discussion

The demonstration that *LePPCK2* was subject to alternative splicing was a surprising discovery. However the functional significance of this observation was unclear. The different ratios of three transcripts produced from *LePPCK2* seen in different tissues of wild-type Alicante tomato cultivar could reflect control of *LePPCK2* splicing and therefore suggest functional significance. Analysis of *LePPCK2* expression in the *gf* mutant revealed some differences in the ratios of the three transcripts in certain tissues compared to the wild-type plants. This would suggest that the altered metabolism of the *gf* plants impacts on the control of *LePPCK2* splicing. Bortolotti *et al.* (2003) have demonstrated that enzymes involved in the primary metabolism of glutamate are regulated differently in *gf* fruit during ripening. The requirement of a splicing factor essential for the production of transcript 1 could act as a control point for correct expression of *LePPCK2*. Control of the production of such a splicing factor, modulated by developmental and/or metabolic signals, is one potential hypothesis.

The presence of alternative splicing in other members of the Solanaceae and the conservation of the intron sequence are further lines of evidence to suggest a potential functional role. The occurrence of an in-frame stop codon leading to premature truncation of incorrectly spliced transcripts is seen in all four sequences

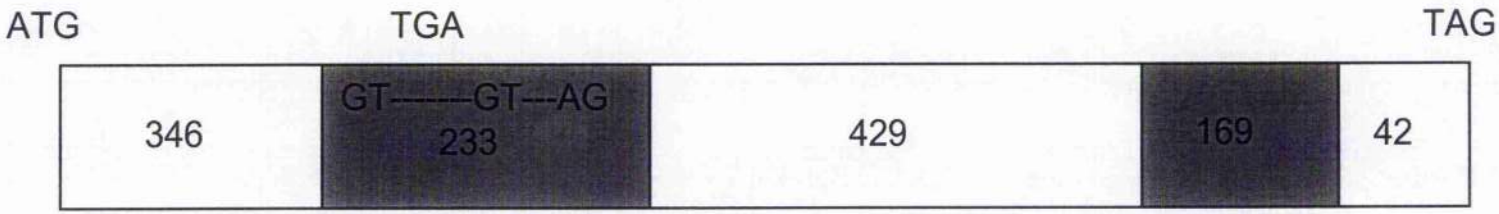
(Fig. 5.5). Whether the truncated protein product of transcripts 2 and 3 accumulates *in vivo*, or the mRNA is unstable and rapidly degraded, is unknown. The truncated protein would extend just beyond the N-terminal ATP-binding domain of PEPc kinase; it may be able to fold stably, to bind ATP, and/or interact with other proteins such as PEPc. The production of antibodies specific to the truncated LePPPCK2 polypeptide would be required to demonstrate *in vivo* accumulation, although the low abundance of PEPc kinase would hinder such an experiment.

The identification of the *PPCK* genes present in rice also revealed an unusual feature in the sequence of one of the genes; *OsPPCK2* contained three in-frame methionine codons. A long-standing question has been the identity of the 37 kDa PEPc kinase isoform revealed to exist in several different plant species, including soybean, by studies using "in-gel" kinase assays (Chollet *et al.*, 1996; Vidal and Chollet, 1997 and references therein). The *OsPPCK2* sequence is the first *PPCK* sequence identified to date that could potentially encode a larger PEPc kinase isoform. Although function was demonstrated for the products of *OsPPCK 1*, *2* and *3* cDNAs, experiments using *in vitro* translation and kinase assays were unable to provide unequivocal evidence to show whether the 'long' form of OsPPCK2 could phosphorylate PEPc. This was because translation could start from the other methionine codons. Our attempts to detect full-length *OsPPCK2* transcripts from rice cDNA were unsuccessful, but work in the laboratory of Dr Hiroshi Fukayama has revealed the presence of both the 'long' and 'normal' *OsPPCK2* transcripts. Furthermore, there is preliminary evidence that the production of these two transcript types is differentially regulated (H. Fukayama, unpublished). However, similar to the situation in tomato, it is important to demonstrate that the 37 kDa isoform of OsPPCK2 accumulates *in vivo*.

The identification of *OsPPCK2* prompted us to examine whether any of the *PPCK* genes identified in soybean contained in-frame upstream methionine codons, as a 37 kDa PEPc kinase isoform was previously identified in protein extracts from soybean root nodules (Zhang and Chollet, 1997). Analysis of the genomic sequences revealed no further methionine codons before the presence of in-frame stop codons. This would suggest that the 37 kDa PEPc kinase isoforms are not the products of translation from alternative methionine residues. However, the possibility that an unidentified *PPCK* gene encoding a larger PEPc kinase isoform exists in the soybean genome cannot be excluded.

Figure 5.1 Organisation of *LePPCK2*

The white boxes indicate exons, the shaded boxes introns. Numbers show the lengths of exons and introns in base pairs. The sequences around the splice sites are shown in full; the intron start GT and end AG sequences are underlined.



Exon 1| Intron 1 | Exon 2 | Intron 2 |Exon 3

...CATGGTACTCTT...GAAAGTAAATTG...TTTTATAGGTAC... CTCAGTATGAT...ATAAACAGGGCA...

<---------138 -------><---------- 95 ---------->

Figure 5.2 Expression patterns of *LePPCK* genes

This shows the products obtained from RT-PCR analysis. Ubiquitin was used as a control for equal loading.

Lanes are: 1, expansion phase 1 fruit; 2, expansion phase 2 fruit; 3, green fruit; 4, breaker fruit; 5, pink fruit; 6, red-ripe fruit; 7, young leaves; 8, mature leaves; 9, flowers; 10, seedlings in light; 11, seedlings in dark and 12. Taken from Marsh (2002)

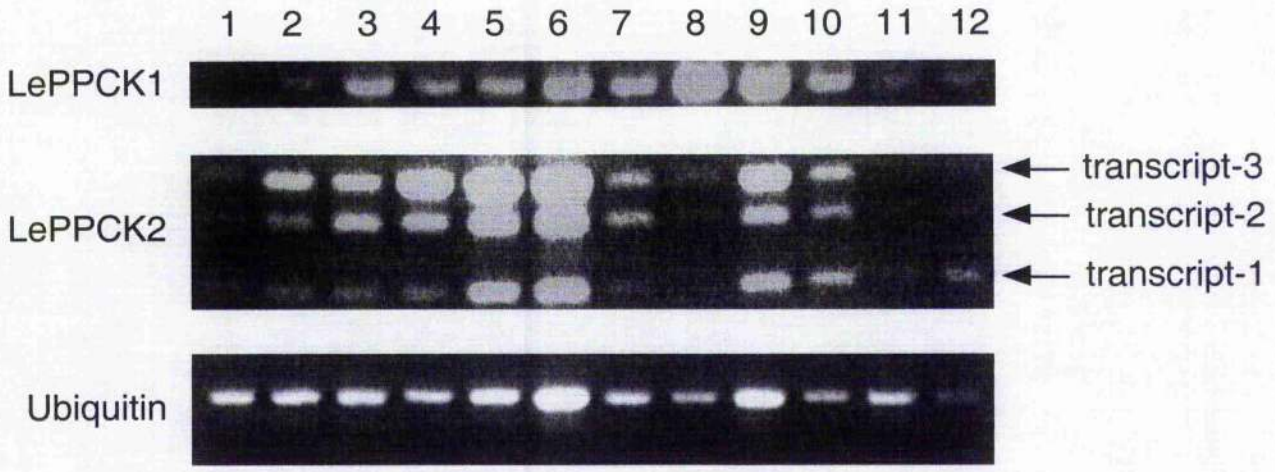

1 2 3 4 5 6 7 8 9 10 11 12
LePPCK1
LePPCK2
transcript-3
transcript-2
transcript-1
Ubiquitin

Figure 5.3 Expression pattern of *LePPCK2* in fruit compartments

This shows the products obtained from RT-PCR analysis from ripe red fruit. Ubiquitin was used as a control for equal loading. Taken from Marsh (2002)

Skin
Outer pericarp
Inner pericarp
Gel
Seeds
LePPCK1
LePPCK2
Ubiquitin

Figure 5.4 Expression patterns of *LePPCK* genes in the tomato *gf* mutant

This shows the products obtained from RT-PCR analysis. Ubiquitin was used as a control for equal loading. 30 cycles of amplification were used for all genes except for detection of transcripts in locule tissue where 35 cycles were used.

Mature Leaf
Young Leaf
Root
Flower
Skin
Pericarp
Locule
LePPCK1
LePPCK2
Ubiquitin

Figure 5.5 Sequence analysis of the unusual *PPCK2* intron in members of the Soalanaceae

Alignment of the nucleotide sequences from tomato (Le), potato (St), aubergine (Sm) and tobacco (Nt). The in-frame stop codons are in bold. The sequences surrounding the internal splice site are underlined, with the intron start GT in italics.

```
Le    GTACTCTTCC AGATCACCAA AATCTCACTC GTTGCTGATC AAAATGGGGA AATTTT-ACT
St    GTACTCTTCC AGATCACCAA AATCTCACTC GTTGCTGATC AAAATGGGGA AATTTTTACT
Sm    GTACTCTTCC AGATTACCAA AATCTCACTC GCTGCTGATC ATAAAAGGGG AATTTT-ACT
Nt    GTACTCTTCC AGATCACCAC AATCTCACTC GTTGCTGATC AAAAAGGGGA AATTTT-ACT

Le    TTTGCTTCGT AAAATGGCTT GAATCAGGAA CGATCTACGA TGCAGCACGT TGTTGAATTT
St    TTTGCTTCGT AAAATGGCTT GAATCAGGAG CGATCTACGA AGCAGCACGT TGTTGAATTT
Sm    TTTCCTTCGT TAAATGGCTT GAATCAGGAA CGATCTACGA AGGAGCATCT TGTTGAATTT
Nt    TTTGCTTCGT ATAATGGCTT GAATCAGGAA CGATCTACGA AGCAGCACCT TGTTGAATTT

Le    TCGAAAATTG GAGGAGAAAG TAAATTGTAA AGCTAAGCGA AGTAGAATTC ATAATCGGCT
St    TCGAAAATTG GAGGAGAAAG TAAATTGTAA AGCTAAGCGA AGTAGAATTC ATAATCGGCT
Sm    TGGAAAATTG GAGGAGGAAG TAA-TTGTAA AGCTAAGCGA AGTAGAATTT ATAATCGGCT
Nt    TCGAAAATTG GAGGAGAAAG TAAAT------ ---------- ---AGAATTC ATAATCGGCT

Le    TTT--ATTTT AGTTTTTCGT CATTTCTGAT AAATTTTATG CTTAATGTTTT TATAG
St    TTT--ATTTT AGTTTTTCGC CATGTCTGAT AATTTTTCTG CCTAATGTTTC TATAG
Sm    TTTAT-TTTA GGTTTTCGGT CATCTCTGAC AA-TTTTCTG TATAATGTTTG TACAG
Nt    TTTTTATTTT AGTTTTTCGT AATTACTGAT AAATTTTCTG CATAATGTATG TATAG
```

Figure 5.6 Analysis of *NtPPCK2* transcripts in tobacco

This shows the products obtained from RT-PCR analysis and PCR of genomic DNA. Ubiquitin was used as a control for equal loading. 35 cycles of amplification were used for NtPPCK2 and 30 cycles for Ubiquitin.

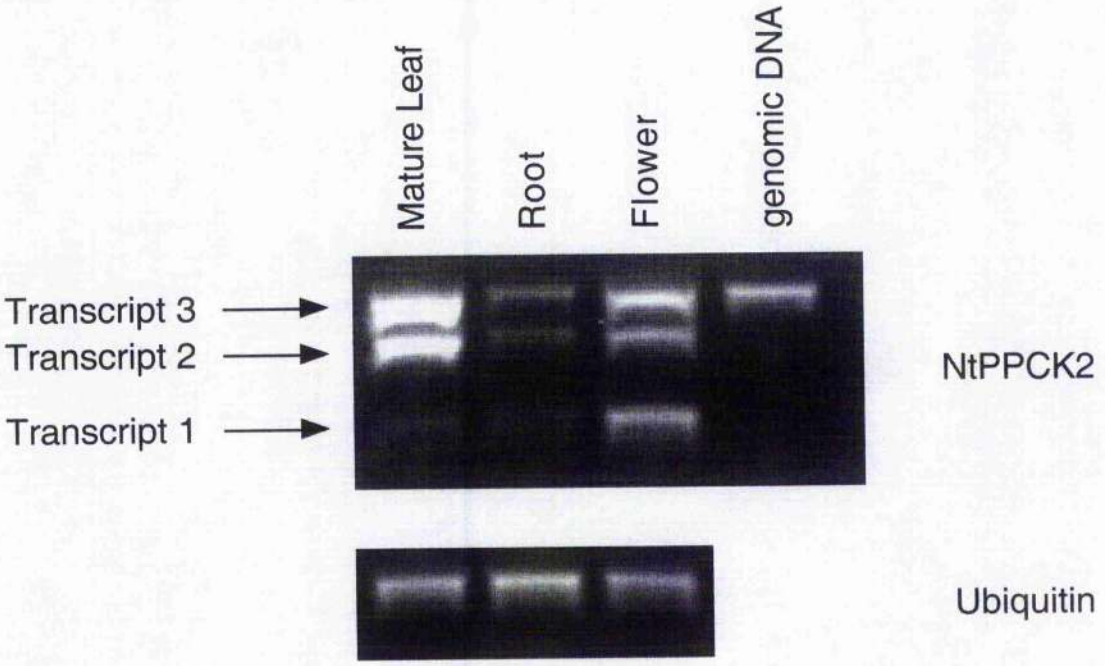
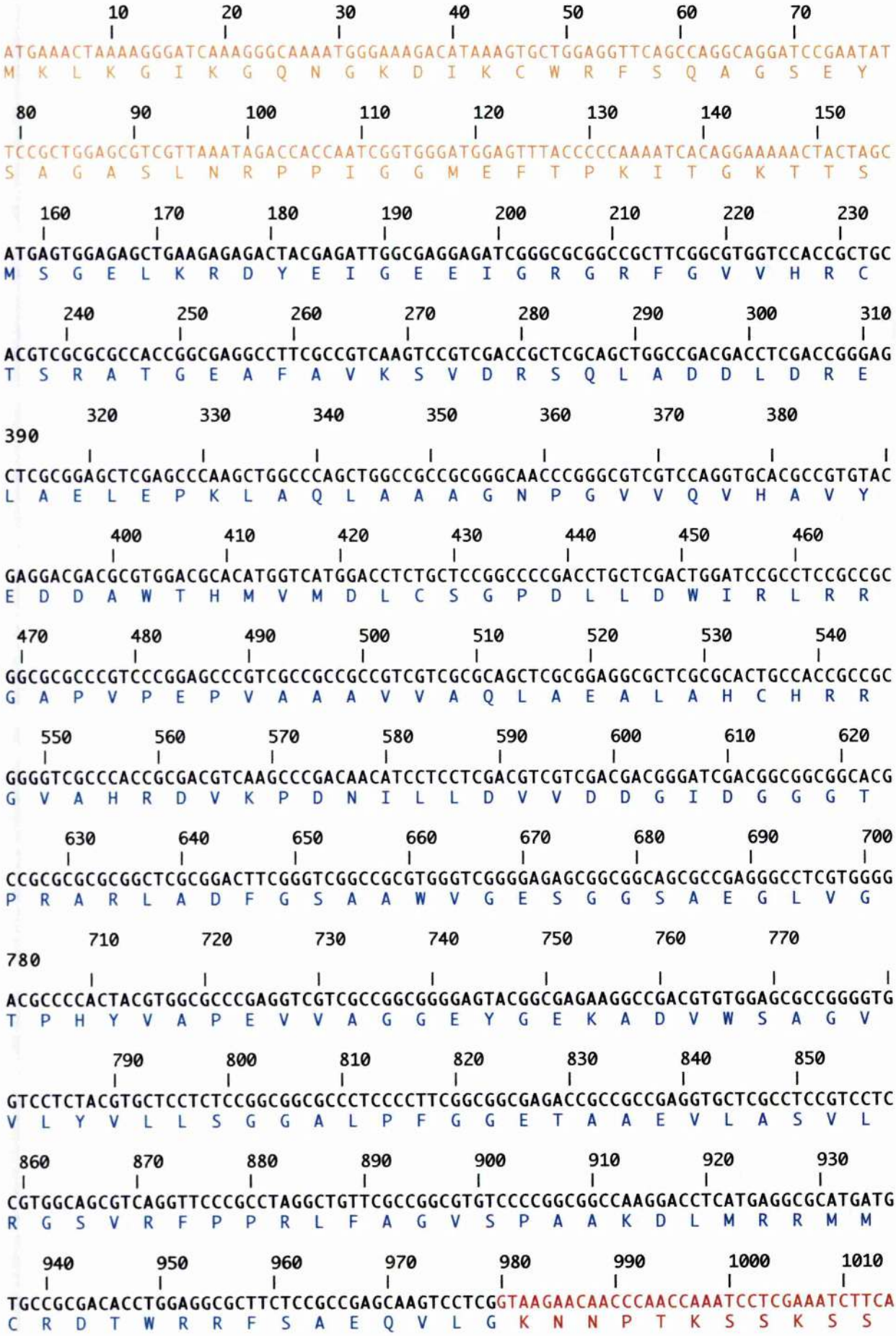

