Comparative enzymology of the convergent co-option of phosphoenolpyruvate carboxylase for C_4 photosynthesis

by

Nicholas Raymond Moody

Department of Chemistry

University of Sheffield

Submitted in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

August 2018

CONTENTS

Acknowledgments ..... **iii**

Abstract ..... **iv**

Chapter 1: General Introduction **1**

Calvin-Benson cycle 2

Photorespiration 4

C_4 photosynthesis 7

Phosphoenolpyruvate carboxylase enzyme overview 12

Functional residues of PEPC 17

Evolution of C_4 PEPC 20

Thesis Aims and Structure **35**

Chapter 2: Changes in bicarbonate specificity during the evolution of C_4 PEPC in *Flaveria* **37**

Abstract 38

Results 41

Discussion 47

Experimental Procedure 50

References 54

Supplementary Data 57

Chapter 3: Comparative biochemistry reveals greater adaptation of PEPC in older C_4 lineages **67**

Abstract 68

Results 72

Discussion 79

Experimental Procedure 83

References 88

Supplementary Data 91

Chapter 4: Amino acids of PEPC positively selected for C_4 photosynthesis have various functions 103

Abstract 104

Results 108

Discussion 114

Experimental Procedure 118

References 122

Supplementary Data 126

General Discussion 137

Insight into the metabolic changes of the C_4 cell 139

Convergent evolution of the C_4 PEPC 141

Greater adaptation of PEPC after full establishment of C_4 143

Acquisition rates of C_4 specific properties in C_4 PEPC 145

Conclusion 147

References 149

Appendix 155

Molecular Cloning of PEPC 156

PEPC Expression Optimization 160

PEPC Purification Optimization 162

PEPC Quantification 163

PEPC Assay Design 164

Experimental Procedure 166

References 175

ACKNOWLEDGEMENTS

I would first like to thank my supervisors Jim Reid and Pascal-Antoine Christin. Jim and Pascal-Antoine have provided me with their time and their expertise throughout my project. Even in the most pressing of circumstances, they have been able to assist me. I have greatly benefited from Jim's knowledge of cloning, protein purification, enzyme kinetics and data analysis. Pascal-Antoine has been a font of wisdom in bioinformatics, phylogenetics and evolution. Their combined knowledge provided illuminating guidance through a complex web of overlapping science that has meshed together to make this thesis. They have both put in a tremendous amount of effort helping me, sometimes at very short notice that I cannot thank them enough for.

I would like to thank Dr Chatchawal Phansopa and Antonia Tuberville in the Reid Group. Chat has been always helpful with practical first-hand knowledge of techniques and an enthusiasm for research that was always appreciated. He was particularly helpful in some of the particularly trying days of cloning. Antonia has been a great help for discussing ideas with. We have worked closely on enzyme kinetics and purification, and the disparity between our projects gave Antonia a point of view which gave clarity to my own work. Combined they have been valuable companions and a source of ideas and advice. Their help has been essential for working in a chemistry laboratory environment. Their moral support has made an enjoyable and productive work environment. I would like to thank Ben Ambrose for writing a program that expedited protein analysis, it was much appreciated.

I am grateful to Dr Luke Dunning, Dr Marjorie Lundgren, Dr Jose Moreno-Villena, and Dr Jill Oloffson for their advice and assistance. Luke and Jose helped produce a detailed phylogeny, informed by their data which I am incredibly grateful for. Marj has been essential navigating and identifying the variety of plant species. Luke, Marj and Jill have been a combined force for good, greatly aiding my understanding of the complexities of evolution and gene adaption.

I am indebted to Natalie Edwards who has provided me with companionship and support throughout my time in Sheffield. Natalie has provided essential proofreading of everything I have written. Without her support encouragement and biting wit, I would not have been able to finish the project.

I would like to thank Matheus Bianconi, Dr Daniel Wood, and the Christin and Osborne groups who have been helpful and accommodating. Their weekly lab meetings have helped my get to grips with plant cell biology and the wider context of my research. I am grateful for their patience with all my questions and I wish them well in their future endeavours after finishing their PhDs. I would also like to thank the members of the Staniland group and the members of office E81 who have been great company and better conversation. Thanks also to the Nick Williams group who were gracious enough to let me use their Cary spectrometers.

I would like to acknowledge my fellow cohort of Grantham Scholars, my comrades, who supported me throughout my time in Sheffield. They could fully appreciate the challenges of acclimating to multiple disciplines. I hope that interdisciplinary research has treated them as well as it has treated me.

I am very grateful to Professor Peter Westhoff, who gifted me the plasmids for the *Flaveria trinervia* and *Flaveria pringlei* PEPC genes, which have been essential for my project.

I am grateful for the Grantham Centre for funding this research, and the University of Sheffield for funding this research and hosting me.

ABSTRACT

Approximately a third of the world population relies on rice every day; however, increases rice yields are not increasing to match the predicted increase in world population. Engineering the C_4 photosynthetic carbon dioxide concentrating mechanism in rice may reduce inefficiencies in rice and increase yield. C_4 photosynthesis uses the enzyme phosphoenolpyruvate carboxylase (PEPC) to fix atmospheric carbon dioxide in a four-carbon acid which is shuttled to a separate compartment where the carbon dioxide is released and concentrated around the key enzyme of the Calvin-Benson cycle. Plants that have evolved the C_4 cycle co-opted PEPC from a background role. The expression patterns and kinetic properties of PEPC are adapted to the demands of carbon fixation. The changes in gene expression have been studied previously through high-throughput sequencing techniques but the biochemical changes remain largely unexplored. Understanding how PEPC adapted is a key part of engineering C_4 photosynthesis. In this thesis I looked at the kinetic changes in C_4 PEPC from one of the youngest C_4 species in the genus *Flaveria* which showed that the C_4 PEPC has a higher specificity for bicarbonate, a lower specificity for phosphoenolpyruvate, and a decreased sensitivity to inhibitors when compared to the non-C_4 *Flaveria* PEPC. I then compared the kinetic properties of PEPCs from species in the genus *Panicum*, an early and successful C_4 origin. Comparison showed a convergence in kinetic properties of C_4 PEPCs in *Panicum* and *Flaveria*. However, the changes seen in C_4 *Panicum* PEPC are quantitively greater showing further adaptation. C_4 specific changes resulting from specific amino acids changes were investigated. It was shown that the same C_4 specific mutation was responsible for similar reduction in magnitude of PEP specificity in both *Flaveria* and *Panicum* C_4 PEPC. Other investigated C_4 specific amino acids were shown not to contribute to major kinetic properties. This surprising result suggested other selection forces act in the evolution of C_4 PEPC. The evolution of C_4 PEPC involved adaption increase in the specificity for bicarbonate. It was shown the that bicarbonate plays a part in the evolution of PEPC but adaptations towards inhibitor and PEP specificity are selected preferentially.

General Introduction

One hectare of rice can feed 27 people, however by 2050 this same amount of land will need to feed 43 to cope with the increasing human population (Hibberd, Sheehy and Langdale, 2008). Through plant breeding and inorganic fertiliser, the latter half of the 20th century experienced a 'green revolution', in which crop yields dramatically increased (Evenson and Gollin, 2003). However, in recent years rice yield growth has stagnated and may decrease with climate change (Dawe, 2007). Rice is a C_3 species that is hampered by inefficiencies in carbon dioxide capture. The C_4 pathway is an adaptation that increases the productivity of plants in warm and tropical environments, and its introduction into rice is predicted to increase its yield (Edwards, 1999; Hibberd, Sheehy and Langdale, 2008; Sage, Sage and Kocacinar, 2012; Leegood, 2013). C_4 species have a much higher solar, water and nitrogen efficiency, and C_4 crops have higher yield (Sage, 2004).

In C_3 plants, which represent the ancestral state, atmospheric CO_2 is fixed into organic compounds directly by ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and other enzymes, in the first step of the Calvin-Benson cycle. By contrast, C_4 plants fix atmospheric carbon dioxide via the coupled action of the enzymes carbonic anhydrase (CA) and phosphoenolpyruvate carboxylase (PEPC), which produce a four-carbon acid. This acid is shuttled to a separate compartment, where the Calvin-Benson cycle is segregated. Carbon dioxide is released therein, increasing the concentration of carbon dioxide around RuBisCO (von Caemmerer and Furbank, 2003). An efficient C_4 concentrating mechanism requires the coordinated action of numerous anatomical and biochemical components (Hatch, 1987). Despite this complexity, the C_4 process has evolved many times in flowering plants, across both monocots and eudicots, and ranks amongst the most convergent complex traits (Sage, Christin and Edwards, 2011; Christin *et al.*, 2013). The differences between C_3 and C_4 plants have been addressed over the past 50 years using comparative anatomy and physiology, and more recently comparative transcriptomics and genomics (Bräutigam *et al.*, 2011, 2014; Dunning *et al.*, 2017; Lauterbach *et al.*, 2017; Moreno-Villena *et al.*, 2018). While the C_4 trait consists primarily in the synchronized action

of multiple enzymes, comparisons of enzymatic properties between C_3 and C_4 plants remain relatively sparse. There is therefore a need to evaluate the enzymatic changes involved in the transition to a C_4 type to understand the evolutionary mechanisms underlying the repeated origins of the C_4 trait.

Calvin-Benson Cycle

During the light-independent phase of photosynthesis, the Calvin-Benson cycle or reductive pentose phosphate cycle, uses the energy fixed during the light-dependent phase to fix of carbon dioxide to produce triose sugars, the pre-requisite for carbohydrates in plant. The key enzyme in the cycle is ribulose-1,5-bisphosphate carboxylase/ oxygenase (RuBisCO). The fixation of carbon dioxide by this enzyme produces phosphoglycerate (3PG), the prerequisite for triose sugars (Calvin, 1962). The abridged cycle catalysed by RuBisCO and other enzymes, is shown in Figure 1. The process requires ATP and NADH which are generated by the light reactions of photosynthesis (Mathis and Paillotin, 1981). The cycle is essential for all higher life on earth and RuBisCO makes up about 50% of soluble leaf protein (Ellis, 1979).

In C_3 species, atmospheric carbon dioxide reaching the photosynthetic cells by diffusion is fixed directly by RuBisCO. Exchange between the cells and the atmosphere is controlled by stomata, leaf pores that can be opened or closed depending on the conditions (Cowan and Troughton, 1971). RuBisCO most likely evolved 2.7-2.9 billion years ago in bacteria, when atmospheric concentrations of oxygen were effectively zero and CO_2 concentrations were several magnitudes higher than today (Nisbet *et al.*, 2007; Nisbet and Nisbet, 2008; Christin and Osborne, 2013). Probably by chance, RuBisCO evolved with a tendency to confuse the O_2 and CO_2 substrates. Both molecules are featureless, the dioxygen molecule and carbon dioxide molecule are very similar in size and shape; these gaseous molecules share a point group D∞h, as well having electron distribution, focused at the two terminal areas. While this did not represent a problem under the high-CO_2 environment in which it evolved, the dual affinity of RuBisCO created a challenge for plants following the oxygenation of Earth's atmosphere and the continued decreases of CO_2 concentrations. In

O_2-rich atmospheres that have prevailed prior to the 30 million years, RuBisCO fixes atmospheric oxygen at a rate that can compete with carbon dioxide fixation in warm, arid and saline environments typically found across the tropics and sub-tropics (Ehleringer and Björkman, 1977; Skillman, 2007). The binding of oxygen generates compounds that have few metabolic uses and become toxic in high concentrations. These compounds are broken down and reincorporated into metabolic cycles using a process called photorespiration (Heber and Krause, 1980).

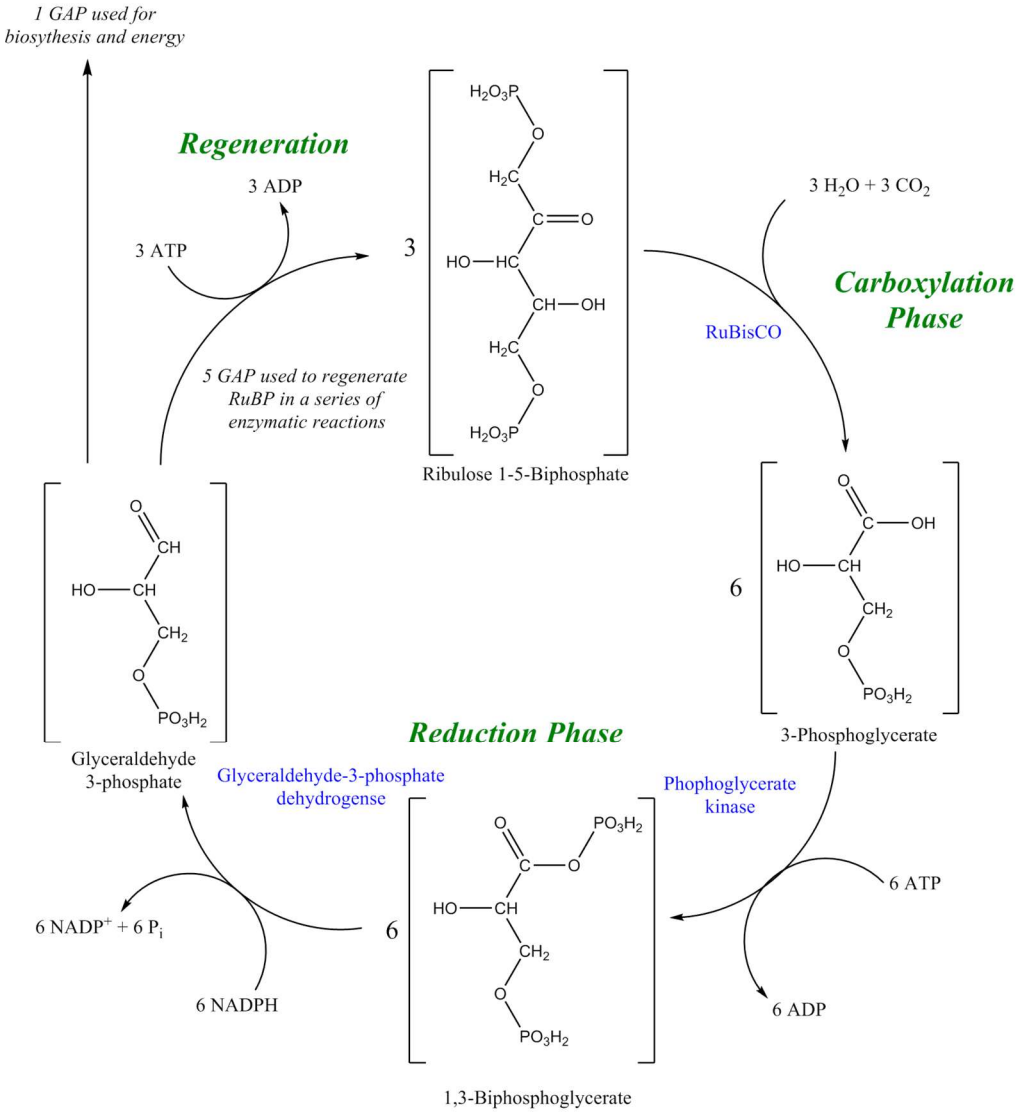


Figure 1: The abridged Calvin cycle indicating the three important steps. The carboxylation of RuBP, the reduction of 3PG and the regeneration of RuBP from 5 molecules of glyceraldehyde-3-phosphate (GAP). For every six molecules of GAP produced from 3 molecules of RuBP, only one molecule is used for biosynthesis and energy storage. The process has a net use of 3 H_2O, 3 CO_2, 3 ATP and 6 NADPH molecules (Calvin, 1962; Mathis and Paillotin, 1981).

Photorespiration

The RuBisCO oxygenase activity net product is one molecule of 3PG and one molecule of 2-phosphoglycerate (2PG) from one molecule of RuBP, as opposed to the RuBisCO carboxylase activity which results in two molecules of 3PG (Heber and Krause, 1980; Ogren, 1984; Wingler *et al.*, 2000). The molecule 2PG cannot be used in the Calvin Benson cycle, so oxygenase activity reaction represents a net loss in carbon for the plant. Further, 2PG has no known metabolic use and may inhibits some photosynthetic enzymes (Anderson, 1971; Kelly and Latzko, 1976). Plant species utilise photorespiration to convert 2GP to 3PG (Ogren, 1984). 3PG is converted into glycolate, and transamination then converts it into glycine. A hydroxyl group is added to form serine, which is then deaminated to form hydroxypyruvate. It is then reduced to glycerate, which is then phosphorylated to regenerate 3-phosphoglycerate (Siedow and Day, 2000). The full cycle is displayed in Figure 2. Some species segregated glycine formation and the release of carbon dioxide in different cells to create a weak carbon dioxide concentrating mechanism, with glycine as the carbon shuttle (Monson and Rawsthorne, 2000). These species, originally called Type I C_3-C_4 intermediate or 'C_2' plants, are often seen as a possible intermediate step for the evolution of C_4 photosynthesis (Rumpho *et al.*, 1984; Moore *et al.*, 1988; Monson *et al.*, 2008; Sage, Sage and Kocacinar, 2012).

Photorespiration primarily results from RuBisCO's inability to effectively distinguish between O_2 and CO_2 (Ehleringer *et al.*, 1991). As mentioned before, the ability to discriminate O_2 and CO_2 was not important in the CO_2-rich environment where RuBisCO evolved, and evolution later evolved versions of the enzyme with increased specificity for CO_2. However, increases of specificity come at the expense of the turnover rate of the enzyme, so that more specific RuBisCO are slower (Tcherkez, Farquhar and Andrews, 2006). Due to the essential role RuBisCO plays in the metabolism of plants, any reduction in enzyme velocity would have a catastrophic effect on the plant. RuBisCO evolved at the during the Archean period, when atmospheric carbon dioxide was considerably higher, and oxygen much lower (Nisbet and Nisbet, 2008; Christin and Osborne, 2013).

While the relative CO_2:O_2 concentration depends on the atmosphere composition, the ratio of these gases within the plant depend on further factors. Firstly, temperature affects the CO_2 available to RuBisCO. At equilibrium at 25°C there is 500-fold more oxygen than carbon dioxide dissolved in water (Griffiths, 2006), and CO_2 solubility decreases faster than O_2 solubility with an increase in temperature (Ku and Edwards, 1977). Combined with decreases of RuBisCO specificity at higher temperatures, the balance between the competing pathways shifts more towards the oxygenase activity in warm conditions (Ehleringer and Björkman, 1977). The relationship between substrate availability and specificity is described in Equation 1. Hence C_3 plants in warm environments suffer from oxygenation of RuBP. In addition, the internal CO_2:O_2 ratio depends on the rate of exchange with the atmosphere, and therefore on the level of stomata aperture. Stomatal closure, which can be forced by aridity or salinity, leads to CO_2 depletion within the leaf, so that photorespiration is exacerbated by aridity/salinity. Counterintuitively, CO_2 is also depleted in warm aquatic environments. In all these environments, increases of RuBisCO specificity were not sufficient to avoid high levels of photorespiration, and some lineages evolved elaborate mechanisms to increase the relative concentration of CO_2 before its fixation by RuBisCO. Of these mechanisms, C_4 photosynthesis is present in many terrestrial and aquatic flowering plants (Salvucci and Bowes, 1981, 1983; Sage, 2004; Sage, Christin and Edwards, 2011).

Equation 1: Ratio of carboxylase activity to oxygenase activity. V_c and V_o are the maximum velocities (V_{max}) for the respective reactions and K_c and K_o are the Michaelis-Menten constants. (Ogren, 1984).

$$\frac{v_c}{v_o} = \left(\frac{V_C}{K_C}\right) \times \left(\frac{K_0}{V_0}\right) \times \left(\frac{[CO_2]}{[O_2]}\right)$$

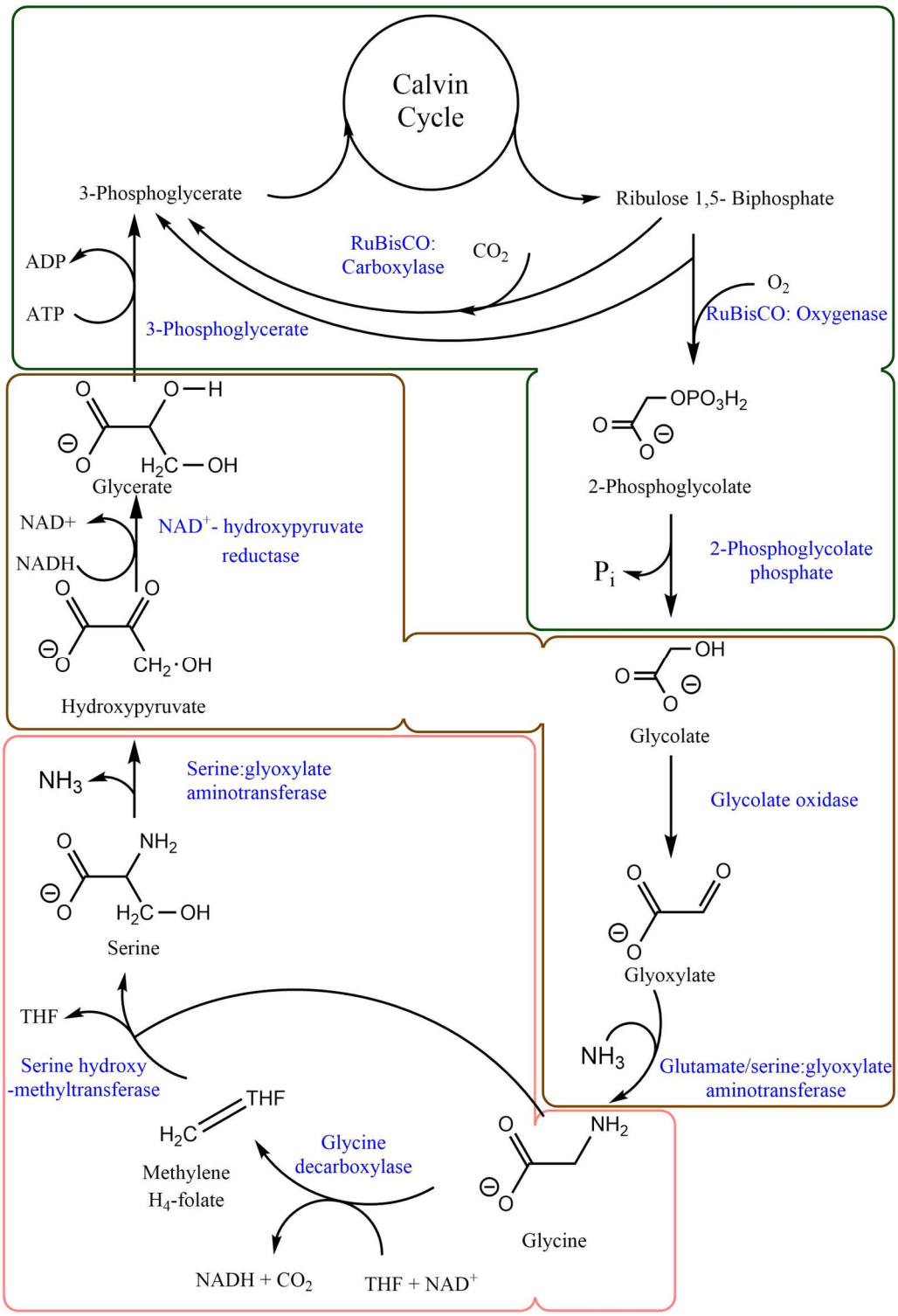


Figure 2: The full photorespiration cycle, responsible for converting 2-phosphoglycoate into 3PG. Displayed is the organelle compartmentalisation of full cycle, green indicated chloroplasts, brown peroxisomes; and pink mitochondria. Indicated in blue are the key enzymes for photorespiration (Ogren, 1984).

C_4 Photosynthesis

C_4 photosynthesis requires a complex reorganisation of leaf anatomy, metabolism, gene expression and change in enzyme kinetics that together create a carbon concentration mechanism. By increasing the relative CO_2:O_2 ratio around RuBisCO, C_4 photosynthesis almost completely suppresses the enzyme oxygenase activity and therefore photorespiration (Hatch, 1987; Sage, Sage and Kocacinar, 2012). It provides an advantage in all conditions that promote photorespiration, and C_4 plants are especially abundant in warm habitats, but also arid, saline and aquatic environments. Phylogenetic analyses indicate that the C_4 physiology evolved more than 60 times independently in flowering plants (Sage, Christin and Edwards, 2011). All these origins are clustered in the last 30 million years, which coincides with low CO_2 levels in the atmosphere that likely made photorespiration significant in some of Earth environment, thereby providing the selective impetus for C_4 evolution.

As in C_3 plants, CO_2 reaches the cells of C_4 plants via diffusion through the stomata. In C_4 plants however, RuBisCO is absent in the cells with direct contact to the atmosphere, and the atmospheric carbon is fixed as bicarbonate into oxaloacetic acid (OAA) by PEPC. OAA is rapidly converted into a four-carbon acid, usually malate or aspartate (Bräutigam *et al.*, 2014), and moves to another compartment where RuBisCO is segregated, generally the bundle sheath cells (Figure 3). This transfer occurs via plasmodesmata – active-transport channels that traverse the cell walls (Weiner *et al.*, 1988). The C_4 acid is then decarboxylated in the second compartment, releasing carbon dioxide around the RuBisCO enzyme. This process creates a concentration of CO_2 ten times higher than in the atmosphere, which increases the efficiency of the Calvin-Benson cycle (von Caemmerer and Furbank, 2003).

To facilitate the carbon concentrating mechanism, in its classical version, the anatomy of the leaf changes to a wreath like structure, often termed Kranz anatomy (Brown, 1975; Hatch, 1987; Sage, 2004; Sage, Sage and Kocacinar, 2012). In C_3 species, the majority of chloroplasts are located in the mesophyll cell. The bundle sheath cells surround leaf veins

and facilitate the transport of metabolites to the plant (Leegood, 2007). In C_4 plants, the chloroplasts move from the mesophyll to the bundle sheath cell which are enlarged to accommodate them. The intervein distance decreases to accommodate more bundle sheath cells (McKown and Dengler, 2010).

While the segregation of carbon fixation by PEPC and its reduction by the Calvin-Benson cycle usually occurs among distinct cells, mono-cellular C_4 carbon concentrating mechanisms have been observed in aquatic and terrestrial plants (Freitag and Stichler, 2000; Voznesenskaya, Franceschi and Kiirats, 2001; von Caemmerer *et al.*, 2014). In the desert species *Bienertia sinusperisci,* plant cells have two types of chloroplast that function analogous to mesophyll and bundle sheath cells (Offermann, Okita and Edwards, 2011). In the aquatic single cell C_4 species *Hydrilla verticillate*, PEPC and RuBisCO are specially separated at the extremities of the cell, separated by the vacuole (Edwards, Franceschi and Voznesenskaya, 2004).

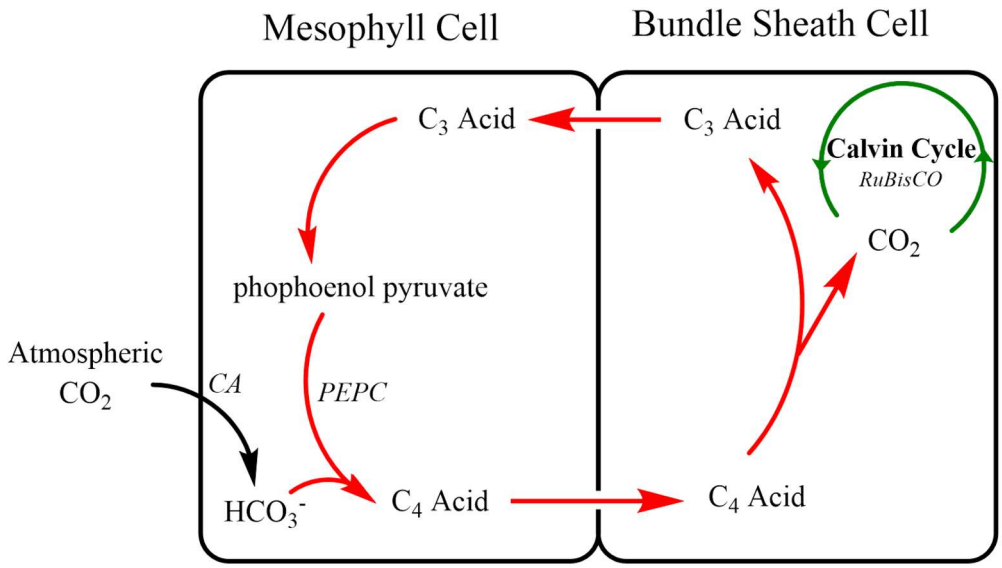


Figure 3: Simplified C_4 carbon concentrating mechanism. CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase (Hatch, 1987; Sage, 2004).

While some aspects of the CO_2-concentrating mechanism are shared by all C_4 plants, the exact components used to achieve them vary among C_4 species. Firstly, C_4 photosynthesis can be performed in leaves or in stems and using different cell types within the leaf (Lundgren, Osborne and Christin, 2014). Secondly, the details of the biochemical pathway

also vary. While it is now recognized that different C_4 pathways represent a gradient of involvement of various enzymes (Furbank, 2011; Wang *et al.*, 2014), three subtypes of C_4 photosynthesis have been classically defined according to the decarboxylating enzyme. In all subtypes, OAA is produced by the coupled action of CA and PEPC in the cytosol of the mesophyll cells. The typical C_4 subtypes are described below, together with their energetic cost. For comparison, the cost of fixing one molecule of carbon dioxide is 3 ATP and 2 NADH in C_3 photosynthesis (Kanai and Edwards, 1999).

NADP-ME Subtype

In the NADP-ME subtype, OAA is converted to malate by NADP-malate dehydrogenase (NADP-MDH) in the mesophyll cytosol, and malate is then transported into the bundle sheath cell. Malate is then decarboxylated by NADP-malic enzyme (NADP-ME) in the chloroplast of the bundle sheath cell, which produces pyruvate in addition to CO_2. The pyruvate is then transported back to the mesophyll cell where it is converted into phosphoenolpyruvate by pyruvate orthophosphate dikinase. Each carbon dioxide molecule fixed costs 5 ATP and 2 NADPH in the NADP-ME cycle (Hatch and Slack, 1966; Kanai *et al.*, 1999; Malkin, R., 2000; Furbank, 2011; Ren *et al.*, 2014; Figure 4A).

NAD-ME Subtype

In the NAD-ME subtype, OAA is reduced to aspartate by aspartate-transaminase in the cytosol of the mesophyll cell. The aspartate is then transported into the bundle sheath cell where it is converted back into OAA by aspartate-transaminase. OAA is then converted into malate by NADP-MDH in the chloroplast of the bundle sheath cell. The CO_2 is released from malate by the NAD-malic enzyme (NAD-ME) in the mitochondria and pyruvate is produced. Alanine-transaminase then converts the pyruvate into alanine, which is transported back into the mesophyll cell. Alanine-transaminase then converts the alanine back into pyruvate which is then phosphorylated by pyruvate orthophosphate dikinase to regenerate PEP. Conversion to alanine and aspartate acts as a nitrogen shuttle. Each carbon dioxide molecule fixed costs 5 ATP and 2 NADPH in the NAD-ME cycle (Kanai *et al.*, 1999; Furbank, 2011; Wang *et al.*, 2014).

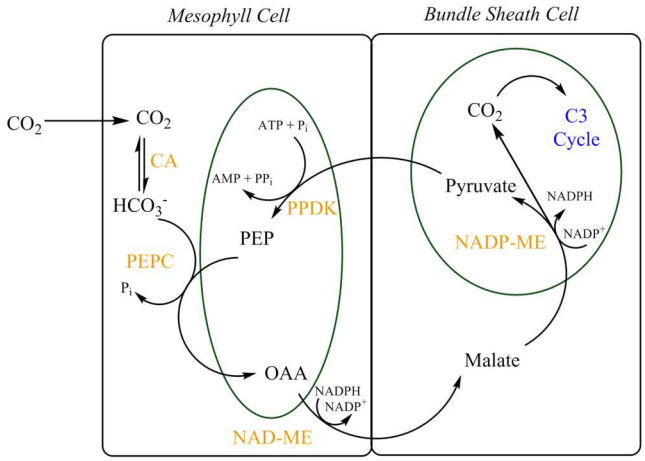
PEPCK Subtype

In the PEPCK cycle OAA is converted into aspartate by aspartate-transaminase in the cytosol of the mesophyll cell and then transported into the bundle sheath cells. It is then converted back into OAA by aspartate aminotransferase and decarboxylated by phosphoenolpyruvate carboxykinase (PEPCK). This also produces PEP which is either transported into the mesophyll cell or converted into alanine by alanine amino transferase. Alanine is transported and then converted back to PEP in the mesophyll cell by alanine amino transferase. Each CO_2 molecule fixed is estimated to cost a minimum of 3.5 ATP and 2.25 NADPH per fixed carbon dioxide molecule (Kanai *et al.*, 1999; Furbank, 2011; Wang *et al.*, 2014).

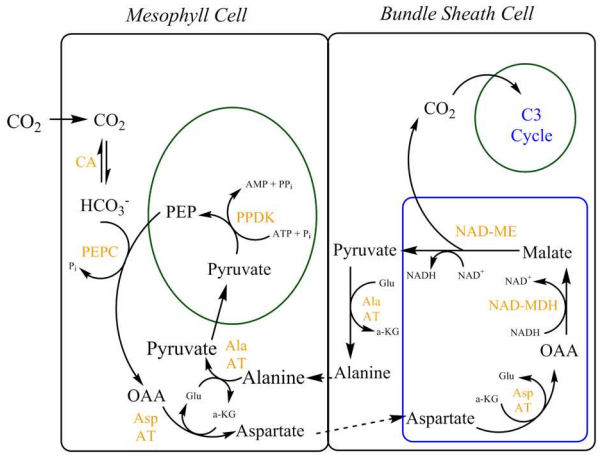
The distinctions between the cycles are not clear cut, and many C_4 species combine multiple C_4 sub-cycles. A pure PEPCK-type has not been observed, presumably due to the imbalance in amino groups in shuttling between the two cells (Weber and Bräutigam, 2013). Species are considered to have a mix of PEPCK and NADP-ME or a mix of PEPCK and NAD-ME (Furbank, 2011; Wang *et al.*, 2014).