Mature Leaf
Root
Flower
genomic DNA
Transcript 3
Transcript 2
Transcript 1
NtPPCK2
Ubiquitin

Figure 5.7 Translation map of the genomic sequence of *OsPPCK2*

DNA Sequence in black denotes exon sequence; sequence in red denotes intron sequence. Sequence in orange denotes the sequence upstream of the 'typical' methionine residue.


10 20 30 40 50 60 70
ATGAAACTAAAAGGGATCAAAGGGCAAAATGGGAAAGACATAAAGTGCTGGAGGTTCAGCCAGGCAGGATCCGAATAT
M K L K G I K G Q N G K D I K C W R F S Q A G S E Y
80 90 100 110 120 130 140 150
TCCGCTGGAGCGTCGTTAAATAGACCACCAATCGGTGGGATGGAGTTTACCCCCAAAATCACAGGAAAAACTACTAGC
S A G A S L N R P P I G G M E F T P K I T G K T T S
160 170 180 190 200 210 220 230
ATGAGTGGAGAGCTGAAGAGAGACTACGAGATTGGCGAGGAGATCGGGCGCGGCCGCTTCGGCGTGGTCCACCGCTGC
M S G E L K R D Y E I G E E I G R G R F G V V H R C
240 250 260 270 280 290 300 310
ACGTCGCGCGCCACCGGCGAGGCCTTCGCCGTCAAGTCCGTCGACCGCTCGCAGCTGGCCGACGACCTCGACCGGGAG
T S R A T G E A F A V K S V D R S Q L A D D L D R E
320 330 340 350 360 370 380
390
CTCGCGGAGCTCGAGCCCAAGCTGGCCCAGCTGGCCGCCGCGGGCAACCCGGGCGTCGTCCAGGTGCACGCCGTGTAC
L A E L E P K L A Q L A A A G N P G V V Q V H A V Y
400 410 420 430 440 450 460
GAGGACGACGCGTGGACGCACATGGTCATGGACCTCTGCTCCGGCCCCGACCTGCTCGACTGGATCCGCCTCCGCCGC
E D D A W T H M V M D L C S G P D L L D W I R L R R
470 480 490 500 510 520 530 540
GGCGCGCCCGTCCCGGAGCCCGTCGCCGCCGCCGTCGTCGCGCAGCTCGCGGAGGCGCTCGCGCACTGCCACCGCCGC
G A P V P E P V A A A A V V V A Q L A E A L A H C H R R
550 560 570 580 590 600 610 620
GGGGTCGCCCACCGCGACGTCAAGCCCGACAACATCCTCCTCGACGTCGTCGACGACGGGATCGACGGCGGCGGCACG
G V A H R D V K P D N I L L D V V D D G I D G G G T
630 640 650 660 670 680 690 700
CCGCGCGCGCGGCTCGCGGACTTCGGGTCGGCCGCGTGGGTCGGGGAGAGCGGCGGCAGCGCCGAGGGCCTCGTGGGG
P R A R L A D F G S A A W V G E S G G S A E G L V G
710 720 730 740 750 760 770
780
ACGCCCCACTACGTGGCGCCCGAGGTCGTCGCCGGCGGGGAGTACGGCGAGAAGGCCGACGTGTGGAGCGCCGGGGTG
T P H Y V A P E V V A G G E Y G E K A D V W S A G V
790 800 810 820 830 840 850
GTCCTCTACGTGCTCCTCTCCGGCGGCGCCCTCCCCTTCGGCGGCGAGACCGCCGCCGAGGTGCTCGCCTCCGTCCTC
V L Y V L L S G G A L P F G G E T A A E V L A S V L
860 870 880 890 900 910 920 930
CGTGGCAGCGTCAGGTTCCCGCCTAGGCTGTTCGCCGGCGTGTCCCCGGCGGCCAAGGACCTCATGAGGCGCATGATG
R G S V R F P P R L F A G V S P A A K D L M R R M M
940 950 960 970 980 990 1000 1010
TGCCGCGACACCTGGAGGCGCTTCTCCGCCGAGCAAGTCCTCGGTAAGAACAACCCAACCAAATCCTCGAAATCTTCA
C R D T W R R F S A E Q V L G K N N P T K S S K S S

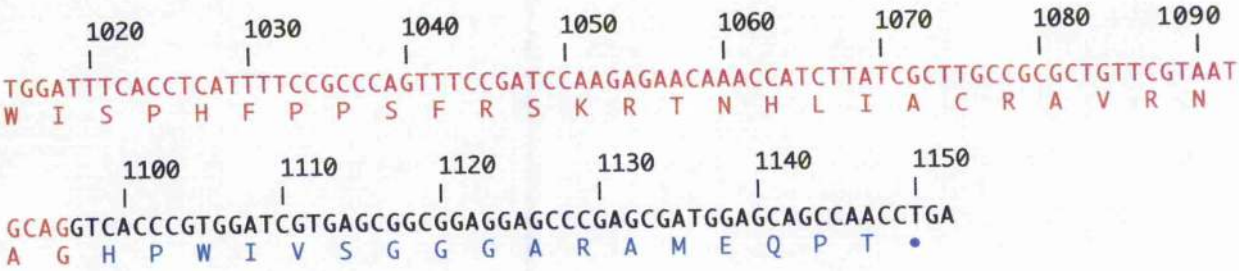

1020 1030 1040 1050 1060 1070 1080 1090
TGGATTTCACCTCATTTTCCGCCCAGTTTCCGATCCAAGAGAACAAACCATCTTATCGCTTGCCGCGCTGTTCGTAAT
W I S P H F P P S F R S K R T N H L I A C R A V R N
1100 1110 1120 1130 1140 1150
GCAGGTCACCCGTGGATCGTGAGCGGCGGAGGAGCCCGAGCGATGGAGCAGCCAACCTGA
A G H P W I V S G G G A R A M E Q P T •

Figure 5.8 Sequence analysis of the rice *PPCK* genes

Alignment of deduced PPCK amino acid sequences from rice (OsPPCK1, OsPPCK2 and OsPPCK3), maize (ZmPPCK2), soybean (GmPPCK2) and *Kalanchoë fedtschenkoi* (KfPPCK).