All enzymes of the C_4 pathway also exist in C_3 plants, where they are responsible for different non-photosynthetic functions (Aubry, Brown and Hibberd, 2011; Christin *et al.*, 2013). The evolution of C_4 photosynthesis requires their co-option for the C_4 pathway, which often involved their upregulation. In some cases, the co-option of enzymes was followed by adaptation of their kinetics. The possibility to co-opt different enzymes means that the kinetic modifications are likely to vary among subtypes. The one step that is shared by all C_4 plants is the primary carbon fixation enzyme PEPC, which defines the C_4 trait (Hatch, 1987). Understanding the properties that are required for C_4-specific PEPC is essential to predict how to engineer the biochemistry of C_4 photosynthesis.

A. NADP-ME



B. NAD-ME



C. PEPCK

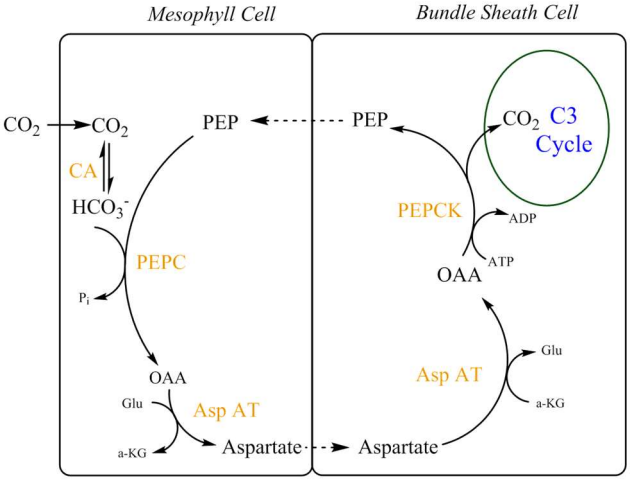


Figure 4: Summary of the subtypes of C_4 photosynthesis carbon concentrating mechanism. Metabolites indicated in black, enzymes indicated in orange, green boxes indicating mesophyll, blue boxes indicating mitochondria. CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; MDH, malate dehydrogenase; NADP-ME, NADP malic enzyme; PPDK, pyruvate phosphate dikinase; AspAT, aspartate amino transferase NADP-MDH, NADP malate dehydrogenase; AlaAT, alanine amino transferase; PEPCK phosphoenolpyruvate carboxykinase. OAA, oxaloacetate; PEP, phosphoenolpyruvate; Glu, glutamate; a-KG, α-ketoglutarate. Diagrams adapted from (Furbank, 2011; Wang *et al.*, 2014).

Phosphoenolpyruvate Carboxylase Enzyme Overview

Phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) catalyses the carboxylation of phosphoenolpyruvate (PEP) with the substrate bicarbonate. The reaction produces one molecule of oxaloacetic acid (OAA) and a phosphate ion, and is highly exergonic with a $\Delta G^{\circ\prime}$ of -30 $kJmol^{-1}$, a full reaction mechanism is shown in Figure 5 (O'Leary, 1982; Chollet, Vidal and O'Leary, 1996; Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004). Essential for activity of the enzyme is a divalent metal ion, magnesium (Mg^{2+}) in plants and manganese (Mn^{2+}) in bacteria (Chollet, Vidal and O'Leary, 1996). PEPC is present in all photosynthetic organisms, including plants, cyanobacteria, algae, archaea as well as most bacteria, however is absent in animals and fungi, including yeast (Ettema *et al.*, 2004; O'Leary, Park and Plaxton, 2011).

PEPC plays a housekeeping role in plants, bacteria and algae regenerating the oxaloacetate for key metabolic pathways, including the Krebs cycle (Day and Hanson, 1977; Edwards *et al.*, 1998; O'Leary, Park and Plaxton, 2011). Other roles in plants include carbon storage, cell expansion, seed germination and development, energy supply, and stress acclimation (Sangwan, Singh and Plaxton, 1992; Delgado *et al.*, 1993; Osuna *et al.*, 1996; Dobrota, 2006) . The enzyme also plays a key role in carbon fixation in the C_4 and CAM carbon concentrating mechanism. Plant PEPCs belong to a small multigene family encoding several 'plant-type' PEPCs and at least one distantly related 'bacterial-type' PEPC. The latter have more structural similarities to the *E. coli* PEPC (Sánchez and Cejudo, 2003; O'Leary *et al.*, 2009; Gowik and Westhoff, 2011; O'Leary, Park and Plaxton, 2011).

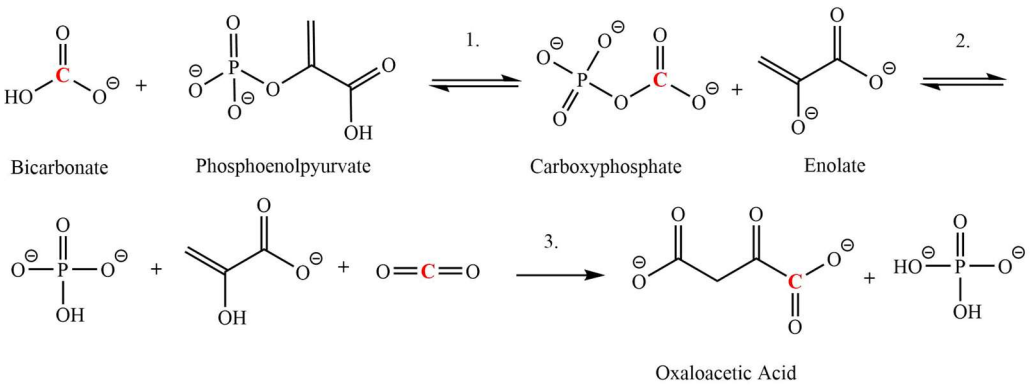


Figure 5: Reaction catalysed by the enzyme PEPC. PEPC catalyses the carboxylation of one molecule of phosphoenolpyruvate to produce one molecule of oxaloacetic acid and a phosphate ion. The atmospheric carbon dioxide carbon highlighted in red (González and Andreo, 1989; Kai, Matsumura and Izui, 2003).

Detailed studies have suggested that the reaction of PEPC proceeds through a three-step mechanism (O'Leary, 1982; Chollet, Vidal and O'Leary, 1996). First, the carboxyphosphate and enolate of pyruvate are formed by a reversible reaction. The enolate isomerises and the carboxyphosphate cleaves into phosphate ion (P_i) and CO_2. Finally, CO_2 makes an electrophilic attack on the enolate to form oxaloacetic acid. The full proposed reaction mechanism is presented in Figure 6 (Janc, O'Leary and Cleland, 1992; Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004).

From crystal structures of the *E. coli* and *Zea mays* C_4 PEPC it was shown the enzyme is tetrameric. Each monomer is composed of 42 α-helices arranged around an 8-stranded β -barrel or 'TIM' like structure with no clear subdomains in the structure, *Escherichia coli* PEPC has an overall size of 130 × 120 × 70 Å (Kai *et al.*, 1999; Kai, Matsumura and Izui, 2003). The tetramer is arranged in a dimer of dimers. In the *Zea mays* PEPC the dimer interface has a surface area 3000 $Å^2$ and the dimer- dimer interface has a surface area of 450 $Å^2$ (Kai, Matsumura and Izui, 2003). A representation of the tetrameric and monomeric structures of PEPC are shown in Figure 7.

The *E. coli* form of PEPC has been crystallised in the presence of the inhibitor aspartate and the *Zea mays* PEPC was crystallised in the absence of inhibitor. When the monomers were superimposed, the surface of the *Z. mays* PEPC is rotated 10° clockwise with respect to the *E. coli* PEPC. This suggests a T/R conformation change, and allosteric

inhibition is mediated by large structural movement between states, the active state being associated with the R conformation. This T/R transition appears to result in the active site opening (Matsumura *et al.*, 1999, 2002; Kai, Matsumura and Izui, 2003).

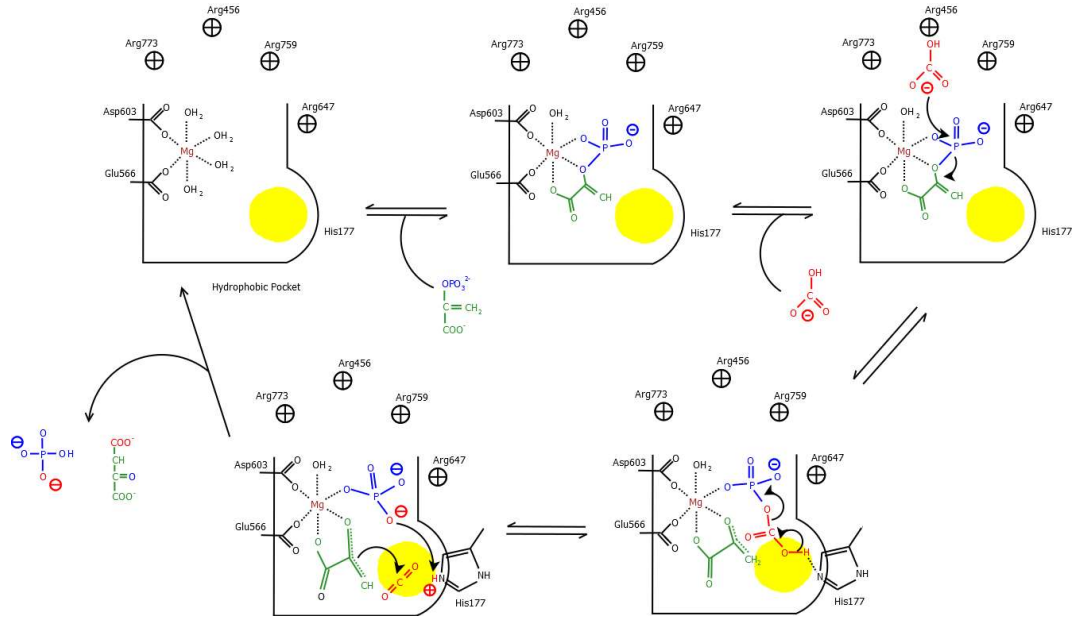


Figure 6: Proposed reaction mechanism of the carboxylation of phosphoenolpyruvate by the enzyme PEPC. The active site magnesium ion is indicated in brown, the phosphate ion (P_i) in blue, the carbon component of the PEP molecule in green and the bicarbonate ion in red. In step one of the proposed mechanisms the magnesium – PEP complex is bound in the active site. In the second step the bicarbonate enters the active site, the ion is stabilized by the positive residues in the 'hinged lid' structure of the enzyme. The bicarbonate ion attacks the PEP molecule, forming an enolate complexed to the magnesium ion and a carboxy-phosphate intermediate. The intermediate is stabilized by the hydrophobic region of the active site, indicated in yellow. The catalytic histidine attacks the carboxy-phosphate intermediate and produces a phosphate ion and carbon dioxide. The enolate attacks the carbon dioxide, forming oxaloacetate. The oxaloacetate and P_i are released (Janc, O'Leary and Cleland, 1992; Kai, Matsumura and Izui, 2003).

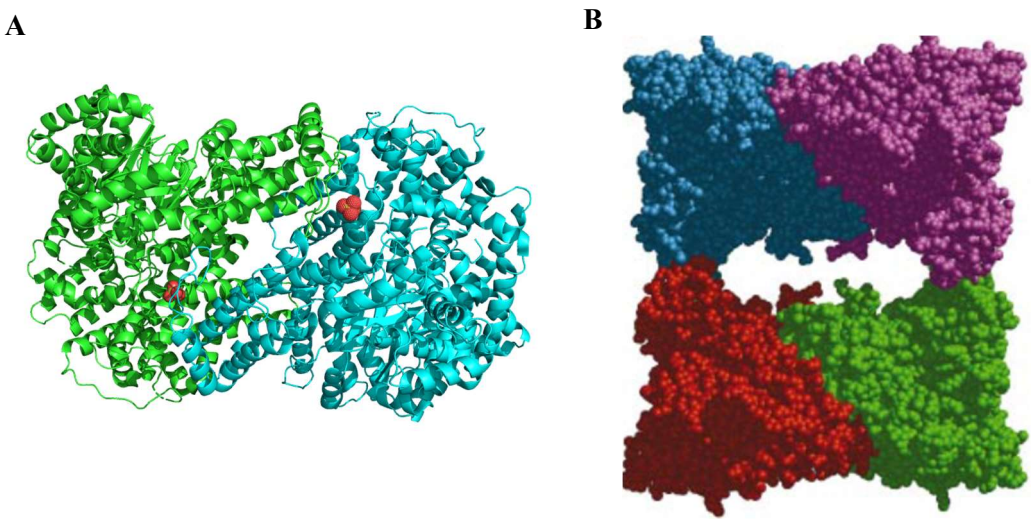


Figure 7: Structures of the PEPC enzyme. A Cartoon of dimer structure of *Zea mays* PEPC, monomers coloured in green and blue, with sulfate moiety bound to each monomer indicated in pink (Matsumura *et al.*, 2002). **B** Tetrameric structure of *E. coli* PEPC, each monomer is indicated by a different colour. (Matsumura *et al.*, 2002).

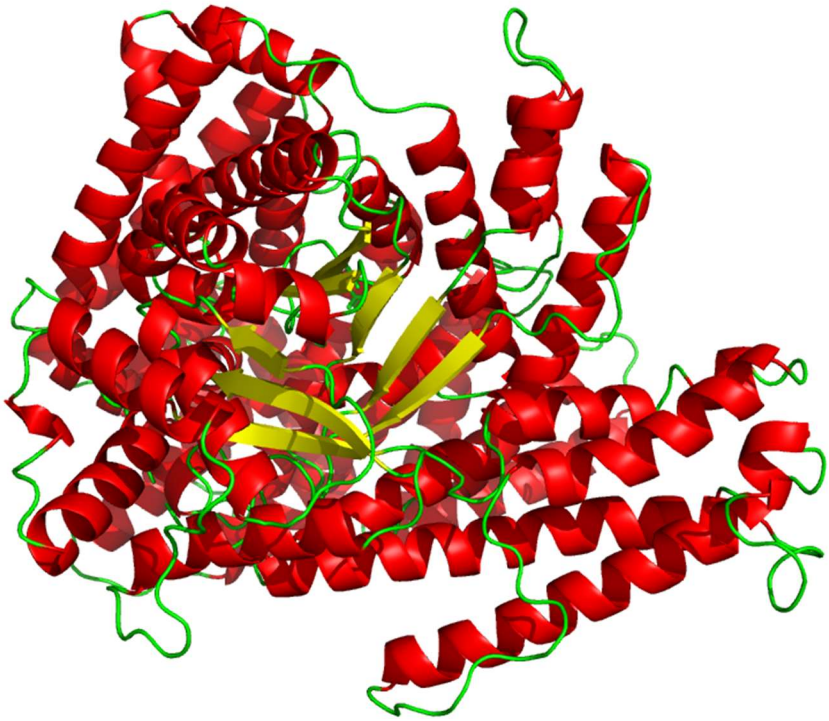


Figure 8: Monomer structure of PEPC. α-helices indicated in red, β-sheets indicated in yellow, and mobile loops indicated in green. The structure is orientated with the active site, at the top of the β barrel, is at the front (Matsumura *et al.*, 1999).

All forms of the enzyme that have been explored have been inhibited by malic acid and aspartic acid (Izui *et al.*, 1981; Matsumura *et al.*, 1999; Paulus, Schlieper and Groth, 2013). These four carbon acids are products of reactions with oxaloacetic acid in the C_4 cycle as well as other metabolic cycles and are thus feedback inhibitors. PEPC is also subject to activation. Bacterial and plant forms of the enzyme have been shown to be activated by sugar phosphates such as glucose-6-phosphate (Doncaster and Leegood, 1987; Wedding, Black and Meyer, 1989; Woo and Xu, 1996; Tovar-Méndez, Mújica-Jiménez and Muñoz-Clares, 2000), and monocot C_4 PEPCs, such as *Z. mays*, have been shown to be further activated by neutral side chain amino acids such as glycine and alanine (Wong and Davies, 1973; González-Segura *et al.*, 2018). The *E. coli* form is also activated by acetyl-CoA, long-chain fatty acids and guanosine 3'5'-bisphosphate (Izui *et al.*, 1981).

The activity of C_4 and CAM carbon fixing PEPCs are modulated by phosphorylation. PEPC is phosphorylated in the day in C_4 plants, and night in CAM plants which is when most of the carbon fixation occurs for the respective mechanism (Lepiniec *et al.*, 1994; Nimmo, 2000, 2003). PEPC is phosphorylated by the specific kinase, phosphoenolpyruvate carboxylase kinase (PEPCK; Nimmo *et al.*, 1987). The N terminal S15 (*Zea mays* numbering, accession number NM_001161348.2) has been identified as the major phosphorylation point, however secondary phosphorylation positions have been identified (Chollet, Vidal and O'Leary, 1996; Nimmo, 2000). Phosphorylated PEPC shows increased specificity for PEP without changing k_{cat}, decreased sensitivity to inhibitors and increased sensitivity to glucose-6-phosphate (Doncaster and Leegood, 1987; Nimmo *et al.*, 1987; Echevarria *et al.*, 1994; Duff *et al.*, 1995; Chollet, Vidal and O'Leary, 1996; Nimmo, 2003). PEPC kinase expression is controlled by the circadian rhythm, being upregulated in C_4 plants in daylight (Hartwell *et al.*, 1999; Leegood and Walker, 2003). PEPC is dephosphorylated by protein phosphatase 2A (Carter *et al.*, 1990).

Functional residues of PEPC

The active site of PEPC is located near the C terminal part of the protein, at the end of the β5 and β6 strand of the TIM barrel. The metal cofactor was present in this location in all resolved structures (Matsumura *et al.*, 1999, 2002; Paulus, Schlieper and Groth, 2013; Schlieper *et al.*, 2014; González-Segura *et al.*, 2018). The metal is bound to the oxygen atoms of the carboxyl groups of the conserved E506 and D543 (*Z. mays* numbering). The configuration is like other PEP utilizing enzymes such as pyruvate kinase (Muirhead *et al.*, 1986). R773, R456, R759 and R647 form a positive pocket which is thought to stabilise the electron rich bicarbonate ion; R647 is essential for catalytic activity. This positive pocket is above the metal ion. H177 is an essential catalytic base (Terada and Izui, 1991), it plays a role in stabilising the carboxyphosphate intermediate by removing a proton from the carboxyl group (Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004).

A loop of amino acids covers the face of the β-barrel with the sequence $K^{762}RRPGGG^{768}$ (Matsumura *et al.*, 1999). Mutations at K762, R763 and R764 affected the enzyme's ability to interact with bicarbonate. These mutations increase the side reactions of PEPC that result in pyruvate and orthophosphate (Terada and Izui, 1991). This suggests this loop acts as somewhat of a 'hinged lid' above the catalytic site that interacts with the bicarbonate ion. Hinged lid motifs have been observed in other proteins and act to prevent water interfering with catalysis (Sun and Sampson, 1999). A hydrophobic pocket has been identified in the active site composed of W248, L504 and M253 (Matsumura *et al.*, 1999). A putative reaction mechanism illustrating the roles of the active site amino acids has been described and can be seen in Figure 6 (Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004).

Crystal structures of PEPC bound to the inhibitor aspartic acid have determined the site of allosteric inhibition (Matsumura *et al.*, 1999; Paulus, Schlieper and Groth, 2013). No structures with malic acid have been resolved as the molecule has been observed to inhibit PEPC crystallisation. The binding site for aspartic acid was located *ca.* 20 Å from the active site. Aspartate directly interacts with hydrogen bonding to the amino acids R647, K835,

R894, N968 and R878 (Matsumura *et al.*, 1999; Paulus, Niehus and Groth, 2013). Mutations at K835 and R894 cause a marked desensitization to both inhibitors (Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004).

The amino acid R647 is located in the aspartate inhibition site of PEPC but the guanidino group of the sidechain is also though to interact with PEP in the active site of the *Z mays* PEPC (Matsumura *et al.*, 2002). This residue at position 647 is essential for activity (Yano *et al.*, 1995). The loop on which the site occurs is thought to be mobile and held away from the active site on inhibitor binding as a component of the inhibition mechanism (Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004).

The structure of Matsumura *et al.*, 2002 *Z. mays* C_4-PEPC is thought to be in the activated form due to the presence of ethylene glycol in the crystallisation of the enzyme. A sulfate group is bound 15 Å from the active site, in a hydrophobic pocket, a region that is large enough to accommodate a glucosyl moiety. The positively charged arginine residues in this region are highly conserved and replacement of these residues with glutamine results in desensitisation to glucose-6-phosphate (G6P) (Takahashi-Terada *et al.*, 2005). However, these positions are on the dimerization interface and mutations here may have a great effect destabilising the enzyme. Structures have been resolved of the *Flaveria trinervia* PEPC bound to G6P in the active site, this structure also has a sulfate moiety bound ion the same location as the *Z. mays* PEPC suggesting this not an activation site (Schlieper *et al.*, 2014). This indicates that G6P acts as a competitive activator and interacts with W283, R450, M592, D767 and R768 (*Flaveria* numbering). Ethylene glycol was also found to bind in the same location in the *Flaveria trinervia* PEPC structure (Schlieper *et al.*, 2014).

Phosphorylation occurs at a conserved serine residue near the N-terminus, S15 (*Z. mays* numbering). The surrounding region is highly conserved. The N terminal region may change conformation on phosphorylation, however, the structure of this region has not been visualised in any crystal structure. Mutants insensitive to glucose-6-phosphate no longer respond to phosphorylation suggesting that the R residues involved in glucose-6-phosphate binding are also necessary for activation by phosphorylation (Kai, Matsumura and Izui, 2003; Izui *et al.*, 2004).

Evolution of C_4 PEPC

The gene encoding the C_4 isoform of PEPC evolved from a non-photosynthetic PEPC gene. The role played by C_4 PEPC is drastically different to that of its ancestor, as the C_4 cycle is characterized by high concentrations of metabolites and high fluxes (Stitt and Zhu, 2014). Expression data has clearly shown that the expression level of PEPC has been massively increased during the evolution of C_4 photosynthesis (Bräutigam *et al.*, 2011, 2014; Lauterbach *et al.*, 2017; Moreno-Villena *et al.*, 2018). Changes to the enzymes are less well understood, but fundamental comparative work has revealed C_4-specific PEPC properties within some taxonomic groups.

Flaveria PEPC

The genus *Flaveria* has been used as a model system to observe the adaptation of the kinetic properties of C_4 PEPC (Svensson, Bläsing and Westhoff, 1997; Engelmann *et al.*, 2003). The genus contains closely related C_4, intermediate and C_3 species as well as several species that have been classed as intermediate photosynthetic types (Powell, 1978; McKown, Moncalvo and Dengler, 2005). The species *Flaveria trinervia*, which is C_4, expresses a C_4 PEPC (*Ft* PEPC) in the leaf tissue. The C_3 congener *Flaveria pringlei* expresses an orthologous non-C_4 PEPC (*Fp* PEPC), which is thought to be similar to the ancestor of the *Flaveria trinervia* C_4 PEPC. The *Ft* and *Fp* PEPC display stark differences in kinetic properties (Figure 11). However, these two proteins display a 96% amino acid sequence similarity and have a broad similarity in 3D structures, Figure 10 (Paulus, Schlieper and Groth, 2013).

Ft PEPC displays a K_m^{PEP} an order of 10 larger than *Fp* PEPC, suggesting that the C_4 PEPC has 10 times decrease in specificity for PEP (Svensson, Bläsing and Westhoff, 1997). The reason for this decrease is unknown, but it has been suggested that might represent a sacrifice made to increase specificity for bicarbonate (Jacobs *et al.*, 2008; Gowik and Westhoff, 2011). Analysis of chimeric *Ft*/*Fp* mutants determined that two regions of the C_4 protein, the first region from position 296 to 437 and the second from 645 to 966, confer a decrease in specificity for PEP (Blasing, Westhoff and Svensson, 2000). Within the second

region, position 774 (780 in *Z. mays* numbering) was determined to be the key C_4 determinant. Indeed, when the C_4 site was substituted into the *Fp* enzyme, the K_m (PEP) significantly increased. Position 774 is located above and very close (within 20 Å) to the active site in the crystal structures (Matsumura *et al.*, 1999).

The C_4 *Ft* PEPC exhibits a decreased sensitivity to the inhibitor malic acid compared to *Fp* PEPC. The IC_{50} for malic acid of the C_4 form is three times higher than *Fp* PEPC (Svensson, Bläsing and Westhoff, 1997). C_4 PEPC is subject to much greater concentrations of feedback inhibitor than the non-orthologue as the C_4 cycle generates high concentrations of its metabolites. The inhibition sensitivity is therefore likely to decrease in the C_4-specific forms of PEPC to allow high activity despite the presence of high concentrations of inhibitors. Analyses of chimeric proteins identified three regions of importance for malate inhibition: the regions encompassing positions 1 to 296, positions 297 to 437, and positions 646 to 966. The latter was the strongest determinant of inhibitor sensitivity (Jacobs *et al.*, 2008). While two of these regions have been observed to be important in the specificity for PEP, investigation of position 774 mutants show no change in sensitivity, suggesting the decrease in specificity for PEP is unrelated to the decrease in sensitivity for malate (Jacobs *et al.*, 2008). Crystal structures were determined for both *Ft* and *Fp* PEPC with a molecule of aspartate bound. It was shown that position 884 (890 *Z. mays* numbering), which is a glycine in the C_4 *Ft* PEPC and an arginine in *Fp* PEPC, reduces the number of hydrogen bonding interactions with aspartate thus reducing the enzyme sensitivity to inhibitors (Paulus, Schlieper and Groth, 2013; Schlieper *et al.*, 2014).

With respect to activation, *Ft* PEPC also displays greater sensitivity to G6P compared to *Fp* PEPC (Svensson, Bläsing and Westhoff, 1997). The C_4 enzyme shows a 5-fold activation, where the non C_4 enzyme shows a 1.5-fold activation (Blasing, Westhoff and Svensson, 2000).

```
Fp (C3)   1    MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQEC  60
Ft (C4)        MANRNVEKLASIDAQLRLLVPGKVSEDDKLVEYDALLLDKFLDILQDLHGEDLKEAVQQC

Fp (C3)   61   YELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAKAFSHMLNLANLAEEVQIAYRRRIK  120
Ft (C4)        YELSAEYEGKHDPKKLEELGSLLTSLDTGDSIVIAKAFSHMLNLANLAEELQIAYRRRIK

Fp (C3)   121  LKRGDFADEANATTESDIEETFKKLVLKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRS  180
Ft (C4)        LKSGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVELVLTAHPTQSVRRS

Fp (C3)   181  LLQKHGRIRNCLAQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGM  240
Ft (C4)        LLQKHGRIRNCLAQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGM

Fp (C3)   241  SYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDRDGNPRVTPEVTR  300
Ft (C4)   241  SYFHETIWKGVPKFLRRVDTALKNIGINERFPYNAPLIQFSSWMGGDRDGNPRVTPEVTR

Fp (C3)   301  DVCLLARMMASNMYFSQIEDLMFEMSMWRCNSELRVRAEELYRTARRDVKHYIEFWKQVP  360
Ft (C4)   301  DVCLLARMMTSNMYFSQIEDLMIEMSMWRCNSELRVRAEELYRTARKDVKHYIEFWKRIP

Fp (C3)   361  PTEPYRVILGDVRDKLYNTRERSRHLLAHGISDIPEEAVYTNVEQFLEPLELCYRSLCDC  420
Ft (C4)   361  PNQPYRVILGDVRDKLYNTRERSRHLLVDGKSDIPDEAVYTNVEQLLEPLELCYRSLCDC

Fp (C3)   421  GDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR  480
Ft (C4)   421  GDHVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTEVLDAITQHLGIGSYREWSEEKR

Fp (C3)   481  QEWLLAELSGKRPLFGSDLPKTEEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAV  540
Ft (C4)   481  QEWLLAELSGKRPLIGPDLPKTEEVKDCLDTFKVLAELPSDCFGAYIISMATSTSDVLAV  540

Fp (C3)   541  ELLQRECHVKHPLRVVPLFEKLADLEAAPAAMARLFSIDWYRNRIDGKQEVMIGYSDSGK  600
Ft (C4)   541  ELLQREYHIKHPLRVVPLFEKLADLEAAPAAMTRLFSMDWYRNRIDGKQEVMIGYSDSGK  600

Fp (C3)   601  DAGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTIHG  660
Ft (C4)   601  DAGRFSAAWQLYKTQEQIVKIAKEFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTING  660

Fp (C3)   661  SLRVTVQGEVIEQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATE  720
Ft (C4)   661  SLRVTVQGEVIEQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATE  720

Fp (C3)   721  EYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRAIPWIFAWTQTRF  780
Ft (C4)   721  EYRSVVFKEPRFVEYFRLATPELEFGRMNIGSRPSKRKPSGGIESLRAIPWIFSWTQTRF  780

Fp (C3)   781  HLPVWLGFGAAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDK  840
Ft (C4)   781  HLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDK  840

Fp (C3)   841  LLVSEDLWPFGESLRANYEETKDYLLKIAGHRDLLEGDPYLKQRIRLRDSYITTLNVCQA  900
Ft (C4)   841  LLVSEDLRPFGESLRANYEETKNYLLKIAGHKDLLEGDPYLKQGIRLRDPYITTLNVCQA  900

Fp (C3)   901  YTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA  960
Ft (C4)   901  YTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA  960

Fp (C3)   961  GMQNTG  966
Ft (C4)   961  GMQNTG  966
```

Figure 9: Amino acid alignment of C4 and Non-C4 PEPCs showing key C4 amino acids positions. Amino acids that are involved in essential catalytic activity are coloured yellow. Amino acids key for tetramer formation are coloured red. Amino acids key for inhibition are coloured grey. The phosphorylation of site is coloured in pink. C4 specific sites are indicated in green. Adapted from (Kai, Matsumura and Izui, 2003; Jacobs *et al.*, 2008).

Grass PEPC

Flaveria is one of the most recent C_4 origins. Fully C_4 species in the group emerged 1 to 2 million years ago (Christin *et al.*, 2011).). The genus *Alternanthera* represents an older C_4 origin in a different group of eudicots (Christin *et al.*, 2011). Interestingly, some of the same changes in K_m^{PEP} were observed on the C_4 PEPC of *Flaveria* and *Alternanthera* (Figure 11; Gowik *et al.*, 2006). Comparisons of *Flaveria* and *Alternanthera* suggests that the biochemical evolution of C_4 PEPC is a convergent process, but most of the C_4 lineages have not been considered with the same amount of details. Most of the C_4 species belong to monocots, including sedges and mainly grasses (Sage, Christin and Edwards, 2011). The grass family alone includes 60% of all C_4 species, which are clustered in 22-24 independent C_4 lineages (Grass Phylogeny Working Group II, 2012). Some of these C_4 origins rank among the earliest, having occurred between 15 and 35 million years ago (Christin *et al.*, 2008). C_4 grasses dominate most open biomes in tropical and subtropical conditions (Cerling *et al.*, 1997; Sage, 2004; Beerling and Osborne, 2006; Osborne and Freckleton, 2009). They rank among the most productive plants, and include important crops, such as maize, sugarcane, and sorghum.

Kinetic investigations of the adaptation of C_4 PEPC in monocots have been limited. Work on the kinetics of *Panicum* PEPCs has shown trends broadly similar to the eudicots with respect to K_m^{PEP} and malate sensitivity (Ting and Osmond, 1973; Ting and Osmond, 1973; Holaday and Black, 1981), These early studies did however not account for proteolysis of the N terminal of PEPC during purification, which has been shown to affect the protein activity (Chollet, Vidal and O'Leary, 1996). Comparison between the root form of PEPC and the C_4 form of PEPC in *Zea mays* has shown similar differences in kinetics as in the eudicots, the C_4 form having a lower specificity for PEP and a higher sensitivity to activators (Dong *et al.*, 1998). The enzyme in this study was expressed in *E. coli* and cleaved with enterokinase, which can also cleave the N terminal region of PEPC (Gasteiger *et al.*, 2005). Further, it is important to note that the compared enzymes are not encoded by

orthologous genes, but by paralogs that diverged long before the photosynthetic types emerged (Christin *et al.*, 2007).

Despite the low number of species analysed, the comparison of C_4 monocot PEPCs has suggested specific properties when compared to C_4 eudicots PEPCs. The C_4 *Z. mays* PEPC is sensitive to activation by neutral amino acids such as glycine, a sensitivity not observed in *Flaveria* (Figure 11; Wong and Davies, 1973; González-Segura *et al.*, 2018). Analysis of genes encoding C_4 and non-C_4 PEPCs in grasses and sedges have shown that adaptive amino acid transitions at 22 sites accompanied the evolution of C_4 photosynthesis (Christin *et al.*, 2007; Besnard *et al.*, 2009). Two of these positions are 780 and 884, which have been identified to play key roles in C_4 specific properties (Blasing, Westhoff and Svensson, 2000; Paulus, Niehus and Groth, 2013). The serine 780 in particular was observed across most C_4 lineages in both monocots and eudicots (Christin *et al.*, 2007; Besnard *et al.*, 2009). However, the few exceptions clearly show that this residue is not necessary for the C_4 unction (Rao, Reiskind and Bowes, 2008; Rosnow, Edwards and Roalson, 2014). The function of the other amino acid substitutions remains unknown, and the lack of dedicated comparisons between the kinetics of C_4 and non-C_4 orthologs in monocots, means that the biochemical history of C_4 evolution in the group remains largely unexplored. Filling in this gap is necessary to determine whether the observed convergent amino acid replacements are linked to convergent biochemical changes. Conversely, it has been observed that while amino acid changes were extremely convergent within each of the plant family with C_4 origins, different families tended to fix different amino acid substitutions (Besnard *et al.*, 2009). This might suggest that each family reaches a different biochemical solution to the C_4 challenge, or that the same solution is achieved via unique mutations. Testing these hypotheses requires establishing the direction and magnitude of C_4-specific modifications in multiple groups. In addition, all kinetic parameters need to be considered. In particular, the bicarbonate specificity has been measured in a range of C_4 and bacterial PEPC isoforms

(Bauwe, 1986; Janc, O'Leary and Cleland, 1992), but differences in bicarbonate specificity between closely-related C_4 and non-C_4 forms of PEPC remain undescribed.

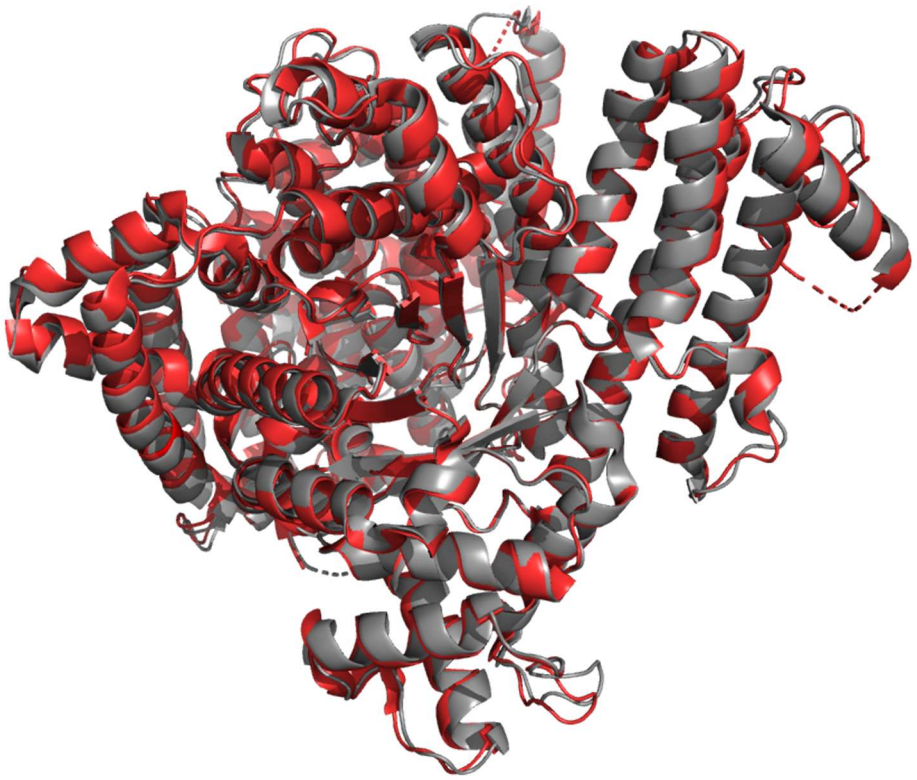


Figure 10: Superimposition of PEPC from *Flaveria pringeli* (C_3) in grey and *Flaveria trinervia* (C_4) in red. Structures are orientated with the active site towards the front. (Paulus, Schlieper and Groth, 2013).

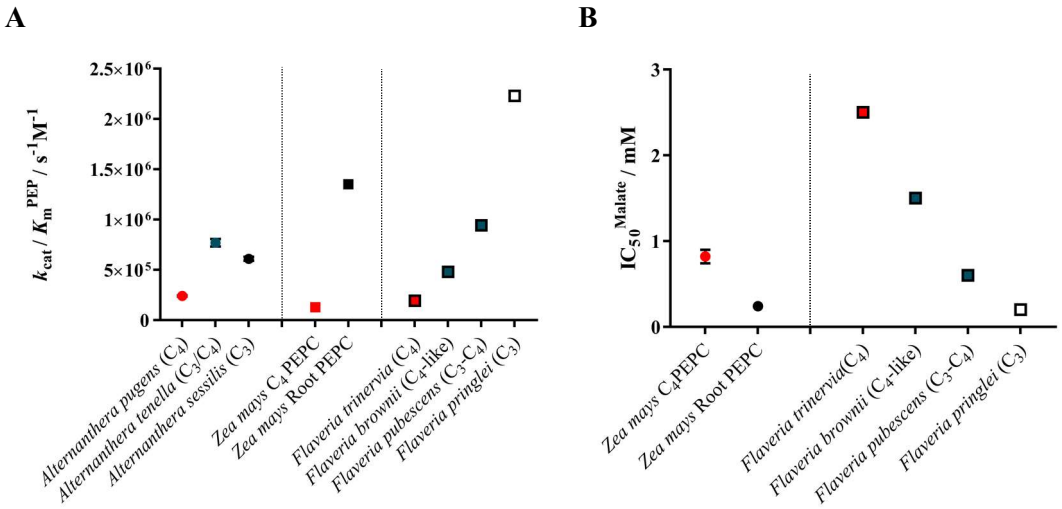


Figure 11: Summary of kinetic parameters of closely related C_4 and C_3 PEPCs. A Specificity for PEP for PEPCs. **B** Inhibition values for PEPCs. Values for *Alternanthera* taken from (Gowik *et al.*, 2006). Values for *Zea mays* taken from (Dong *et al.*, 1998). Values of *Flaveria* taken from (Engelmann *et al.*, 2003).

REFERENCES

Anderson, L. E. (1971) Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases, *Biochimica et Biophysica Acta (BBA) - Enzymology*, 235(1), pp. 237–244.

Aubry, S., Brown, N. J. and Hibberd, J. M. (2011) The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway, *Journal of Experimental Botany*, 62(9), pp. 3049–3059.

Bauwe, H. (1986) An efficient method for the determination of K_m values for HCO_3^- of phosphoenolpyruvate carboxylase, *Planta*, 169(3), pp. 356–360.

Beerling, D. J. and Osborne, C. P. (2006) The origin of the savanna biome, *Global Change Biology*, 12(11), pp. 2023–2031.

Besnard, G., Muasya, a. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. (2009) Phylogenomics of C_4 Photosynthesis in Sedges (Cyperaceae): Multiple Appearances and Genetic Convergence, *Molecular Biology and Evolution*, 26(8), pp. 1909–1919.

Blasing, O. E., Westhoff, P. and Svensson, P. (2000) Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria- a conserved serine residue in the carboxyterminal part of the enzyme is a major determinant for C_4-specific characteristics, *Journal of Biological Chemistry*, 275(36), pp. 27917–23.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., Carr, K. M., Gowik, U., Mass, J., Lercher, M. J., Westhoff, P., Hibberd, J. M. and Weber, A. P. M. (2011) An mRNA Blueprint for C_4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C_3 and C_4 Species, *Plant Physiology*, 155(1), pp. 142–156.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P. and Weber, A. P. M. (2014) Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species, *Journal of experimental botany*, 65(13), pp. 3579–3593.

Brown, W. V. (1975) Variations in anatomy, associations, and origins of Kranz tissue, *American Journal of Botany*, 62(4), pp. 395–402.

von Caemmerer, S., Edwards, G. E., Koteyeva, N. and Cousins, A. B. (2014) Single cell C_4 photosynthesis in aquatic and terrestrial plants: A gas exchange perspective, *Aquatic Botany*. Elsevier B.V., 118, pp. 71–80.

von Caemmerer, S. and Furbank, R. T. (2003) The C_4 pathway: an efficient CO_2 pump, *Photosynthesis Research*, 77(2/3), pp. 191–207.

Calvin, M. (1962) The Path of Carbon in Photosynthesis: The carbon cycle is a tool for exploring chemical biodynamics and the mechanism of quantum conversion, *Science*, 135(3507), pp. 879–889.

Carter, P. J., Nimmo, H. G., Fewson, C. A. and Wilkins, M. B. (1990) Bryophyllum fedtschenkoi protein phosphatase type 2A can dephosphorylate phosphoenolpyruvate carboxylase, *FEBS Letters*, 263(2), pp. 233–236.

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V. and Ehleringer, J. R. (1997) Global vegetation change through the Miocene/Pliocene boundary, *Nature*, 389(6647), pp. 153–158.

Chollet, R., Vidal, J. and O'Leary, M. H. M. (1996) Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants, *Annual review of plant biology*, 47(1), pp. 273–298.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V. and Salamin, N. (2008) Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses, *Current Biology*, 18(1), pp. 37–43.

Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J. and Osborne, C. P. (2013) Parallel Recruitment of Multiple Genes into C_4 Photosynthesis, *Genome Biology and Evolution*, 5(11), pp. 2174–2187.

Christin, P.-A. and Osborne, C. P. (2013) The recurrent assembly of C_4 photosynthesis, an evolutionary tale, *Photosynthesis Research*, 117(1–3), pp. 163–175.

Christin, P.-A., Osborne, C. P., Sage, R. F., Arakaki, M. and Edwards, E. J. (2011) C_4 eudicots are not younger than C_4 monocots, *Journal of experimental botany*, 62(9), pp. 3171–3181.

Christin, P., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. (2007) C_4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes, *Current Biology*, 17(14), pp. 1241–1247.

Cowan, I. R. and Troughton, J. H. (1971) The relative role of stomata in transpiration and assimilation, *Planta*, 97(4), pp. 325–336.

Dawe, D. (2007) Agricultural research, poverty alleviation, and key trends in Asia's rice economy, in *Charting New Pathways to C4 Rice*, pp. 37–53.

Day, D. A. and Hanson, J. B. (1977) Pyruvate and Malate Transport and Oxidation in Corn Mitochondria, *Plant Physiology*, 59(4), pp. 630–635.

Delgado, M. J., Garrido, J. M., Ligero, F. and Lluch, C. (1993) Nitrogen fixation and carbon metabolism by nodules and bacteroids of pea plants under sodium chloride stress, *Physiologia Plantarum*, 89(4), pp. 824–829.

Dobrota, C. (2006) Energy dependant plant stress acclimation, *Reviews in Environmental Science and Bio/Technology*, 5(2–3), pp. 243–251.

Doncaster, H. D. and Leegood, R. C. (1987) Regulation of phosphoenolpyruvate carboxylase activity in maize leaves, *Plant physiology*, 84(1987), pp. 82–87. doi: 10.1104/pp.84.1.82.

Dong, L.-Y., Masuda, T., Kawamura, T., Hata, S. and Izui, K. (1998) Cloning, Expression, and Characterization of a Root-Form Phosphoenolpyruvate Carboxylase from Zea mays: Comparison with the C_4-Form Enzyme, *Plant and Cell Physiology*, 39(8), pp. 865–873.

Duff, S. M. G., Andreo, C. S., Pacquit, V., Lepiniec, L., Sarath, G., Condon, S. A., 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(1), pp. 92–95.

Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P. and Christin, P.-A. (2017) Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives, *Evolution*, 71(6), pp. 1541–1555.

Echevarria, C., Pacquit, V., Bakrim, N., Osuna, L., Delgado, B., Arriodupont, M. and Vidal, J. (1994) The Effect of pH on the Covalent and Metabolic Control of C_4 Phosphoenolpyruvate Carboxylase from Sorghum Leaf', *Archives of Biochemistry and Biophysics*, 315(2), pp. 425–430.

Edwards, G. (1999) 'Tuning up crop photosynthesis', *Nature Biotechnology*, 17(1), pp. 22–23.

Edwards, G. E., Franceschi, V. R. and Voznesenskaya, E. V. (2004) Single-Cell C4 Photosynthesis Versus the Dual-Cell (Kranz) Paradigm, *Annual Review of Plant Biology*, 55(1), pp. 173–196. doi: 25.

Edwards, S., Nguyen, B.-T. T., Do, B. and Roberts, J. K. M. M. (1998) Contribution of malic enzyme, pyruvate kinase, phosphoenolpyruvate carboxylase, and the Krebs cycle to respiration and biosynthesis and to intracellular pH regulation during hypoxia in maize root tips observed by nuclear magnetic resonance imaging and gas, *Plant Physiology*, 116(3), pp. 1073–1081.

Ehleringer, J. and Björkman, O. (1977) Quantum yields for CO_2 uptake in C_3 and C_4 plants, *Plant Physiology*, 59(577), pp. 86–90.

Ehleringer, J. R., Sage, R. F., Flanagan, L. B. and Pearcy, R. W. (1991) Climate change and the evolution of C_4 photosynthesis, *Trends in ecology & evolution*, 6(3), pp. 95–9.

Ellis, R. J. (1979) The most abundant protein in the world, *Trends in Biochemical Sciences*, 4(11), pp. 241–244.

Engelmann, S., Bläsing, O. E., Gowik, U., Svensson, P. and Westhoff, P. (2003) Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus Flaveria- a gradual increase from C_3 to C_4 characteristics., *Planta*, 217(5), pp. 717–25.

Ettema, T. J. G., Makarova, K. S., Jellema, G. L., Gierman, H. J., Koonin, E. V, Huynen, M. a, de Vos, W. M. and van der Oost, J. (2004) Identification and Functional Verification of Archaeal-Type Phosphoenolpyruvate Carboxylase, a Missing Link in Archaeal Central Carbohydrate Metabolism, *Journal of Bacteriology*, 186(22), pp. 7754–7762.

Evenson, R. E. and Gollin, D. (2003) Assessing the Impact of the Green Revolution, 1960 to 2000, *Science*, 300 (May), pp. 758–762.

Freitag, H. and Stichler, W. (2000) A Remarkable New Leaf Type With Unusual Photosynthetic Tissue in a Central Asiatic Genus of Chenopodiaceae, *Plant Biology*, 2(2), pp. 154–160.

Furbank, R. T. (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types?, *Journal of Experimental Botany*, 62(9), pp. 3103–3108.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) *The Proteomics Protocols Handbook*, *The Proteomics Protocols Handbook*. Edited by J. M. Walker. Totowa, NJ: Humana Press.

González-Segura, L., Mújica-Jiménez, C., Juárez-Díaz, J. A., Güémez-Toro, R., Martinez-Castilla, L. P. and Muñoz-Clares, R. A. (2018) Identification of the allosteric site for neutral amino acids in the maize C_4 isozyme of phosphoenolpyruvate carboxylase: The critical role of Ser-100, *Journal of Biological Chemistry*, 293(26), pp. 9945–9957.

González, D. H. and Andreo, C. S. (1989) The use of substrate analogues to study the active-site structure and mechanism of PEP carboxylase, *Trends in Biochemical Sciences*, 14(1), pp. 24–27.

Gowik, U., Engelmann, S., Bläsing, O. E., Raghavendra, and Westhoff, P. (2006) Evolution of C_4 phosphoenolpyruvate carboxylase in the genus Alternanthera: Gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 Alternantheras, *Planta*, 223, pp. 359–368.

Gowik, U. and Westhoff, P. (2011) C_4 Phosphoenolpyruvate Carboxylase, in Raghavendra, A. S. and Sage, R. F. (eds) *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*. 2011 ed. Dordrecht, Netherlands: Springer, pp. 257–275.

Grass Phylogeny Working Group II (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins, *New Phytologist*, 193(2), pp. 304–312.

Griffiths, H. (2006) Designs on Rubisco, *Nature*, 441(7096), pp. 940–941.

Hartwell, J., Gill, A., Nimmo, G. A., Wilkins, M. B., Jenkins, G. I. and Nimmo, H. G. (1999) Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression, *The Plant Journal*, 20(3), pp. 333–342.

Hatch, M. D. (1987) C_4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure, *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 895(2), pp. 81–106.

Hatch, M. and Slack, C. (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation, *Biochemical Journal*, 101(1), pp. 103–111.

Heber, U. and Krause, G. H. (1980) What is the physiological role of photorespiration?, *Trends in Biochemical Sciences*, 5(2), pp. 32–34.

Hibberd, J. M., Sheehy, J. E. and Langdale, J. a. (2008) Using C_4 photosynthesis to increase the yield of rice - rationale and feasibility, *Current Opinion in Plant Biology*, 11(2), pp. 228–231.

Holaday, A. S. and Black, C. C. (1981) Comparative Characterization of Phosphoenolpyruvate Carboxylase in C_3, C_4, and C_3-C_4 Intermediate Panicum Species, *Plant Physiology*, 67(2), pp. 330–334.

Izui, K., Matsumura, H., Furumoto, T. and Kai, Y. (2004) Phosphoenolpyruvate Carboxylase: A New Era of Structural Biology, *Annual review of plant biology*, 55, pp. 69–84.

Izui, K., Taguchi, M. K., Morikawa, M. and Katsuki, H. (1981) Regulation of *Escherichia co*li Phosphoenolpyruvate Carboxylase by Multiple Effectors In Vivo. Kinetic Studies with a Reaction System Containing Physiological Concentrations of Ligands, *The Journal of Biochemistry*, 90(5), pp. 1321–1331.

Jacobs, B., Engelmann, S., Westhoff, P. and Gowik, U. (2008) Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*: Determinants for high tolerance towards the inhibitor L-malate, *Plant, Cell and Environment*, 31, pp. 793–803.

Janc, J. W., O'Leary, M. H. and Cleland, W. W. (1992) A kinetic investigation of phosphoenolpyruvate carboxylase from *Zea mays*, *Biochemistry*, 31(28), pp. 6421–6426.

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 Sciences of the United States of America*, 96(February), pp. 823–828.

Kai, Y., Matsumura, H. and Izui, K. (2003) Phosphoenolpyruvate carboxylase: three-dimensional structure and molecular mechanisms, *Archives of Biochemistry and Biophysics*, 414(2), pp. 170–179.

Kanai, R. and Edwards, G. E. (1999) The Biochemistry of C_4 Photosynthesis, in *C_4 Plant biology*, pp. 49–87.

Kanai, R., Edwards, G. E., Sage, R. F. and Monson, R. K. (1999) C_4 plant biology. Academic Press, San Diego, CA.

Kelly, G. J. and Latzko, E. (1976) Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate, *FEBS Letters*, 68(1), pp. 55–58.

Ku, S. B. and Edwards, G. E. (1977) Oxygen Inhibition of Photosynthesis, *Plant Physiology*, 59(5), pp. 991–999.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. and Gowik, U. (2017) C_3 cotyledons are followed by C_4 leaves: intra-individual transcriptome analysis of Salsola soda (*Chenopodiaceae*), *Journal of Experimental Botany*, 68(2), pp. 161–176.

Leegood, R. C. (2007) Roles of the bundle sheath cells in leaves of C_3 plants, *Journal of Experimental Botany*, 59(7), pp. 1663–1673.

Leegood, R. C. (2013) Strategies for engineering C_4 photosynthesis, *Journal of Plant Physiology*. Elsevier GmbH., 170(4), pp. 378–388.

Leegood, R. C. and Walker, R. P. (2003) Regulation and roles of phosphoenolpyruvate carboxykinase in plants, *Archives of Biochemistry and Biophysics*, 414(2), pp. 204–210.

Lepiniec, L., Vidal, J., Chollet, R., Gadal, P. and Crétin, C. (1994) Phosphoenolpyruvate carboxylase: structure, regulation and evolution, *Plant Science*, 99(2), pp. 111–124.

Lundgren, M. R., Osborne, C. P. and Christin, P.-A. (2014) Deconstructing Kranz anatomy to understand C_4 evolution, *Journal of Experimental Botany*, 65(13), pp. 3357–3369.

Malkin, R. and Niyogi, K. (2000) Photosynthesis ages, in Kluwer, D. (ed.) *Biochemistry & Molecular Biology of Plants*, pp. 413–429.

Mathis, P. and Paillotin, G. (1981) Primary Processes of Photosynthesis, in Hatch, M. and Boardman, N. K. (eds) *Photosynthesis*. Academic Press, pp. 97–161.

Matsumura, H., Terada, M., Shirakata, S., Inoue, T., Yoshinaga, T., Izui, K. and Kai, Y. (1999) Plausible phosphoenolpyruvate binding site revealed by 2.6 Å structure of Mn^{2+} - bound phosphoenolpyruvate carboxylase from *Escherichia coli*, *FEBS Letters*, 458(2), pp. 93–96.

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*, 10(12), pp. 1721–1730.

McKown, A. D. and Dengler, N. G. (2010) Veining Patterning and Evolution in C_4 Plants, *Botany*, 88(9), pp. 775–786.

McKown, A. D., Moncalvo, J.-M. and Dengler, N. G. (2005) Phylogeny of *Flaveria* (*Asteraceae*) and inference of C_4 photosynthesis evolution, *American Journal of Botany*, 92(11), pp. 1911–1928.

Monson, R. K., Edwards, G. E., Ku, (2008) C_3-C_4 Intermediate Photosynthesis Plants, 34(9), pp. 563–566.

Monson, R. K. and Rawsthorne, S. (2000) CO_2 Assimilation in C_3-C_4 Intermediate Plants, in Leegood, R. C., Sharkey, T. D., and von Caemmerer, S. (eds) *Photosynthesis: Physiology and Metabolism*. Dordrecht: Springer Netherlands, pp. 533–550.

Moore, B. d., Monson, R. K., Ku, M. S. B. and Edwards, G. E. (1988) Activities of Principal Photosynthetic and Photorespiratory Enzymes in Leaf Mesophyll and Bundle Sheath Protoplasts from the C_3-C_4 Intermediate *Flaveria ramosissima*, *Plant Cell Physiology*, 29(6), pp. 999–1006.

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. and Christin, P. A. (2018) Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis, *Molecular Biology and Evolution*, 35(1), pp. 94–106.

Muirhead, H., Clayden, D. A., Barford, D., Lorimer, C. G., Fothergill-Gilmore, L. A., Schiltz, E. and Schmitt, W. (1986) The structure of cat muscle pyruvate kinase, *The EMBO journal*, 5(3), pp. 475–81.

Nimmo, G. A., McNaughton, G. A. L., Fewson, C. A., Wilkins, M. B. and Nimmo, H. G. (1987) Changes in the kinetic properties and phosphorylation state of phospho enol pyruvate carboxylase in *Zea mays* leaves in reponse to light and dark, *FEBS Letters*, 213(1), pp. 18–22.

Nimmo, H. G. (2000) The regulation of phosphoenolpyruvate carboxylase in CAM plants, *Trends in Plant Science*, 5(2), pp. 75–80.

Nimmo, H. G. (2003) Control of the phosphorylation of phosphoenolpyruvate carboxylase in higher plants, *Archives of Biochemistry and Biophysics*, 414(2), pp. 189–196.

Nisbet, E. G., Grassineau, N. V., Howe, C. J., Abell, P. I., Regelous, M. and Nisbet, R. E. R. (2007) The age of Rubisco: the evolution of oxygenic photosynthesis, *Geobiology*, 5(4), pp. 311–335.

Nisbet, E. G. and Nisbet, R. E. R. (2008) Methane, oxygen, photosynthesis, rubisco and the regulation of the air through time, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1504), pp. 2745–2754.

O'Leary, B., Park, J. and Plaxton, W. C. (2011) The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs, *Biochemical Journal*, 436(1), pp. 15–34.

O'Leary, B., Rao, S. K., Kim, J. and Plaxton, W. C. (2009) Bacterial-type Phosphoenolpyruvate Carboxylase (PEPC) Functions as a Catalytic and Regulatory Subunit of the Novel Class-2 PEPC Complex of Vascular Plants, *Journal of Biological Chemistry*, 284(37), pp. 24797–24805.

O'Leary, M. H. (1982) Phosphoenolpyruvate Carboxylase: An Enzymologist's View, *Plant Physiology*, 33, pp. 297–315.

Offermann, S., Okita, T. W. and Edwards, G. E. (2011) Resolving the Compartmentation and Function of C_4 Photosynthesis in the Single-Cell C_4 Species B*ienertia sinuspersici*, *Plant Physiology*, 155(4), pp. 1612–1628.

Ogren, W. L. (1984) Photorespiration: Pathways, Regulation, and Modification, *Annual Review of Plant Physiology*, 35(1), pp. 415–442.

Osborne, C. P. and Freckleton, R. P. (2009) 'Ecological selection pressures for C_4 photosynthesis in the grasses', *Proceedings of the Royal Society: Biological Sciences*, 276(1663), pp. 1753–1760.

Osuna, L., Gonzalez, M. C., Cejudo, F. J., Vidal, J. and Echevarria, C. (1996) 'In Vivo and in Vitro Phosphorylation of the Phosphoenolpyruvate Carboxylase from Wheat Seeds during Germination', *Plant Physiology*, 111(2), pp. 551–558.

Paulus, J. K., Niehus, C. and Groth, G. (2013) Evolution of C_4 Phosphoenolpyruvate Carboxylase: Enhanced Feedback Inhibitor Tolerance Is Determined by a Single Residue, *Molecular Plant*, 6(6), pp. 1996–1999.

Paulus, J. K., Schlieper, D. and Groth, G. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution, *Nature communications*, 4, p. 1518.

Powell, A. M. (1978) Systematics of *Flaveria* (*Flaveriinae-Asteraceae*), *Annals of the Missouri Botanical Garden*, 65(2), p. 590.

Rao, S. K., Reiskind, J. B. and Bowes, G. (2008) Kinetic analyses of recombinant isoforms of phosphoenolpyruvate carboxylase from *Hydrilla verticillata* leaves and the impact of substituting a C_4-signature serine, *Plant Science*, 174(4), pp. 475–483.

Ren, C. G., Li, X., Liu, X. L., Wei, X. D. and Dai, C. C. (2014) Hydrogen peroxide regulated photosynthesis in C_4-PEPC transgenic rice, *Plant Physiology and Biochemistry*, 74, pp. 218–229.

Rosnow, J. J., Edwards, G. E. and Roalson, E. H. (2014) Positive selection of Kranz and non-Kranz C_4 phosphoenolpyruvate carboxylase amino acids in *Suaedoideae* (*Chenopodiaceae*)., *Journal of experimental botany*, 65(13), pp. 3595–607.

Rumpho, M. E., Ku, M. S. B., Cheng, S.-H. and Edwards, G. E. (1984) Photosynthetic Characteristics of C_3-C_4 Intermediate *Flaveria* Species : III. Reduction of Photorespiration by a Limited C_4 Pathway of Photosynthesis in *Flaveria ramosissima*, *Plant Physiology*, 75(4), pp. 993–996.

Sage, R. F. (2004) The evolution of C_4 photosynthesis, *New Phytologist*, 161(2), pp. 341–370.

Sage, R. F., Christin, P.-A. and Edwards, E. J. (2011) The C_4 plant lineages of planet Earth, *Journal of experimental botany*, 62(9), pp. 3155–3169.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) Photorespiration and the Evolution of C_4 Photosynthesis, *Annual Review of Plant Biology*, 63(1), pp. 19–47.

Salvucci, M. E. and Bowes, G. (1981) Induction of Reduced Photorespiratory Activity in Submersed and Amphibious Aquatic Macrophytes, *Plant Physiology*, 67(2), pp. 335–340.

Salvucci, M. E. and Bowes, G. (1983) Two Photosynthetic Mechanisms Mediating the Low Photorespiratory State in Submersed Aquatic Angiosperms, *Plant Physiology*, 73(2), pp. 488–496.

Sánchez, R. and Cejudo, F. (2003) Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from *Arabidopsis* and rice, *Plant physiology*, 132(June), pp. 949–957.

Sangwan, R. S., Singh, N. and Plaxton, W. C. (1992) Phosphoenolpyruvate Carboxylase Activity and Concentration in the Endosperm of Developing and Germinating Castor Oil Seeds, *Plant Physiology*, 99(2), pp. 445–449.

Schlieper, D., Förster, K., Paulus, J. K. and Groth, G. (2014) Resolving the Activation Site of Positive Regulators in Plant Phosphoenolpyruvate Carboxylase, *Molecular Plant*, 7(2), pp. 437–440.

Siedow, J. N. and Day, D. A. (2000) Respiration and photorespiration*, Biochemistry and molecular biology of plants*. American Society of Plant Physiologists, Rockville, MD 676-728.

Skillman, J. B. (2007) Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark, *Journal of Experimental Botany*, 59(7), pp. 1647–1661.

Stitt, M. and Zhu, X.-G. (2014) The large pools of metabolites involved in intercellular metabolite shuttles in C_4 photosynthesis provide enormous flexibility and robustness in a fluctuating light environment, *Plant, Cell & Environment*, 37(9), pp. 1985–1988.

Sun, J. and Sampson, N. S. (1999) Understanding Protein Lids: Kinetic Analysis of Active Hinge Mutants in Triosephosphate Isomerase, *Biochemistry*, 38(35), pp. 11474–11481.

Svensson, P., Bläsing, O. E. and Westhoff, P. (1997) Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase- a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3)., *European journal of biochemistry*, 246(2), pp. 452–60.

Takahashi-Terada, A., Kotera, M., Ohshima, K., Furumoto, T., Matsumura, H., Kai, Y. and Izui, K. (2005) Maize Phosphoenolpyruvate Carboxylase: Mutations at the Putative Binding Site for Glucose 6-Phosphate Caused Desensitization and Abolished Responsiveness to Regulatory Phosphorylation, *Journal of Biological Chemistry*, 280(12), pp. 11798–11806.

Tcherkez, G. G. B., Farquhar, G. D. and Andrews, T. J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized, *Proceedings of the National Academy of Sciences*, 103(19), pp. 7246–7251.

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(3), pp. 797–803.

Ting, I. P. and Osmond, C. B. (1973) Multiple Forms of Plant Phosphoenolpyruvate Carboxylase Associated with Different Metabolic Pathways, *Plant Physiology*, 51(3), pp. 448–453.

Ting, I. P. and Osmond, C. B. (1973) Photosynthetic Phosphoenolpyruvate Carboxylases, *Plant physiology*, 51(3), pp. 439–447.

Tovar-Méndez, A., Mújica-Jiménez, C. and Muñoz-Clares, R. a (2000) Physiological Implications of the Kinetics of Maize Leaf Phosphoenolpyruvate Carboxylase, *Plant Physiology*, 123(1), pp. 149–160.

Voznesenskaya, E., Franceschi, V. and Kiirats, O. (2001) Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis, *Nature*, 480(1999), pp. 1999–2002.

Wang, Y., Bräutigam, A., Weber, A. P. M. and Zhu, X.-G. (2014) Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis, *Journal of Experimental Botany*, 65(13), pp. 3567–3578.

Weber, A. P. and Bräutigam, A. (2013) The role of membrane transport in metabolic engineering of plant primary metabolism, *Current Opinion in Biotechnology*, 24(2), brpp. 256–262.

Wedding, R. T., Black, M. K. and Meyer, C. R. (1989) Activation of higher plant phosphoenolpyruvate carboxylases by glucose-6-phosphate., *Plant Physiology*, 90, pp. 648–652.

Weiner, H., Burnell, J. N., Woodrow, I. E., Heldt, H. W., Hatch, M. D. and Sauer, F. (1988) Metabolite diffusion into bundle sheath cells from C_4 plants: relation to C_4 photosynthesis and plasmodesmatal function, *Plant Physiology*, 88(3), pp. 815–22.

Wingler, A., Lea, P. J., Quick, W. P. and Leegood, R. C. (2000) Photorespiration: metabolic pathways and their role in stress protection, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1402), pp. 1517–1529.

Wong, K. F. and Davies, D. D. (1973) Regulation of phosphoenolpyruvate carboxylase of *Zea mays* by metabolites, *Biochemical Journal*, 131(3), pp. 451–458.

Woo, K. and Xu, S. (1996) Metabolite Regulation of Phosph enol pyruvate Carboxylase in Legume Root Nodules, *Australian Journal of Plant Physiology*, 23(4), p. 413.

Yano, M., Terada, K., Umiji, K. and Izui, K. (1995) Catalytic Role of an Arginine Residue in the Highly Conserved and Unique Sequence of Phosphoenolpyruvate Carboxylase', *The Journal of Biochemistry*, 117(6), pp. 1196–1200.

THESIS AIMS AND STRUCTURE

This project aims to explore C_4 phosphoenolpyruvate carboxylase (PEPC), and the kinetic properties acquired by the enzyme when it is co-opted for a role as a primary carbon fixing enzyme. This was done by exploring the kinetic properties of a carbon fixing PEPC from a C_4 species and the PEPC encoded by the orthologous gene from a closely related species that has not evolved the C_4 trait. These proteins were expressed in *E. coli* and purified, allowing analysis of the homogenous protein. Comparing between a recently diverged C_4 PEPC and the non-C_4 PEPC from one of the earliest origins can shed light on the important of kinetic properties and the flexibility in the adaption of the enzyme.

In the genus *Flaveria*, the PEPCs from the C_4 species *Flaveria trinervia* and C_3 species *Flaveria pringlei* were used as a model for the evolution of C_4 PEPC. These PEPCs were investigated to expand on work described in previous investigations, by looking at bicarbonate specificity, phosphoenolpyruvate (PEP) specificity and sensitivity to the feedback inhibitors aspartate and malate (Chapter 2). The results show that an increase in specificity for bicarbonate was selected for in the evolution of *Flaveria* C_4 PEPC. When the two enzymes were compared, the C_4 PEPC has a factor of ten lower specificity for PEP. Malate was shown to inhibit PEPC at limiting and saturating PEP; aspartate was shown to inhibit PEPC only at limiting PEP. The C_4 PEPC was a factor of five times less sensitive to competitive inhibition by malate and a factor of ten less sensitive to aspartate. The C_4 PEPC was also shown to be ten times less sensitive to non-competitive inhibition by malate. These PEPCs have a *ca.* 96% amino acid sequence similarity, however, the enzymes exhibit stark kinetic differences.

The carbon fixing PEPCs from the C_4 grass *Panicum queenslandicum* and the PEPC from C_3 species *Panicum pygmaeum* were investigated (Chapter 3). Grasses have been established as one of the earliest lineages of C_4 species. The C_4 PEPC from *P. queensladicum* shown a similar change in kinetic properties to that seen in *F. trinervia* when compared to the non-C_4 PEPC; the C_4 PEPC has a higher specificity for bicarbonate, a lower specificity for bicarbonate, less sensitive to competitive inhibition by malate and aspartate,

and less sensitive to non-competitive inhibition by malate. This shows convergence in properties of C_4 PEPCs from distantly related species. The changes observed however are quantitively greater in *Panicum* than in *Flaveria.* This suggests that longer evolutionary period in grasses resulted in greater adaptations exhibited by the C_4 PEPC.

C_4 specific amino acid sites were then explored using site directed mutagenesis. It has been established that at least 20 specific amino acids have been selected for in C_4 grass and sedge PEPCs. Three of these C_4 specific amino acids were investigated, two of which are convergent in *Flaveria*, the other site being C_4 grass PEPC specific (Chapter 4). Mutant enzymes were generated by introducing the C_4 specific site in the non-C_4 enzyme and vice versa. Position 780 (*Zea mays* numbering) is serine in C_4 PEPCs, and alanine in the non-C_4 PEPCs. Investigation of position 780 showed that this site is important for the C_4 PEP specificity, having the same magnitude of effect when changed in grass and *Flaveria*; this position has no effect on inhibition sensitivity or bicarbonate specificity. Position 761 is serine in *P. queenslandicum*, and alanine in *P. pygmaeum* and *Flaveria* PEPC. Investigation of this position showed no effect on PEP specificity, bicarbonate specificity or inhibitor sensitivity. Bioinformatic analysis suggested that the mutation at position 761 might prevent erroneous post-translational modification. Position 665 is asparagine in C_4 PEPCs and histidine in non-C_4 PEPCs. It was shown this amino acid site does not affect the kinetic properties explored. However, this mutation might reduce the cost of expression.