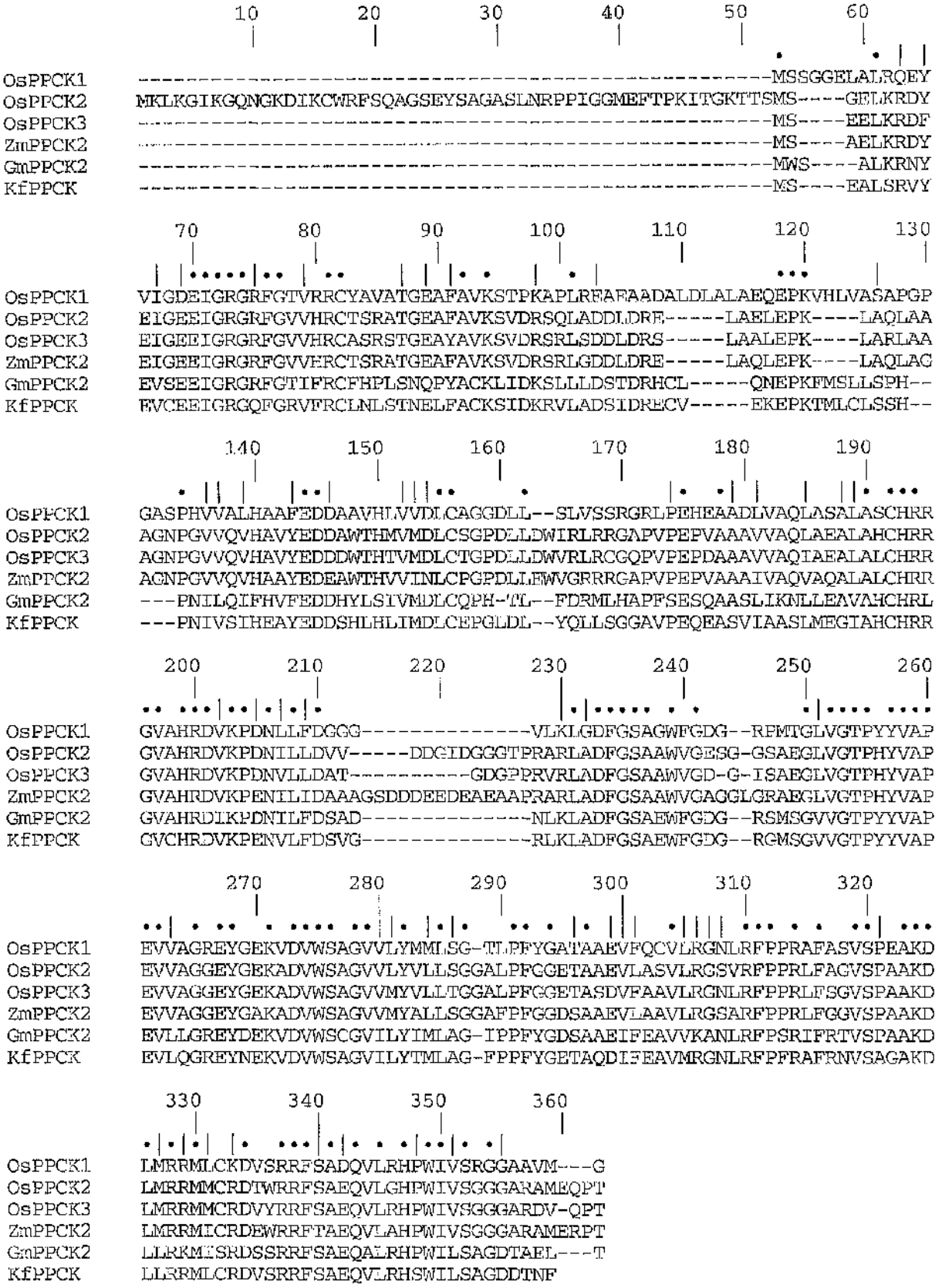

10 20 30 40 50 60
OsPPCK1 ----------MSSGGELALRQEY
OsPPCK2 MKLKGIKGQNGKDIKCWRFSQAGSEYSAGASLNRPPIGGMEFTPKITGKTTSMS----GELKRDY
OsPPCK3 ----------MS----EELKRDF
ZmPPCK2 ----------MS----AELKRDY
GmPPCK2 ----------MWS----ALKRNY
KfPPCK ----------MS----EALSRVY
70 80 90 100 110 120 130
OsPPCK1 VIGDEIGRGRFGTVRRCYAVATGEAFAVKSTPKAPLRFAFAADALDLALAEQEPKVHLVASAPGP
OsPPCK2 EIGEEIGRGRFGVVHRCTSRATGEAFAVKSVDRSQLADDLDRF-----LAELEPK----LAQLAA
OsPPCK3 EIGEEIGRGRFGVVHRCASRSTGEAYAVKSVDRSRLSDDLDRS-----LAALEPK----LARLAA
ZmPPCK2 EIGEEIGRGRFGVVERCTSRATGEAFAVKSVDRSRLGDDLDRE-----LAQLEPK----LAQLAG
GmPPCK2 EVSEEIGRGRFGTIFRCFHPLSNQPYACKLIDKSLLLDSTDRHCL-----QNEPKFMSLLSPH--
KfPPCK EVCEEIGRGQFGRVFRCLNLSTNELFACKSIDKRVLADSIDRECV-----EKEPKTMLCLSSH--
140 150 160 170 180 190
OsPPCK1 GASPHVVALHAAFEDDAAVHLVVDLCAGGDLL--SLVSSRGRLPEHEAADLVAQLASALASCHRR
OsPPCK2 AGNPGVVQVHAVYEDDAWTHMVMDLCSGPDLLDWIRLRRGAPVPEPVAAAVVAQLAEALAHCHRR
OsPPCK3 AGNPGVVQVHAVYEDDDWTHTVMDLCTGPDLLDWVRLRCGQPVPEPDAAAVVAQIAEALALCHRR
ZmPPCK2 AGNPGVVQVHAAYEDEAWTHVVINLCPGPDLLEWVGRRRGAPVPEPVAAAIVAQVAQALALCHRR
GmPPCK2 ---PNILQIFHVFEDDHYLSIVMDLCQPH-TL--FDRMLHAPFSESQAASLIKNLLEAVAHCHRL
KfPPCK ---PNIVSIHEAYEDDSHLHLIMDLCEPGLDL--YQLLSGGAVPEQEASVIAASLMEGIAHCHRR
200 210 220 230 240 250 260
OsPPCK1 GVAHRDVKPDNLLFDGGG----------VLKLGDFGSAGWFGDG--RPMTGLVGTPYYVAP
OsPPCK2 GVAHRDVKPDNILLDVV-----DDGIDGGGTPRARLADFGSAAWVGESG-GSAEGLVGTPHYVAP
OsPPCK3 GVAHRDVKPDNVLLDAT----------GDGPPRVRLADFGSAAWVGD-G-ISAEGLVGTPHYVAP
ZmPPCK2 GVAHRDVKPENILIDAAAGSDDDEEDEAEAAPRARLADFGSAAWVGAGGLGRAEGLVGTPHYVAP
GmPPCK2 GVAHRDIKPDNILFDSAD----------NLKLADFGSAEWFGDG--RSMSGVVGTPYYVAP
KfPPCK GVCHRDVKPENVLFDSVG----------RLKLADFGSAEWFGDG--RGMSGVVGTPYYVAP
270 280 290 300 310 320
OsPPCK1 EVVAGREYGEKVDVWSAGVVLYMMLSG-TLPFYGATAAEVFQCVLRGNLRFPPRAFASVSPEAKD
OsPPCK2 EVVAGGEYGEKADVWSAGVVLYVLLSGGALPFGGETAAEVLASVLRGSVRFPPRLFAGVSPAAKD
OsPPCK3 EVVAGGEYGEKADVWSAGVVMYVLLTGGALPFGGETASDVFAAVLRGNLRFPPRLFSGVSPAAKD
ZmPPCK2 EVVAGGEYGAKADVWSAGVVMYALLSGGAFPFGGDSAAEVLAAVLRGSARFPPRLFGGVSPAAKD
GmPPCK2 EVLLGREYDEKVDVWSCGVILYIMLAG-IPPFYGDSAAEIFEAVVKANLRFPSRIFRTVSPAAKD
KfPPCK EVLQGREYNEKVDVWSAGVILYTMLAG-FPPFYGETAQDIFEAVMRGNLRFPFRAFRNVSAGAKD
330 340 350 360
OsPPCK1 LMRRMLCKDVSRRFSADQVLRHPWIVSRGGAAVM---G
OsPPCK2 LMRRMMCRDTWRRFSAEQVLGHPWIVSGGGARAMEQPT
OsPPCK3 LMRRMMCRDVYRRFSAEQVLRHPWIVSGGGARDV-QPT
ZmPPCK2 LMRRMICRDEWRRFTAEQVLAHPWIVSGGGARAMERPT
GmPPCK2 LLRKMISRDSSRRFSAEQALRHPWILSAGDTAEL---T
KfPPCK LLRRMLCRDVSRRFSAEQVLRHSWILSAGDDTNF

Figure 5.9 *In vitro* transcription of rice *PPCK* clones

Formaldehyde-MOPS gel showing the products from in vitro transcript reactions. Lanes are: 1, OsPPCK1 cDNA; 2, OsPPCK2 cDNA; 3, OsPPCK2 genomic clone; 4, OsPPCK3 cDNA and 5, OsPPCK2 long cDNA.

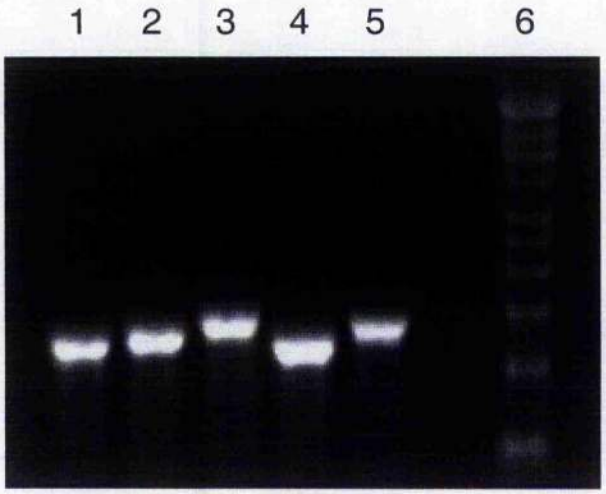

1
2
3
4
5
6

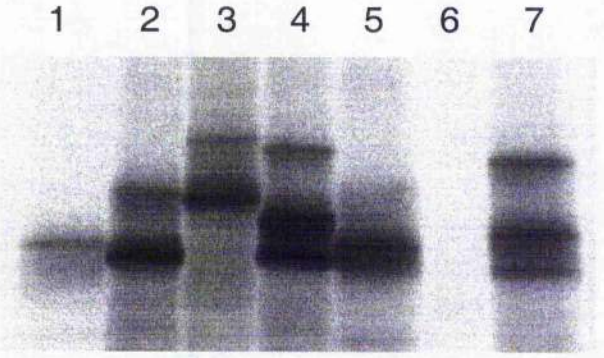
Figure 5.10 Functional analysis of full-length *OsPPCK* clones

(A) Phosphor images of [^{35}S] Met-labelled products from *in vitro* translation of RNA samples separated on 12.5% SDS-PAGE gel.

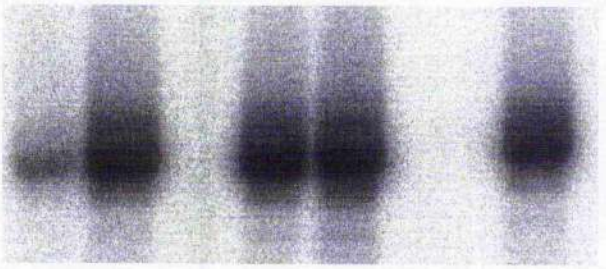
(B) Phosphor images of immunoprecipitated, ^{32}P-labelled PEPc from assays of the PEPc kinase activity of translation products separated on an 8% SDS-PAGE gel.

Lanes are: 1, OsPPCK1 cDNA; 2, OsPPCK2 cDNA; 3, OsPPCK2 genomic clone; 4, OsPPCK2 long cDNA; 5, OsPPCK3 cDNA; 6, no RNA control and 7, OsPPCK2 long cDNA (+KCl).

A



B



Chapter 6

General Discussion

This work has broadly answered the questions posed at the outset. The relation between PEPcs and PEPc kinases in soybean is considerably more complex than first imagined. There are at least 4 *PPCK* genes, 2 of which are nodule-enhanced and likely regulated by photosynthate, but these are also expressed in other tissues. Another *PPCK* gene shows the very unusual property of tissue specific circadian expression. The implications of these findings have been discussed in the relevant chapters. In this general discussion I use phylogenetic analysis of *PPCK* and *PEPC* genes to throw light on the functions of different gene products and identify possible areas of future research.

Since the first cloning of a *PPCK* cDNA (Hartwell *et al.*, 1999a) many more sequences have now been identified from a wide range of plant species. In a recent review, Nimmo (2003) lists 14 full-length PEPc kinase sequences identified from 10 different plant species. The work presented in this thesis includes the identification of four *PPCK* genes from soybean and three *PPCK* genes from the genome sequence of rice. Together with sequences identified from genomic and EST sequence databases (TIGR and NCBI) by similarity to known *PPCK* genes, sequences published in the literature and unpublished work from our laboratory, a dataset consisting of over 30 sequences from 21 species has been constructed. In order to gain insight into the evolution of PEPc kinase and to examine how the various isoforms in soybean relate to each other, a phylogenetic analysis was performed and the resulting tree is shown in Fig. 6.1. The deduced amino acid sequences used in the analysis are either (putative) full-length or close to being putative full-length cDNA sequences (denoted by an asterisk).

In the tree shown in Fig 6.1 several plant CDPK sequences were designated as the outgroup. The only PPCK sequence from a gymnosperm, *Pinus taeda*, falls on a separate branch to all the angiosperm PPCK sequences. The angiosperms sequences fall into two main groups: the eudicots and the non-eudicots (following the naming of Savolainen *et al.*, 2000). The non-eudicot group consists mainly of members of the grass family (Poaceae), together with *Allium cepa* and *Amborella trichopoda*. The placement of *A. trichopoda* at the base of the non-eudicot clade is in agreement with the results of Savolainen *et al.* (2000). Qiu *et al.* (1999) recently identified *Amborella*

as the root of the angiosperms in a combined analysis of genes from the mitochondrial, plastid and nuclear genomes.

The groupings seen in the Poaceae do not reflect taxonomic relationships but rather the groupings of different PPCK isoforms. Work in our lab has identified four maize (*Zea mays*) *PPCK* genes (Shenton *et al.*, unpublished). The *Z. mays* PPCK1 sequence is more similar to a PPCK sequence from rice (*O. sativa* PPCK1) and a sequence from wheat (*T. aestivum* PPCK1) than to the three other maize PPCK sequences. It is tempting to speculate that these groupings may reflect the functions of particular isoforms. Knowledge of the relationships of the different PPCK isoforms could help direct future research. For example, unpublished results from our lab show that the *Z. mays* PPCK1 gene is the light-responsive C_4 *PPCK* isoform (Shenton *et al.*, unpublished). Therefore, it would be interesting to investigate whether the *PPCK1* isoforms of *O. sativa* and *T. aestivum*, both C_3 plants, are also light responsive. However a detailed analysis of the *PPCK* genes in these species has yet to be published.

Within the eudicot clade relationships consisting of groups of taxonomically related species as well as groups of more distantly related species are seen. The phylogenetic tree shows GmPPCK1-3 group with the PPCK sequences from the two other legume species, *L. japonicus* and *M. truncatula*. The Lotus *PPCK* gene has been shown to have nodule-enhanced expression (Nakagawa *et al.*, 2003a). The expression pattern of the Medicago *PPCK* gene has yet to be investigated, but two corresponding ESTs were isolated from nodule libraries demonstrating that it is expressed in nodules. It may be that this group represents the nodule-enhanced *PPCK* isoforms for these plants. The isolation and characterisation of other *PPCK* isoforms from Lotus and Medicago would be required to confirm this and would greatly aid the understanding of *PPCK* evolution in the legumes.

The placement of GmPPCK4 away from the other legume sequences is surprising. GmPPCK4 is more closely related to the PPCK2 isoforms from members of the Solanaceae (*N. benthamiana*, *S. tuberosum* and *L. esculentum*) than it is to other soybean PPCKs. The two kinase isoforms identified in the solanaceous plants are also separate from each other in the tree topology. These relationships suggest that a gene duplication event occurred early during the evolution of the flowering plants. However, Arabidopsis contains only two closely related PPCK genes, implying either that the above relationships are spurious or that gene loss has occurred. The addition

of further sequences, in particular sequences of different PPCK isoforms from the same plant, will help to address these questions.

In contrast to PEPc kinase, PEPc sequences have been used in phylogenetic analyses due to their presence in both prokaryotic microorganisms and plants (Gehrig *et al.*, 1998; Gehrig *et al.*, 2001). Gehrig *et al.* (2001) produced the most extensive analysis to date using 143 partial amino acid sequences of PEPc from bacteria, vascular and non-vascular plants. The PEPc sequence from *E. coli* was used as the outgroup for the analysis and the authors suggest that an ancestor of the plant PEPcs arose from the γ proteobacterial lineage. The existence of the plant 'bacterial-type' PEPc genes were unknown at this time. Sanchez and Cejudo (2003) first identified these novel PEPc isoforms from the genome sequences of Arabidopsis and rice. They produced an unrooted phylogenetic tree based on the analysis of 22 bacterial and 21 plant full-length amino acid sequences (Sanchez and Cejudo, 2003). The tree displayed two major clades, one consisting of only bacterial PEPcs, the other containing all the plant PEPcs and the PEPcs from the γ proteobacteria. Although all the plant PEPc sequences formed a single group the two bacterial-type PEPc sequences from Arabidopsis (Atppc4) and rice (Osppc-b) diverged from the rest of the plant PEPcs very early during the evolution from a common ancestral gene similar to the PEPc genes of the γ proteobacteria.

To examine the placements of the soybean PEPc genes identified in this work and the recently published sequences from the green algae *Chlamydomonas reinhardtii*, a phylogenetic analysis of PEPc amino acid sequences was performed. Fig. 6.2 shows an unrooted radial phylogenetic tree resulting from the analysis of 40 bacterial, 64 plant and 2 algal PEPc sequences. The same two major clades identified in Sanchez and Cejudo (2003) are visible, with the addition of the *C. reinhardtii* sequences in the plant and γ proteobacteria group. The plant PEPc sequences, excluding the bacterial-type isoforms, form a tight cluster demonstrating their relative homogeneity in comparison to the more divergent bacterial sequences. The *C. reinhardtii* PEPc1 sequence is the most similar to the plant PEPc clade, indicating a common ancestor. As expected from the results presented in chapter 3, the soybean GmPEPc17 sequence groups with the bacterial-type Atppc4 and Osppc-b sequences. The *C. reinhardtii* PEPc2 sequence is the most similar sequence to these novel PEPc

isoforms. This indicates that the divergence of these two classes of PEPcs in plants occurred in a common ancestor of green algae and land plants. With the phylogenetic analysis indicating that the occurrence of two distinct groups of PEPc isoforms in plants is evolutionarily related to the two classes of PEPcs identified in algae (Rivoal *et al.*, 1996; Rivoal *et al.*, 1998; Rivoal *et al.*, 2001), it is tempting to speculate that the bacterial-type PEPcs in plants may also form large heteromeric complexes. The results of Blonde and Plaxton (2003) on castor plants (*Ricinus communis*) support this hypothesis. The results from the phylogenetic analysis suggest that antibodies raised against algal Class 2 PEPcs (Rivoal *et al.*, 1996) may also recognise the plant bacterial-type PEPcs, facilitating identification of these polypeptides in plants. The identification of further members of this group and their characterisation should be prove to be an exciting area of future research.

To examine the relationships within the plant PEPc clade a phylogenetic analysis including only one bacterial PEPc sequence, *E. coli*, was performed (Fig 6.3). As seen in the radial tree (Fig. 6.2) the plant bacterial-type PEPc sequences form a group with *C. reinhardtii* PEPc2, and the rest of the plant PEPcs form a single group with *C. reinhardtii* PEPc1 as the sister group. As seen in the PPCK tree (Fig. 6.1), the plant sequences fall into eudicot and non-eudicot groupings, with the exception of certain sequences. The two gymnosperm sequences, *Welwitschia mirablis* and *Pinus abies*, are placed within the angiosperm non-eudicot clade. Furthermore, two eudicot sequences, *M. crystallinum* PEPc1 and *Amaranthus hypochondriacus*, are also embedded within the non-eudicot group. *M. crystallinum* PEPc1 is the CAM photosynthetic PEPc isoform and the *A. hypochondiacus* sequence is the C_4 photosynthetic isoform. The placement of these sequences may be a reflection of their function; they are placed between a group of sequences encoding the C_4 photosynthetic PEPc isoforms from several species of grasses (*Sorghum* PEPc3, *Saccharum spontaneum, Z. mays* PEPc1, *Chloris gayana* and *Seteria italica*), and the rest of the plant PEPc sequences. There are several possible explanations for the unusual positions of *M. crystallinum* PEPc1 and *A. hypochondiacus*. Firstly, it could be an artefact of the tree due to reasons such as uneven sampling. However, the bootstrap support for the position of *A. hypochondiacus* is high at 99%, and reasonable for *M. crystallinum* PEPc1 at 59%. If this topology is accepted then one explanation is that the ancestor of monocots and dicots contained two PEPc genes, one of which has been lost in the dicotyledonous plants that do not fix CO_2

photosynthetically with PEPc. An alterative hypothesis would be that it is the result of convergent evolution to provide PEPc with the properties needed for CO_2 fixation in photosynthesis. However, it is clear that this has not occurred in Flaveria, where the C_4 photosynthetic PEPc isoforms are closely related to the non-photosynthetic isoforms.