Overall this work shows that an increase specificity for bicarbonate is a property selected for in the evolution C_4 PEPC. Investigation of the kinetic adaptations of C_4 PEPC are convergent in *Flaveria* and *Panicum*, the changes observed in C_4 grass PEPC are quantitatively greater, reflecting the relative time from divergence. Investigation of C_4 specific amino acids has indicated that efficiency of expression and post-translational modification may also play a driving force in the evolution of C_4 PEPC. This project sheds light on the nature of enzyme evolution, and which biochemical properties are selected for in the evolution of an enzyme.

Changes in bicarbonate specificity during the evolution of C_4 PEPC in *Flaveria*

Nicholas R Moody[1], Pascal-Antoine Christin[2] and James D. Reid[1]

[1]Department of Chemistry, University of Sheffield, Brook Hill, Sheffield, S3 7HF.

[2]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN.

Personal Contribution: I generated and analysed all the data and wrote the paper with the help of my co-authors. As of the time of thesis submission, this paper was in draft form.

ABSTRACT

C_4 photosynthesis is a complex assemblage of anatomical and biochemical components that act together to concentrate CO_2 within the leaf and boost productivity in tropical conditions. This complex trait evolved independently many times, resulting in various realizations of the phenotypes, but in all C_4 plants the primary fixation of atmospheric carbon is catalysed by the enzyme phosphoenolpyruvate carboxylase (PEPC). This enzyme existed before the C_4 path evolved, and its co-option for the C_4 pathway was followed by massive upregulation. Previous comparisons of C_4 and non-C_4 PEPC from closely related species of *Flaveria* indicated that the enzyme was modified to meet the demands of the C_4 metabolic cycle by a ten-fold decrease in specificity to phosphoenolpyruvate and a decrease in sensitivity to inhibitor the inhibitor malate. However, the enzyme specificity for one of its substrates, bicarbonate, has not been assessed in this model system. In this work, we assessed the kinetic properties of closely related C_4 and non-C_4 of PEPC from *Flaveria*. The k_{cat}/K_m for bicarbonate of the C_4-specific enzyme is $6.92 \times 10^5 \pm 4.17 \times 10^4$ s^{-1} M^{-1}, compared with the non-C_4 isoform at $4.43 \times 10^5 \pm 2.17 \times 10^4$ s^{-1} M^{-1}, a one-third increase. We conclude that the adaptation of PEPC for the C_4 context involved increases of the affinity for bicarbonate, potentially as a because bicarbonate becomes limiting in the high flux systems.

C_4 photosynthesis is a complex trait that boosts productivity in tropical conditions (Atkinson *et al.*, 2016). It relies on the spatial segregation of photosynthetic reactions among cell types to concentrate CO_2 around ribulose-bisphosphate carboxylase oxygenase (RuBisCO), the enzyme responsible for the incorporation of CO_2 into the Calvin-Benson cycle (Hatch, 1987; Sage, 2004; Sage, Sage and Kocacinar, 2012). RuBisCO evolved billions of years ago with a dual affinity for CO_2 and O_2 (Tcherkez, Farquhar and Andrews, 2006; Nisbet *et al.*, 2007). In conditions of CO_2 depletion, including high temperature, salinity and aridity, increased O_2 fixation reduces photosynthetic efficiency (Ehleringer and Björkman, 1977; Skillman, 2007). The C_4 pathway solves this problem by fixing

atmospheric CO_2 in the form of bicarbonate via the enzyme phosphoenolpyruvate carboxylase (PEPC), an enzyme without affinity for O_2 (Hatch, 1987). PEPC produces the four carbon acid oxaloacetate, which is then converted into another more stable acid, usually malate or aspartate (Bräutigam *et al.*, 2014). This acid is then shuttled into a leaf compartment isolated from the atmosphere, where RuBisCO is localized in C_4 plants. The CO_2 is released, increasing the CO_2 concentration by up to 10 times relative to the atmosphere (von Caemmerer and Furbank, 2003).

All enzymes of the C_4 pathway existed in non-C_4 ancestors, and C_4 evolution therefore consists of the co-option of multiple genes followed by adaptation of the expression patterns and kinetic properties (Aubry, Brown and Hibberd, 2011; Christin *et al.*, 2013). The changes in gene expression are being increasingly studied through high-throughput sequencing techniques (Bräutigam *et al.*, 2014; Lauterbach *et al.*, 2017; Moreno-Villena *et al.*, 2018), but the biochemical changes remain largely unexplored. Previous efforts have focused on PEPC, which is known to be massively upregulated during the evolution of C_4 photosynthesis (Moreno-Villena *et al.*, 2018). The kinetic changes linked to the evolution of C_4-specific PEPC have been assessed using as a model system *Flaveria* (Engelmann *et al.*, 2003; Svensson, Bläsing and Westhoff, 2003; Westhoff, 2004), a genus of eudicots that contains closely-related C_4 and non-C_4 species (McKown, Moncalvo and Dengler, 2005). Previous efforts have shown C_4-specific increases of the K_m for phosphoenolpyruvate (PEP) (Svenssonz and Westhoff, 1997), with decreases of sensitivity to malate and increased sensitivity to the activator glucose-6-phosphate (Wedding, Black and Meyer, 1990; Jacobs *et al.*, 2008; Paulus, Schlieper and Groth, 2013).

The selective driver of increased K_m^{PEP} of C_4 PEPC remain however elusive. One possibility is that this change was directly selected to avoid depleting other biochemical cycles. Alternatively, this change might have happened as side-effect of other, undetected protein adaptations (Gowik and Westhoff, 2011a). The specificity for bicarbonate may also have changed during C_4 evolution, either under direct selection or potentially as a

consequence of the amino acid changes required to produce an increase in K_m^{PEP}. While the bicarbonate specificity has been measured in a range of C_4 and bacterial PEPC isoforms (Bauwe, 1986; Janc, O'Leary and Cleland, 1992), any change in bicarbonate specificity that happened during C_4 evolution remains undescribed.

In this work, we use the well characterized *Flaveria* model system to test the hypothesis that the evolution of C_4-specific PEPC produced a change in specificity for bicarbonate. In addition, we describe the different sensitivities to inhibitors, including the previously assessed malate, but also aspartate, another metabolite of the C_4 cycle. Our data distinguishes between inhibition at limiting and saturating concentrations PEP, shedding new light on the role of metabolic conditions in feedback regulation of C_4 photosynthesis. Overall, our work provides a detailed understanding of the complex differences in behaviour between a related C_3 and C_4 enzyme from species in a small genus that evolved the trait recently.

RESULTS

DNA Cloning

The *ppcA* genes for PEPC (*ppc-1E2* in Christin *et al.*, 2015) from *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3) as described in (Svensson, Bläsing and Westhoff, 1997) were sub-cloned into the pET-1B plasmid encoding an N-terminal poly-histidine tag and a TEV cleavage site. Expressed protein was purified to > 95% purity by SDS PAGE with a single immobilised metal column (Supp. Figure 1). Assays at saturating bicarbonate and variable concentrations of PEP showed that both proteins behaved similarly to untagged proteins previously described (Svenssonz, and Westhoff, 1997; Bläsing, Westhoff and Svensson, 2000; Jacobs et al., 2008). This indicates that the presence of an N-terminal polyhistidine tag does not adversely affect the activity of the protein.

Controlled Bicarbonate Assay System

The specificity for bicarbonate of both enzymes was determined using a gas-tight assay system. Background bicarbonate was reduced to *ca.* 50 µM by sparging with nitrogen gas. Assays were performed at five PEP concentrations, while varying the concentration of bicarbonate (Figure 1). Analysis of secondary plots allowed determination of the steady-state kinetic parameters, k_{cat}/K_m^{HCO3-} and k_{cat}/K_m^{PEP}, and k_{cat}. The specificity for bicarbonate of the C_4 PEPC was determined as $6.92 \times 10^5 \pm 0.42 \times 10^5$ s^{-1} M^{-1}, (Figure 1C) which is approximately one-third higher than that of the C_3 PEPC which was determined at $4.43 \times 10^5 \pm 0.22 \times 10^5$ s^{-1} M^{-1} (Figure 1D). The kinetic parameters are summarised in Table 1.

Inhibition of PEPC by C_4 metabolites

For both C_3 and C_4 enzymes, we investigated inhibition by the two feedback inhibitors, malate (Figure 2) and aspartate (Figure 3) at both limiting and saturating PEP. These two closely structurally related inhibitors show very different kinetic characteristics; inhibition by malate is much less sensitive to the concentration of PEP than inhibition by aspartate. Under all conditions, with both inhibitors, the C_4 form of PEPC is less sensitive to inhibition. The non-competitive inhibition constant (K_{iu}) was determined by the secondary

plot of k_{cat}^{app} against inhibitor concentration. The competitive inhibition constant (K_{ic}) was determined by the secondary plot of k_{cat}^{app}/K_m^{app} against inhibitor concentration.

The C_4 *Flaveria trinervia* PEPC is inhibited by malate at both limiting and saturating concentrations of PEP, this mixed inhibition can be characterised at limiting PEP by K_{ic}^{Malate} = 10.96 ± 1.55 mM (Figure 2A) and at saturating PEP by K_{iu}^{Malate} = 40.72 ± 4.59 mM (Figure 2C). In contrast, aspartate is a competitive inhibitor characterised by $K_{ic}^{Aspartate}$ = 40.02 ± 6.49 mM (Figure 3A). No inhibition by aspartate at saturating PEP was observed (Figure 3C).

Likewise, the C_3 *Flaveria pringlei* PEPC is inhibited at both limiting and saturating concentrations of PEPC, this mixed inhibition can be characterised at limiting PEP by K_{ic}^{Malate} = 2.14 ± 0.62 mM (Figure 2B) and at saturating PEP by K_{iu}^{Malate} = 4.56 ± 1.17 mM (Figure 2D). As with the C_4 PEPC, aspartate is a competitive inhibitor characterised by $K_{ic}^{Aspartate}$ = 4.13 ± 0.60 mM (Figure 3B). No inhibition at saturating PEP was observed (Figure 3D).

Comparing the two enzymes, the C_4 isoform is five times less sensitive to inhibition by malate at limiting PEP and ten times less sensitive to malate at saturating PEP when compared to the C_3 isoform. *Flaveria trinervia* PEPC is ten times less sensitive to competitive inhibition by aspartate than *Flaveria pringeli* PEPC (Table 2).

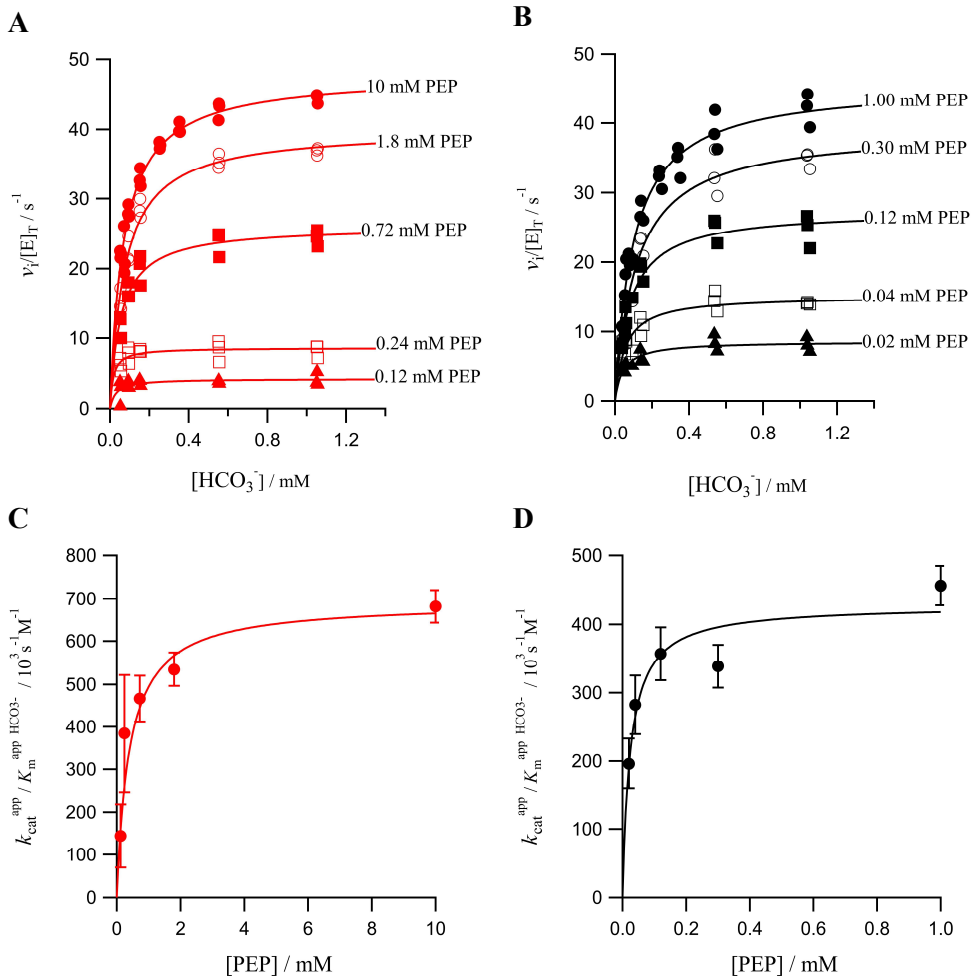


Figure 1: Rate of oxaloacetic acid formation, catalysed by PEPC, varying the concentration of bicarbonate. Assays conditions were 50 mM Tricine.KOH, pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 U μl^{-1} malate dehydrogenase and 50 nM PEPC. **A** Markers represent individual data points from *Flaveria trinervia* PEPC. The lines are described by equation 1, the kinetic parameters are shown in Supp. Figure 3A and in panel C. **B** Markers represent individual data points from *Flaveria trinervia* PEPC. The lines are described by equation 1, the kinetic parameters are shown in Supp. Figure 3B and in panel D. **C** Secondary plot of $k_{cat}^{app}/K_m^{app\ HCO3-}$ from *Flaveria trinervia* PEPC, the line is described by equation 1, characterised by the parameters k_{cat} = 53.89 ± 4.12 s^{-1}, K_m^{HCO3-} = 0.070 ± 0.005 mM, K_m^{PEP} = 0.84 ± 0.02 mM and k_{cat}/K_m^{HCO3-} = 6.92 × 10^5 ± 4.17× 10^4 s^{-1} M^{-1}, error bars represent standard errors. **D** Secondary plot of $k_{cat}^{app}/K_m^{app\ HCO3-}$ from *Flaveria pringlei* PEPC, the line is described by equation 1, characterised by the parameters k_{cat} = 51.01 ± 0.01 s^{-1}, K_m^{HCO3-} = 0.99 ± 0.007 mM, K_m^{PEP} = 0.0245 ±0.007 mM and k_{cat}/K_m^{HCO3-} = 4.43 × 10^5 ± 2.17 × 10^4 s^{-1} M^{-1}, error bars represent standard errors.

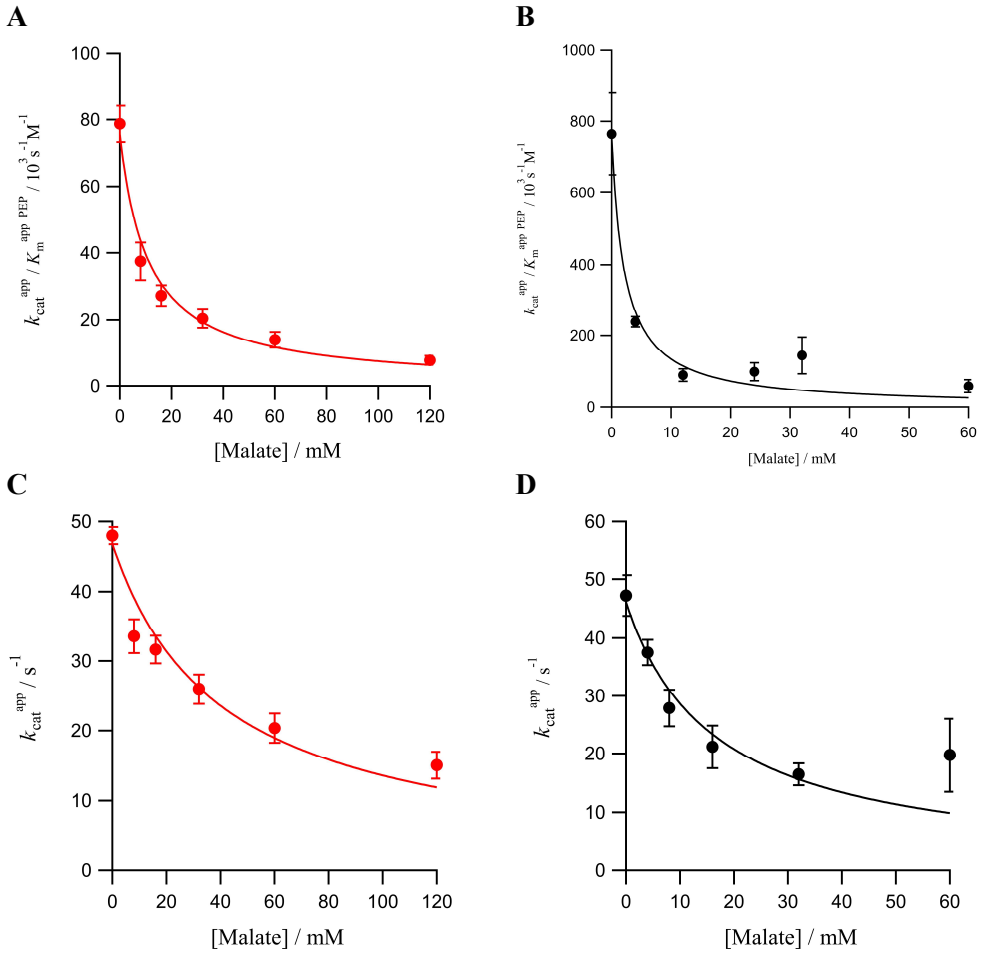


Figure 2: Plots of malate inhibition values for the PEPC enzymes. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 U μl^{-1} malate dehydrogenase, 10 mM $KHCO_3$ and 10 nM of *Flaveria trinervia* PEPC or 5 nM *Flaveria pringlei* PEPC. Markers are derived from Supp. Figure 4A for *Flaveria trinervia* PEPC and Supp. Figure 4B for *Flaveria pringeli* PEPC, error bars represent standard errors. Secondary plot lines are characterised by equation 2. **A** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against malate concentration for the enzyme *Flaveria trinervia* PEPC characterised by K_{ic}^{Malate} = 10.96 ± 1.55 mM. **B** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against malate concentration for the enzyme *Flaveria pringlei* PEPC characterised by K_{ic}^{Malate} = 2.14 ± 0.62 mM. **C** Secondary plot of k_{cat}^{app} against malate concentration for the enzyme *Flaveria trinervia* PEPC characterised by K_{iu}^{Malate} = 40.72 ± 4.59 mM. **D** Secondary plot of k_{cat}^{app} against malate concentration for the enzyme *Flaveria pringlei* PEPC characterised by K_{iu}^{Malate} = 4.56 ± 1.72 mM.

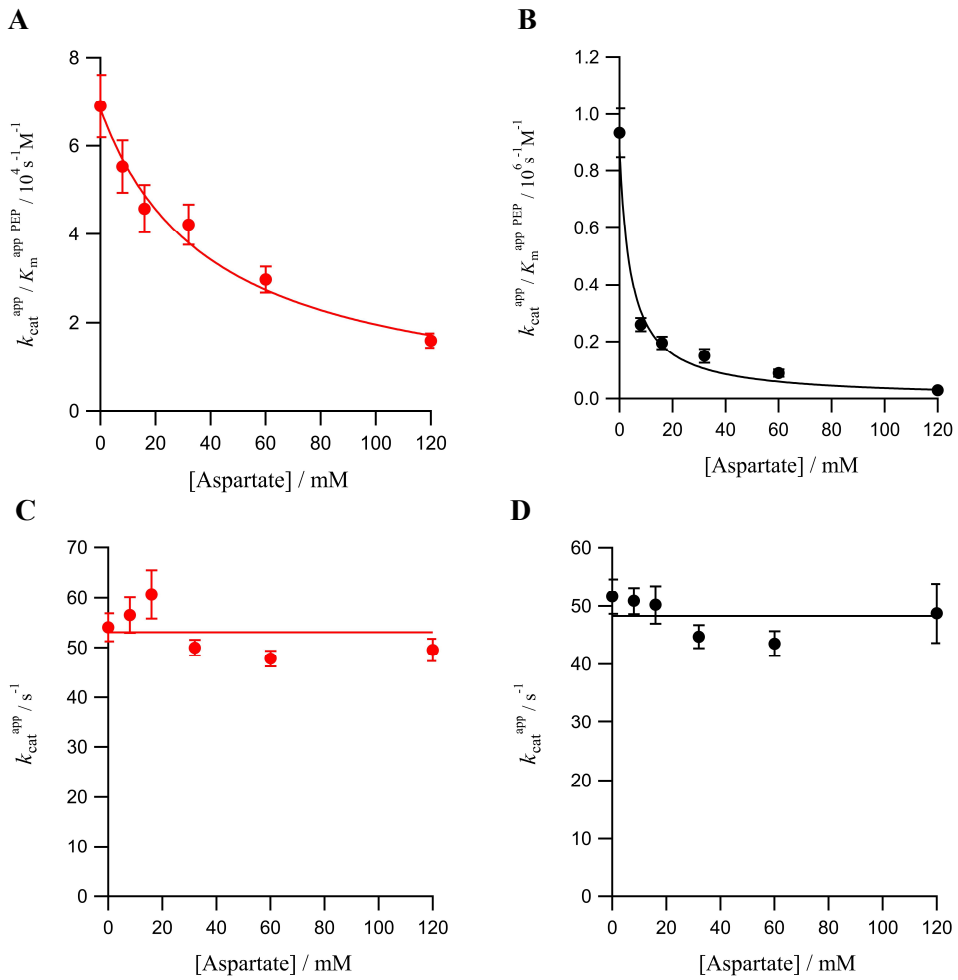


Figure 3: Plots of aspartate inhibition values for the PEPC enzymes. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 U μl^{-1} malate dehydrogenase, 10 mM $KHCO_3$ and 10 nM of *Flaveria trinervia* PEPC or 5 nM *Flaveria pringlei* PEPC. Markers are derived from Supp. Figure 5A for *Flaveria trinervia* PEPC and Supp. Figure 5B for *Flaveria pringeli* PEPC, error bars represent standard errors. Secondary plot lines are characterised by equation 2. **A** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against aspartate concentration for the enzyme *Flaveria trinervia* PEPC characterised by $K_{ic}^{Aspartate} = 40.02 \pm 6.49$ mM. **B** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against aspartate concentration for the enzyme *Flaveria pringlei* PEPC characterised by $K_{ic}^{Aspartate} = 4.31 \pm 0.60$ mM. **C** Secondary plot of k_{cat}^{app} against aspartate concentration showing no aspartate competitive inhibition of *Flaveria trinervia* PEPC. **D** Secondary plot of k_{cat}^{app} against aspartate concentration showing no aspartate competitive inhibition of *Flaveria pringlei* PEPC.

Table 1: Summary of kinetic parameters found in this study. Standard errors are given, based on fitted theoretical curves.

Species	k_{cat} / s^{-1}	K_m^{PEP} / mM	k_{cat}/K_m^{PEP} / s^{-1} M^{-1}	K_m^{HCO3-} / mM	k_{cat}/K_m^{HCO3-} / s^{-1} $^{s}M^{-1}$
Flaveria trinervia (C_4)	47.99 ± 1.21	0.60 ± 0.05	$7.87 \times 10^4 \pm 5.43 \times 10^3$	0.065 ± 0.007	$6.92 \times 10^5 \pm 4.17 \times 10^4$
Flaveria pringlei (C_3)	52.65 ± 1.37	0.056 ± 0.001	$9.35 \times 10^5 \pm 8.49 \times 10^4$	0.099 ± 0.007	$4.43 \times 10^5 \pm 2.17 \times 10^4$

Table 2: Summary of inhibition parameters found in this study. Standard errors are given, based on fitted theoretical curves.

Species	K_{ic}^{Malate} / mM	K_{iu}^{Malate} / mM	$K_{ic}^{Aspartate}$ / mM
Flaveria trinervia (C_4)	10.96 ± 1.55	40.72 ± 4.59	40.02 ± 6.49
Flaveria pringlei (C_3)	2.14 ± 0.62	4.56 ± 1.72	4.13 ± 0.60

DISCUSSION

Adaptation of kinetics involved opposite changes in specificities of the two substrates

The specificity for bicarbonate of the C_4 PEPC is approximately one-third higher than that of the non-C_4 PEPC (Figure 1). This difference in bicarbonate specificity suggests that the adaptation of *Flaveria* PEPC for the C_4 context was in part driven by the availability of bicarbonate. The concentration of bicarbonate is determined by the equilibrium between dissolved carbon dioxide and bicarbonate in the cytosol. The enzyme carbonic anhydrase speeds up the interconversion of carbon dioxide and bicarbonate, and high levels are found in the cytosol of mesophyll cells in C_4 leaves (Tetu *et al.*, 2007). Carbonic anhydrase, of course, cannot influence the equilibrium position, only the speed at which it is reached. In the C_4 pathway, the high activity of PEPC also requires a high carbonic anhydrase activity to prevent depletion of bicarbonate (Hatch and Burnell, 1990), although C_4 plants can function without the carbonic anhydrase if CO_2 concentrations are high enough (Studer *et al.*, 2014). At higher temperatures or in saline environments, typical of C_4 species, the availability of bicarbonate decreases, which suggests that an increase in specificity for bicarbonate of PEPC is necessary for high flux demands of the C_4 cycle. Because of their involvement in different cycles, the non-C_4 and C_4 PEPC enzymes are likely to be differentially limited by the PEP and bicarbonate substrates.

The observed different of bicarbonate specificity contrasts with the change in specificity for PEP of the C_4 PEPC, which is ten times lower than that of the non-C_4 PEPC (Figure 1). The primary function of the non-C_4 PEPC is anaplerotic, replenishing oxaloacetate for key metabolic pathways such the citric acid cycles (O'Leary, Park and Plaxton, 2011). In this context, the enzyme deals with relatively low substrate concentrations, but the function is likely to require a fast response to small concentration changes. It is therefore likely that high specificity for PEP is strongly selected for in non-C_4 PEPC. Because the overall activity of non-C_4 PEPC is low, the background bicarbonate concentration is likely to be sufficient for anaplerotic reactions, potentially relaxing pressures on bicarbonate specificity. Following its co-option for the C_4 cycle, selective pressures on PEPC are likely altered. Indeed, C_4

PEPC plays a central role in the high-flux C_4 cycle (Svensson, Bläsing and Westhoff, 2003; Stitt and Zhu, 2014). The activity of carbonic anhydrase is necessary for the C_4 cycle at low-CO_2 concentrations (Osborn et al. 2017), demonstrating that bicarbonate is a limiting factor. We therefore suggest that the specificity of PEPC towards bicarbonate is increased during C_4 evolution to boost the rate of the pathway.

In contrast to the observed change in bicarbonate affinity, the specificity of PEPC towards PEP is substantially lower in the C_4 enzymes than the C_3 forms (Svensson, Bläsing and Westhoff, 1997; Blasing, Westhoff and Svensson, 2000; Jacobs *et al.*, 2008). This presents an interesting conundrum; why on adaption to a high flux pathway have the kinetic properties of PEPC moved in a direction expected to reduce the flux through the pathway? One potential explanation arises from the high intracellular concentration of PEPC. In the *Zea mays* leaf, the concentration of C_4 PEPC is *ca.* 0.14 mM, and is therefore likely to be greater than this in the mesophyll (McNaughton *et al.*, 1989; Jiao and Chollet, 1991). This enzyme concentration is comparable to the concentration of PEP in mesophyll cells, at *ca.* 0.38 mM in the C_4 species *Zea mays* (Arrivault et al., 2017). At these concentrations, a C_3 type K_m^{PEP} (*i.e.* 0.06 mM, Table 1) would result in a substantial fraction of the intracellular PEP being bound to the enzyme, while the order-of-magnitude larger C_4 type K_m^{PEP} results in much more of the PEP pool being unbound and thus available to other enzymes and metabolic pathways (Gowik and Westhoff, 2011). The magnitudes of changes in affinity for PEP and bicarbonate are very different, therefore we proposed that the two are decoupled, and results from distinct selective pressures on enzymes of the C_4 pathway.

The C_4 form is tightly controlled by C_4 metabolites

It is well established that malate acts as an inhibitor of PEPC and comparative work in *Flaveria* has shown that the C_4-specific form of the enzyme is less sensitive to malate than the C_3 form (Svenssonz, and Westhoff, 1997; Paulus, Schlieper and Groth, 2013). Previous work has shown that malate acts as a mixed inhibitor towards the C_4 *Zea mays* PEPC, inhibiting the enzyme at both saturating PEP and limiting PEP concentrations, and that the

form of the inhibition varies with the source and storage of the enzyme (Wedding, Black and Meyer, 1990). As a result, the both the type and the extent of inhibition may have changed on the transition between a C_3-type and a C_4-type form of PEPC. To investigate this possibility, we have determined the inhibition behaviour of two critical four-carbon inhibitors of PEPC, malate and aspartate, towards both C_3 and C_4 forms of the enzyme from the *Flaveria* model system.

Our investigations have shown that malate exhibits mixed inhibition of both isoforms of PEPC, but inhibition differs quantitatively among the two enzymes (Figure 2). The C_4 isoform is five times less sensitive to competitive inhibition and ten times less sensitive to non-competitive inhibition when compared to the C_3 isoform. So, while the balance between the two forms of inhibition differs between the C_3 and C_4 isoforms, both retain the same general pattern that malate will inhibit to some extent at all PEP concentrations. In the case of aspartate, again the type of inhibition remains the same in both *Flavaria* enzymes; aspartate competitively inhibits both the C_4 and C_3 PEPC forms (Figure 3). In general, C_4 evolution seems to have driven a decrease of sensitivity towards both of these inhibitors.

Sensitivity to both malate (Figure 2) and aspartate inhibition (Figure 3) is reduced in the C_4 PEPC as compared to the non-C_4 isoform. The high sensitivity of the non-C_4 enzyme likely allows tight control by relatively low concentrations of feedback inhibitor. On the other hand, the C_4-specific isoform functions in a context where metabolite concentrations can be high; the concentrations of malate and aspartate have been estimated at 11.18 mM and 1.17 mM respectively in *Zea mays* (Arrivault *et al.*, 2017), requiring that PEPC has a reduced sensitivity to these inhibitors.

EXPERIMENTAL PROCEDURES

Unless otherwise stated, reagents and components were from Sigma. For purification, unless otherwise stated the equipment used procured from GE Healthcare. unless otherwise stated enzymes and *E. coli* strains were from NEB.

DNA Preparation

Plasmids that encode the *Flaveria trinervia* PEPC gene and the *Flaveria pringlei* PEPC gene in the pTrc 99A plasmid were gifted by Peter Westhoff. The PEPC genes were sub cloned into the pET-1B His6 TEV LIC vector plasmid, gifted by Scott Gradia (University of California, Berkeley, Addgene plasmid # 29653). Genes were copied with PCR using the Q5 polymerase and the Flv1BFor and Flv1BRev primers (Primers synthesised by Sigma). Genes were sub cloned using the ligation independent cloning method with Q5 DNA polymerase and T4 DNA polymerase (NEB). Cloned plasmids were isolated using a Miniprep DNA kit (Qiagen). Plasmids were Sanger sequenced (GATC Biotech) using the T7 promotor, T7 terminator, Flav_1303_Seq_For and Flav_1832_Seq_Rev primers (Primers summarised in Table 3).

Protein Expression

For protein expression, BL21λ(DE3) strain *E. coli* (NEB) was used. Chemically competent *E. coli* cells were transformed with each of the plasmids. Eight litres of cultures were grown in LB medium at 37°C to OD_{600} 0.8. Cultures were cooled to 4°C for one hour prior to recombinant protein induction with 0.5 mM IPTG (Fischer). Cultures were then incubated at 18°C for 18 hours. Cells were harvested by centrifugation at 5,422 × *g* for 25 minutes and stored at -80°C.

Protein Purification

Cells were suspended with immobilized metal affinity column (IMAC) buffer (25 mM Tris, 0.5 M NaCl, 0.3 M glycerol, 20 mM imidazole (Acros Scientific), 10 ml per 2 L of culture with 50 µl of 50mg ml^{-1} DNase I and 100 µl of 100 mg ml^{-1} Pefabloc. Cells were passed twice through a cell disruptor (Constant Systems) before centrifugation at 26,902 × *g* for 40 minutes. The supernatant was passed through a 0.45 µm pore filter (Elkay Labs.).

PEPC was separated from soluble protein with a prepacked 1 ml nickel affinity column using an ÄKTA™ Pure 25 L Chromatography System. The loaded column was washed with 50 column volumes of IMAC buffer, then 50 column volumes of IMAC buffer containing 150 mM imidazole. Pure PEPC was eluted with 10 column volumes of IMAC buffer containing 400 mM imidazole.

Protein eluted from IMAC purification was loaded onto a Sephadex G50 desalting column (Amersham Biosciences) and rebuffered in storage buffer (20 mM Tris, 5% v/v glycerol, 150 mM KCl, 1 mM DTT (AnaSpec. Inc). Protein was concentrated to *ca.* 12-15 µM with a Vivaspin 20 MWCO 3000 (Sartorius), aliquoted and frozen at -80°C until use.

Enzyme Quantification

PEPC enzyme concentration was quantified by absorption at 280 nm. Enzyme extinction coefficient was calculated using the ExPASy protein parameter tool and corrected by determining the absorbance of the protein denatured in 6 M guanidine hydrochloride (Gill and von Hippel, 1989). The difference between the denatured and folded protein at 280 nm was used to adjust the extinction coefficient of the protein. The extinction coefficient for *Flaveria trinervia* PEPC was determined to be 120480 M^{-1} cm^{-1}, the extinction coefficient for *Flaveria pringlei* PEPC was determined to be 117030 M^{-1} cm^{-1}. A difference of -0.5% and -6.7% between predicted absorbance coefficients for *Flaveria trinervia* and *Flaveria pringlei* PEPC were observed respectively. It is assumed that all enzyme used to initialise the assay was active.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE) Analysis

Total protein concentration for purification efficiency was determined using the BCA Pierce quantification kit (Thermo Scientific). Concentration was determined using a standard curve performed with bovine serum albumin, over a concentration range 0 – 2.0 mg ml^{-1}.