In the eudicot clade taxonomic grouping are observed, such as the group of sequences from members of the Fabaceae (legume family) and the group containing sequences from members of the Solanaceae. However, subgroups comprising different PEPc isoforms are also seen; the *L. japonicus* PEPc1 and PEPc2 sequences fall on different branches within the group of sequences from leguminous plants and a similar pattern is seen in the group of solanaceous plant sequences. The *G. max* PEPc4 sequence does not group with the other legume PEPc sequences but instead forms a group with a PEPc sequence from cotton (*G. hirsutum* PEPc) and one of the Arabidopsis PEPc sequences (*A. thaliana* PEPc2). The placement of GmPEPC4 is reminiscent of the placement of GmPPCK4 in the PPCK tree (Fig. 6.1). The significance of this observation is unknown, however, as seen in chapters 3 and 4 there are no obvious relationships between specific isoforms of PEPc and PEPc kinase.

The phylogenetic trees produced from the analysis of PEPc and PEPc kinase sequences both show recognisable taxonomic groupings, suggesting that, even though the sampling distribution of different plant families is very uneven, a signal reflecting evolutionary relationships is detected. However, the aim of these analyses was not to determine the usefulness of these genes for evolutionary reconstruction but to examine the relationships between the various isoforms and identify any functional groups of sequences, in particular the novel PEPc kinase sequences identified in this work. For PPCK genes is has not been possible to assign function to a particular gene based upon its position within a phylogenetic tree or its similarity to other PPCK genes. Some isoform-specific groups are seen in the trees, the best example being the relationships seen in the grass family, but presently not enough information on the expression patterns or functions of these genes is known to determine whether these isoforms are performing similar roles in the different plant species. The production of phylogenetic trees has proved a useful tool in the identification of new sequences, particularly in the case of PEPc kinase where short EST sequences can have significant similarity to both PEPc kinase and CDPKs. They are also useful in

developing new hypotheses and highlighting potentially interesting areas of future research.

Similarly, examination of the expression patterns of *PPCK* genes in C_3 plants, at least at the organ/tissue level, has not allowed easy allocation of specific isoforms to specific tasks. In soybean *GmPPCK2* and *GmPPCK3* are preferentially expressed in nodules, while in tomato *LePPCK2* is most heavily expressed in ripe fruit. However, transcripts of these genes are also clearly detectable, at various levels, in other tissues elsewhere in the plant. *GmPPCK4* is also expressed in nodules, but its expression is not controlled by photosynthate supply from the shoots and therefore it is not thought to be responsible for the control of the phosphorylation status of the nodule-enhanced PEPc. The circadian control of *GmPPCK4* in leaves but not roots suggests it may be fulfilling different functions in these tissues. Clearly a more detailed examination of expression patterns, via *in situ* hybridisation or the use of transgenic plants expressing promoter-reporter constructs, may reveal patterns in the types of cells or tissues which different isoforms are expressed in and therefore provide a better understanding of their roles.

While further characterisation of *PPCK* genes at the transcriptional level is desirable, especially in the dissection of the signalling pathways controlling *PPCK* transcript abundance, the lack of information at the protein level, particularly in C_3 plants, is a major barrier to understanding PPCK *in vivo*. Work by Hartwell *et al.* (1996,1999a) demonstrated the close relationship between *PPCK* transcript levels, PEPc kinase activity and PEPc phosphorylation status in the leaves of *K. fedtschenkoi*. The low abundance of PEPc kinase, and to a lesser extent PEPc, protein in C_3 plants make similar experiments technically difficult. However, the ability to relate differences in the spatial and/or temporal abundance of transcripts to changes at the protein and kinase activity levels is required. For example, in the case of *GmPPCK4* it is vital to know whether the circadian fluctuations in transcript abundance result in changes at the protein and activity levels.

While the results presented in this work do not support the concept of *PEPC-PPCK* expression partners as suggest in Xu *et al.* (2003) and Echevarria and Vidal (2003), the question of whether a specific PEPc kinase is equally effective at phosphorylating all PEPcs from the same plant has yet to be addressed. The recombinant production of all PPCKs and PEPcs from a single plant species would be required to allow pairwise comparisons. Recent work from our laboratory has

demonstrated the possibilty of producing soluble and active recombinant PEPc kinase using a *Pichia methanolica* expression system, suggesting that such a set of experiments may soon be possible.

Further work arising from the unusual regulation of *GmPPCK4* would be to identify components of the signalling pathway which link the circadian clock to *GmPPCK4* expression in a tissue dependent manner. One method would be to create a *GmPPCK4* promoter: luciferase expressing transgenic Arabidopsis line allowing the characterisation of GmPPCK4 expression via automatic imaging. A mutagenised population of transgenic plants could then be screened for loss of and/or changes in the circadian expression of luciferase. Such a strategy would require the GmPPCK4 promoter to drive luciferase expression in Arabidopsis in a tissue-dependent circadian manner as it does in soybean, and that luciferase is expressed at detectable levels in the leaves. Other experiments could investigate regulatory regions of the GmPPCK4 promoter using gel shift assays. Using nuclear extracts from leaf and root cells the possibility that specific DNA binding factors are found in leaves but not roots could be investigated. The abundance of these putative factors in leaves would be expected to vary over a 24 h period.

One potential applied aspect of this work stems from the identification of a bacterial-type PEPc gene expressed in soybean. PEPc has been proposed to play important roles in seed development. In seeds of the leguminous plant *Vicia faba,* PEPc is thought to be involved in the synthesis of storage proteins by redirecting carbon flux towards amino acid synthesis (Golombek *et al.*, 1999). In the oilseed plant *R. communis*, different PEPc isoforms are proposed to play roles in both storage protein and lipid accumulation (Blonde and Plaxton, 2003). Recently, Rolletschek *et al.* (2004) generated transgenic bean plants (*Vicia narbonensis*) expressing a PEPc gene from the bacterium *Corynebacterium glutamicum*. The *C. glutamicum* PEPc was under the control of the legumin B4 promoter to confer seed-specific expression. Analysis of seeds from the transgenic plants showed that, compared to wild type seeds, they contained 15 – 25% more protein on a per gram basis and individual dry weight increased by 20 – 30%. However, there was a decrease in seed number per pod by 20 – 30% leading to no significant change in total seed weight per plant.

Soybean is a major source of protein and vegetable oil for both animal and human nutrition. Soybean seeds accumulate high levels of both protein and oil, with a typical seed containing 40% protein and 20% oil by weight (Stacey *et al.*, 2004).

Presently there is no evidence for a role of the soybean bacterial-type PEPc gene, GmPEPC17, in seed development. An investigation into the expression of *GmPEPC17* during seed development and maturation would answer whether this gene is expressed in seeds and whether expression coincides with any particular developmental stages. However, even if the plant bacterial-type PEPcs are not involved in seed development they could prove a useful resource for plant biotechnology. Biochemical characterisation of these novel plant PEPcs has yet to be published, but it would appear likely that they will be regulated differently to the previously identified plant PEPcs considering the lack of the conserved phosphorylatable N-terminal sequence. The work from bacterial and algal PEPcs would suggest that the plant bacterial-type PEPc proteins will also have very different allosteric regulation. Directed expression of these proteins in seeds may allow similar results to those described by Rolletschek *et al.* (2004) to be achieved using the plant's own genetic material.

In another recent paper, Chen *et al.* (2004) introduced the PEPc gene from the cyanobacterium *Synechococcus vulcans* into Arabidopsis driven by the constitutive cauliflower mosaic virus 35S promoter. Visible phenotypes observed in the transgenic plants included a block in true leaf development and partially bleached leaves. These phenotypes could be rescued by growth on media supplemented with aromatic amino acids. Analysis of the transgenic plants revealed significantly decreased levels of aromatic amino acids, compared to wild-type plants, and an increase in the levels of amino acids of the aspartic and glutamic families, such as asparagine, arganine and glutamine. These results suggest that as a result of increased flux through PEPc there was a reduced amount of PEP available for synthesis of aromatic amino acids via the shikimate pathway.

The work of Rolletschek *et al.* (2004) and Chen *et al.* (2004) demonstrate the usefulness of a PEPc protein with reduced sensitivity to metabolic feedback inhibition in engineering plant metabolism. They achieved their results using by introducing genes of bacterial origin into plants. Public concern over the introduction of foreign DNA into food crops forms the basis of some of the objections to transgenic crops, hindering the development of transgenic products for human consumption (Rommens, 2004). It is therefore becoming advantageous to produce transgenic products containing only native DNA. Nielsen (2003) proposed that transgenic organisms should be classified on the basis of the genetic distance between the target organism

and the source of the new trait. Plants containing genetic material from within the same sexual compatibility group would be termed intragenic. Rommens *et al.* (2004) recently reported the first example of genetically engineered plants containing only native DNA. The identification of bacterial-type plant PEPc genes may be a useful tool in the engineering of plant metabolism without the incorporation of foreign DNA.

Figure 6.1 Phylogenetic analysis of plant PPCK genes based on deduced amino acid sequences.

Branch numbers indicate bootstrap support (1000 replicates). Non full-length sequences are indicated by an asterisk.

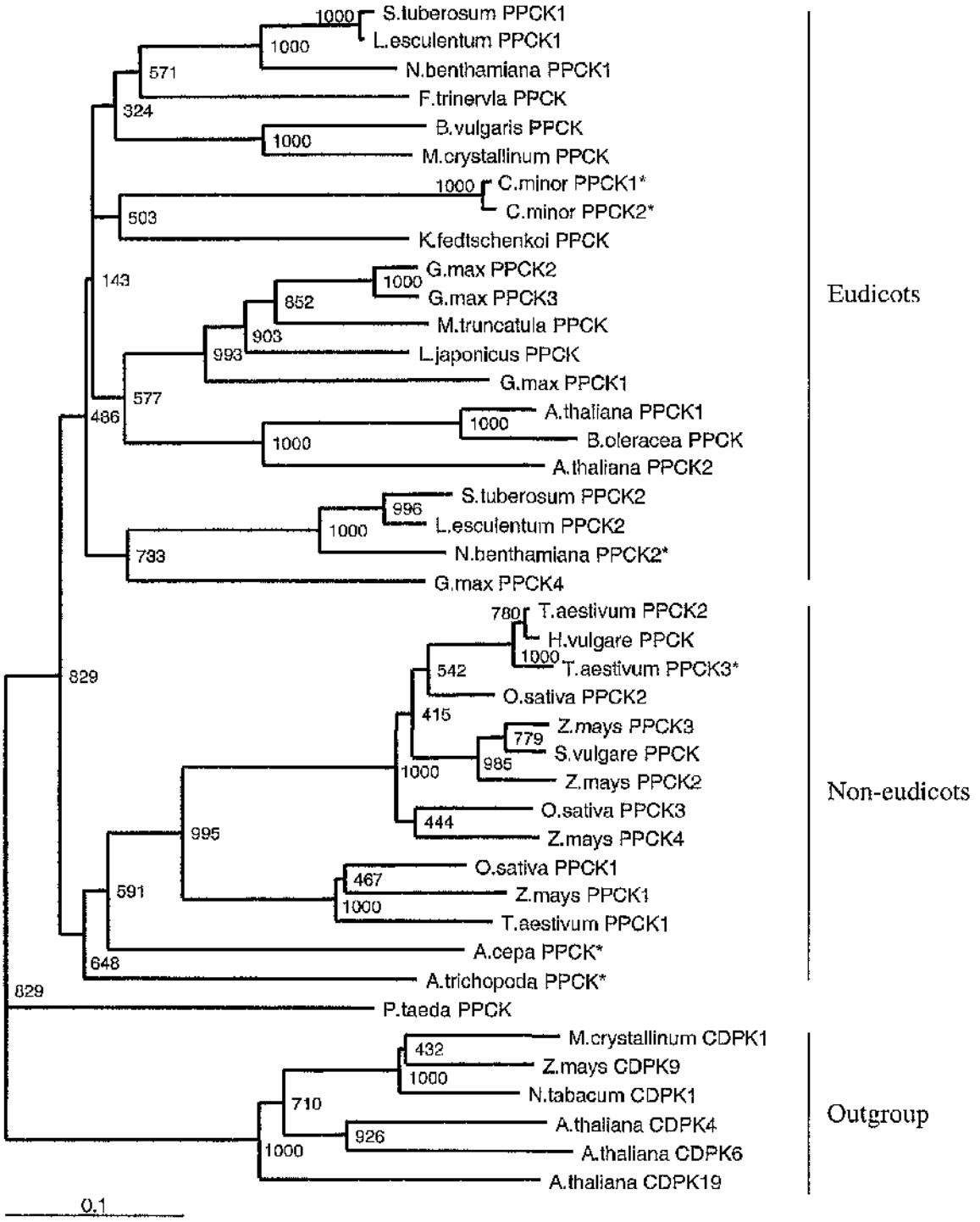

S.tuberosum PPCK1
L.esculentum PPCK1
N.benthamiana PPCK1
F.trinervia PPCK
B.vulgaris PPCK
M.crystallinum PPCK
C.minor PPCK1*
C.minor PPCK2*
K.fedtschenkoi PPCK
G.max PPCK2
G.max PPCK3
M.truncatula PPCK
L.japonicus PPCK
G.max PPCK1
A.thaliana PPCK1
B.oleracea PPCK
A.thaliana PPCK2
S.tuberosum PPCK2
L.esculentum PPCK2
N.benthamiana PPCK2*
G.max PPCK4
T.aestivum PPCK2
H.vulgare PPCK
T.aestivum PPCK3*
O.sativa PPCK2
Z.mays PPCK3
S.vulgare PPCK
Z.mays PPCK2
O.sativa PPCK3
Z.mays PPCK4
O.sativa PPCK1
Z.mays PPCK1
T.aestivum PPCK1
A.cepa PPCK*
A.trichopoda PPCK*
P.taeda PPCK
M.crystallinum CDPK1
Z.mays CDPK9
N.tabacum CDPK1
A.thaliana CDPK4
A.thaliana CDPK6
A.thaliana CDPK19
0.1
Eudicots
Non-eudicots
Outgroup

Figure 6.2 Phylogenetic analysis of PEPC genes based on deduced amino acid sequences.