Protein samples were analysed for purity using SDS PAGE analysis. Protein samples were quantified using the BCA Peirce method, 25 µg of cell lysis and 5 µg of pure protein elutions were denatured in 2 × SDS PAGE loading dye (200 mM Tris.HCl pH 6.8, 2

% SDS, 20 % Glycerol, 0.01% Bromophenol blue (BDH Laboratory Supplies) and 7 % β-mercaptoethanol). Protein was loaded onto an 8% acrylamide SDS gel with 2 µl of Blue Prestained Protein Standard Broad Range (11-190 kDa) (NEB). Gels were run for 50 minutes at 200 V with 1 × Tris/Glycine/SDS running buffer (Geneflow). Gels were stained with InstantBlue (Expedeon) and imaged with a ChemiDoc™ MP (BioRad).

Enzyme Assays

PEPC activity was measured spectroscopically at 340 nm by coupling to NADH-malate dehydrogenase. Assays with a high fixed concentration of bicarbonate were observed using a FLUOstar plate reader (BMG Labtech) using the 340 nm ± 5 nm absorbance filter (BMG Labtech). Plate reader assays were conducted in a reaction volume of 150 µl at 25°C. Typical reaction mixture contained 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$ (Fluka), 5 mM $KHCO_3$. 0.2 mM NADH (Fischer) and 0.1 Uμl^{-1} malate dehydrogenase. Assays were initiated with the addition of PEPC enzyme. Rates were calculated with a NADH calibration curve.

Assays at a range of bicarbonate concentrations were observed with a Cary 300 Bio spectrophotometer (Agilent Technologies) in the same reaction buffer, in a total reaction volume of 600 µl. In bicarbonate assays, the water and tricine buffer were sparged with nitrogen for 18 hours prior to use in assays. Bicarbonate assays were constructed under a nitrogen flow. Assays were performed in a sealed cuvette. The reaction was initiated with the addition of 50 nM PEPC, delivered with a gastight syringe (Hamilton). Bicarbonate concentrations were controlled with the addition of freshly prepared potassium bicarbonate. Background bicarbonate was determined using an endpoint assay with no potassium bicarbonate, run for 30 minutes. Rates were calculated using the Cary analysis software.

Data Analysis

Kinetic parameters were evaluated by non-linear regression analysis in Igor Pro (Version 7.0.8.1; Wavemetrics Inc., Lake Oswego Orgeon). The following equations were used:

(Equation 1)

$$v_i/[E_T] = \frac{k_{cat} \times [S]}{K_m + [S]}$$

Equation 1, where $v_i/[E_T]$ is the steady state rate divided by the total enzyme concentration, k_{cat} is the first order rate constant, K_m is the Michaelis constant, and S is the substrate concentration.

(Equation 2)

$$k^{app} = \frac{k}{1 + [i]/K_i}$$

Equation 2, where k^{app} is the apparent rate constant, k is the uninhibited constant, i is the inhibitor concentration, and K_i is the inhibition constant.

REFERENCES

Arrivault, S., Obata, T., Szecówka, M., Mengin, V., Guenther, M., Hoehne, M., Fernie, A. R. and Stitt, M. (2017) Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation., *Journal of experimental botany*, 68(2), pp. 283–298.

Atkinson, R. R. L., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M. and Osborne, C. P. (2016) C_4 photosynthesis boosts growth by altering physiology, allocation and size, *Nature Plants*. Nature Publishing Group, 2(5), p. 16038.

Aubry, S., Brown, N. J. and Hibberd, J. M. (2011) The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway, *Journal of Experimental Botany*, 62(9), pp. 3049–3059.

Bauwe, H. (1986) An efficient method for the determination of K_m values for HCO_3^- of phosphoenolpyruvate carboxylase, *Planta*, 169(3), pp. 356–360.

Blasing, O. E., Westhoff, P. and Svensson, P. (2000) Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria- a conserved serine residue in the carboxyterminal part of the enzyme is a major determinant for C_4-specific characteristics, *Journal of Biological Chemistry*, 275(36), pp. 27917–23.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P. and Weber, A. P. M. (2014) Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species, *Journal of experimental botany*, 65(13), pp. 3579–3593.

von Caemmerer, S. and Furbank, R. T. (2003) The C_4 pathway: an efficient CO_2 pump, *Photosynthesis Research*, 77(2/3), pp. 191–207.

Christin, P.-A., Arakaki, M., Osborne, C. P. and Edwards, E. J. (2015) Genetic Enablers Underlying the Clustered Evolutionary Origins of C_4 Photosynthesis in Angiosperms, *Molecular Biology and Evolution*, 32(4), pp. 846–858.

Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J. and Osborne, C. P. (2013) Parallel Recruitment of Multiple Genes into C_4 Photosynthesis, *Genome Biology and Evolution*, 5(11), pp. 2174–2187.

Ehleringer, J. and Björkman, O. (1977) Quantum yields for CO_2 uptake in C_3 and C_4 plants, *Plant Physiology*, 59(577), pp. 86–90.

Engelmann, S., Bläsing, O. E., Gowik, U., Svensson, P. and Westhoff, P. (2003) Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria*- a gradual increase from C_3 to C_4 characteristics., *Planta*, 217(5), pp. 717–25.

Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Analytical Biochemistry*, 182(2), pp. 319–326.

Gowik, U. and Westhoff, P. (2011a) C_4 Phosphoenolpyruvate Carboxylase, in Raghavendra, A. S. and Sage, R. F. (eds) *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*. 2011 ed. Dordrecht, Netherlands: Springer, pp. 257–275.

Gowik, U. and Westhoff, P. (2011b) The path from C_3 to C_4 photosynthesis., *Plant Physiology*, 155(1), pp. 56–63.

Hatch, M. D. (1987) C_4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure, *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 895(2), pp. 81–106.

Hatch, M. D. and Burnell, J. N. (1990) Carbonic Anhydrase Activity in Leaves and Its Role in the First Step of C_4 Photosynthesis, *Plant Physiology*, 93(2), pp. 825–828.

Jacobs, B., Engelmann, S., Westhoff, P. and Gowik, U. (2008) Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*: Determinants for high tolerance towards the inhibitor L-malate, *Plant, Cell and Environment*, 31, pp. 793–803.

Janc, J. W., O'Leary, M. H. and Cleland, W. W. (1992) A kinetic investigation of phosphoenolpyruvate carboxylase from *Zea mays*, *Biochemistry*, 31(28), pp. 6421–6426.

Jiao, J. -a. and Chollet, R. (1991) Posttranslational Regulation of Phosphoenolpyruvate Carboxylase in C_4 and Crassulacean Acid Metabolism Plants, *Plant Physiology*, 95(4), pp. 981–985.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. and Gowik, U. (2017) C_3 cotyledons are followed by C_4 leaves: intra-individual transcriptome analysis of *Salsola soda* (*Chenopodiaceae*), *Journal of Experimental Botany*, 68(2), pp. 161–176.

McKown, A. D., Moncalvo, J.-M. and Dengler, N. G. (2005) Phylogeny of *Flaveria* (*Asteraceae*) and inference of C_4 photosynthesis evolution, *American Journal of Botany*, 92(11), pp. 1911–1928.

McNaughton, G. A. L., Fewson, C. A., Wilkins, M. B. and Nimmo, H. G. (1989) Purification, oligomerization state and malate sensitivity of maize leaf phosphoenolpyruvate carboxylase, *The biochemical journal*, 261(2), pp. 349–55.

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. and Christin, P. A. (2018) Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis, *Molecular Biology and Evolution*, 35(1), pp. 94–106.

Nisbet, E. G., Grassineau, N. V., Howe, C. J., Abell, P. I., Regelous, M. and Nisbet, R. E. R. (2007) The age of Rubisco: the evolution of oxygenic photosynthesis, *Geobiology*, 5(4), pp. 311–335.

O'Leary, B., Park, J. and Plaxton, W. C. (2011) The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs, *Biochemical Journal*, 436(1), pp. 15–34.

Paulus, J. K., Schlieper, D. and Groth, G. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution, *Nature communications*, 4, p. 1518.

Sage, R. F. (2004) The evolution of C_4 photosynthesis, *New Phytologist*, 161(2), pp. 341–370.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) Photorespiration and the Evolution of C_4 Photosynthesis, *Annual Review of Plant Biology*, 63(1), pp. 19–47.

Skillman, J. B. (2007) Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark, *Journal of Experimental Botany*, 59(7), pp. 1647–1661.

Stitt, M. and Zhu, X.-G. (2014) The large pools of metabolites involved in intercellular metabolite shuttles in C_4 photosynthesis provide enormous flexibility and robustness in a fluctuating light environment, *Plant, Cell & Environment*, 37(9), pp. 1985–1988.

Studer, A. J., Gandin, A., Kolbe, A. R., Wang, L., Cousins, A. B. and Brutnell, T. P. (2014) 'A Limited Role for Carbonic Anhydrase in C_4 Photosynthesis as Revealed by a ca1ca2 Double Mutant in Maize', *Plant Physiology*, 165(2), pp. 608–617.

Svensson, P., Bläsing, O. E. and Westhoff, P. (1997) 'Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase- a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3)., *European journal of biochemistry*, 246(2), pp. 452–60.

Svensson, P., Bläsing, O. E. and Westhoff, P. (2003) Evolution of C_4 phosphoenolpyruvate carboxylase, *Archives of Biochemistry and Biophysics*, 414(2), pp. 180–188.

Tcherkez, G. G. B., Farquhar, G. D. and Andrews, T. J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized, *Proceedings of the National Academy of Sciences*, 103(19), pp. 7246–7251.

Tetu, S. G., Tanz, S. K., Vella, N., Burnell, J. N. and Ludwig, M. (2007) The *Flaveria* bidentis beta-Carbonic Anhydrase Gene Family Encodes Cytosolic and Chloroplastic Isoforms Demonstrating Distinct Organ-Specific Expression Patterns, *Plant Physiology*, 144(3), pp. 1316–1327.

Wedding, R. T., Black, M. K. and Meyer, C. R. (1990) Inhibition of Phosphoenolpyruvate Carboxylase by Malate, *Plant Physiology*, 92(2), pp. 456–461.

Westhoff, P. (2004) Evolution of C_4 Phosphoenolpyruvate Carboxylase. Genes and Proteins: a Case Study with the Genus *Flaveria*, *Annals of Botany*, 93(1), pp. 13–23.

SUPPLEMENTARY DATA

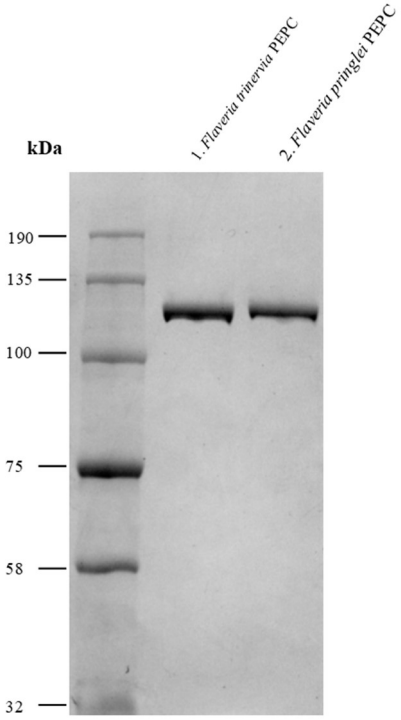


Figure S 1: 10% acrylamide SDS PAGE analysis of *Flaveria trinervia* PEPC and *Flaveria pringlei* PEPC. Lane one contains *ca.* 6 µg of *Flaveria trinervia* PEPC protein eluted from a nickel IMAC column. Lane two contains *ca.* 6 µg of *Flaveria pringlei* PEPC protein eluted from a nickel IMAC column. A *ca.* 120 kDa band is seen in each lane corresponding to PEPC, no other bands are detected.

Kinetic Assay Results

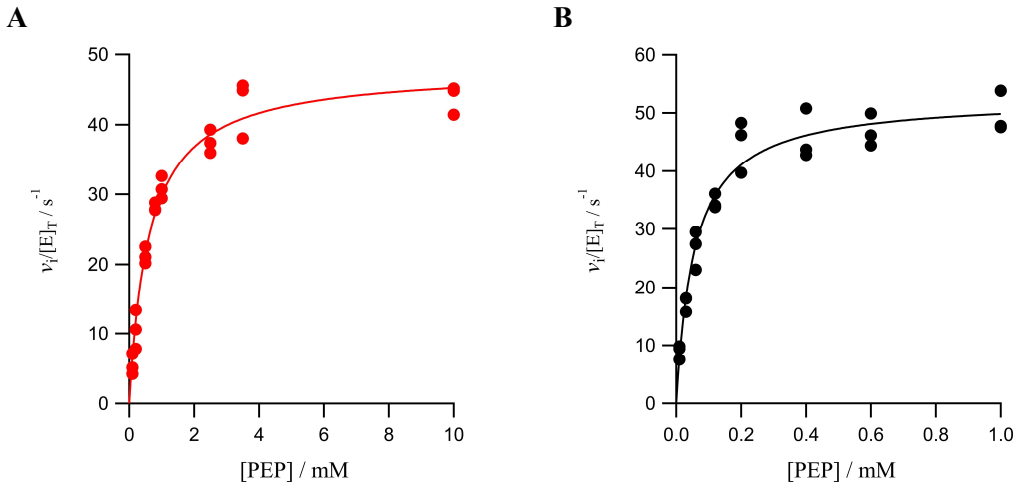


Figure S 2: Rate of oxaloacetic acid formation, catalysed by PEPC, varying the concentration of phosphoenolpyruvate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM $KHCO_3$, 0.15 mM NADH and 0.01 Uμl^{-1} malate dehydrogenase. Assays were repeated (n = 3) for each concentration of PEP. **A** Filled circles represent experimental data points for *Flaveria trinervia*. The line is described by equation 1, with characterising parameters k_{cat} = 47.99 ± 1.21 s^{-1}, K_m^{PEP} = 0.60 ± 0.05 mM, and k_{cat}/K_m^{PEP} = 78735 ± 5430 $s^{-1}M^{-1}$. **B** Filled circles represent experimental data points for *Flaveria pringlei.* The line is described by equation 1, with characterising parameters k_{cat} = 52.65 ± 1.37 s^{-1}, K_m^{PEP} = 0.056 ± 0.0006 mM and k_{cat}/K_m^{PEP} = 934550 ± 84900 s^{-1} M^{-1}.

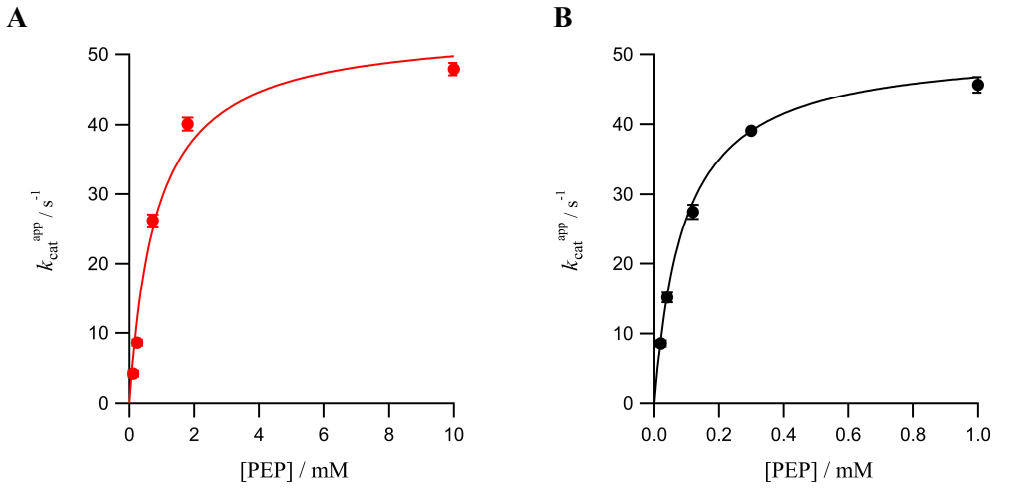


Figure S 3: Secondary plot of the k_{cat}^{app} parameter from bicarbonate assays in Figure 1. A Filled circles represent the k_{cat}^{app} parameter from Figure 1A for the enzyme the *Flaveria trinervia* PEPC, error bars represent standard errors from fit of lines. The line described by equation 1, characterised by the parameters k_{cat} = 53.89 ± 4.12 s^{-1} and K_m^{PEP} = 0.84 ± 0.02 mM. **B** Filled circles represent the k_{cat}^{app} parameter from Figure 1C for the enzyme *Flaveria pringlei* PEPC, error bars represent standard errors from fit of lines, the line is described by equation 1, characterised by the parameters k_{cat} = 51.01 ± 0.05 s^{-1} K_m^{PEP} = 0.077 ± 0.007 mM.

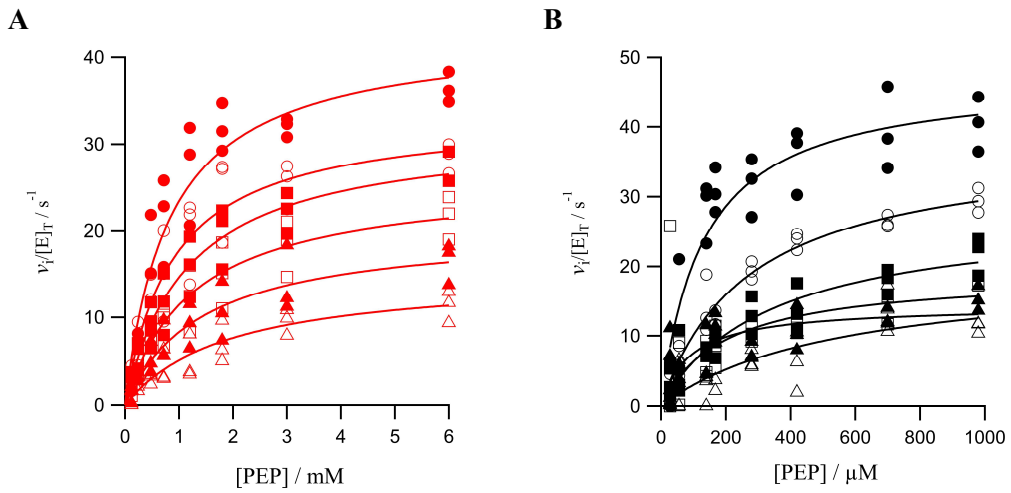


Figure S 4: Primary plot of PEPC inhibited by malate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM $KHCO_3$ 0.2 mM NADH and 0.01 Uμl^{-1} malate dehydrogenase. The lines are described by equation 1. **A** Points indicate experimental data run for *Flaveria trinervia* PEPC, filled circles indicates no inhibitor, open circles indicate the presence of 8 mM malate, filled squares indicate the presence of 16 mM malate, open squares indicate the presence of 32 mM malate, filled triangles indicate the presence of 60 mM malate and open triangles indicate the presence of 120 mM malate. **B** Points indicate experimental data runs, filled circles indicates no inhibitor, open circles indicate the presence of 4 mM malate, filled squares indicate the presence of 12 mM malate, open squares indicate the presence of 24 mM malate, filled triangles indicate the presence of 32 mM malate and open triangles indicate the presence of 60 mM malate.

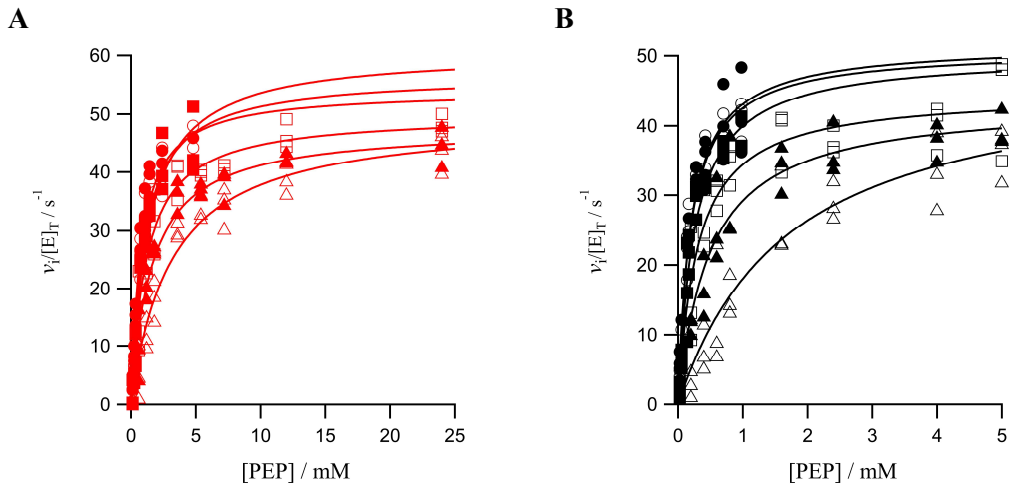


Figure S 5: Primary plot of PEPC inhibited by aspartate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM $KHCO_3$, 0.2 mM NADH and 0.01 Uμl^{-1} malate dehydrogenase. The lines are described by equation 1. **A** Points indicate experimental data run for *Flaveria trinervia* PEPC, filled circles indicates no inhibitor, open circles indicate the presence of 8 mM aspartate, filled squares indicate the presence of 16 mM aspartate, open squares indicate the presence of 32 mM aspartate, filled triangles indicate the presence of 60 mM aspartate and open triangles indicate the presence of 120 mM aspartate. **B** Points indicate experimental data run for *Flaveria pringlei* PEPC, open circles indicate the presence of 8 mM aspartate, filled squares indicates the presence of 16 mM aspartate, open squares indicate the presence of 32 mM aspartate, filled triangles indicate the presence of 60 mM aspartate and open triangles indicate the presence of 120 mM aspartate.

Primers

Table 3: Summary of the primers used in this study for cloning and sequencing.

Primer	Sequence, 5' to 3'
FlvFor1B	TACTTCCAATCCAATGCAATGGCTAACCGGAAT
FlvRev1B	TTATCCACTTCCAATGTTATTACTAACCGGTGTTCTGC
Flav_1303_Seq_For	AGACAAGTGTCGACTT
Flav_1832_Seq_Rev	TTGTAGAGCTGCCATG
T7 Promotor	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG

Plasmid Sequences

>pET-1B_*Flaveria_trinervia*_PEPC

AACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGAT
GCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGC
TGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCC
CGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGA
GAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTT
CCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTG
TGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCG
GCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAG
CAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCAT
GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATT
AGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGG
CGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGG
CGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
GGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATT
GTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGTTCTTCTC
ACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCAATGGCTAACCGGAATGTGGAGAAATTAGCA
TCGATCGATGCTCAGTTGAGGCTTTTAGTCCCTGGGAAAGTTTCTGAGGATGATAAGCTTGTTGAGTATGATGCT
TTGCTTTTGGATAAGTTTCTGGATATTCTTCAGGATTTGCATGGGGAAGATCTCAAGGAAGCGGTTCAACAATGC
TATGAGCTATCTGCTGAATATGAAGGAAAACATGACCCGAAGAAGCTGGAGGAGCTTGGAAGTCTGTTGACAAG
TTTAGATACAGGGGATTCCATTGTCATTGCAAAAGCCTTTTCTCACATGCTTAACTTAGCCAATCTGGCTGAAGA
ACTTCAGATTGCTTACCGCCGAAGAATCAAACTGAAGAGTGGTGATTTTGCTGACGAGGCTAACGCAACAACTG
AATCAGATATTGAAGAAACTTTCAAGAGACTTGTGCATAAGCTTAACAAGTCCCCTGAAGAGGTTTTTGATGCAC
TGAAGAATCAAACTGTTGAGTTGGTCTTGACTGCTCATCCAACTCAATCCGTCCGTAGATCTTTGCTTCAAAAGC
ATGGAAGGATTCGTAATTGTCTGGCCCAGTTGTATGCCAAAGACATCACTCCTGATGATAAGCAGGAACTCGATG
AGGCTTTGCATAGAGAAATTCAAGCTGCATTCCGTACTGATGAAATCAGAAGGACTCCACCAACACCACAAGAT
GAAATGAGAGCAGGAATGAGTTACTTCCATGAAACAATCTGGAAGGGTGTTCCAAAATTCTTACGTCGTGTTGA
CACTGCCCTAAAGAATATTGGAATTAATGAACGTTTTCCCTATAATGCACCTCTAATTCAATTCTCTTCATGGATG
GGTGGTGACCGTGATGGCAATCCGAGGGTTACTCCTGAGGTAACAAGGGATGTTTGTTTGCTAGCCAGAATGAT
GACGTCAAACATGTACTTTTCTCAGATAGAGGATCTTATGATTGAGATGTCCATGTGGCGTTGTAATAGTGAATT
ACGTGTTCGAGCAGAAGAACTGTACAGAACAGCAAGAAAAGATGTGAAGCACTACATAGAGTTTTGGAAACGG
ATTCCTCCCAATCAACCTTATCGTGTAATTCTTGGTGATGTAAGGGACAAATTATATAATACACGTGAACGATCT
CGTCATTTATTGGTCGATGGGAAATCTGACATCCCAGACGAAGCTGTTTATACCAATGTTGAACAGCTCTTGGAA
CCACTGGAGCTATGCTACAGATCACTATGTGACTGTGGTGACCATGTGATTGCTGATGGAAGCCTTCTTGATTTTC
TAAGACAAGTGTCGACTTTTGGACTCTCACTTGTAAAACTTGATATAAGACAAGAATCTGACCGTCACACTGAAG
TCCTTGATGCAATCACTCAACATTTAGGAATTGGGTCCTATCGTGAGTGGTCTGAAGAAAAACGCCAAGAATGGC
TTCTAGCTGAACTCAGTGGAAAACGTCCTCTTATTGGTCCAGACCTTCCAAAAACTGAGGAAGTTAAGGATTGTT
TAGACACGTTTAAGGTTTTAGCAGAACTCCCGTCTGACTGTTTCGGTGCTTACATCATCTCAATGGCCACATCAAC
TTCTGATGTCCTTGCTGTTGAGCTTCTCCAGCGTGAATACCATATAAAACATCCGTTACGCGTGGTCCCCTTATTT
GAAAAACTTGCTGACCTGGAGGCGGCCCCTGCGGCCATGACCCGCCTTTTCTCAATGGATTGGTACAGAAACCG
AATTGATGGTAAACAAGAAGTCATGATTGGGTACTCTGATTCAGGAAAAGATGCAGGCCGGTTCTCTGCTGCAT
GGCAGCTCTACAAAACTCAAGAACAGATTGTTAAAATTGCAAAAGAGTTTGGAGTCAAACTTGTTATATTTCATG
GGCGTGGTGGAACTGTTGGTAGAGGTGGTGGGCCCACACATCTGGCTCTTCTCTCTCAACCACCGGACACCATTA
ACGGGTCTTTAAGAGTGACAGTTCAGGGTGAGGTCATAGAGCAGTCGTTTGGTGAGGAACATTTGTGCTTTAGAA
CACTTCAGAGATTTTGTGCAGCTACACTTGAGCATGGGATGAACCCACCAATCTCACCACGACCCGAGTGGCGTG
AACTTATGGACCAGATGGCTGTTGTTGCAACCGAGGAGTACCGTTCTGTTGTGTTCAAGGAACCACGTTTTGTGG
AGTATTTCCGGCTTGCAACACCTGAACTGGAGTTCGGGCGTATGAATATTGGAAGTCGCCCATCAAAAAGAAAA
CCGAGTGGTGGCATTGAATCACTCAGAGCCATTCCATGGATCTTTTCATGGACTCAGACCAGGTTCCATCTCCCA
GTTTGGCTTGGGTTTGGGGCGGCGTTCAAACACGCCATCCAAAAAGACAGCAAGAATCTCCAAATGCTTCAAGA
AATGTACAAAACATGGCCTTTCTTTCGGGTCACCATTGATTTAGTTGAAATGGTGTTTGCTAAAGGTAACCCAGG
CATTGCTGCCCTGAATGACAAGCTCCTTGTTTCTGAAGATCTAAGGCCCTTTGGAGAATCTTTGAGAGCAAACTA
TGAAGAAACCAAAAATTATCTTCTCAAGATTGCTGGACATAAGGACCTTCTAGAGGGTGATCCCTACTTGAAACA
AGGAATCAGGCTGCGTGATCCGTACATCACAACCTTGAATGTATGCCAAGCTTATACCCTAAAGAGGATCCGTG
ACCCGAACTATCATGTGACATTAAGGCCTCATATTTCTAAAGAATATGCCGCTGAGCCGAGCAAACCAGCTGATG
AGCTTATCCACCTGAACCCAACCAGCGAGTACGCACCCGGTTTGGAGGACACGCTCATCTTGACCATGAAAGGG
ATTGCTGCTGGAATGCAGAACACCGGTTAGTAATAACATTGGAAGTGGATAACGGATCCGAATTCGAGCGCCGT
CGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGG
AAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG
GGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

>pET-1B_*Flaveria_pringlei*_PEPC

AACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGAT
GCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGC
TGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCC
CGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGA
GAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTT
CCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTG
TGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCG
GCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAG
CAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCAT
GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATT
AGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGG
CGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGG
CGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
GGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATT
GTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGTTCTTCTC
ACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCAATGGCTAACCGGAATTTGGAGAAATTAGCAT
CGATCGATGCTCAGTTGAGGCTTTTAGTCCCTGGGAAAGTTTCTGAGGATGATAAGCTTATTGAGTATGATGCTT
TGCTTTTGGATAAGTTTCTGGATATTCTTCAAGATTTGCATGGGGAAGATCTCAAGGAAGCGGTTCAAGAATGCT
ATGAGCTATCTGCTGAATATGAAGGAAAACATGACCCGAAGAAGCTGGAGGAGCTTGGAAGTGTGTTGACAAGT
TTAGATCCAGGGGATTCCATTGTCATTGCAAAAGCTTTTTCTCACATGCTTAACTTAGCCAATCTGGCTGAAGAA
GTTCAGATTGCTTACCGCCGAAGAATCAAACTGAAGAGAGGTGATTTTGCTGATGAGGCTAATGCAACAACTGA
ATCAGATATTGAAGAAACTTTCAAGAAACTTGTGCTTAAGCTTAACAAGTCCCCTGAAGAGGTTTTTGATGCACT
CAAGAATCAAACTGTTGACTTGGTCTTGACTGCTCATCCAACTCAATCCGTCCGCAGATCTTTGCTTCAAAAGCA
TGGAAGGATTCGCAACTGTCTGGCCCAGTTGTATGCCAAAGACATCACTCCTGATGATAAGCAGGAATTAGATG
AGGCTTTGCATAGAGAAATTCAAGCTGCATTTCGTACTGATGAAATCAGGAGGACCCCACCAACACCACAAGAT
GAAATGAGAGCAGGAATGAGTTACTTCCATGAAACAATCTGGAAGGGTGTTCCTAAATTCTTACGTCGTGTTGAC
ACCGCCCTAAAGAATATAGGGATTAATGAACGTGTTCCCTATAATGCACCTCTAATTCAATTCTCTTCATGGATG
GGTGGTGACCGTGATGGCAATCCGAGGGTAACTCCTGAGGTAACGAGGGATGTTTGTTTGCTAGCCAGAATGAT
GGCTTCAAACATGTACTTTTCTCAGATAGAGGATCTTATGTTTGAGATGTCCATGTGGCGTTGTAATAGTGAACT
ACGTGTTCGAGCAGAAGAACTATATAGAACAGCAAGAAGAGATGTGAAGCACTACATAGAGTTCTGGAAACAG
GTTCCTCCCACTGAACCTTATCGTGTAATTCTTGGTGATGTAAGGGACAAATTATATAATACACGTGAACGATCT
CGCCATTTATTAGCCCATGGGATATCTGACATCCCAGAAGAAGCTGTTTATACCAATGTTGAACAGTTCTTGGAA
CCACTGGAGCTATGCTACAGATCACTATGTGACTGTGGTGACCGTGTGATTGCTGATGGAAGCCTTCTTGATTTTC
TAAGACAAGTGTCGACTTTTGGACTCTCACTTGTAAAACTTGATATAAGACAAGAATCTGACCGTCACACTGACG
TCCTTGATGCAATCACTCAACATTTAGAAATTGGGTCCTACCGTGAGTGGTCTGAAGAAAAACGCCAAGAATGG
CTTCTAGCTGAACTCAGTGGAAAACGTCCTCTTTTCGGTTCAGACCTTCCAAAAACTGAGGAAGTTAAGGATGTT
TTAGACACGTTTAATGTTTTAGCAGAACTCCCATCTGACTGTTTCGGTGCTTACATCATCTCAATGGCCACATCAC
CTTCTGATGTCCTTGCTGTTGAGCTTCTCCAACGTGAATGCCATGTAAAACATCCGTTACGCGTGGTCCCCCTATT
TGAAAAACTTGCTGACCTAGAGGCGGCCCCTGCGGCCATGGCCCGCCTTTTCTCAATCGATTGGTACAGAAATCG
GATCGACGGTAAACAAGAAGTCATGATTGGGTACTCTGATTCAGGAAAAGATGCAGGCCGGTTTTCTGCTGCAT
GGCAGCTCTACAAAGCTCAAGAAGAGATTATTAAAGTTGCAAAAGAGTTTGGGGTCAAACTTGTTATATTTCATG
GGCGTGGGGGGACTGTTGGTAGAGGTGGCGGGCCCACACATTTAGCTATCCTCTCTCAACCACCAGACACCATTC
ACGGGTCGTTAAGAGTCACGGTTCAGGGTGAGGTCATAGAGCAGTCGTTTGGTGAGGAACATTTGTGTTTTAGAA
CACTTCAGAGATTTTGTGCAGCTACACTTGAGCATGGGATGAACCCACCAATTTCACCACGGCCTGAGTGGCGTG
AACTTATGGACCAGATGGCTGTTGTTGCAACCGAGGAGTACCGTTCTATTGTGTTTAAGGAACCACGTTTTGTGG
AGTATTTCCGCCTTGCAACACCTGAATTGGAGTACGGGCGTATGAATATTGGAAGTCGCCCATCAAAAAGAAAA
CCTAGTGGTGGCATTGAATCACTCAGAGCCATTCCATGGATCTTTGCATGGACTCAGACCAGGTTCCATCTCCCA
GTTTGGCTTGGGTTTGGAGCGGCATTCAAACATGCCATTAAAAAAGACAGCAAGAATCTTCAAATGCTTCAAGA
AATGTACAAAACATGGCCTTTCTTTCGGGTCACCATTGATTTAGTTGAAATGGTGTTTGCTAAAGGAGACCCAGG
CATTGCTGCCTTGAATGACAAACTCCTTGTTTCTGAAGATCTATGGCCTTTTGGAGAATCTTTGAGAGCAAACTAT
GAAGAAACCAAAGATTATCTTCTCAAGATTGCTGGACACAGGGACCTTCTAGAGGGTGATCCCTACTTAAAACA
AAGAATCAGGCTGCGTGATTCATACATCACAACCTTAAATGTATGTCAAGCTTATACCCTAAAGCGGATCCGCGA
CCCGAACTATCATGTGACATTAAGGCCTCATATTTCCAAAGAATACGCCGCCGAGCCGAGCAAACCAGCTGACG
AGCTTATCCACCTGAACCCAACCAGTGAATACGCACCCGGTTTGGAGGACACGCTCATCTTGACCATGAAAGGG
ATTGCTGCTGGAATGCAGAACACCGGTTAGTAATAACATTGGAAGTGGATAACGGATCCGAATTCGAGCGCCGT
CGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGG
AAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG
GGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

Protein Amino Acid Sequence
>pET-1B *Flaveria trinervia* PEPC amino acid sequence

MGSSHHHHHHENLYFQSNAMANRNVEKLASIDAQLRLLVPGKVSEDDKLVEYDAL
LLDKFLDILQDLHGEDLKEAVQQCYELSAEYEGKHDPKKLEELGSLLTSLDTGDSIV
IAKAFSHMLNLANLAEELQIAYRRRIKLKSGDFADEANATTESDIEETFKRLVHKLN
KSPEEVFDALKNQTVELVLTAHPTQSVRRSLLQKHGRIRNCLAQLYAKDITPDDKQ
ELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALK
NIGINERFPYNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLARMMTSNMYFSQIE
DLMIEMSMWRCNSELRVRAEELYRTARKDVKHYIEFWKRIPPNQPYRVILGDVRDK
LYNTRERSRHLLVDGKSDIPDEAVYTNVEQLLEPLELCYRSLCDCGDHVIADGSLLD
FLRQVSTFGLSLVKLDIRQESDRHTEVLDAITQHLGIGSYREWSEEKRQEWLLAELS
GKRPLIGPDLPKTEEVKDCLDTFKVLAELPSDCFGAYIISMATSTSDVLAVELLQREY
HIKHPLRVVPLFEKLADLEAAPAAMTRLFSMDWYRNRIDGKQEVMIGYSDSGKDA
GRFSAAWQLYKTQEQIVKIAKEFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTIN
GSLRVTVQGEVIEQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMA
VVATEEYRSVVFKEPRFVEYFRLATPELEFGRMNIGSRPSKRKPSGGIESLRAIPWIFS
WTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFA
KGNPGIAALNDKLLVSEDLRPFGESLRANYEETKNYLLKIAGHKDLLEGDPYLKQGI
RLRDPYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEY
APGLEDTLILTMKGIAAGMQNTG*

>pET-1B *Flaveria pringlei* PEPC amino acid sequence

MGSSHHHHHHENLYFQSNAMANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDAL
LLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIV
IAKAFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKKLVLKLN
KSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCLAQLYAKDITPDDKQ
ELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALK
NIGINERVPYNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLARMMASNMYFSQIE
DLMFEMSMWRCNSELRVRAEELYRTARRDVKHYIEFWKQVPPTEPYRVILGDVRD
KLYNTRERSRHLLAHGISDIPEEAVYTNVEQFLEPLELCYRSLCDCGDRVIADGSLLD
FLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKRQEWLLAELS
GKRPLFGSDLPKTEEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRE
CHVKHPLRVVPLFEKLADLEAAPAAMARLFSIDWYRNRIDGKQEVMIGYSDSGKD
AGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTI
HGSLRVTVQGEVIEQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQM
AVVATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRAIPWIF
AWTQTRFHLPVWLGFGAAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVF
AKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDYLLKIAGHRDLLEGDPYLKQ
RIRLRDSYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSE
YAPGLEDTLILTMKGIAAGMQNTG*

Comparative biochemistry reveals greater adaptation of PEPC in older C_4 lineages

Nicholas R. Moody[1], Pascal-Antoine Christin[2] and James D. Reid[1]

[1]Department of Chemistry, University of Sheffield, Brook Hill, Sheffield, S3 7HF.