Names of the plant PEPc sequences were omitted for clarity.

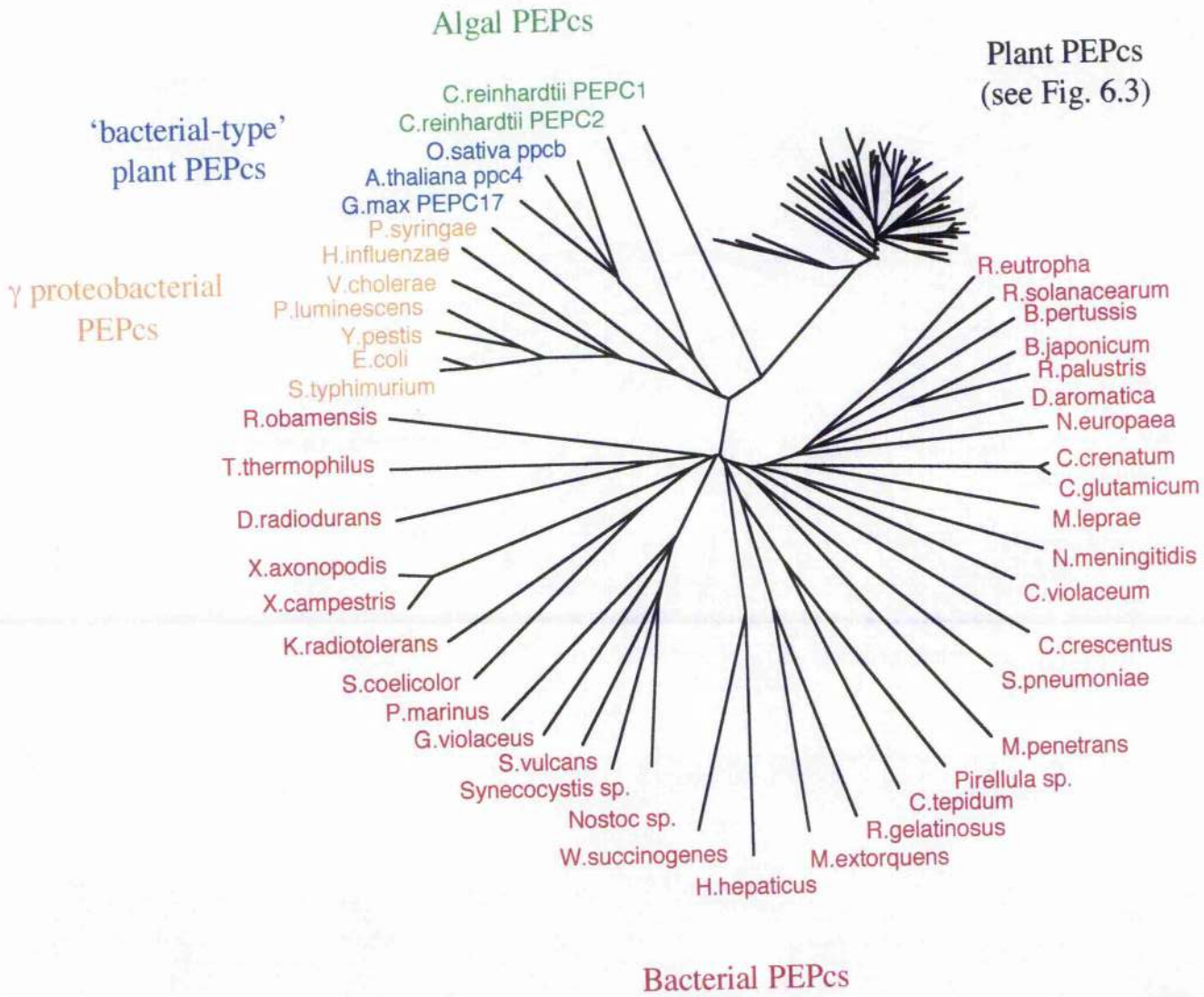
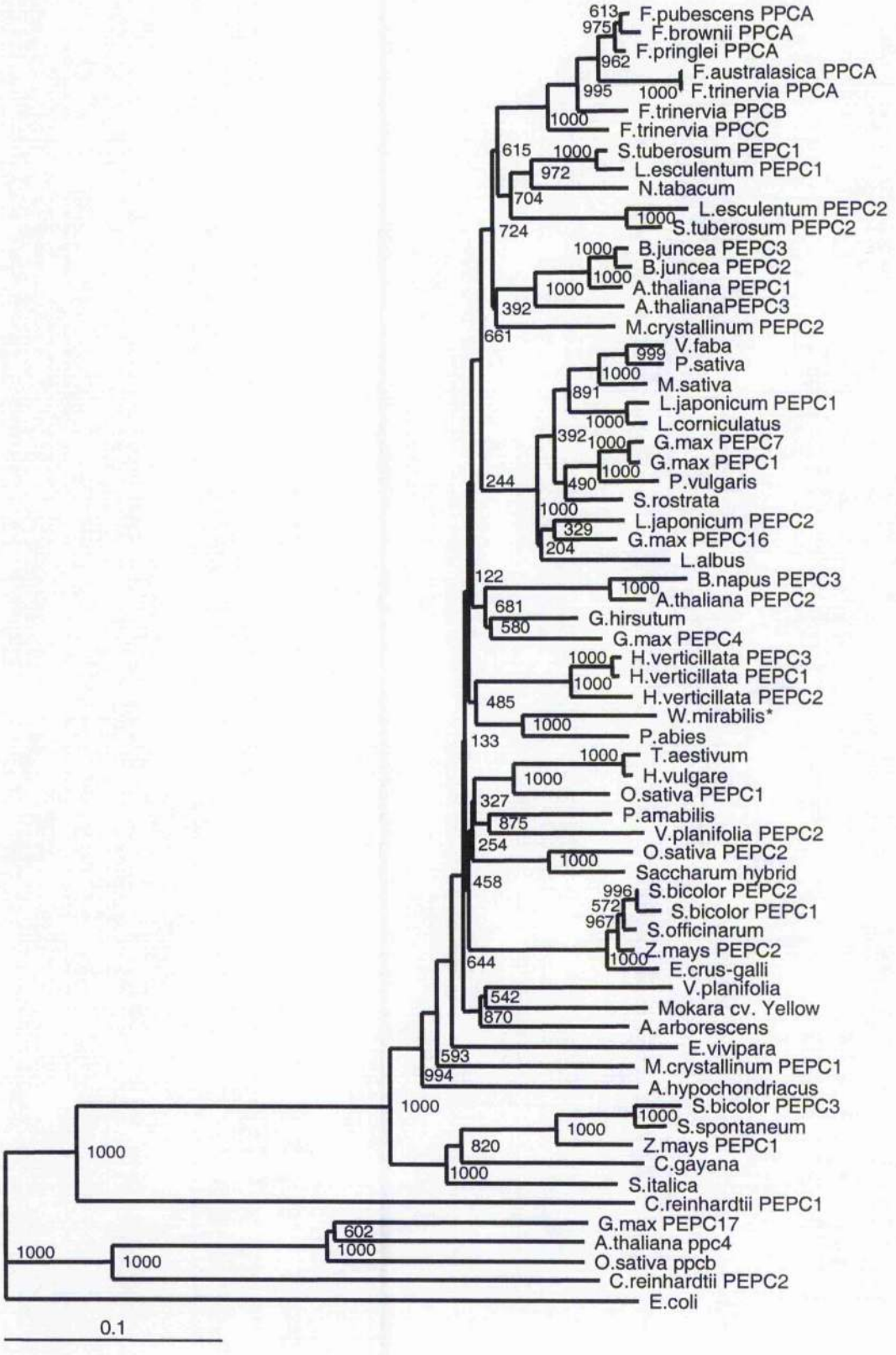

Algal PEPcs
Plant PEPcs
(see Fig. 6.3)
'bacterial-type'
plant PEPcs
γ proteobacterial
PEPcs
C.reinhardtii PEPC1
C.reinhardtii PEPC2
O.sativa ppcb
A.thaliana ppc4
G.max PEPC17
P.syringae
H.influenzae
V.cholerae
P.luminescens
Y.pestis
E.coli
S.typhimurium
R.obamensis
T.thermophilus
D.radiodurans
X.axonopodis
X.campestris
K.radiotolerans
S.coelicolor
P.marinus
G.violaceus
S.vulcans
Synecocystis sp.
Nostoc sp.
W.succinogenes
H.hepaticus
M.extorquens
R.gelatinosus
C.tepidum
Pirellula sp.
M.penetrans
S.pneumoniae
C.crescentus
C.violaceum
N.meningitidis
M.leprae
C.glutamicum
C.crenatum
N.europaea
D.aromatica
R.palustris
B.japonicum
B.pertussis
R.solanacearum
R.eutropha
Bacterial PEPcs

Figure 6.3 Phylogenetic analysis of plant PEPC genes based on deduced amino acid sequences.

Branch numbers indicate bootstrap support (1000 replicates). Non full-length sequences are indicated by an asterisk. *E. coli* was defined as the outgroup.


F.pubescens PPCA
F.brownii PPCA
F.pringlei PPCA
F.australasica PPCA
F.trinervia PPCA
F.trinervia PPCB
F.trinervia PPCC
S.tuberosum PEPC1
L.esculentum PEPC1
N.tabacum
L.esculentum PEPC2
S.tuberosum PEPC2
B.juncea PEPC3
B.juncea PEPC2
A.thaliana PEPC1
A.thalianaPEPC3
M.crystallinum PEPC2
V.faba
P.sativa
M.sativa
L.japonicum PEPC1
L.corniculatus
G.max PEPC7
G.max PEPC1
P.vulgaris
S.rostrata
L.japonicum PEPC2
G.max PEPC16
L.albus
B.napus PEPC3
A.thaliana PEPC2
G.hirsutum
G.max PEPC4
H.verticillata PEPC3
H.verticillata PEPC1
H.verticillata PEPC2
W.mirabilis*
P.abies
T.aestivum
H.vulgare
O.sativa PEPC1
P.amabilis
V.planifolia PEPC2
O.sativa PEPC2
Saccharum hybrid
S.bicolor PEPC2
S.bicolor PEPC1
S.officinarum
Z.mays PEPC2
E.crus-galli
V.planifolia
Mokara cv. Yellow
A.arborescens
E.vivipara
M.crystallinum PEPC1
A.hypochondriacus
S.bicolor PEPC3
S.spontaneum
Z.mays PEPC1
C.gayana
S.italica
C.reinhardtii PEPC1
G.max PEPC17
A.thaliana ppc4
O.sativa ppcb
C.reinhardtii PEPC2
E.coli
0.1

References

Akhtar MS, Goldschmidt EE, John I, Rodoni S, Matile P and Grierson D (1999) Altered patterns of senescence and ripening in *gf*, a stay-green mutant of tomato (*Lycopersicon esculentum* Mill.). *Journal of Experimental Botany* **50**: 1115-1122

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P and Kay SA (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* **293**: 880-883

Alvarez R, Garcia-Maurino S, Feria AB, Vidal J and Echevarria C (2003) A conserved 19-amino acid synthetic peptide from the carboxy terminus of phosphoenolpyruvate carboxylase inhibits the *in vitro* phosphorylation of the enzyme by the calcium-independent phosphoenolpyruvate carboxylase kinase. *Plant Physiology* **132**: 1097-106

Ambasht PK and Kayastha AM (2002) Plant pyruvate kinase. *Biologia Plantarum* **45**: 1-10

Andreo CS, Gonzalez DH and Iglesias AA (1987) Higher plant phosphoenolpyruvate carboxylase. *Febs Letters* **213**: 1-8

Appleby CA (1984) Leghemoglobin and *Rhizobium* respiration. *Annual Review of Plant Physiology* **35**: 443-478

Atkins CA, Fernando M, Hunt S and Layzell DB (1992) A metabolic connection between nitrogenase activity and the synthesis of ureides in nodulated soybean. *Plant Physiology* **84**: 441-447

Bakrim N, Brulfert J, Vidal J and Chollet R (2001) Phosphoenolpyruvate carboxylase kinase is controlled by a similar signaling cascade in CAM and C_4 plants. *Biochemical and Biophysical ResearchCommunications* **286**: 1158-62

Bergersen FJ and Turner GL (1975) Leghemoglobin and the supply of O_2 to nitrogen fixing root nodule bacteroids: presence of two oxidase systems and ATP production at low free O_2 concentrations. *Journal of Genaral Microbiology* **91**: 345-354

Blonde JD and Plaxton WC (2003) Structural and kinetic properties of high and low molecular mass phosphoenolpyruvate carboxylase isoforms from the endosperm of developing castor oilseeds. *Journal Of Biological Chemistry* **278**: 11867-73

Borland AM and Griffiths H (1997) A comparative study on the regulation of C_3 and C_4 carboxylation processes in the constitutive Crassulacean acid metabolism (CAM) plant *Kalanchoe daigremontiana* and the C_3-CAM intermediate *Clusia minor*. *Planta* **201**: 368-378

Borland AM, Hartwell J, Jenkins GI, Wilkins MB and Nimmo HG (1999) Metabolite Control Overrides Circadian Regulation of Phosphoenolpyruvate Carboxylase Kinase and CO_2 Fixation in Crassulacean Acid Metabolism. *Plant Physiology* **121**: 889-896

Bortolotti S, Boggio SB, Delgado L, Orellano EG and Valle EM (2003) Different induction patterns of glutamate metabolizing enzymes in ripening fruits of the tomato mutant *green flesh*. *Physiologia Plantarum* **119**: 384-391

Brewin NJ (1991) Development of the legume root nodule. *Annual Review of Cell Biology* **7**: 191-226

Budde RJA and Chollet R (1986) *In vitro* phosphorylation of maize leaf phosphoenolpyruvate carboxylase. *Plant Physiology* **82**: 1107-1114

Carter PJ, Nimmo HG, Fewson CA and Wilkins MB (1990) *Bryophyllum fedtschenkoi* protein phosphatase type 2A can dephosphorylate phosphoenolpyruvate carboxylase. *Febs Letters* **263**: 233-236

Carter PJ, Nimmo HG, Fewson CA and Wilkins MB (1991) Circadian rhythms in the activity of a plant protein kinase. *EMBO Journal* **10**: 2063-8

Cavalar M, Moller C, Offermann S, Krohn NM, Grasser KD and Peterhansel C (2003) The interaction of DOF transcription factors with nucleosomes depends on the positioning of the binding site and is facilitated by maize HMGB5. *Biochemistry* **42**: 2149-2157

Chattopadhyay S, Ang LH, Puente P, Deng XW and Wei N (1998) Arabidopsis bZIP protein HY5 directly interacts with light- responsive promoters in mediating light control of gene expression. *Plant Cell* **10**: 673-683