[2]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN.

Personal Contribution: I generated and analysed all the data and wrote the paper with the help of my co-authors. As of the time of thesis submission, this paper was in draft form..

ABSTRACT

C_4 photosynthesis is a complex trait that evolved repeatedly to remove the inefficiencies of carbon fixation in C_3 photosynthesis. In the C_4 carbon fixation cycle, the initial carbon fixation is catalyzed by phosphoenolpyruvate carboxylase (PEPC). PEPC is encoded by a multigene family, with multiple isoforms present in all plants and bacteria. The C_4-specific PEPC evolved from the co-option of an ancestral C_3 form, which was massively upregulated to sustain the high fluxes of the C_4 pathway. Previous work has suggested that its kinetic properties had been adapted to the demands of the metabolic cycle, but whether similar kinetic modifications occurred in all C_4 lineages remains unknown. In this work, we assess the kinetic differences between PEPC of a C_4 and a C_3 grass and contrast them to those of the distantly related *Flaveria*. Despite their evolutionary distance, the kinetic behaviour was modified in the same direction in the two groups, although differences are markedly stronger in grasses. This is likely explained by the age of the C_4 groups, which is about 16 million years ago for the grass as opposed to less than two in *Flaveria*. We suggest that the longer evolutionary period, potentially coupled with stronger selective pressure, lead to greater adaption of the grass PEPC to the demands of the C_4 cycle, as suggested based on comparisons of amino acid sequences. We conclude that the outcome of convergent origins of biochemical pathways depend on the length of the subsequent period of adaptation of the trait.

C_4 photosynthesis is a CO_2-concentrating mechanism that boosts productivity in tropical conditions (Atkinson *et al.*, 2016). The higher efficiency of C_4 plants results from the increased concentration of CO_2 around ribulose-bisphosphate carboxylase oxygenase (RuBisCO), the central enzyme of the Calvin-Benson cycle (Sage, Sage and Kocacinar, 2012). RuBisCO has a tendency to confuse CO_2 and O_2, and the fixation of the latter produces toxic compounds that need to be processed in the energetically costly photorespiration pathway (Tcherkez, Farquhar and Andrews, 2006; Nisbet *et al.*, 2007). In C_3 plants, RuBisCO is in direct contact with atmospheric gases, and photorespiration can

become consequential in conditions that decrease the relative concentration of CO_2, including high temperature, aridity and salinity (Ehleringer and Björkman, 1977; Skillman, 2007). C_4 plants tackle this problem by segregating primary carbon fixation from the enzyme RuBisCO into two cell types (Hatch, 1987; Sage, 2004; Sage, Sage and Kocacinar, 2012). In C_4 plants, atmospheric CO_2 is initially fixed by the enzyme phosphoenolpyruvate carboxylase (PEPC) in the form of bicarbonate, thus preventing oxygen binding (Hatch, 1987). PEPC produces the four carbon acid oxaloacetate, which is rapidly converted into the more stable acids malate or aspartate (Bräutigam *et al.*, 2014). The four carbon acids are shuttled to a cell isolated from the atmosphere in which RuBisCO is localised, and CO_2 is released therein. The biochemical pumping of CO_2 into cells containing RuBisCO leads to an increase of the relative concentration of CO_2 by a factor of 10 when compared to a non-C_4 cell, and a consequence dramatically increase of photosynthetic efficiency at high temperature (von Caemmerer and Furbank, 2003).

The C_4 photosynthetic mechanism is a classic example of convergent evolution, which has evolved more than 60 times independently in various groups of flowering plants (Sage, Christin and Edwards, 2011). Because all known C_4 enzymes exist in C_3 plants, the evolution of C_4 involved the co-option of genes and proteins essential for the cycle followed by adaption of their expression levels and, at least in some cases, their kinetic properties (Aubry, Brown and Hibberd, 2011; Christin *et al.*, 2013). In particular, the transcription levels of PEPC are massive increased in all C_4 lineages so far screened (Svensson, Bläsing and Westhoff, 2003; Marshall *et al.*, 2007; Bräutigam *et al.*, 2014; Christin *et al.*, 2015; Moreno-Villena *et al.*, 2018). By contrast, the kinetic behaviour of the enzymes has received less attention, and have been investigated mainly using the model genus *Flaveria*, in which there are closely related C_4, non-C_4 and intermediate species (Chapter 2; McKown, Moncalvo and Dengler, 2005). It has been shown that the *Flaveria* C_4 PEPC has a ten-fold decrease in specificity for phosphoenolpyruvate (PEP), an increased sensitivity to activators such as glucose-6-phosphate, and a decreased sensitivity to feedback inhibition from malate

and aspartate (Chapter 2; Svensson, Bläsing and Westhoff, 1997, 2003; Engelmann *et al.*, 2003; Westhoff, 2004; Paulus, Schlieper and Groth, 2013). Comparison of PEPCs from intermediate species in the genus *Flaveria* further suggested that C_4 properties of the enzyme were gradually acquired during the diversification of the genus (Engelmann *et al.*, 2003). Similar comparative efforts have been conducted in *Alternanthera*, a distantly related genus of eudicots, which have shown that PEPC kinetics evolved convergently in the C_4 *Alternanthera* and *Flaveria* with respect to PEP (Svensson, Bläsing and Westhoff, 1997; Engelmann *et al.*, 2003; Gowik *et al.*, 2006). Both *Alternanthera* and *Flaveria* were investigated because they contain closely related C_3, C_4 and intermediate species. Whether the observed patterns extend to more ancient C_4 groups consequently remains unknown.

Based on molecular dating, the multiple origins of C_4 photosynthesis are spread throughout the last 35 million years, a period when the atmospheric CO_2 levels were constantly low (Christin *et al.*, 2008, 2011). *Flaveria* represents one of the most recent C_4 origins, its different photosynthetic types having diverged in the last 3 million years, with the emergence of fully C_4 plants 1-2 million years ago (Christin *et al.*, 2011). *Alternanthera* represents a slightly older C_4 group (Christin *et al.*, 2011), but some of the earliest origins of C_4 are observed in grasses, from 15 to 35 million years ago (Christin *et al.*, 2008). The C_4 monocots, including grasses but also sedges, are among the most productive plants and dominate most open biomes in tropical and subtropical regions (Cerling *et al.*, 1997; Sage, 2004; Osborne and Beerling, 2006; Osborne and Freckleton, 2009). From a genetic point of view, monocots and eudicots co-opted different genes for C_4 evolution (Christin *et al.* 2015). Genes encoding C_4-specific PEPC evolved under positive selection in several C_4 groups, but the identity and quantity of fixed amino acid changes varies among families (Besnard *et al.*, 2009). In particular, more of these changes are observed in grasses than in *Flaveria* (Christin *et al.*, 2007), which might result from the longest evolutionary time. Alternatively, the genes co-opted for C_4 in grasses might have been less fit for the C_4 function, requiring therefore more adaptive changes (Christin, Weinreich and Besnard, 2010). The lack of kinetic data,

and comparable protein preparations when data have been determined, hampers comparative analyses. Indeed, grass PEPC have previously been compared, but efforts focused on distinct isoforms with the same species (Dong et al. 1998), which diverged long before the photosynthetic types (Christin *et al.*, 2007). There is therefore a need for comparisons of C_4 and non-C_4 PEPC representing the divergence of photosynthetic carbon fixing mechanisms, which can be obtained by comparing species of the *Panicoideae* subfamily of grasses that represent different photosynthetic types (Moreno-Villena *et al.*, 2018).

In this work, we characterised orthologous genes encoding PEPC from the C_4 grass *Panicum queenslandicum* and its C_3 relative *Panicum pygmaeum* to test the hypotheses that (i) despite very different starting points, qualitatively similar changes happened in *Flaveria* and grass C_4 PEPCs, and (ii) the kinetic changes differ more between C_4 and non-C_4 PEPC in grass than in *Flaveria* due to an expanded period of adaptive evolution. We describe the changes in specificity for both substrates (bicarbonate and PEP) as well as the nature of inhibition by aspartate and malate. Overall, out work sheds new light on the impacts of evolutionary time and distance on the convergent evolution of enzyme kinetics.

RESULTS

DNA Cloning and protein purification

The *ppc-1E2* gene for the C_4 PEPC (Christin *et al.*, 2015) from the species *Panicum queenslandicum* was isolated from cDNA generated from RNA extracted from leaf tissue. The isozyme from the non-C_4 species *Panicum pygmaeum* was fully sequenced from amplification from cDNA, however, the yield was too low to clone. The *P. pygmaeum* PEPC gene was synthesised from the sequence. These genes were confirmed to be the orthologous genes encoding the C_4 and non-C_4 PEPC in each species by a maximum likelihood tree (Supp. Figure 1). These genes were cloned into plasmids for expression in *E. coli.* Expressed protein was purified to > 95% by SDS PAGE with a single immobilised metal column (Supp. Figure 3) as in Chapter 1, .

Kinetic analysis of PEPC varying both Bicarbonate and PEP

The specificity for bicarbonate of both enzymes was determined using a gas-tight assay system (Chapter 2). Assays were performed at five PEP concentrations, while varying the concentration of bicarbonate. Analysis of secondary plots allowed determination of k_{cat}/ K_m^{HCO3-}, k_{cat}/ K_m^{PEP}, and k_{cat}. The specificity for bicarbonate of the C_4 *P. queenslandicum* PEPC was determined as $1.09 \times 10^6 \pm 8.88 \times 10^4$ s^{-1} M^{-1} (Figure 1C). The specificity for bicarbonate for the C_3 *P. pygmaeum* PEPC was determined as $5.99 \times 10^5 \pm 2.93 \times 10^4$ s^{-1} M^{-1} (Figure 1D). The C_4 PEPC has nearly two thirds higher specificity for bicarbonate. The specificity for PEP of the C_4 *P. queenslandicum* PEPC was determined as $1.04 \times 10^4 \pm 1.08 \times 10^3$ s^{-1} M^{-1} (Supp. Figure 4A). The specificity for PEP of the C_3 *P. pygmaeum* PEPC was determined as $5.01 \times 10^5 \pm 2.44 \times 10^4$ s^{-1} M^{-1} (Supp. Figure 4B). The C_4 PEPC has 50 times lower specificity for PEP (Summarised in Table 1).

PEPC inhibition by malate and aspartate

The nature of inhibition of grass PEPCs was also explored, looking at the two main feedback inhibitors, malate and aspartate. As in the genus *Flaveria* (Chapter 2), both isoforms of PEPC exhibit mixed inhibition in the presence of malate (Figure 2), and competitive inhibition in the presence of aspartate (Figure 3), at pH 8.0. Under all conditions, with both inhibitors, the C_4 form of PEPC is less sensitive to inhibition. The non-competitive inhibition constant (K_{iu}) was determined by the secondary plot of the k_{cat}^{app} against inhibitor concentration. The competitive inhibition constant (K_{ic}) was determined by the secondary plot of the k_{cat}^{app}/K_m^{app} against inhibitor concentration.

Panicum queenslandicum PEPC is inhibited by malate at both limiting and saturating concentrations of PEP, this mixed inhibition can be characterised at limiting PEP by K_{ic}^{Malate} = 7.51 ± 1.17 mM (Figure 2A), and at saturating PEP by K_{iu}^{Malate} = 146.08 ± 20.40 mM (Figure 2C). Aspartate is a competitive inhibitor characterised by $K_{ic}^{Aspartate}$ = 49.44 ± 7.86 mM (Figure 3A). No inhibition at saturating PEP was observed (Figure 3C).

Panicum pygmaeum PEPC is inhibited by malate at both limiting and saturating concentrations of PEP, this mixed inhibition can be characterised at limiting PEP by K_{ic}^{Malate} = 0.52 ± 0.22 mM (Figure 2B), and at saturating PEP by K_{iu}^{Malate} = 31.23 ± 0.65 mM (Figure 2D). Aspartate is a competitive inhibitor characterised by $K_{ic}^{Aspartate}$ = 2.31 ± 0.63 mM (Figure 3A). No inhibition at saturating PEP was observed for aspartate (Figure 3D).

Comparing the two enzymes, the C_4 *P. queenslandicum* PEPC is 15 times less sensitive to malate at limiting PEP, and five times less sensitive to malate at saturating PEP when compared to the C_3 *P. pygmaeum* PEPC. The C_4 PEPC is 20 times less sensitive to competitive inhibition by aspartate than the C_3 ortholog for aspartate (Summarised in Table 2).

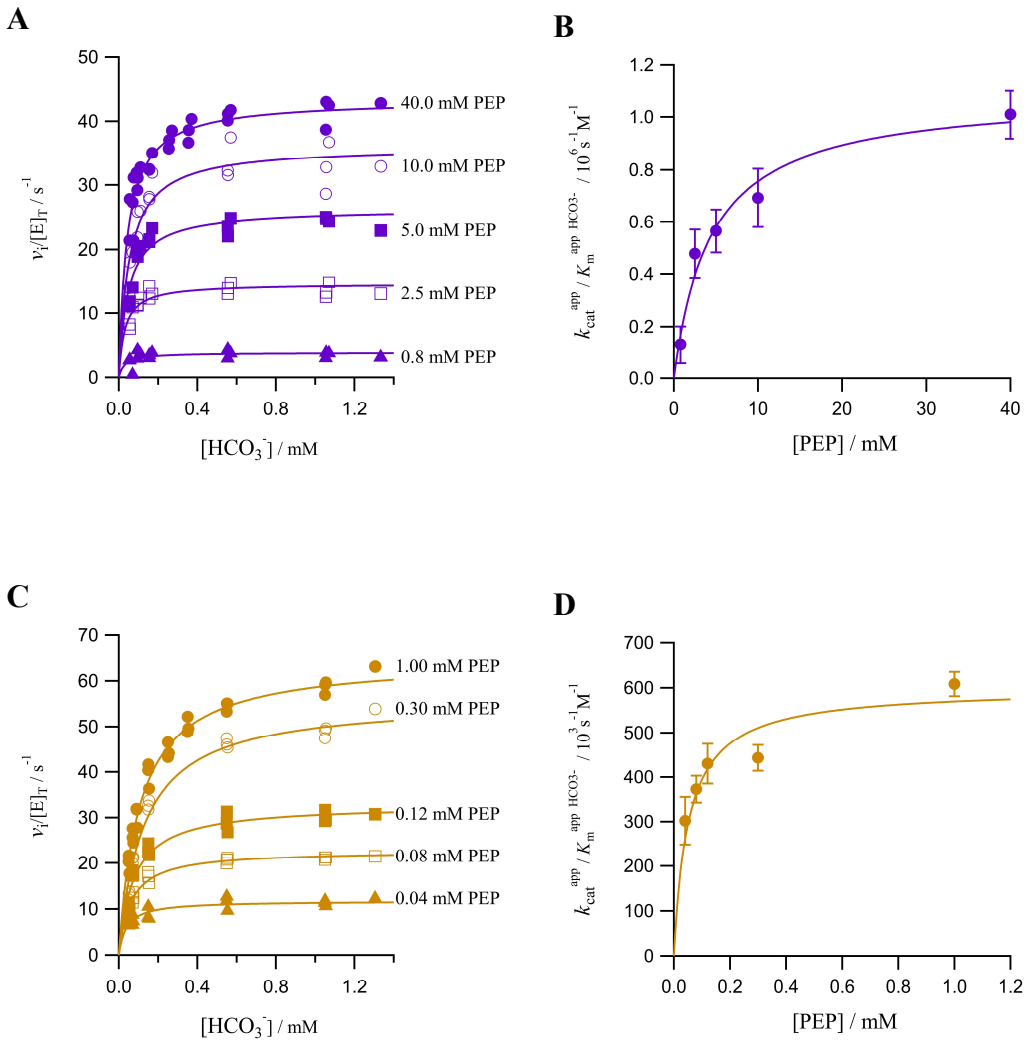


Figure 1: Rate of oxaloacetic acid formation, catalysed by PEPC, varying the concentration of bicarbonate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 U μl^{-1} malate dehydrogenase and 50 nM PEPC**. A** Markers represent the experimental data points for *Panicum queenslandicum* PEPC. The lines equation 1, kinetic parameters are shown in Supp. Figure 5A and panel B. **B** Secondary plots $k_{cat}^{app}/K_m^{appHCO3-}$ from *P. queenslandicum* PEPC, the line is described by equation 1, characterised by the parameters k_{cat} = 46.96 ± 1.71 s^{-1}, K_m^{HCO3-} = 0.036 ± 0.02 mM, K_m^{PEP} = 4.39 ± 1.10 mM and k_{cat}/K_m^{HCO3-} = 1.09 × 1.06 ± 8.88 × 10^4 s^{-1} M^{-1}. **C** Markers represent the experimental data points for *Panicum pygmaeum* PEPC. The lines equation 1, kinetic parameters are shown in Supp. Figure 5B and panel D. **D** Secondary plots $k_{cat}^{app}/K_m^{appHCO3-}$ from *P. pygmaeum* PEPC, the line is described by equation 1, characterised by the parameters k_{cat} = 65.59 ± 1.74 s^{-1}, K_m^{HCO3-} = 0.122 ± 0.015 mM, K_m^{PEP} = 0.17 ± 0.05 mM and k_{cat}/K_m^{HCO3-} = 5.99 × 10^5 ± 2.93 × 10^4 s^{-1} M^{-1}.

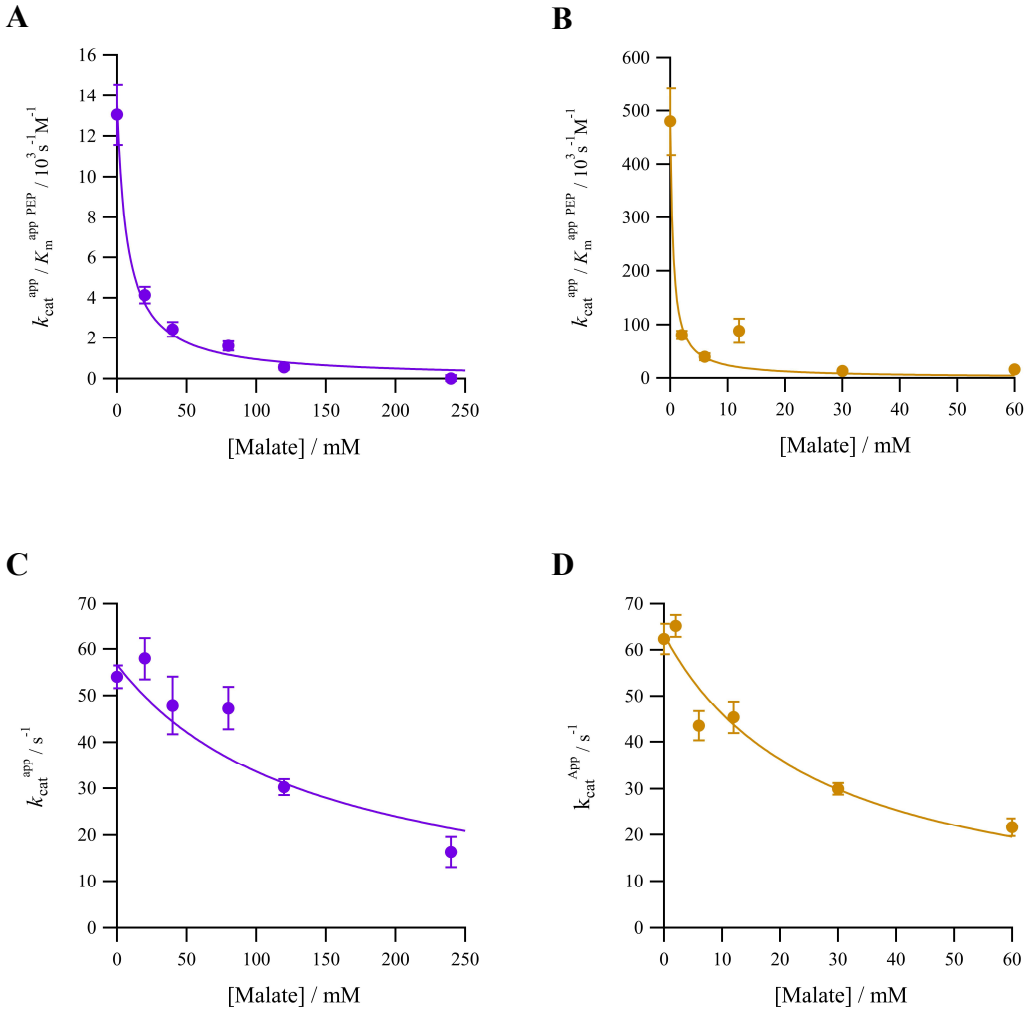


Figure 2: Plots of malate inhibition values for the PEPC enzymes. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 $U\mu l^{-1}$ malate dehydrogenase, 10 mM $KHCO_3$ and 10 nM of *Panicum queenslandicum* PEPC or 5 nM *Panicum pygmaeum* PEPC. Markers are derived from Supp. Figure 6A for *P. queenslandicum* PEPC and Supp. Figure 6B for *P. pygmaeum* PEPC, error bars represent standard errors. Secondary plot lines are characterised by equation 2. **A** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against malate concentration for the enzyme *P. queenslandicum* PEPC characterised by $K_{ic}^{Malate} = 7.51 \pm 1.17$ mM. **B** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against malate concentration for the enzyme *P. pygmaeum* PEPC. characterised by $K_{ic}^{Malate} = 0.52 \pm 0.22$ mM. **C** Secondary plot of k_{cat}^{app} against malate concentration for the enzyme *P. queenslandicum* characterised by $K_{iu}^{Malate} = 146.08 \pm 20.40$ mM. **D** Secondary plot of k_{cat}^{app} against malate concentration for the enzyme *P. pygmaeum* characterised by $K_{iu}^{Malate} = 31.23 \pm 0.65$ mM.

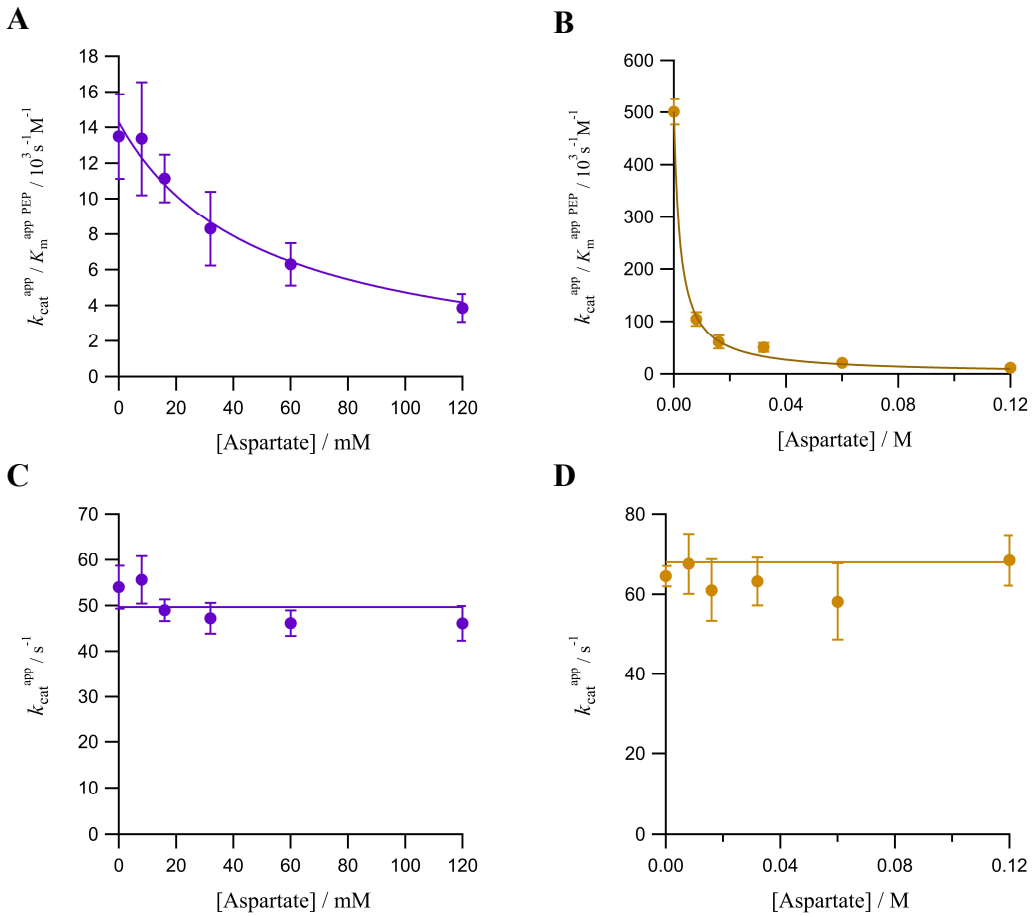


Figure 3: Plots of aspartate inhibition values for the PEPC enzymes. Assay conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 U μl^{-1} malate dehydrogenase, 10 mM $KHCO_3$ and 10 nM of *Panicum queenslandicum* PEPC or 5 nM *Panicum pygmaeum* Markers are derived from Supp. Figure 7A for *P. queenslandicum* PEPC and Supp. Figure 7B for *P. pygmaeum* PEPC, error bars represent standard errors. Secondary plot lines are characterised by equation 2. **A** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against aspartate concentration for the enzyme *P. queenslandicum* PEPC characterised by $K_{ic}^{Aspartate} = 49.44 \pm 7.86$ mM. **B** Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against aspartate concentration for the enzyme *P. pygmaeum* PEPC characterised by $K_{ic}^{Aspartate} = 2.27 \pm 0.02$ mM. **C** Secondary plot of k_{cat}^{app} against aspartate concentration showing no aspartate competitive inhibition of *P. queenslandicum* PEPC. **D** Secondary plot of k_{cat}^{app} against aspartate concentration showing no aspartate competitive inhibition of *P. pygmaeum* PEPC.

Table 1: Summary of kinetic parameters found in this study. Standard errors are given, based on fitted theoretical curves.

PEPC Species	k_{cat} / s^{-1}	K_m^{PEP} / mM	k_{cat}/K_m^{PEP} / s^{-1} M^{-1}	K_m^{HCO3-} / mM	k_{cat}/K_m^{HCO3-} / s^{-1} M^{-1}
Panicum queenslandicum (C_4)	46.96 ± 1.71	4.17 ± 0.30	1.04×10^4 ± 1.08×10^3	0.036 ± 0.02	1.09×10^6 ± 8.88×10^4
Panicum pygmaeum (C_3)	65.59 ± 1.74	0.17 ± 0.05	5.01×10^5 ± 2.44×10^4	0.122 ± 0.015	5.99×10^5 ± 2.93×10^4
Flaveria trinervia (C_4)	47.99 ± 1.21	0.60 ± 0.05	7.87×10^4 ± 5.43×10^3	0.065 ± 0.007	6.92×10^5 ± 4.17×10^4
Flaveria pringlei (C_3)	52.65 ± 1.37	0.056 ± 0.001	9.35×10^5 ± 8.49×10^4	0.099 ± 0.007	4.43×10^5 ± 2.17×10^4

Table 2: Summary of inhibition parameters found in this study. Standard errors are given, based on fitted theoretical curves. IC_{50} calculated at saturating PEP (40 mM PEP for *Pqu*, 5 mM PEP for *Ft*, 1 mM PEP for *Fp,* and 2 mM PEP for *Ppy*). Mixed inhibition IC_{50} values were calculated with equation 3, competitive inhibition IC_{50} values were calculated with equation 4.

PEPC Species	K_{ic}^{Malate} / mM	K_{iu}^{Malate} / mM	$K_{ic}^{Aspartate}$ / mM	IC_{50}^{Malate} / mM	$IC_{50}^{Aspartate}$ / mM
Panicum queenslandicum (C_4)	7.51 ± 1.17	146.08 ± 20.40	49.44 ± 7.86	53.3	523.7
Panicum pygmaeum (C_3)	0.52 ± 0.22	31.23 ± 0.65	2.27 ± 0.02	5.6	29.5
Flaveria trinervia (C_4)	10.96 ± 1.55	40.72 ± 4.59	40.02 ± 6.49	31.6	380.1
Flaveria pringlei (C_3)	2.14 ± 0.62	4.56 ± 1.72	4.13 ± 0.60	4.3	77.9

Table 3: Summary of kinetic parameters found in the literature. Values denoted with (*) for *Alternanthera* are taken from (Gowik *et al.*, 2006), values denoted with (†) for *Zea mays* are taken from (Dong *et al.*, 1998) k_{cat} values converted from Units mg^{-1}, values denoted with (‡) for *Zea mays* are taken from (Janc, O'Leary and Cleland, 1992) k_{cat} values converted from Units mg^{-1}.