Chen L-M, Li K-Z, Miwa T and Izui K (2004) Overexpression of a cyanobacterial phosphoenolpyruvate carboxylase with diminished sensitivity to feedback inhibition in Arabidopsis changes amino acid metabolism. *Planta* **219**: 440-449

Chollet R, Vidal J and O'Leary MH (1996) Phosphoenolpyruvate carboxylase: A ubiquitous, highly regulated enzyme in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 273-298

Coker GT and Schubert KR (1981) Carbon dioxide fixation in soybean roots and nodules.1. Characterization and comparison with N_2 fixation and composition of xylem exudate during early nodule development. *Plant Physiology* **67**: 691-696

Colombo SL, Pairoba CF and Andreo CS (1996) Inhibitory effect of shikimic acid on PEP carboxylase activity. *Plant Cell and Physiology* **37**: 870-872

Cotelle V, Pierre J-N and Vavasseur A (1999) Potential strong regulation of guard cell phosphoenolpyrvate carboxylase through phosphorylation. *Journal of Experimental Botany* **50**: 777-783

Coursol S, Giglioli-Guivarc'h N, Vidal J and Pierre JN (2000) An increase in phosphoinositide-specific phospholipase C activity precedes induction of C_4 phosphoenolpyruvate carboxylase phosphorylation in illuminated and NH_4Cl-treated protoplasts from Digitaria sanguinalis. *Plant Journal* **23**: 497-506

Craig J, Barratt P, Tatge H, Dejardin A, Handley L, Gardner CD, Barber L, Wang T, Hedley C, Martin C and Smith AM (1999) Mutations at the *rug4* locus alter the carbon and nitrogen metabolism of pea plants through an effect on sucrose synthase. *Plant Journal* **17**: 353-362

Curioni PMG, Hartwig JN and Schuller KA (1999) Glycolytic flux is adjusted to nitrogenase activity in nodules of detoppped and argon-treated alfalfa plants. *Plant Physiology* **119**: 445-453

Cushman JC and Bohnert HJ (1992) Salt stress alters A/T-rich DNA-binding factor interactions within the phosphoenolpyruvate carboxylase promoter from *Mesembryanthemum crystallinum*. *Plant Molecular Biology* **20**: 411-424

Cushman JC, Meyer G, Michalowski CB, Schmitt JM and Bohnert HJ (1989) Salt stress leads to differential expression of two isogenes of phosphoenolpyruvate carboxylase during Crassulacean acid metabolism induction in the common ice plant. *The Plant Cell* **1**: 715-725

de Lima M, Oresnik IJ, Fernando SM, Hunt S, Smith R, Turpin DH and Layzell DB (1994) The relationship between nodule adenylates and the regulation of nitrogenase activity by oxygen in soybean. *Physiologia Plantarum* **91**: 687-695

Dean RM, Rivers RL, Zeidel ML and Roberts DM (1999) Purification and functional reconstitution of soybean nodulin 26: An aquaporin with water and glycerol transport properties. *Biochemistry* **38**: 347-353

Denison RF, Hunt S and Layzell DB (1992) Nitrogenase activity, nodule respiration and O2 permeability following detopping of alfalfa and birdsfoot trefoil. *Plant Physiology* **98**: 894-900

Deroche ME and Carrayol E (1988) Nodule phosphoenolpyruvate carboxylase: a review. *Physiologia Plantarum* **74**: 775-782

Dolgikh EA, Fedorova MYu, Filatov AA and Tikhonovich IA (1998) Phosphoenolpyruvate carboxylase from pea root nodules. *Russian Journal Of Plant Physiology* **45**: 342-352

Doncaster HD and Leegood RC (1987) Regulation of phosphoenolpyruvate carboxylase in plant metabolism. *Plant Physiology* **84**: 82-87

Dong L, Ermolova NV and Chollet R (2001) Partial purification and biochemical characterization of a heteromeric protein phosphatase 2A holoenzyme from maize (Zea mays L.) leaves that dephosphorylates C_4 phosphoenolpyruvate carboxylase. *Planta* **213**: 379-89

Doyle JJ (1998) Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in Plant Science* **3**: 473-478

Du Z, Aghoram K and Outlaw WH, Jr. (1997) *In vivo* phosphorylation of phosphoenolpyruvate carboxylase in guard cells of *Vicia faba* L. is enhanced by fusicoccin and suppressed by abscisic acid. *Archives of Biochemistry and Biophysics* **337**: 345-50

Duff SMG, Andreo CS, Pacquit V, Lepiniec L, Sarath G, Condon SA, Vidal J, Gadal P and Chollet R (1995) Kinetic analysis of the non-phosphorylated, *in vitro* phosphorylated, and phosphorylation-site mutant (Asp8) forms of intact recombinant C_4 phosphoenolpyruvate carboxylase from *Sorghum*. *European Journal Of Biochemistry* **228**: 92-95

Duff SMG and Chollet R (1995) *In vivo* regulation of wheat-leaf phosphoenolpyruvate carboxylase by reversible phosphorylation. *Plant Physiology* **107**: 775-782

Duff SMG, Giglioli-Guivarc'h N, Pierre J-N, Vidal J, Condon SA and Chollet R (1996) *In-situ* evidence for the involvement of calcium and bundle-sheath-derived photosynthetic metabolites in the C_4 phosphoenolpyruvate-carboxylase kinase signal-transduction chain. *Planta* **199**: 467-474

Ermolova NV, Ann Cushman M, Taybi T, Condon SA, Cushman JC and Chollet R (2003) Expression, purification, and initial characterization of a recombinant

form of plant PEP-carboxylase kinase from CAM-induced *Mesembryanthemum crystallinum* with enhanced solubility in *Escherichia coli*. *Protein Expression and Purification* **29**: 123-31

Fedorova M, Tikhonovich IA and Vance CP (1999) Expression of C-assimilating enzymes in pea (*Pisum sativum* L.) root nodules. *In situ* localization in effective nodules. *Plant Cell and Environment* **22**: 1249-1262

Fontaine V, Hartwell J, Jenkins GI, Nimmo GA and Nimmo HG (2002) *Arabidopsis thaliana* contains two phosphoenolpyruvate carboxylase kinase genes with different expression patterns. *Plant Cell and Environment* **25**: 115-122

Fukayama H, Hatch MD, Tamai T, Tsuchida H, Sudoh S, Furbank RT and Miyao M (2002) Activity regulation and physiological impacts of the maize C_4-specific phosphoenolpyruvate carboxylase overproduced in transgenic rice plants. *Photosynthesis Research* **77**: 227-239

Gálvez S, Hirsch AM, Wycoff KL, Hunt S, Layzell DB, Kondorosi A and Crespi M (2000) Oxygen regulation of a nodule-located carbonic anhydrase in alfalfa. *Plant Physiology* **124**: 1059-1068

Gehrig H, Heute V and Kluge M (2001) New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. *Molecular Phylogenetics and Evolution* **20**: 262-74

Gehrig HH, Heute V and Kluge M (1998) Toward a better knowledge of the molecular evolution of phosphoenolpyruvate carboxylase by comparison of partial cDNA sequences. *Journal of Molecular Evolution* **46**: 107-14

Giglioli-Guivarc'h N, Pierre JN, Brown S, Chollet R, Vidal J and Gadal P (1996) The Light-Dependent Transduction Pathway Controlling the Regulatory Phosphorylation of C_4 Phosphoenolpyruvate Carboxylase in Protoplasts from *Digitaria sanguinalis*. *Plant Cell* **8**: 573-586

Giovannoni J (2001) Molecular biology of fruit maturation and ripening. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**: 725-749

Golombek S, Heim U, Horstmann C, Wobus U and Weber H (1999) Phosphoenolpyruvate carboxylase in developing seeds of *Vicia faba*: gene expression and metabolic regulation. *Planta* **208**: 66-720

Gonzalez MC, Osuna L, Echevarria C, Vidal J and Cejudo FJ (1998) Expression and localization of phosphoenolpyruvate carboxylase in developing and germinating wheat grains. *Plant Physiology* **116**: 1249-58

Gordon AJ, Minchin FR, James CL and Komina O (1999) Sucrose synthase in legume nodules is essential for nitrogen fixation. *Plant Physiology* **120**: 867-877

Guillet C, Just D, Bénard N, Destrac-Irvine A, Baldet P, Hernould M, Raymond P and Rothan C (2002) A fruit-specific phosphoenolpyruvate carboxylase is related to rapid growth in tomato fruit. *Planta* **214**: 717-726

Gupta SK, Ku MSB, Lin JH, Zhang D and Edwards GE (1994) Light/dark modulation of phosphoenolpyruvate carboxylase in C_3 and C_4 species. *Photosynthesis Research* **42**: 133-143

Hahn NJ (1966) The congo red reaction in bacteria and its usefulness in the identification of rhizobia. *Canadian Journal of Microbiology* **12**: 725-733

Hall A, Kozma-Bognar L, Toth R, Nagy F and Millar AJ (2001) Conditional circadian regulation of PHYTOCHROME A gene expression. *Plant Physiology* **127**: 1808-1818

Hanks SK and Hunter T (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification based on phylogentic analysis. In DG Hardie, SK Hanks, eds *The Protein Kinase Factsbook.* Academic Press, London,

Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA and Kay SA (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science* **290**: 2110-3

Hartwell J, Gill A, Nimmo GA, Wilkins MB, Jenkins GI and Nimmo HG (1999a) Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant Journal* **20**: 1-10

Hartwell J, Jenkins GI, Wilkins MB and Nimmo HG (1999b) The light induction of maize phosphoenolpyruvate carboxylase kinase translatable mRNA requires transcription but not translation. *Plant, Cell and Environment* **22**: 883-889

Hartwell J, Smith LH, Wilkins MB, Jenkins GI and Nimmo HG (1996) Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *Plant Journal* **10**: 1071-1078

Hartwig U, Boller B and Nosberger J (1987) Oxygen supply limits nitrogenase activity of clover nodules after defoliation. *Annals of Botany* **59**: 285-291

Hata S, Izui K and Kouchi H (1998) Expression of a soybean nodule-enhanced phosphoenolpyruvate carboxylase gene that shows striking similarity to another gene for a house-keeping isoform. *Plant Journal* **13**: 267-273

Hausler RE, Hirsch HJ, Kreuzaler F and Peterhansel C (2002) Overexpression of C_4-cycle enzymes in transgenic C_3 plants: a biotechnological approach to improve C_3-photosynthesis. *Journal of Experimental Botany* **53**: 591-607

Hausler RE, Rademacher T, Li J, Lipka V, Fischer KL, Schubert S, Kreuzaler F and Hirsch HJ (2001) Single and double overexpression of C_4-cycle genes had differential effects on the pattern of endogenous enzymes, attenuation of photorespiration and on contents of UV protectants in transgenic potato and tobacco plants. *Journal of Experimental Botany* **52**: 1785-803

Horvath B, Heidstra R, Lados M, Moerman M, Spaink HP, Promé JC, van Kammen A and Bisseling T (1993) Lipo-oligosaccharides of *Rhizobium* induce infection-related early nodulin gene expression in pea root hairs. *Plant Journal* **4**: 727-733

Hrabak EM, Chan CW, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, Thomas M, Walker-Simmons K, Zhu JK and Harmon AC (2003) The Arabidopsis CDPK-SnRK superfamily of protein kinases. *Plant Physiology* **132**: 666-80

Hunt S, King BJ and D.B. L (1989) Effects of gradual increases in oxygen on nodule activity in soybean. *Plant Physiology* **91**: 315-321

Huppe HC and Turpin DH (1994) Integration of carbon and nitrogen metabolism in plant and algal cells. *Annual Review of Plant Physiology and Plant Molecular Biology* **45**: 577-607

Israel DW and Jackson WA (1982) Ion balance, uptake, and transport processes in N_2-fixing and nitrate- and urea-dependent soybean plants. *Plant Physiology* **69**: 171-178

Izui K, Ishijima S, Yamaguchi Y, Katagiri F, Murata T, Katsuya S, Sugiyama T and Katsuki H (1986) Cloning and sequence analysis of cDNA encoding active phosphoenolpyruvate carboxylase of the C_4-pathway from maize. *Nucleic Acids Research* **14**: 1615-1628

Izui K, Matsumara H, Furumoto T and Kai Y (2004) Phosphoenolpyruvate carboxylase: a new era of structural biology. *Annual Reviews of Plant Biology* **55**: 69-84

Jeanneau M, Vidal J, Gousset-Dupont A, Lebouteiller B, Hodges M, Gerentes D and Perez P (2002) Manipulating PEPC levels in plants. *Journal of Experimental Botany* **53**: 1837-45

Jiao JA, Vidal J, Echevarria C and Chollet R (1991) *In vivo* regulatory phosphorylation site in C_4-leaf phosphoenolpyruvate carboxylase from maize and *Sorghum*. *Plant Physiology* **96**: 297-301

Johnson LN, Noble MEM and Owen DJ (1996) Active and inactive protein kinases. *Cell* **85**: 149-158

Kai Y, Matsumura H, Inoue T, Terada K, Nagara Y, Yoshinaga T, Kihara A, Tsumura K and Izui K (1999) Three-dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition. *Proceedings of the National Academy of Science, U S A* **96**: 823-8

Kai Y, Matsumura H and Izui K (2003) Phosphoenolpyruvate carboxylase: three-dimensional structure and molecular mechanisms. *Archives of Biochemistry and Biophysics* **414**: 170-9

Kalamajka R, Hahnen S, Cavalar M, Töpsch S, Weier D and Peterhänsel C (2003) Restriction accessibility in isolated nuclei reveals light-induced chromatin reorganisation at the PEPC promoter in maize. *Plant Molecular Biology* **52**: 669-678

Kaldis AD, Kousidis P, Kesanopoulos K and Prombona A (2003) Light and circadian regulation in the expression of LHY and Lhcb genes in *Phaseolus vulgaris*. *Plant Molecular Biology* **52**: 981-997

Kausch AP, Owen TPJ, Zachwieja SJ, Flynn AR and Sheen J (2001) Mesophyll-specific, light and metabolic regulation of the C_4 PPCZm1 promoter in transgenic maize. *Plant Molecular Biology* **45**: 1-15

King BJ, Layzell DB and Canvin DT (1986) The role of dark carbon dioxide fixation in root nodules of soybean. *Plant Physiology* **81**: 200-205

Koizumi N, Toyota K, Kitajima S, Yamada Y and Sato F (2002) Genomic structure and promoter analysis of phosphoenolpyruvate carboxylase in a C_3 plant, *Nicotiana sylvestris*. *Bioscience, Biotechnology, and Biochemistry* **66**: 1691-1696

Ku MS, Agarie S, Nomura M, Fukayama H, Tsuchida H, Ono K, Hirose S, Toki S, Miyao M and Matsuoka M (1999) High-level expression of maize

phosphoenolpyruvate carboxylase in transgenic rice plants. *Nature Biotechnology* **17**: 76-80