PEPC Species	k_{cat} / s^{-1}	K_m^{PEP} / mM	k_{cat}/K_m^{PEP} / $s^{-1}M^{-1}$	K_m^{HCO3-} / mM	k_{cat}/K_m^{HCO3-} / $s^{-1}M^{-1}$	IC_{50}^{Malate} / mM
Alternanthera pugens (C_4)*	38 ± 0.5	0.157 ± 0.05	$2.4 \pm 0.06 \times 10^5$	-	-	-
Alternanthera tenella (C_3/C_4)*	33 ± 0.7	0.042 ± 0.01	$7.7 \pm 0.38 \times 10^5$	-	-	-
Alternanthera sessilis (C_3)*	22 ± 0.6	0.036 ± 0.02	$6.1 \pm 0.20 \times 10^5$	-	-	-
Zea mays C_4 PEPC†	41.9	0.59 ± 0.06	1.29×10^5	0.10 ± 0.03	7.63×10^5	0.82 ± 0.08
Zea mays Root PEPC†	54.1	0.04 ± 0.004	1.35×10^6	0.05 ± 0.013	1.08×10^6	0.24 ± 0.03
Zea mays C_4 PEPC‡	36.4	3.6 ± 0.6	1.01×10^4	0.18 ± 0.04	2.02×10^5	-

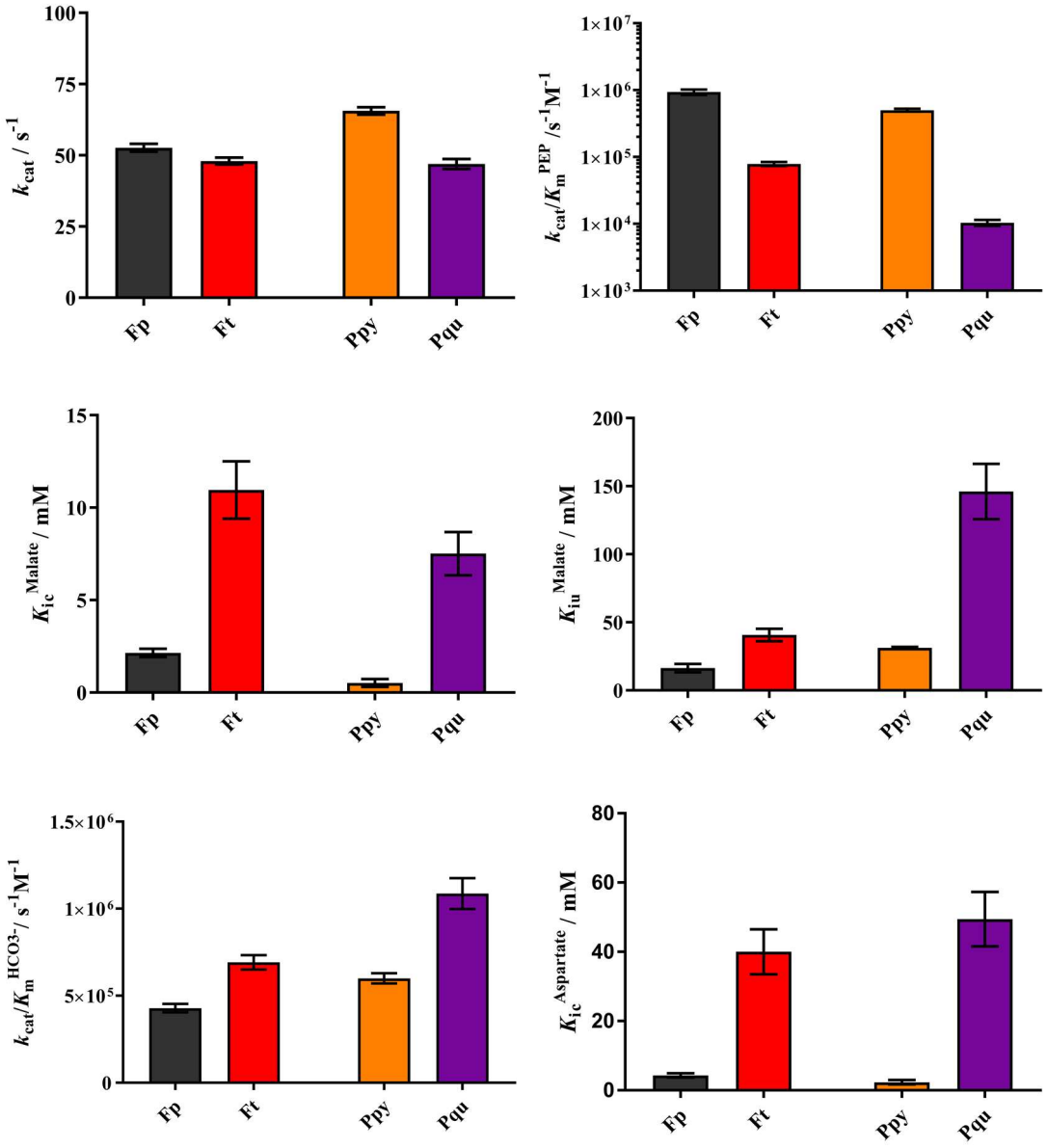


Figure 4: Summary of kinetic and inhibition parameters found in this study. Errors quoted are standard errors from the fitted theoretical lines. Values for the C_3 *Flaveria pringlei* PEPC are displayed in grey, for C_4 *Flaveria trinervia* PEPC in red, for *C_3 Panicum pygmaeum* in yellow, and for C_4 *Panicum queenslandicum* in purple.

DISCUSSION

Convergent kinetic changes across C_4 flowering plants

Genes encoding the non-C_4 PEPC of the C_3 *Panicum pygmaeum* and *Flaveria pringlei* diverged about 150 million years, and each underwent multiple subsequent gene duplications (Christin *et al.*, 2007, Christin *et al.*, 2015). Each consequently accumulated numerous mutations since their divergence (Supp. Figure 1; Christin *et al.*, 2007). While their exact functions are not known, they are expressed at similarly moderate levels (Moreno-Villena *et al.*, 2018), but numerous amino acid differences are present. Our investigation however shows that the two non-C_4 enzymes exhibit overall similar kinetic characteristics (Figure 4), including high sensitivity to competitive inhibition by malate and aspartate. The *P. pygmaeum* PEPC shows a slightly higher specificity for bicarbonate, a lower specificity for PEP, a lower sensitivity to non-competitive inhibition by malate, and a higher sensitivity to competitive inhibition by malate (Table 1 and Table 2). The variation among kinetic properties of PEPC between C_3 *Flaveria* and C_3 grasses is not known with confidence; assuming that the kinetics of the non-C_4 PEPCs did not differ drastically from their last common ancestor with the respective C_4, evolution of C_4-specific PEPC started at similar kinetic points in grasses and *Flaveria*. While fewer parameters were measured, the PEPC of C_3 *Alternanthera* seems moreover to have similar kinetics (Gowik *et al.*, 2006), which might indicate limited kinetic diversification of PEPC before C_4 evolution. This is moreover supported by the kinetic similarity of the distant root homolog of *Zea mays* (Dong *et al.*, 1998). We therefore conclude that the evolution of PEPC adapted for the C_4 pathway started at similar points in distant groups of angiosperms.

Analyses of amino acid sequences show that most modifications happened in the C_4-specific PEPC, with relative conservation among non-C_4 orthologs (Christin *et al.*, 2007; Besnard *et al.*, 2009), so that differences between closely-related C_4 and non-C_4 enzymes are likely the result of C_4 evolution. In grasses, the PEPC of the C_4 *Panicum queenslandicum* shows increased specificity for bicarbonate and decreased specificity for PEP when compared to the PEPC of the C_3 *P. pygmaeum* (Table 1). The direction of the kinetic

changes was therefore similar in grasses and *Flaveria* (Table 1). The affinity for PEP moreover changed in the same direction in *Alternanthera* (Gowik *et al.*, 2006) and in *Zea* (Dong *et al.*, 1998), which represents an independent C_4 origin within grasses. We therefore show that the evolution of C_4 PEPC involves qualitatively convergent changes in kinetic properties, which explains the previously observed convergent amino acid sequence (Christin *et al.*, 2007; Besnard *et al.*, 2009). It also suggests that an increase in specificity for bicarbonate and a decrease in specificity for PEP are important for the C_4 function of PEPC. The primary function of the non-C_4 PEPC is replenishing oxaloacetate for key metabolic pathways and likely to require a fast response to small concentration changes, it is therefore likely that high specificity for PEP is strongly selected for in non-C_4 PEPC. C_4 PEPC plays a central role in the high-flux C_4 cycle (Svensson, Bläsing and Westhoff, 2003; Stitt and Zhu, 2014). In a high flux C_4 system, bicarbonate is a limiting factor, therefore the specificity of PEPC towards bicarbonate is increased during C_4 evolution to boost the rate of the pathway (Chapter 2).

When compared to the C_3 *P. pygmaeum*, the PEPC of the C_4 *P. queenslandicum* PEPC shows markedly decreased sensitivity to both malate and aspartate (Table 2). The changes are qualitatively, similar to those observed in *Flaveria* (Chapter 2; Table 2). We therefore conclude that the same direction of changes happened independently in C_4 eudicots and monocots, for sensitivity to inhibitors in addition to kinetic parameters. This supports the importance of decreasing the sensitivity to malate and aspartate in enzymes that act in the C_4 pathway where all metabolites are abundant (Arrivault *et al.*, 2017).

***The differences in enzyme behaviour are quantitatively more important in grasses than in* Flaveria.**

While the direction of changes is similar in monocots and eudicots, differences between C_3 and C_4 species are more marked in grasses than in *Flaveria* (Fig. 7), and in *Alternanthera* summarized in Table 3 (Gowik et al. 2006).

The PEPC of the C_4 *P. queenslandicum* has a greater increase in bicarbonate specificity. This strong specificity for bicarbonate might make the C_4 pathway of grasses less reliant on the activity of carbonic anhydrase, explaining why this enzyme is less important in grasses than in eudicots at ambient CO_2 levels (Studer *et al.*, 2014). The PEPC of *P. queenslandicum* shows the lowest specificity for PEP of the four enzymes compared (Table 1). The disconnection between specificities for PEP and bicarbonate reinforces the conclusion that low specificity for PEP in C_4 was independently selected and does not result from adaptation on another property of the enzyme (Chapter 2). While numbers are difficult to compare among studies under different conditions, the specificity for PEP measured in *P. queenslandicum* is comparable to that previously reported for *Zea mays* (Janc, O'Leary and Cleland, 1992), while the ones of the C_4 *Flaveria* fall in between those of *Alternanthera* and the C_4 grasses (Table 1; Dong et al. 1998; Gowik et al. 2006).

When compared to *Flaveria* and *Zea* (Dong *et al.*, 1998), the PEPC of the C_4 grass also shows a greater reduction in sensitivity inhibition by malate at saturating PEP (Table 3) although the inhibition by malate and aspartate at limiting PEP are comparable in magnitude between *P. queenslandicum* and *Flaveria trinervia* PEPC (Table 2, Figure 4).

The quantitative differences between *Flaveria* and grasses might be linked to the contrast between the length of time spent as C_4 in each lineage, from more than 16 million years for *P. queenslandicum* to less than 3 for *Flaveria* (Christin *et al.*, 2008, 2011). Indeed, the kinetic properties observed in the PEPC of extant taxa result from adaptive changes accumulated since the initial origin of C_4 photosynthesis. According to the current model, an initial C_4 pathway can evolve via enzyme upregulation and limited modifications of the proteins (Sage, Sage and Kocacinar, 2012; Heckmann *et al.*, 2013), as observed in C_3-C_4 intermediates (Svensson, Bläsing and Westhoff, 2003; Dunning *et al.*, 2017). Once a C_4 pathway is in place, selection will act to improve its efficiency. The remarkable convergence of sensitivity to inhibitors in *F. trinervia* and *P. queenslandicum* might suggest that these

parameters are the first target of selection. This conclusion is further supported by previous reports that changes in sensitivity in *Flaveria* involved a key amino acid substitution (Paulus, Niehus and Groth, 2013; Paulus, Schlieper and Groth, 2013), which is observed in many C_4 lineages (Besnard *et al.*, 2009). Modifications of the specificities for PEP and bicarbonate might represent later evolutionary modifications, which continued within C_4 lineages leading to stronger differences in *P. queenslandicum*. The largest amount of adaptive amino acid substitutions is observed in grasses and sedges (Christin *et al.*, 2007; Besnard *et al.*, 2009) and would therefore indicate longest periods of sustained enzyme adaptation, or stronger selection in the monocots. This scenario leads us to predict that the properties that vary quantitatively between *Flaveria* and *P. queenslandicum* will also vary within C_4 lineages of grasses, as the fingerprint of continuous adaptation. Additional species will need to be screened to test this hypothesis.

Here, we present the first biochemical characterization of a PEPC from a C_3 grass encoded by an orthologous gene to those recurrently co-opted for C_4. Coupled with a detailed characterization of a C_4 grass PEPC, we were able to show that the direction of changes was similar among distant C_4 origins, indicating convergent biochemical adaptation for the C_4 catalytic context. The magnitude of changes was, however, more marked for some parameters in the C_4 grass, and we suggest that extended evolutionary periods, potentially coupled with stronger selective pressures, lead to enzymes that are better adapted for the C_4 pathway in some C_4 plants. We conclude that gene co-option and initial enzyme adaptation, including decreased sensitivity to inhibitors, are necessary steps that are consequently shared by all C_4 lineages, and the length of the subsequent period of adaptation of the trait has effect on the magnitude of these changes. Other, facultative enzyme adaptations are restricted to some C_4 plants, leading to a variety of C_4 biochemical phenotypes across flowering plants.

EXPERIMENTAL PROCEDURES

Unless otherwise stated, reagents and components were from Sigma. For purification, unless otherwise stated the equipment used procured from GE Healthcare. unless otherwise stated enzymes and *E. coli* strains were from NEB.

DNA Preparation

Leaf samples from the species *Panicum queenslandicum* and *Panicum pygmaeum* were taken at midday in full daylight and were flash frozen in liquid nitrogen. Leaf samples were homogenised with a pestle and mortar in liquid nitrogen. RNA was extracted from ground leaves using the RNeasy Kit (Qiagen). Libraries of cDNA were generated with SuperScript II Reverse Transcriptase (Thermo Fisher Scientific). The PEPC for *P. queenslandicum* was amplified using the primers PquFor1B and PquRev1B, and Q5 polymerase. The amplified gene was Sanger sequenced (GATC Biotech) with the PCR primers and with the primers Pqu_1323_Seq_For and Pqu_1752_Seq_Rev (Primers synthesised by Sigma, summarised in Table 4). The gene was the cloned into the pET-1B His6 TEV LIC vector plasmid, gifted by Scott Gradia (University of California, Berkeley, Addgene plasmid # 29653).

The PEPC for *P. pygmeaum* was amplified using the primers PpyFor1B and PpyRev1B. The amplified gene was sequenced with the PCR primers and with the primers Ppy_1291_Seq_For and Ppy_1791_Seq_Rev. The gene was then synthesised by GenArt Gene Synthesis in the pTRCC Plasmid. The synthesised gene was sub-cloned into the pET-1B His6 TEV LIC vector plasmid and sequenced with the PCR primers and with the primers Ppy_1291_Seq_For and Ppy_1791_Seq_Rev.

Genes were sub cloned using the ligation independent cloning method with Q5 DNA polymerase (NEB) and T4 DNA polymerase (NEB). Cloned plasmids were isolated using a Miniprep DNA kit (Qiagen). Plasmids were Sanger sequenced using the T7 promotor, T7 terminator and respective primers.

Protein Expression

For protein expression, BL21λ(DE3) strain of *E. coli* (NEB) was used. Chemically competent *E. coli* cells were transformed with each of the plasmids. Eight litres of cultures were grown in LB medium at 37°C to OD_{600} 0.8. Cultures were cooled to 4°C for one hour prior to recombinant protein induction with 0.5 mM IPTG (Fischer). Cultures were then incubated at 18°C for 18 hours. Cells were harvested by centrifugation at 5,422 × *g* for 25 minutes and stored at -80°C.

Protein Purification

Cells were suspended in IMAC buffer (25 mM Tris, 0.5 M NaCl, 0.3 M glycerol, 20 mM imidazole (Acros Scientific)), 10 ml per 2 L of culture with 50 µl of 50mg ml^{-1} DNase I and 100 µl of 100 mg ml^{-1} Pefabloc. Cells were passed twice through a cell disruptor (Constant Systems) before centrifugation at 26,902 × *g* for 40 minutes. The supernatant was passed through a 0.45 µm pore filter (Elkay Labs.). PEPC was separated from soluble protein with a prepacked 1 ml nickel affinity column using an ÄKTA™ Pure 25 L Chromatography System. The loaded column was washed with 50 column volumes of IMAC buffer, then 50 column volumes of IMAC buffer containing 150 mM imidazole. Pure PEPC was eluted with 10 column volumes of IMAC buffer containing 400 mM imidazole.

Protein eluted from IMAC purification was loaded onto a Sephadex G50 desalting column (Amersham Biosciences) and rebuffered in storage buffer (20 mM Tris, 5% v/v glycerol, 150 mM KCl, 1 mM DTT (AnaSpec. Inc.)). Protein was aliquoted and frozen at -80°C until use.

Enzyme Quantification

PEPC enzyme concentration was quantified by absorption at 280 nm. Enzyme extinction coefficient was calculated using the ExPASy protein parameter tool and corrected by determining the absorbance of the protein denatured in 6 M guanidine hydrochloride (Gill and von Hippel, 1989). The difference between the denatured and folded protein at 280 nm was used to adjust the extinction coefficient of the protein. The extinction coefficient for *Panicum queenslandicum* PEPC was determined to be 105805 M^{-1} cm^{-1}, the extinction coefficient for *Panicum pygmaeum* PEPC was determined to be 111514 M^{-1} cm^{-1}. Difference

of 8.3 % and 4.6 % between predicted absorbance coefficients for *P. queenslandicum* and *P. pygmaeum* PEPC were observed respectively. It is assumed that all enzyme used in the assay was active.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE) Analysis

Total protein concentration for purification efficiency was determined using the BCA Pierce quantification kit (Thermo Scientific). Concentration was determined using a standard curve performed with bovine serum albumin, over a concentration range 0 – 2.0 mg ml^{-1}.

Protein samples were analysed for purity using SDS PAGE analysis. Protein samples were quantified using the BCA Pierce method, 25 µg of cell lysate and 5 µg of pure protein were denatured in 2 × SDS PAGE loading dye (200 mM Tris.HCl pH 6.8, 2 % SDS, 20 % Glycerol, 0.01% Bromophenol blue (BDH Laboratory Supplies) and 7 % β-mercaptoethanol). Protein was loaded onto an 8% acrylamide SDS gel with 2 µl of Blue Prestained Protein Standard Broad Range (11-190 kDa) (NEB). Gels were run for 50 minutes at 200 V with 1 × Tris/Glycine/SDS running buffer (Geneflow). Gels were stained with InstantBlue (Expedeon) and imaged with a ChemiDoc™ MP (BioRad).

Enzyme Assays

PEPC activity was measured spectroscopically at 340 nm by coupling to NADH-malate dehydrogenase. Assays with a high fixed concentration of bicarbonate were observed using a FLUOstar plate reader (BMG Labtech) using the 340 nm ± 5 nm absorbance filter (BMG Labtech). Plate reader assays were conducted in a reaction volume of 150 µl at 25°C. Typical reaction mixtures contained 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$ (Fluka), 5 mM $KHCO_3$. 0.2 mM NADH (Fisher) and 0.1 U μl^{-1} malate dehydrogenase. Assays were initiated with the addition of PEPC enzyme. Rates were calculated with a NADH calibration curve.

Assays at a range of bicarbonate concentrations were observed with a Cary Bio 300 spectrophotometer (Agilent Technologies) in the same reaction buffer, in a total reaction volume of 600 µl. In bicarbonate assays, the water and tricine buffer were sparged with nitrogen for 18 hours prior to use in assays. Bicarbonate assays were constructed under a

nitrogen flow. Assays were performed in a sealed cuvette. The reaction was initiated with the addition of 50 nM PEPC, delivered with a gastight syringe (Hamilton). Bicarbonate concentrations were controlled with the addition of freshly prepared potassium bicarbonate. Background bicarbonate was determined using an endpoint assays with no potassium bicarbonate, run for 30 minutes. Rates were calculated using the Cary analysis software.

Data Analysis

Kinetic parameters were evaluated by non-linear regression analysis in Igor Pro (Version 7.0.8.1; Wavemetrics Inc., Lake Oswego Orgeon). Kinetic parameters were analysed with bar graphs in GraphPad Prism 7 for Windows (Version 7.04, GraphPad Software, Inc.). The following equations were used:

(Equation 1)

$$v_i/_{[E_T]} = \frac{k_{cat} \times [S]}{K_m + [S]}$$

Equation 1, where $v_i/[E_T]$ is the steady state rate divided by the total enzyme concentration, k_{cat} is the first order rate constant, K_m is the Michaelis constant, and S is the substrate concentration.

(Equation 2)

$$k^{app} = \frac{k}{1 + {[i]}/{K_i}}$$

Equation 2, where k^{app} is the apparent rate constant, k is the uninhibited constant, i is the inhibitor concentration, and K_i is the inhibition constant.

(Equation 3)

$$IC_{50}^{Mixed} = \frac{K_m + a}{\frac{K_m}{K_{ic}} + \frac{a}{K_{iu}}}$$

Equation 3, where IC^{Mixed} is the IC_{50} for mixed inhibition, K_m is the uninhibited Michaelis constant, K_{ic} is the competitive inhibition constant, K_{iu} is the uncompetitive inhibition constant and a is the substrate concentration

(Equation 4)

$$IC_{50}^{Comp.} = K_{ic}\left(1 + \frac{a}{K_m}\right)$$

Equation 4, where $IC^{Comp.}$ is the IC_{50} for competitive inhibition, K_m is the uninhibited Michaelis constant, K_{ic} is the competitive inhibition constant, and a is the substrate concentration.

REFERENCES

Arrivault, S., Obata, T., Szecówka, M., Mengin, V., Guenther, M., Hoehne, M., Fernie, A. R. and Stitt, M. (2017) Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation., *Journal of experimental botany*, 68(2), pp. 283–298.

Atkinson, R. R. L., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M. and Osborne, C. P. (2016) C_4 photosynthesis boosts growth by altering physiology, allocation and size, *Nature Plants*. Nature Publishing Group, 2(5), p. 16038.

Aubry, S., Brown, N. J. and Hibberd, J. M. (2011) The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway, *Journal of Experimental Botany*, 62(9), pp. 3049–3059.

Besnard, G., Muasya, a. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. (2009) Phylogenomics of C_4 Photosynthesis in Sedges (*Cyperaceae*): Multiple Appearances and Genetic Convergence, *Molecular Biology and Evolution*, 26(8), pp. 1909–1919.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P. and Weber, A. P. M. (2014) Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species, *Journal of experimental botany*, 65(13), pp. 3579–3593.

von Caemmerer, S. and Furbank, R. T. (2003) The C_4 pathway: an efficient CO_2 pump, *Photosynthesis Research*, 77(2/3), pp. 191–207.

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V. and Ehleringer, J. R. (1997) Global vegetation change through the Miocene/Pliocene boundary, *Nature*, 389(6647), pp. 153–158.

Christin, P.-A., Arakaki, M., Osborne, C. P. and Edwards, E. J. (2015) Genetic Enablers Underlying the Clustered Evolutionary Origins of C_4 Photosynthesis in Angiosperms, *Molecular Biology and Evolution*, 32(4), pp. 846–858.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V. and Salamin, N. (2008) Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses, *Current Biology*, 18(1), pp. 37–43.

Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J. and Osborne, C. P. (2013) Parallel Recruitment of Multiple Genes into C_4 Photosynthesis, *Genome Biology and Evolution*, 5(11), pp. 2174–2187.

Christin, P.-A., Osborne, C. P., Sage, R. F., Arakaki, M. and Edwards, E. J. (2011) C_4 eudicots are not younger than C_4 monocots, *Journal of experimental botany*, 62(9), pp. 3171–3181.

Christin, P.-A., Weinreich, D. M. and Besnard, G. (2010) Causes and evolutionary significance of genetic convergence, *Trends in Genetics*. Elsevier Ltd, 26(9), pp. 400–405.

Christin, P., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. (2007) C_4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes, *Current Biology*, 17(14), pp. 1241–1247.

Dong, L.-Y., Masuda, T., Kawamura, T., Hata, S. and Izui, K. (1998) Cloning, Expression, and Characterization of a Root-Form Phosphoenolpyruvate Carboxylase from *Zea mays*: Comparison with the C_4-Form Enzyme, *Plant and Cell Physiology*, 39(8), pp. 865–873.

Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P. and Christin, P.-A. (2017) Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives, *Evolution*, 71(6), pp. 1541–1555.

Ehleringer, J. and Björkman, O. (1977) Quantum yields for CO_2 uptake in C_3 and C_4 plants, *Plant Physiology*, 59(577), pp. 86–90.

Engelmann, S., Bläsing, O. E., Gowik, U., Svensson, P. and Westhoff, P. (2003) Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria*- a gradual increase from C_3 to C_4 characteristics., *Planta*, 217(5), pp. 717–25.

Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Analytical Biochemistry*, 182(2), pp. 319–326.

Gowik, U., Engelmann, S., Bläsing, O. E., Raghavendra, a. S. and Westhoff, P. (2006) Evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Alternanthera*: Gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 Alternantheras, *Planta*, 223, pp. 359–368.

Hatch, M. D. (1987) C_4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure, *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 895(2), pp. 81–106.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M. and Lercher, M. J. (2013) Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape, *Cell*, 153(7), pp. 1579–1588.

Janc, J. W., O'Leary, M. H. and Cleland, W. W. (1992) A kinetic investigation of phosphoenolpyruvate carboxylase from *Zea mays*, *Biochemistry*, 31(28), pp. 6421–6426.

Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., Sage, R. F. and Hibberd, J. M. (2007) Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C_3 to C_4 photosynthesis, *The Plant Journal*, 51(5), pp. 886–896.

McKown, A. D., Moncalvo, J.-M. and Dengler, N. G. (2005) Phylogeny of *Flaveria* (*Asteraceae*) and inference of C_4 photosynthesis evolution, *American Journal of Botany*, 92(11), pp. 1911–1928.

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. and Christin, P. A. (2018) Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis, *Molecular Biology and Evolution*, 35(1), pp. 94–106.

Nisbet, E. G., Grassineau, N. V., Howe, C. J., Abell, P. I., Regelous, M. and Nisbet, R. E. R. (2007) The age of Rubisco: the evolution of oxygenic photosynthesis, *Geobiology*, 5(4), pp. 311–335.

Osborne, C. P. and Beerling, D. J. (2006) Nature's green revolution: The remarkable evolutionary rise of C_4 plants, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361(1465), pp. 173–194.

Osborne, C. P. and Freckleton, R. P. (2009) Ecological selection pressures for C_4 photosynthesis in the grasses, *Proceedings of the Royal Society: Biological Sciences*, 276(1663), pp. 1753–1760.

Paulus, J. K., Niehus, C. and Groth, G. (2013) Evolution of C_4 Phosphoenolpyruvate Carboxylase: Enhanced Feedback Inhibitor Tolerance Is Determined by a Single Residue, *Molecular Plant*, 6(6), pp. 1996–1999.

Paulus, J. K., Schlieper, D. and Groth, G. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution, *Nature communications*, 4, p. 1518.

Sage, R. F. (2004) The evolution of C_4 photosynthesis, *New Phytologist*, 161(2), pp. 341–370.

Sage, R. F., Christin, P.-A. and Edwards, E. J. (2011) The C_4 plant lineages of planet Earth, *Journal of experimental botany*, 62(9), pp. 3155–3169.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) Photorespiration and the Evolution of C_4 Photosynthesis, *Annual Review of Plant Biology*, 63(1), pp. 19–47.

Skillman, J. B. (2007) Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark, *Journal of Experimental Botany*, 59(7), pp. 1647–1661.

Stitt, M. and Zhu, X.-G. (2014) The large pools of metabolites involved in intercellular metabolite shuttles in C_4 photosynthesis provide enormous flexibility and robustness in a fluctuating light environment, *Plant, Cell & Environment*, 37(9), pp. 1985–1988.

Studer, A. J., Gandin, A., Kolbe, A. R., Wang, L., Cousins, A. B. and Brutnell, T. P. (2014) A Limited Role for Carbonic Anhydrase in C_4 Photosynthesis as Revealed by a ca1ca2 Double Mutant in Maize, *Plant Physiology*, 165(2), pp. 608–617.

Svensson, P., Bläsing, O. E. and Westhoff, P. (1997) Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase- a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3)., *European journal of biochemistry*, 246(2), pp. 452–60.

Svensson, P., Bläsing, O. E. and Westhoff, P. (2003) Evolution of C_4 phosphoenolpyruvate carboxylase, *Archives of Biochemistry and Biophysics*, 414(2), pp. 180–188.

Tcherkez, G. G. B., Farquhar, G. D. and Andrews, T. J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized, *Proceedings of the National Academy of Sciences*, 103(19), pp. 7246–7251.

Westhoff, P. (2004) Evolution of C_4 Phosphoenolpyruvate Carboxylase. Genes and Proteins: a Case Study with the Genus *Flaveria*, *Annals of Botany*, 93(1), pp. 13–23.

SUPPLEMENTARY FIGURES

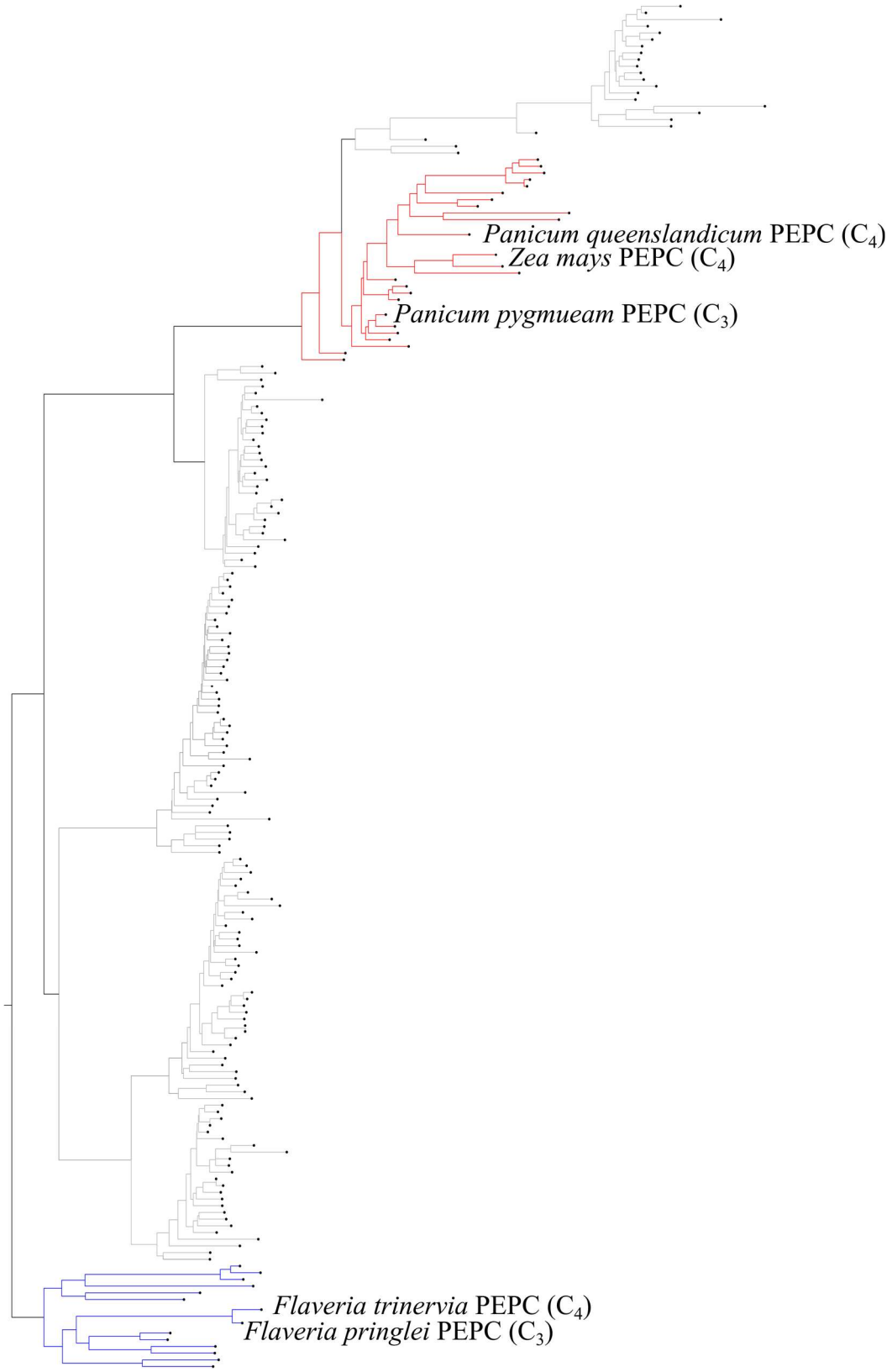


Figure S 1: Maximum likelihood of the PEPC genes in Monocots and Eudicots. In red are the ppc1P3 isoforms. In blue are the ppc1E2 isoforms. Sequences for *Zea mays* C4 taken from (Dong *et al.*, 1998). Sequences for Flaveria from Chapter 2. Tree taken with permission from (Moreno-Villena *et al.*, 2018).

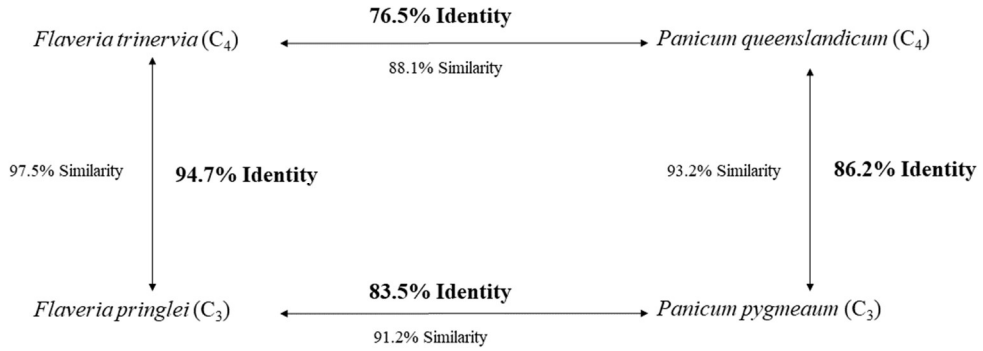


Figure S 2: Similarity and identity comparison between residues of the four PEPCs compared in this study. Similarly, and identity were determined as a percentage of the total number of amino acids.