Kuzma MM, Heike W, Storer P, Oresnik I, Atkins CA and Layzell DB (1999) The site of oxygen limitation in soybean nodules. *Plant Physiology* **119**: 399-407

Kuzma MM, Hunt S and Layzell DB (1993) Role of oxygen in the limitation and inhibition of nitrogenase activity and respiration rate in individual soybean nodules. *Plant Physiology* **106**: 263-270

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685

Law RD and Plaxton WC (1995) Purification and characterization of a novel phosphoenolpyruvate carboxylase from banana fruit. *Biochemical Journal* **307**: 807-16

Layzell DB, Hunt S and Palmer GR (1990) Mechanism of nitrogenase inhibition in soybean nodules. Pulse-modulated spectroscopy indicates that nitrogenase activity is limited by O_2. *Plant Physiology* **92**: 1101-1107

Lee H and Copeland L (1996) Phosphofructokinase from the host fraction of chickpea nodules. *Physiologia Plantarum* **96**: 607-614

Leegood RC (1999) Photosynthesis in C_3 plants: the Benson-Calvin cycle and photorespiration. In PJ Lea, RC Leegood, eds *Plant biochemistry and molecular biology.*, Ed. 2. John Wiley, Chichester,

Leegood RC (2002) C_4 photosynthesis: principles of CO_2 concentration and prospects for its introduction into C_3 plants. *Journal of Experimental Botany* **53**: 581-590

Lepiniec L, Keryer E, Philippe H, Gadal P and Cretin C (1993) Sorghum phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution. *Plant Molecular Biology* **21**: 487-502

Lepiniec L, Vidal J, Chollet R, Gadal P and Cretin C (1994) Phosphoenolpyruvate carboxylase - structure, regulation and evolution. *Plant Science* **99**: 111-124

Leport L, Kandlbinder A, Baur B and Kaiser WM (1996) Diurnal modulation of phosphoenolpyruvate carboxylation in pea leaves and roots as related to tissue malate concentrations and to the nitrogen source. *Planta* **198**: 495-501

Lhuissier FGP, De Ruijter NCA, Sieberer BJ, Esseling JJ and Emons AMC (2001) Time course of cell biological events evoked in legume root hairs by *Rhizobium* Nod factors: state of the art. *Annals of Botany* **87**: 289-302

Li B and Chollet R (1993) Resolution and identification of C_4 phosphoenolpyruvate-carboxylase protein-kinase polypeptides and their reversible light activation in maize leaves. *Archives of Biochemistry and Biophysics* **307**: 416-419

Li B and Chollet R (1994) Salt induction and the partial purification/characterization of phosphoenolpyruvate carboxylase protein-serine kinase from an inducible Crassulacean Acid Metabolism (CAM) plant, *Mesembryanthemum crystallinum* L. *Archives of Biochemistry and Biophysics* **314**: 247-254

Li B, Zhang XQ and Chollet R (1996) Phosphoenolpyruvate Carboxylase Kinase in Tobacco Leaves Is Activated by Light in a Similar but Not Identical Way as in Maize. *Plant Physiology* **111**: 497-505

Lodwig E and Poole P (2003) Metabolism of Rhizobium bacteroids. *Critical Reviews in Plant Sciences* **22**: 37-78

Lodwig EM, Hosie AHF, Bordes A, Findlay K, Allaway D, Karunakaran R, Downie JA and Poole PS (2003) Amino-acid cycling drives nitrogen fixation in the legume - Rhizobium symbiosis. *Nature* **422**: 722-726

Marie C, Broughton WJ and Deakin WJ (2001) Rhizobium type III secretion systems: legume charmers or alarmers? *Current Opinion in Plant Biology* **4**: 336-342

Marsh JT (2002) Phosphoenolpyruvate carboxylase kinase in higher plants. PhD University of Glasgow

Matsumura H, Terada M, Shirakata S, Inoue T, Yoshinaga T, Izui K and Kai Y (1999) Plausible phosphoenolpyruvate binding site revealed by 2.6 A structure of Mn^{2+}-bound phosphoenolpyruvate carboxylase from *Escherichia coli*. *Febs Letters* **458**: 93-6

Matsumura H, Xie Y, Shirakata S, Inoue T, Yoshinaga T, Ueno Y, Izui K and Kai Y (2002) Crystal structures of C_4 form maize and quaternary complex of *E. coli* phosphoenolpyruvate carboxylases. *Structure (Camb)* **10**: 1721-30

Matsuoka M, Kyozuka J, Shimamoto K and 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 Journal* **6**: 311-9

McNaughton GA, MacKintosh C, Fewson CA, Wilkins MB and Nimmo HG (1991) Illumination increases the phosphorylation state of maize leaf phosphoenolpyruvate carboxylase by causing an increase in the activity of a protein kinase. *Biochimica et Biophysica Acta* **1093**: 189-95

Michiels J and Vanderleyden J (1994) Molecular basis of the establishment and functioning of a N_2-fixing root nodule. *World Journal of Microbiology and Biotechnology* **10**: 612-630

Millar AH, Day DA and Bergersen FJ (1995) Microaerobic respiration and oxidative phosphorylation by soybean nodule mitochondria: implications for nitrogen fixation. *Plant, Cell and Environment* **18**: 715-726

Miller SA, Driscoll BT, Gregerson RG, Gantt JS and Vance CP (1998) Alfalfa malate dehydrogenase (MDH): molecular cloning and characterization of five different forms reveals a unique nodule-enhanced MDH. *Plant Journal* **15**: 173-184

Miller SS, Boylan KLM and Vance CP (1987) Alfalfa root nodule CO_2 fixation: III. Immunological studies of nodule phosphoenolpyruvate carboxylase. *Plant Physiology* **84**: 501-508

Miyao M (2003) Molecular evolution and genetic engineering of C_4 photosynthetic enzymes. *Journal of Experimental Botany* **54**: 179-189

Moing A, Rothan C, Svanella L, Just D, Diakou P, Raymond P, Gaudillere JP and Monet R (2000) The role of phosphoenolpuruvate carboxylase in organic acid accumulation during peach fruit development. *Plant Physiology* **108**: 1-10

Moraes TF and Plaxton WC (2000) Purification and characterization of phosphoenolpyruvate carboxylase from *Brassica napus* (rapeseed) suspension cell cultures: implications for phosphoenolpyruvate carboxylase regulation during phosphate starvation, and the integration of glycolysis with nitrogen assimilation. *European Journal Of Biochemistry* **267**: 4465-4476

Moran JF, Sun Z, Sarath G, Arredondo-Peter R, James EK, Becana M and Klucas RV (2002) Characterization and subcellular localization of soybean nodule dihydrolipoamide reductase. *Plant Physiology* **128**: 300-313

Mouritzen P and Rosendahl L (1997) Identification of a transport mechanism for NH_4^+ in the symbiosome membrane of pea root nodules. *Plant Physiology* **115**: 519-526

Munoz T, Escribano MI and Merodio C (2001) Phosphoenolpyruvate carboxylase from cherimoya fruit: properties, kinetics and effects of high CO_2. *Phytochemistry* **58**: 1007-13

Müller-Röber B, Ehrhardt T and Plesch G (1998) Molecular features of stomatal guard cells. *Journal of Experimental Botany* **49**: 293-304

Mylona P, Pawlowski K and Bisseling T (1995) Symbiotic nitrogen fixation. *Plant Cell* **7**: 869-883

Nakagawa T, Izumi T, Banba M, Umehara Y, Kouchi H, Izui K and Hata S (2003a) Characterization and expression analysis of genes encoding phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxylase kinase of *Lotus japonicus*, a model legume. *Molecular Plant-Microbe Interactions* **16**: 281-8

Nakagawa T, Takane K, Sugimoto T, Izui K, Kouchi H and Hata S (2003b) Regulatory regions and nuclear factors involved in nodule-enhanced expression of a soybean phosphoenolpyruvate carboxylase gene: implications for molecular evolution. *Molecular Genetics and Genomics* **269**: 163-72

Nielsen KM (2003) Transgenic organisms - time for conceptual diversification? *Nature Biotechnology* **21**: 227-228

Nimmo GA, McNaughton GAL, Fewson CA, Wilkins MB and Nimmo HG (1987a) Changes in the kinetic properties and phosphorylation state of phosphoenolpyruvate carboxylase in *Zea mays* leaves in response to light and dark. *Federation of European Biochemical Societies Letters* **213**: 18-22

Nimmo GA, Nimmo HG, Fewson CA and Wilkins MB (1984) Diurnal changes in the properties of phosphoenolpyruvate carboxylase in *Bryophyllum* leaves: a possible covalent modification. *Federation of European Biochemical Societies Letters* **178**: 199-203

Nimmo GA, Nimmo HG, Hamilton ID, Fewson CA and Wilkins MB (1986) Purification of the phosphorylated night form and dephosphorylated day form of phosphoenolpyruvate carboxylase from *Bryophyllum fedtshenkoi* . *Biochemical Journal* **239**: 213-220

Nimmo GA, Wilkins MB, Fewson CA and Nimmo HG (1987b) Persistent circadian rhythms in the phosphorylation state of phosphoenolpyruvate carboxylase from *Bryophyllum fedtschenkoi* leaves and in its sensitivity to inhibition by malate. *Planta* **170:** 408-415

Nimmo GA, Wilkins MB and Nimmo HG (2001) Partial purification and characterization of a protein inhibitor of phosphoenolpyruvate carboxylase kinase. *Planta* **213**: 250-7

Nimmo HG (2000) The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends in Plant Science* **5**: 75-80

Nimmo HG (2003) Control of the phosphorylation of phosphoenolpyruvate carboxylase in higher plants. *Archives of Biochemistry and Biophysics* **414**: 189-96

Notton BA and Blanke MM (1993) Phosphoenolpyruvate carboxylase in Avocado fruit - purification and properties. *Phytochemistry* **33**: 1333-1337

O'Leary MH (1982) Phosphoenolpyruvate carboxylase: an enzymologist's view. *Annual Review of Plant Physiology* **33**: 297-315

Ocana A, Cordovilla MdP, Ligero F and Lluch C (1996) Phosphoenolpyruvate carboxylase in root nodules of *Vicia faba*: partial purification and properties. *Physiologia Plantarum* **97**: 724-730

Ogawa N, Okumura S and Izui K (1992) A Ca^{2+}-dependent protein kinase phosphorylates phosphoenolpyruvate carboxylase in maize. *Febs Letters* **302**: 86-8

Ogawa N, Yabuta N, Ueno Y and Izui K (1998) Characterization of a maize Ca^{2+}-dependent protein kinase phosphorylating phosphoenolpyruvate carboxylase. *Plant and Cell Physiology* **39**: 1010-9

Or E, Baybik J, Sadka A and Saks Y (2000) Isolation of mitochondrial malate dehydrogenase and phosphoenolpyruvate carboxylase cDNA clones from grape berries and analysis of their expression pattern throughout berry development. *Journal of Plant Physiology* **157**: 527-534

Osuna L, Coursol S, Pierre JN and Vidal J (2004) A Ca^{2+}-dependent protein kinase with characteristics of protein kinase C in leaves and mesophyll cell protoplasts from Digitaria sanguinalis: possible involvement in the C_4-phosphoenolpyruvate carboxylase phosphorylation cascade. *Biochemical and Biophysical Research Communications* **314**: 428-33

Osuna L, Pierre JN, Gonzalez MC, Alvarez R, Cejudo FJ, Echevarria C and Vidal J (1999) Evidence for a slow-turnover form of the Ca^{2+}-independent phosphoenolpyruvate carboxylase kinase in the aleurone-endosperm tissue of germinating barley seeds. *Plant Physiology* **119**: 511-20

Outlaw JrWH, Du Z, Meng FX, Aghoram K, Riddle KA and Chollet R (2002) Requirements for activation of the signal-transduction network that leads to regulatory phosphorylation of leaf guard-cell phosphoenolpyruvate carboxylase during fusicoccin-stimulated stomatal opening. *Archives of Biochemistry and Biophysics* **407**: 63-71

Outlaw JrWH and Lowry OH (1977) Organic acid accumulation in guard cells during stomatal opening. *Proceeding of the National Academy of Science, U S A* **74**: 4434-4438

Pairoba CF, Colombo SL and Andreo CS (1996) Flavanoids as inhibitors of NADP-malic enzyme and PEP carboxylase from two C_4 plants. *Bioscience, Biotechnology and Biochemistry* **60**: 779-783

Parniske M (2000) Intracellular accommodation of microbes by plants: a common developmental program for symbiosis and disease? *Current Opinion in Plant Biology* **3**: 320-328

Paterson KM and Nimmo HG (2000) Effects of pH on the induction of phosphoenolpyruvate carboxylase kinase in *Kalanchoe fedtschenkoi*. *Plant Science* **154**: 135-141

Pathirana MS, Samac DA, Roeven R, Yoshioka H, Vance CP and Gantt JS (1997) Analyses of phosphoenolpyruvate carboxylase gene structure and expression in alfalfa nodules. *Plant Journal* **12**: 293-304

Peters GA (1991) Azolla and other plant-cyanobacterial symbioses - aspects of form and function. *Plant Soil* **137**: 25-36

Plaxton WC (1996) The organization and regulation of plant glycolysis. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 185-214

Price NPJ, Relic B, Talmont F, Lewin A, Prome D, Pueppke SG, Maillet F, Denarie J, Prome JC and Broughton WJ (1992) Broad-host-range Rhizobium species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are *O*-acetylated or sulphated. *Molecular Microbiology* **6**: 3575-3584

Pühler A, Arlat M, Becker A, Gottfert M, Morrissey JP and O'Gara F (2004) What can bacterial genome research teach us about bacteria-plant interactions? *Current Opinion in Plant Biology* **7**: 137-147

Qiu YL, Lee J, Bernasconi-Quandroni F, Soltis DE, Soltis PS, Zanis M, Chen Z, Savolainen V and Chase MW (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**: 404-407

Raghothama KG (1999) Phosphate acquisition. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**: 665-693

Rajagopalan AV, Devi MT and Raghavendra AS (1994) Molecular biology of C_4 phosphoenolpyruvate carboxylase: Structure, regulation and genetic engineering. *Photosynthesis Research* **39**: 115-135

Rao SK, Magnin NC, Reiskind JB and Bowes G (2002) Photosynthetic and other phosphoenolpyruvate carboxylase isoforms in the single-cell, facultative C_4 system of *Hydrilla verticillata*. *Plant Physiology* **130**: 876-86

Rivoal J, Dunford R, Plaxton WC and Turpin DH (1996) Purification and properties of four phosphoenolpyruvate carboxylase isoforms from the green alga *Selenastrum minutum*: evidence that association of the 102-kDa catalytic subunit with unrelated polypeptides may modify the physical and kinetic properties of the enzyme. *Archives of Biochemistry and Biophysics* **332**: 47-57

Rivoal J, Plaxton WC and Turpin DH (1998) Purification and characterization of high- and low-molecular-mass isoforms of phosphoenolpyruvate carboxylase from *Chlamydomonas reinhardtii*. Kinetic, structural and immunological evidence that the green algal enzyme is distinct from the prokaryotic and higher plant enzymes. *Biochemical Journal* **331 (Pt 1)**: 201-9

Rivoal J, Trzos S, Gage DA, Plaxton WC and Turpin DH (2001) Two unrelated phosphoenolpyruvate carboxylase polypeptides physically interact in the high molecular mass isoforms of this enzyme in the unicellular green alga *Selenastrum minutum*. *Journal Of Biological Chemistry* **276**: 12588-97

Rivoal J, Turpin DH and Plaxton WC (2002) *In vitro* phosphorylation of phosphoenolpyruvate carboxylase from the green alga *Selenastrum minutum*. *Plant and Cell Physiology* **43**: 785-92

Robinson DL, Pathirana SM, Gantt JS and Vance CP (1996) Immunogold localization of nodule-enhanced phosphoenolpyruvate carboxylase in alfalfa. *Plant Cell and Environment* **19**: 602-608

Rolletschek H, Borisjuk L, Radchuk R, Miranda M, Heim U, Wobus U and Weber H (2004) Seed-specific expression of a bacterial phosphoenolpyruvate carboxylase in *Vicia narbonensis* increases protein content and improves carbon economy. *Plant Biotechnology Journal* **2**: 211-219

Rommens CM (2004) All-native DNA transformation: a new approach to plant genetic engineering. *Trends in Plant Science* **9**: 457-464

Rommens CM, Humara JM, Ye J, Yan H, Richael C, Zhang L, Perry R and Swords K (2004) Crop improvement through modification of the plant's own genome. *Plant Physiology* **135**: 421-431

Rosendahl L, Vance CP and Pedersen WB (1990) Products of dark CO_2 fixation in pea root nodules support bacteroid metabolism. *Plant Physiology* **93**: 12-19

Rozen S and Skaletsky HS (2000) Primer3 on the WWW for general users and for biologist programmers. In S Krawetz, S Misener, eds *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365-386

Ryle GJA, Powell CE and Gordon AJ (1978) Effect of source of nitrogen on the growth of Fiskeby soybean: the carbon economy of whole plants. *Annals of Botany* **42**: 637-648

Sambrook J, Fritsch EF and Maniatis T (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanchez R and Cejudo FJ (2003) Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from Arabidopsis and rice. *Plant Physiology* **132**: 949-57

Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, de Bruijn AY, Sullivan S and Qiu YL (2000) Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* **49**: 306-362

Sawhney V, Saharan MR and Singh R (1987) Nitrogen fixing efficiency and enzymes of CO_2 assimilation in nodules of ureide and amide producing legumes. *Plant Physiology* **129**: 201-210

Saze H, Ueno Y, Hisabori T, Hayashi H and Izui K (2001) Thioredoxin-mediated reductive activation of a protein kinase for the regulatory phosphorylation of C_4-form phosphoenolpyruvate carboxylase from maize. *Plant and Cell Physiology* **42**: 1295-302

Schaeffer HJ, Forsthoefel NR and Cushman JC (1995) Identification of enhancer and silencer regions involved in salt-responsive expression of Crassulacean acid metabolism (CAM) genes in the facultative halophyte *Mesembryanthemum crystallinum*. *Plant Molecular Biology* **28**: 205-218

Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carre IA and Coupland G (1998) The late elongated hypocotyl mutation of Arabidopsis disrupts

circadian rhythms and the photoperiodic control of flowering. *Cell* **93**: 1219-1229

Scheible W-R, Krapp A and Stitt M (2000) Reciprocal diurnal changes of phosphoenolpyruvate expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. *Plant, Cell and Environment* **23**: 1155-1167

Schuller KA, Turpin DH and Plaxton WC (1990) Metabolite regulation of partially purified soybean nodule phosphoenolpyruvate carboxylase. *Plant Physiology* **94**: 1429-1435

Schuller KA and Werner D (1993) Phosphorylation of soybean (*Glycine max* L.) nodule phosphoenolpyruvate carboxylase *in vitro* decreases sensitivity to inhibition by L-malate. *Plant Physiology* **101**: 1267-1273

Serraj R, Sinclair TR and Purcell LC (1999) Symbiotic N_2 fixation responce to drought. *Journal of Experimental Botany* **50**: 143-155

Silvente S, Camas A and Lara M (2003) Heterogeneity of sucrose synthase genes in bean (*Phaseolus vulgaris* L.): evidence for a nodule-enhanced sucrose synthase gene. *Journal of Experimental Botany* **54**: 749-755

Smith CR, Knowles VL and Plaxton WC (2000) Purification and characterization of cytosolic pyruvate kinase from *Brassica napus* (rapeseed) suspension cell cultures: implications for the integration of glycolysis with nitrogen assimilation. *European Journal Of Biochemistry* **267**: 4477-4485

Smith LH, Lillo C, Nimmo HG and Wilkins MB (1996) Light regulation of phosphoenolpyruvate carboxylase in barley mesophyll protoplasts is modulated by protein synthesis and calcium and not necessarily correlated with phosphoenolpyruvate carboxylase kinase activity. *Planta* **200**: 174-180

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM and Martin PM (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceeding of the National Academy of Science, U S A* **92**: 2647-2651

Stacey G, Vodkin L, Parrott WA and Shoemaker RC (2004) National science foundation-sponsored workshop report. Draft plan for soybean genomics. *Plant Physiology* **135**: 59-70

Suganuma N, Okada Y and Kanayama Y (1997) Isolation of a cDNA for nodule-enhanced phosphoenolpyruvate carboxylase from pea and its expression in effective and plant-determined ineffective pea nodules. *Journal of Experimental Botany* **48**: 1165-1173

Sugimoto T, Kawasaki T, Kato T, Whittier RF, Shibata D and Kawamura Y (1992) cDNA sequence and expression of a phosphoenolpyruvate carboxylase gene from soybean. *Plant Molecular Biology* **20**: 743-747

Svensson P, Blasing OE and Westhoff P (2003) Evolution of C_4 phosphoenolpyruvate carboxylase. *Archives of Biochemistry and Biophysics* **414**: 180-8

Taiz L (1998) *Plant physiology*. Sinauer Associates, Sunderland, Mass.

Taniguchi M, Izawa K, Ku MS, Lin JH, Saito H, Ishida Y, Ohta S, Komari T, Matsuoka M and Sugiyama T (2000) Binding of cell type-specific nuclear proteins to the 5'-flanking region of maize C_4 phosphoenolpyruvate carboxylase gene confers its differential transcription in mesophyll cells. *Plant Molecular Biology* **44**: 543-557

Taybi T, Patil S, Chollet R and Cushman JC (2000) A minimal serine/threonine protein kinase circadianly regulates phosphoenolpyruvate carboxylase activity in Crassulacean acid metabolism-induced leaves of the common ice plant. *Plant Physiology* **123**: 1471-82

Terada K and Izui K (1991) Site-directed mutagenesis of the conserved histidine residue of phosphoenolpyruvate carboxylase - His138 is essential for the second partial reaction. *European Journal Of Biochemistry* **202**: 797-803

Thain SC, Murtas G, Lynn JR, McGrath RB and Millar AJ (2002) The circadian clock that controls gene expression in Arabidopsis is tissue specific. *Plant Physiology* **130**: 102-10

Thomas M, Cretin C, Vidal J, Keryer E, Gadal P and Monsinger E (1990) Light-regulation of phosphoenolpyruvate carboxylase mRNA in leaves of C_4 plants: evidence for phytochrome control on transcription during greening and rhythmicity. *Plant Science* **69**: 65-78

Thumfort PP, Atkins CA and Layzell DB (1994) Re-evaluation of the role of the infected cell in the control of oxygen diffusion in legume nodules. *Plant Physiology* **105**: 1321-1333

Thummler F and Verma DPS (1987) Nodulin-100 of soybean is the subunit of sucrose synthase regulated by the availability of free heme in nodules. *Journal Of Biological Chemistry* **262**: 14730-14736

Toyota K, Koizumi N and Sato F (2003) Transcriptional activation of phosphoenolpyruvate carboxylase by phosphorus deficiency in tobacco. *Journal of Experimental Botany* **54**: 961-9

Trevaskis B, Colebatch G, Desbrosses G, Wandrey M, Wienkoop S, Saalbach G and Udvardi M (2002) Differentiation of plant cells during symbiotic nitrogen fixation. *Comparative and Functional Genomics* **3**: 151-157

Tsuchida Y, Furumoto T, Izumida A, Hata S and Izui K (2001) Phosphoenolpyruvate carboxylase kinase involved in C_4 photosynthesis in *Flaveria trinervia*: cDNA cloning and characterization. *Febs Letters* **507**: 318-22

Turpin DH, Botha FC, Smith RG, Feil R, Horsey AK and Vanlerberghe GC (1990) Regulation of carbon partitioning to respiration during dark ammonium assimilation by the green alga *Selenastrum minutum*. *Plant Physiology* **93**: 166-175

Udvardi MK and Day DA (1997) Metabolite transport across symbiotic membranes of legume nodules. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**: 493-523

Udvardi MK, Price GD, Gresshoff PM and Day DA (1988) A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. *Febs Letters* **231**: 36-40

Ueno Y, Imanari E, Emura J, Yoshizawa-Kumagaye K, Nakajima K, Inami K, Shiba T, Sakakibara H, Sugiyama T and Izui K (2000) Immunological analysis of the phosphorylation state of maize C_4-form phosphoenolpyruvate carboxylase with specific antibodies raised against a synthetic phosphorylated peptide. *Plant Journal* **21**: 17-26

Uhde-Stone C, Gilbert G, Johnson JM-F, Litjens R, Zinn KE, Temple SJ, Vance CP and Allan DL (2003) Acclimation of white lupin to phosphorous deficiency involves enhanced expression of genes related to organic acid metabolism. *Plant and Soil* **248**: 99-116

Van Quy L, Foyer C and Champigny ML (1991) Effect of light and NO_3^- on wheat leaf phosphoenolpyruvate carboxylase activity - evidence for covalent modulation of the C_3 enzyme. *Plant Physiology* **97**: 1476-1482

Vance CP (1983) *Rhizobium* infection and nodulation: A beneficial plant disease? *Annual Review of Microbiology* **37**: 399-424

Vance CP, Boylan KLM, Maxwell CA, Heichel GH and Hardman LL (1985) Transport and partitioning of CO_2 fixed by root nodules of ureide and amide producing legumes. *Plant Physiology* **78**: 774-778

Vance CP and Gantt JS (1992) Control of Nitrogen and Carbon Metabolism in Root-Nodules. *Physiologia Plantarum* **85**: 266-274

Vance CP and Heichel GH (1991) Carbon in N_2 fixation: Limitation or exquisite adaptation. *Annual Review of Plant Physiology and Plant Molecular Biology* **42**: 373-392

Vance CP and Lamb JFS (2001) Application of biochemical studies to improving nitrogen fixation. *Australian Journal of Experimental Agriculture* **41**: 403-416

Vazquez-Tello A, Whittier RF, Kawasaki T, Sugimoto T, Kawamura Y and Shibata D (1993) Sequence Of a Soybean (Glycine-Max L) Phosphoenolpyruvate Carboxylase Cdna. *Plant Physiology* **103**: 1025-1026

Vessey JK, Walsh KB and Layzell.D.B. (1988) Oxygen limitation of nitrogen fixation in stem-girdled and nitrate-treated soybean. *Plant Physiology* **73**: 113-121

Vidal J and Chollet R (1997) Regulatory phosphorylation of C_4 PEP carboxylase. *Trends in Plant Sciences* **2**: 230-237

Vincent JM (1970) *A manual for practical study of root nodule bacteria.* Blackwell, Oxford

Viprey V, Del Greco A, Golinowski W, Broughton WJ and Peret X (1998) Symbiotic implications of type III protein secretion machinery in *Rhizobium*. *Molecular Microbiology* **28**: 1381-1389

Wadham C, Winter H and Schuller KA (1996) Regulation of soybean nodule phosphoenolpyruvate carboxylase *in vivo*. *Physiologia Plantarum* **97**: 531-535

Wang YH and Chollet R (1993a) In vitro phosphorylation of purified tobacco-leaf phosphoenolpyruvate carboxylase. *Febs Letters* **328**: 215-8

Wang YH and Chollet R (1993b) Partial purification and characterization of phosphoenolpyruvate carboxylase protein-serine kinase from illuminated maize leaves. *Archives of Biochemistry and Biophysics* **304**: 496-502

Wei H, Atkins CA and Layzell DB (2004a) Adenylate gradients and $Ar:O_2$ effects on legume nodules: I. Mathematical models. *Plant Physiology* **134**: 801-812

Wei H, Atkins CA and Layzell DB (2004b) Adenylate gradients and Ar:O2 effects on legume nodules: II. Changes in the subcellular adenylate pools. *Plant Physiology* **134**: 1775-1783

Wilkins MB (1992) Circadian rhythms - their origin and control. *New Phytologist* **121**: 347-375

Witty JF, Minchin FR, Scot L and Revsbech NP (1987) Direct evidence for changes in the resistance of legume root nodules to O_2 diffusion. *Journal of Experimental Botany* **38**: 1129-1140

Woo KC and Xu S (1996) Metabolite regulation of phosphoenolpyruvate carboxylase in legume nodules. *Australian Journal of Plant Physiology* **23**: 413-419

Xu W, Zhou Y and Chollet R (2003) Identification and expression of a soybean nodule-enhanced PEP-carboxylase kinase gene (NE-PpcK) that shows striking up-/down-regulation in vivo. *Plant Journal* **34**: 441-52

Xu Y and Johnson CH (2001) A clock- and light-regulated gene that links the circadian oscillator to LHCB gene expression. *Plant Cell* **13**: 1411-1425

Yanagisawa S (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant Journal* **21**: 281-8

Yanagisawa S and Sheen J (1998) Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *Plant Cell* **10**: 75-89

Yano M, Terada K, Umiji K and Izui K (1995) Catalytic role of an arganine residue in the highly conserved and unique sequence of phosphoenolpyruvate carboxylase. *Journal of Biochemistry (Tokyo)* **117**: 1196-200

Zhang XQ and Chollet R (1997) Phosphoenolpyruvate carboxylase protein kinase from soybean root nodules: partial purification, characterization, and up/down-regulation by photosynthate supply from the shoots. *Archives of Biochemistry and Biophysics* **343**: 260-8

Zhang XQ, Li B and Chollet R (1995) *In Vivo* Regulatory Phosphorylation of Soybean Nodule Phosphoenolpyruvate Carboxylase. *Plant Physiology* **108**: 1561-1568