Enzyme Purification

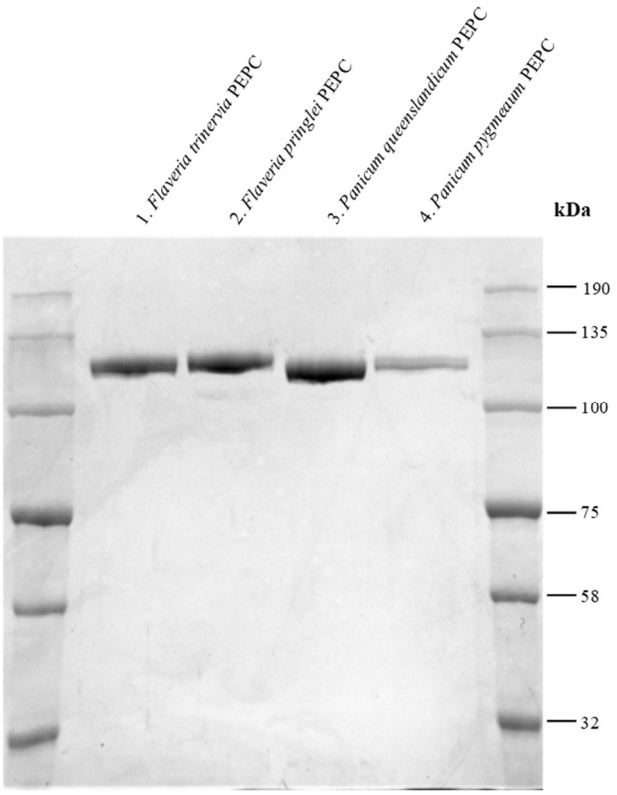


Figure S 3: 8 % acrylamide SDS PAGE analysis of PEPC proteins compared in this study. Lane 1 contains 5 µg of *Flaveria trinervia* PEPC, lane 2 contain 5 µg of *Flaveria pringlei* PEPC, lane 3 contain 5 µg of *Panicum queenslandicum* PEPC, lane 4 contains 5 µg *Panicum pygmaeum* PEPC.

Kinetic Assays Results

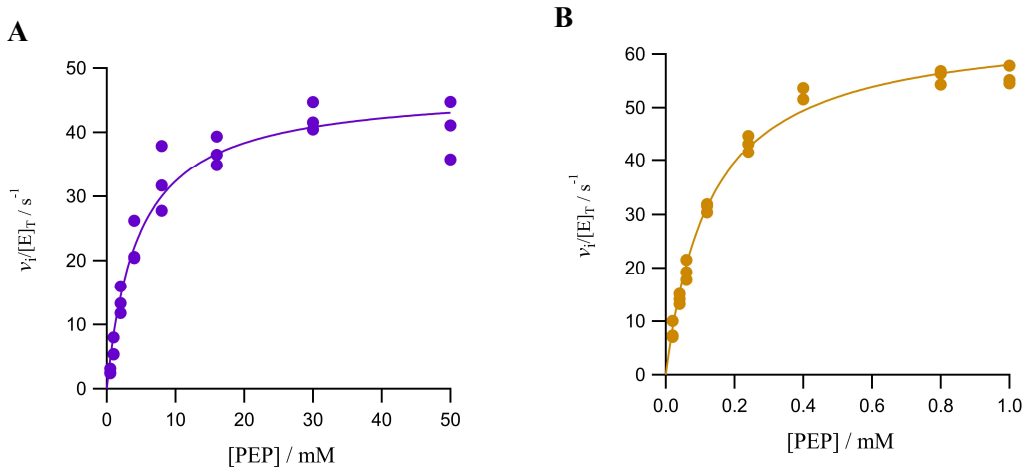


Figure S 4: Rate of oxaloacetic acid formation, catalysed by PEPC, varying the concentration of phosphoenolpyruvate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM $KHCO_3$, 0.15 mM NADH and 0.01 $U\mu l^{-1}$ malate dehydrogenase. Assays were repeated (n = 3) for each concentration of PEP. **A** Filled circles represent experimental data points for *Panicum queenslandicum* PEPC. The line is described by equation 1, with characterising parameters k_{cat} = 46.96 ± 1.74 s^{-1}, K_m^{PEP} = 4.53 ± 0.59 mM and k_{cat}/K_m^{PEP} = 10367 ± 1080 $s^{-1}M^{-1}$. **B** Filled circles represent experimental data points for *Panicum pygmaeum* PEPC. The line is described by equation 1, with characterising parameters k_{cat} = 65.59 ± 1.26 s^{-1}, K_m^{PEP} = 0.131 ± 0.008 mM and k_{cat}/K_m^{PEP} = 500710 ± 24400 $s^{-1}M^{-1}$.

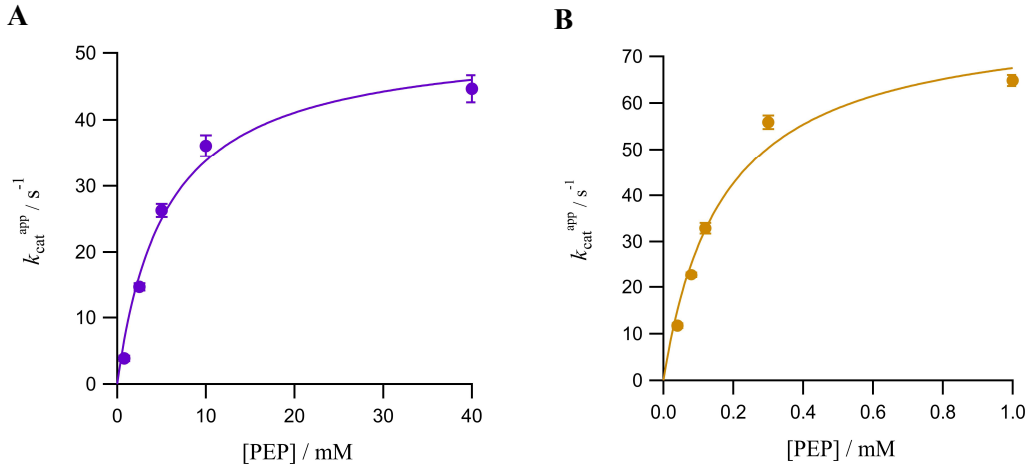


Figure S 5: Secondary plot of the k_{cat}^{app} parameter from bicarbonate assays in Figure 1: A Filled circles represent the k_{cat}^{app} parameter from Figure 1A for the enzyme the *Panicum queenslandicum* PEPC, error bars represent standard errors from fit of lines described by equation 1. The line is theoretical described by equation 1, characterised by the parameters k_{cat} = 52.25 ± 3.72 s^{-1} K_m^{PEP} = 5.46 ± 1.12 mM. **B** Filled circles represent the k_{cat}^{app} parameter from Figure 1C for the enzyme *Panicum pygmaeum* PEPC, error bars represent standard errors from fit of lines. The line is described by equation 1, characterised by the parameters k_{cat} = 79.06 ± 6.64 s^{-1} and K_m^{PEP} = 0.17 ± 0.04 mM.

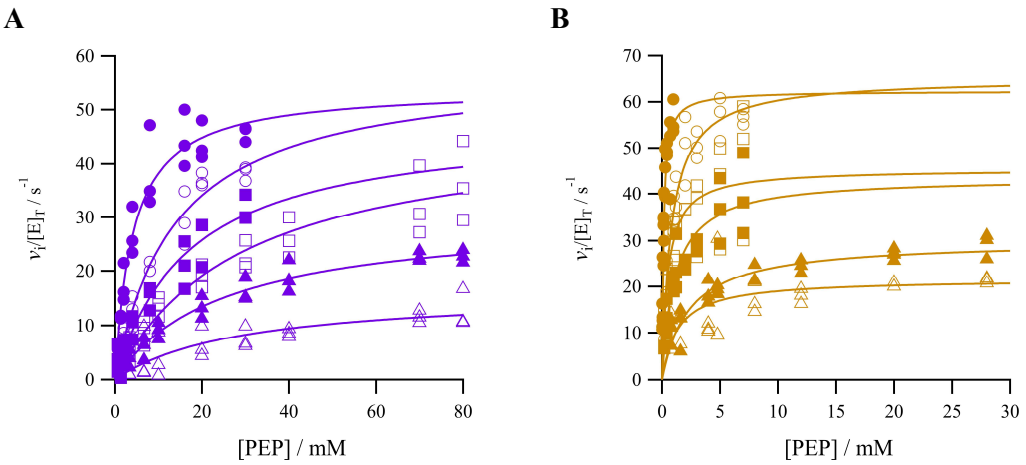


Figure S 6: Primary plot of PEPC inhibited by malate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM potassium bicarbonate, 0.2 mM NADH and 0.01 Uμl^{-1} malate dehydrogenase. The lines are described by equation 1. **A** Points indicated experimental data runs for *Panicum queenslandicum* PEPC, filled circles indicate no inhibitor, open circles indicate the presence of 8 mM malate, filled squares indicate the presence of 16 mM malate, open squares indicate the presence of 32 mM malate, filled triangles indicate the presence of 60 mM malate and open triangles indicate the presence of 120 mM malate. **B** Points indicated experimental data runs for *Panicum pygmaeum*, filled circles indicates no inhibitor, open circles indicate the presence of 4 mM malate, filled squares indicate the presence of 12 mM malate, open squares indicate the presence of 24 mM malate, filled triangles indicate the presence of 32 mM malate and open triangles indicate the presence of 60 mM malate.

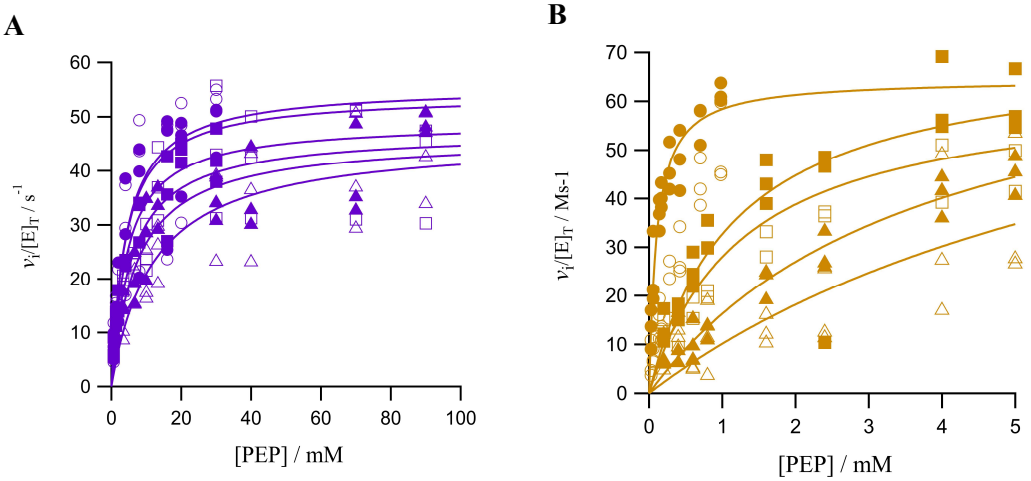


Figure S 7: Primary plot of PEPC inhibited by aspartate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM potassium bicarbonate, 0.2 mM NADH and 0.01 Uμl^{-1} malate dehydrogenase. **A** Points indicated experimental data runs, filled circles indicates no inhibitor, open circles indicate the presence of 8 mM aspartate, filled squares indicate the presence of 16 mM aspartate, open squares indicate the presence of 32 mM aspartate, filled triangles indicate the presence of 60 mM aspartate and open triangles indicate the presence of 120 mM aspartate. The lines are theoretical described by equation 1. **B** Points indicate experimental data runs for *Panicum pygmaeum* PEPC, filled circles indicate no inhibitor, open circles indicate the presence of 8 mM aspartate, filled squares indicate the presence of 16 mM aspartate, open squares indicate the presence of 32 mM aspartate, filled triangles indicate the presence of 60 mM aspartate and open triangles indicate the presence of 120 mM aspartate.

Primers

Table 4: Summary of primers used in this study.

Primer	Sequence, 5' to 3'
PquFor1B	GACGACGACAAGATGGCGTCCTCCGAGCGCCACC
PquRev1B	GAGGAGAAGCCCGGTTAGCCCGTGTTCTGCATGCC
PpyFor1B	TACTTCCAATCCAATGCAATGGCAAGCAG
PpyRev1B	TTATCCACTTCCAATGTTATTATTAACCGGTATTC
Pqu_1323_Seq_For	CGTGAAGCTGGACAT
Pqu_1752_Seq_Rev	ATGACCTGCTGCTTG
Ppy_1291_Seq_For	GATGGTAGTCTGCTGG
Ppy_1791_Seq_Rev	GCTATCGCTATAACCA
T7 Promotor	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG

Plasmid Sequences

>pET-1B *Panicum queenslandicum* PEPC

AACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGAT
GCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGC
TGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCC
CGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGA
GAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTT
CCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTG
TGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCG
GCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAG
CAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCAT
GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATT
AGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGG
CGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGG
CGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
GGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATT
GTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGTTCTTCTC
ACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCAATGGCGTCCTCCGAGCGCCACCACTCCATCG
ACGCGCAGGTCCGGCTCCTGGCCCCCGGCAAGGTCTCCGAGGATGACAAGCTCGTCGAGTACGACGTCCTCCTC
ATGGACCGCTTCCTCGACATCCTCCAGGACCTCCACGGCCCCGGCATCCGCGAATTCGTCCAGGACTGCTACGAG
CTGTCGGCGGAGTACGAGGGCGACCGCAACTCCGCGCGCCTCAAAGACCTCGGGTCCAGGCTCGCAAGCCTGGC
CCCCGCCGACGCCATCCTCGTCGCGGGCTCCATCCAGCACATGCTCAACCTCGCCAACCTCGCCGAGGAGGTGCA
GATCGCGAACCGCCGCAGGAACAAGCTCAAGAGCGGGGACTTCGCCGACGAGGGCTCCGCCACCACCGAGTCC
AACATCGACGAGACGATCAAGCGCCTCGTCGACCTCGGAAAGTCCAAGGAGGAGGTGTTCGAGGCGCTCAAGA
ACCAGAGCGTCGACCTCGTCCTCACCGCGCACCCAACGCAGTCCGTCCGGAGGTCGCTCCTCCAGAAGCACTCCA
GGATCCGGAATTGCCTCACGCAGCTCAATGCAAAGGACATCACGGACGACGAAAAGCAGGAGCTCGACGAGGC
TCTTAGCAGGGAGATCCAAGCAGCCTTCAGAACAGATGAGATTCGGAGAGCACAACCGACCCCACAGGATGAG
ATGCGTTACGGGATGAGCTATATCCACGAAACCATATGGAAGGGCGTTCCAAAGTTTCTGCGTCGTCTGGATACA
GCTCTGAAGAACATCGGGATCGACGAGCGTCTCCCCTACAATGTTCCTCTCATCCAGTTTTGTTCTTGGATGGGTG
GCGACCGTGATGGAAATCCAAGAGTTACGCCGGAGGTGACAAGGGATGTATGCTTGCTGGCAAGAATGATGGCT
GCAAACTTGTACTTCTCTGGGCTAGAAGAACTCATGTTCGAGCTCTCTATGTGGCGCTGCAATGATGAACTCCGT
GCTCGAGCGCAAGAAATTCACAGTGCTCCAAAGAAAGCTGCCAAGCACTACATAGAATTCTGGAAGCAAATCCC
TCTAAGTGAGCCGTATCGCGTGGTGCTTGGTAACGTGAGGGACAAACTGTACAACACACGCGAGCGTGCGCGCC
AACTGCTGACCAATGAATTTTCTGACATTCCGGAGGAATTGGTCTTTAGCAATGTTCAAGAGTTCCTGGAGCCCC
TTGAGCTGTGCTACAAATCACTGTGCGAGTGCGGCGACAAGACCATCGCCGACGGGAGCCTGCTGGACTTCCTTC
GCCAGGTCTCCACGTTCGGGCTCTCCCTCGTGAAGCTGGACATCCGGCAGGAGTCGGAGCGGCACACCGACGTG
ATCGACGCCATCACCACGCACCTCGGCATCGGCTCGTACCGCTCTTGGCCCGAGGACAAGCGCCAGGAGTGGCT
GCTGTCCGAGCTGCGCGGCAAGCGCCCGCTGCTCGCCCCGGACATGCCCCAGACCGAGGAGATCGCCGACGTGC
TCGGGTGCTTCCGCGTCCTCGCCGAGCTGCCCCGCGACAGCTTCGGCCCCTACATCATCTCCATGGCCACGGCGC
CGTCGGACGTCCTCGCCGTCGAGCTCCTGCAGCGGGAGTGCCACGTGCGCGACCCGCTGCCCGTGGTGCCGCTGT
TCGAGAGGCTCGCCGATCTGCAGAACGCGCCCGCGTCCATGGAGCGCCTCTTCTCGGTGGACTGGTACCTGCAGC
GGATCAACGGCAAGCAGCAGGTCATGATCGGCTACTCCGACTCCGGCAAGGACGCCGGGCGCCTGTCCGCGGCG
TGGCAGCTGTACAGGGCGCAGGAGGAGCTCGCGCAGGTGGCCAAGCGCTACGGTGTGAAGCTGACCATGTTCCA
CGGGCGCGGCGGCACCGTCGGACGGGGGGGCGGCCCGTCGCACCTCGCCATCCTGTCGCAGCCGCCGGACACCA
TCAATGGGTCCATCCGCGTGACAATCCAGGGAGAGGTCATCGAGCACTCCTTCGGCGAGGAGCACCTCTGCTTCC
GGACGCTGGAGCGGTTCACGGCCGCCACGCTGGAGCACGGCATGCACCCGCCGGTCTCTCCCAAGCCCGAGTGG
CGCAAGCTCATGGACGAGATGGCCGTCGTGGCCACGGAGGAGTACCGGTCCATCGTCTTCAGAGAGCCGCGCTT
CGTCGAGTACTTCCGATCGGCTACGCCTGAGACGGAGTACGGCAGGATGAACATCGGCAGCCGGCCGGCGAAGA
GGAAGCCCAAGGGCGGCATCGAGTCGCTCCGCGCGATCCCCTGGATCTTCTCGTGGACGCAGACGAGGTTCCAC
CTCCCGGTGTGGCTCGGGGTCGGCGCCGCGTTCCAGTACGCCATTAAGAAGGACAGCAAGAACATCCAGAAGCT
CAAGGACATGTACAAGGAGTGGCCCTTCTTCAGGGTCACCATTGACCTGCTGGAGATGGTCTTCGCCAAGGGGG
ACCCCAGCATCGCCGGCTTGTACGACGAGCTGCTCGTCGCCGCCGACCTCAAGCCCTTCGGGGAGCAGCTGAGG
AACAAGTACCTGGAGACGCAGCAGTTTCTCCTGCAGATCGCTGGGCACAAGGAAATCCTCGAAGGCGATCCCTA
CCTGAAGCAGGGGTTGCGGCTGCGCAACCCCTACATCACGACGCTGAACGTGTTTCAGGCTTACACCCTGAAGCT
GATGAGGGACCCGAGCTTCCAGGTGAAGAAGCAGCCGCCTATGTCCAAGGAGTTCGCCGACGAGAAGAAGCCC
GCCGGGCTGGTGGAGCTGAACCCGGCGAGCGAGTACGCGCCGGGGCTGGAGGACACGCTCATCCTCACCATGAA
GGGTATCGCCGCCGGCATGCAGAACACGGGCTAGTAATAACATTGGAAGTGGATAACGGATCCGAATTCGAGCG
CCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGA
AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTT
GAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

>pET-1B *Panicum pygmaeum* PEPC

AACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGAT
GCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGC
TGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCC
CGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGA
GAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTT
CCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTG
TGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCG
GCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAG
CAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCAT
GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATT
AGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGG
CGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGG
CGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
GGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATT
GTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGTTCTTCTC
ACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCAATGGCAAGCAGCAAAGCACCGGGTCCTGTT
GAACGTCATCAGAGCATTGATGCACAGCTGCGTCTGCTGGCACCGGGTAAAGTTAGCGAAGATGATAAACTGGT
TGAATATGATGCACTGCTGGTTGATCGTTTTCTGGATATTCTGCAGGATCTGCATGGTCCGAGCCTGCGTGAATTT
GTTCAAGAATGTTATGAACTGAGCGCAGAATATGAAGGTGATCGTGATGCAGCACGTCTGGGTGAACTGGGTGA
TCGTCTGACCGGTCTGGCTCCGGCAGATGCTATAGTTGTTGCAAGCAGCTTTAGCCATATGCTGAATCTGGCAAA
TCTGGCCGAAGAAGTTCAGATTGCACATCGTCGTCGTAATAAACTGAAACGTGGTGATTTTGCAGATGAAGCAA
GCGCAACCACCGAAAGCGATATTGAAGAAACCCTGAAACGTCTGGTTAGCGAACTGGGTAAAAGCCGTGAAGA
GGTTTTTGATGCCCTGAAAAATCAGACCGTTGATCTGGTTTTTACCGCACATCCGACACAGAGCATTCGTCGTAG
CCTGCTGCAGAAACATGCACGTATTCGTAATTGTCTGACCCAGCTGTATGCAAAAGATATTACAGCAGATGACAA
ACAAGAACTGGATGAAGCACTGCAGCGTGAAATTCAGGCAGCATTTCGTACCGATGAAATTCGTCGCACCCAGC
CGACACCGCAGGATGAAATGCGTGCAGGTATGAGCTATTTTCACGAAACCATTTGGAAAGGCGTTCCGAAATTT
CTGCGTCGTGTTGATACCGCACTGAAAAACATTGGTATTGATGAACGTCTGCCGTATAATGCACCGCTGATTCAG
TTTAGCAGCTGGATGGGTGGTGACCGTGATGGTAATCCGCGTGTTACACCGGAAGTTACCCGTGATGTTTGTCTG
TTAGCACGTATGATGGCAGCCAATCTGTATTTTAGCCAGATTGAAGAACTGATGTTTGAGCTGAGCATGTGGCGT
TGTAATGATGAACTGCGTGTTCGTGCCGAAGAACTGCATCGTGCAAGCCGTAAAGCAGCCAAACATTATATTGA
ATTTTGGAAGCAGATCCCTCCGAATGAACCGTATCGTGTTATTCTGGGTTATGTTCGCGATAAACTGTATTACACC
CGTGAACGTAGTCGCCATCTGCTGACCACCGGTTTTAGCGAAATTCCGGAAGATAGCGCATTTACCAATGTGGAA
GAATTTCTGGAACCGCTGGAACTGTGTTATCGTAGTCTGTGTGCATGTGGTGATAAAACCATTGCAGATGGTAGT
CTGCTGGATTTTCTGCGCCAGGTTAGCACCTTTGGTCTGAGCCTGGTTAAACTGGATATCCGTCAAGAAAGCGAA
CGTCATACCGATGTTCTGGATGCAATTACCACACATTTAGGTATTGGTAGCTATCGTGAATGGCCTGAAGAAAAA
CGTCAAGAATGGCTGCTGAGCGAGCTGCGTGGTAAACGTCCGCTGCTGGGTCCTGATCTGCCGCAGACCGAAGA
GGTTGCAGATGTGCTGGGCACCTTTCGTGTTCTGGCAGAACTGCCTCCGGATAGCTTTGGTGCATATATCATTAG
CATGGCAACCGCACCGAGTGATGTTCTGACCGTTGAACTGCTGCAGCGCGAATGTCATGTTCGTCATCCGCTGCG
TGTTGTTCCGCTGTTTGAAAAACTGGCAGATCTGGAAGCAGCACCGGCAGCAGTTGCACGTCTGTTTAGTGTTGA
TTGGTATATGGATCGCATCAACGGTAAACAAGAAGTGATGATTGGTTATAGCGATAGCGGTAAAGATGCAGGTC
GTCTGAGTGCAGCATGGCAGCTGTATAAAGCACAAGAAGAACTGGTTCAGGTTGCCAAACGTTATGGTGTTAAA
CTGACCATGTTTCATGGTCGTGGTGGCACCGTTGGTCGCGGTGGTGGTCCGACACATCTGGCCATTCTGAGCCAG
CCACCGGATACCATTCATGGTTCTCTGCGTGTTACCGTTCAGGGTGAAGTTATTGAACATAGTTTTGGCGAAGAA
CATCTGTGTTTTCGTACCCTGCAGCGCTTTACCGCAGCAACCCTGGAACATGGTATGCATCCGCCTGTTAGCCCG
AAACCGGAATGGCGTGCACTGATGGATGAACTGGCAGTTGTTGCCACCGAAGAATATCGTAGCATTGTTTTTAAA
GAACCGCGTTTTGTGGAATATTTTCGTAGCGCAACACCGGAAACCGAATATGGTCGTATGAATATTGGTAGTCGT
CCGAGCAAACGTAAACCGAGCGGTGGTATTGAAAGTCTGCGTGCAATTCCGTGGATTTTTGCATGGACCCAGAC
ACGTTTTCATCTGCCTGTTTGGTTAGGTTTTGGTGCAGCATTTAAACACGCCATGAAAAAAGATATCCGCAACAT
TCAGACCCTGCGCGAAATGTATAATGAATGGCCGTTTTTTCGTGTTACCCTGGATCTGCTGGAAATGGTTTTTGCC
AAAGGTGATCCGGGTATTGCAGGTCTGTATGATGAGCTGGTTGTTGCCGATGATCTGAAACCGTTTGGTGAACAG
CTGCGCAATAACTATGTTGAAACCCAACAGCTGCTGCTGCAGGTTGCAGGTCATAAAGATATTCTGGAAGGCGA
TCCGTATCTGAAACAGCGTCTGCGCCTGCGCGATCCTTATATCACAACCCTGAATGTTTGTCAGGCATATACGCT
GAAACGTATTCGCGATCCGAGCTTTCAGGTGACCGCACAGCGTCCGCTGAGCAAAGAATTTGCGGATGAAAATC
AGCCTGCCGGTCTGGTTAAGCTGAATCCGGCATCAGAATATGCACCTGGTCTGGAAGATACCCTGATTCTGACAA
TGAAAGGTATTGCCGCAGGTATGCAGAATACCGGTTAATAATAACATTGGAAGTGGATAACGGATCCGAATTCG
AGCGCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGC
CCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG
TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

Protein Amino Acid Sequence

>pET-1B *Panicum queensladicum* PEPC amino acid sequence

MGSSHHHHHHENLYFQSNAMASSERHHSIDAQVRLLAPGKVSEDDKLVEYDVLLM
DRFLDILQDLHGPGIREFVQDCYELSAEYEGDRNSARLKDLGSRLASLAPADAILVA
GSIQHMLNLANLAEEVQIANRRRRNKLKSGDFADEGSATTESNIDETIKRLVDLGKSK
EEVFEALKNQSVDLVLTAHPTQSVRRSLLQKHSRIRNCLTQLNAKDITDDEKQELDE
ALSREIQAAFRTDEIRRAQPTPQDEMRYGMSYIHETIWKGVPKFLRRLDTALKNIGID
ERLPYNVPLIQFCSWMGGDRDGNPRVTPEVTRDVCLLARMMAANLYFSGLEELMF
ELSMWRCNDELRARAQEIHSAPKKAAKHYIEFWKQIPLSEPYRVVLGNVRDKLYNT
RERARQLLTNEFSDIPEELVFSNVQEFLEPLELCYKSLCECGDKTIADGSLLDFLRQV
STFGLSLVKLDIRQESERHTDVIDAITTHLGIGSYRSWPEDKRQEWLLSELRGKRPLL
APDMPQTEEIADVLGCFRVLAELPRDSFGPYIISMATAPSDVLAVELLQRECHVRDP
LPVVPLFERLADLQNAPASMERLFSVDWYLQRINGKQQVMIGYSDSGKDAGRLSA
AWQLYRAQEELAQVAKRYGVKLTMFHGRGGTVGRGGGPSHLAILSQPPDTINGSIR
VTIQGEVIEHSFGEEHLCFRTLERFTAATLEHGMHPPVSPKPEWRKLMDEMAVVAT
EEYRSIVFREPRFVEYFRSATPETEYGRMNIGSRPAKRKPKGGIESLRAIPWIFSWTQT
RFHLPVWLGVGAAFQYAIKKDSKNIQKLKDMYKEWPFFRVTIDLLEMVFAKGDPSI
AGLYDELLVAADLKPFGEQLRNKYLETQQFLLQIAGHKEILEGDPYLKQGLRLRNP
YITTLNVFQAYTLKLMRDPSFQVKKQPPMSKEFADEKKPAGLVELNPASEYAPGLE
DTLILTMKGIAAGMQNTG*

>pET-1B *Panicum pygmaeum* PEPC amino acid sequence

MGSSHHHHHHENLYFQSNAMASSKAPGPVERHQSIDAQLRLLAPGKVSEDDKLVE
YDALLVDRFLDILQDLHGPSLREFVQECYELSAEYEGDRDAARLGELGDRLTGLAP
ADAIVVASSFSHMLNLANLAEEVQIAHRRRRNKLKRGDFADEASATTESDIEETLKRL
VSELGKSREEVFDALKNQTVDLVFTAHPTQSIRRSLLQKHARIRNCLTQLYAKDITA
DDKQELDEALQREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRV
DTALKNIGIDERLPYNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLARMMAANL
YFSQIEELMFELSMWRCNDELRVRAEELHRASRKAAKHYIEFWKQIPPNEPYRVILG
YVRDKLYYTRERSRHLLTTGFSEIPEDSAFTNVEEFLEPLELCYRSLCACGDKTIADG
SLLDFLRQVSTFGLSLVKLDIRQESERHTDVLDAITTHLGIGSYREWPEEKRQEWLLS
ELRGKRPLLGPDLPQTEEVADVLGTFRVLAELPPDSFGAYIISMATAPSDVLTVELLQ
RECHVRHPLRVVPLFEKLADLEAAPAAVARLFSVDWYMDRINGKQEVMIGYSDSG
KDAGRLSAAWQLYKAQEELVQVAKRYGVKLTMFHGRGGTVGRGGGPTHLAILSQ
PPDTIHGSLRVTVQGEVIEHSFGEEHLCFRTLQRFTAATLEHGMHPPVSPKPEWRAL
MDELAVVATEEYRSIVFKEPRFVEYFRSATPETEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAMKKDIRNIQTLREMYNEWPFFRVTLDLLE
MVFAKGDPGIAGLYDELVVADDLKPFGEQLRNNYVETQQLLLQVAGHKDILEGDP
YLKQRLRLRDPYITTLNVCQAYTLKRIRDPSFQVTAQRPLSKEFADENQPAGLVKLN
PASEYAPGLEDTLILTMKGIAAGMQNTG*

Amino acids of PEPC positively selected for C_4 photosynthesis have various functions

Nicholas R. Moody[1], Pascal-Antoine Christin[2] and James D. Reid[1]

[1]Department of Chemistry, University of Sheffield, Brook Hill, Sheffield, S3 7HF.

[2]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN.

Personal Contribution: I generated and analysed all the data and wrote the paper with the help of my co-authors. As of the time of thesis submission, this paper was in draft form.

ABSTRACT

The C_4 photosynthetic cycle is an elaborate carbon-concentrating mechanism that improves the efficiency of carbon fixation in tropical conditions. In this cycle, the enzyme phosphoenolpyruvate carboxylase (PEPC) catalyses the initial carbon fixation and is always recruited for this role in the many independent origins of the trait. During the evolution of C_4 PEPC the enzyme has changed with respect to key kinetic properties, such as an increase in specificity for bicarbonate, a decrease in specificity for phosphoenolpyruvate (PEP) and a decrease in sensitivity to inhibitors, including malate and aspartate. Previous work has identified multiple amino acid replacement in C_4-specific PEPC driven by positive selection. Some of these replacements happened independently in distant C_4 origins, but their functional significance remained unknown. Here, we use site-directed mutagenesis to investigate the effect of amino acid mutations associated with C_4 PEPC. We show that the mutation A780S (*Zea mays* numbering) is responsible for part of the change in specificity for PEP in both *Flaveria* and grasses. However, this site does not affect malate inhibition or bicarbonate specificity, showing that these traits are independent. The two other C_4 mutations, H665N and S761A, do not affect any of the studied kinetic parameters, suggesting that these changes adapted non-kinetic aspects of C_4 PEPC. We conclude that the adaptation of PEPC for the C_4 context involved efficiency of protein synthesis and posttranslational modification in addition to kinetic properties.

The carbon concentration mechanism C_4 photosynthesis boosts productivity of plants in tropical conditions (Atkinson *et al.*, 2016). This is achieved by fixing carbon dioxide with the enzyme phosphoenolpyruvate carboxylase (PEPC) in a cell segregated from the enzyme Ribulose-bisphosphate carboxylase/oxygenase (RuBisCO), the enzyme catalysing the entry of inorganic carbon into the Calvin-Benson cycle (Hatch, 1987; Sage, 2004; Sage, Sage and Kocacinar, 2012). In C_3 plants, RuBisCO is in direct contact with atmospheric gases, leading to the oxygen fixation by the enzyme, resulting in toxic products that require an

energetically intensive process called photorespiration to breakdown. In conditions such as high temperatures, aridity or salinity, where the ratio of oxygen to carbon dioxide in the cell is increased, photorespiration can represent a high cost for the plant (Ehleringer and Björkman, 1977; Skillman, 2007). C_4 photosynthesis uses PEPC to fix carbon dioxide in the form of bicarbonate to produce oxaloacetate which is rapidly converted to aspartate or malate (Bräutigam *et al.*, 2014). This acid is shuttled to the RuBisCO containing cell, where atmospheric gas diffusion is limited, and its decarboxylation releases the carbon dioxide, increasing its concentration RuBisCO (von Caemmerer and Furbank, 2003).

The C_4 trait is highly convergent, with over 60 different origins in flowering plants (Sage, Christin and Edwards, 2011). The enzymes of the C_4 cycle exist in C_3 plants, as a result the evolution of the C_4 trait involved the co-option of multiple genes, which were subsequently modified to alter their expression patterns and the kinetic properties of the encoded enzymes (Blasing, Westhoff and Svensson, 2000; Tausta *et al.*, 2002; Aubry, Brown and Hibberd, 2011; Christin *et al.*, 2013). The gene encoding PEPC that has been co-opted for C_4 was especially massively upregulated (Bräutigam *et al.*, 2011; Lauterbach *et al.*, 2017; Moreno-Villena *et al.*, 2018). Changes in PEPC kinetic properties have been investigated in the model genus *Flaveria*, which includes closely-related C_4 and non-C_4 species (Chapter 2; Svensson, Bläsing and Westhoff, 1997, 2003; McKown, Moncalvo and Dengler, 2005), and more recently in grasses (Chapter 3). These efforts have shown a reduction of specificity for phosphoenolpyruvate (PEP), an increase of specificity for bicarbonate and a reduction of sensitivity to the inhibitors malate and aspartate during C_4 evolution (Chapter 3). The magnitude of these changes is higher in grasses than in *Flaveria*, which reflect the longer amount of time the former spent in a C_4 state (Chapter 3). An increase in the sensitivity to the activator glucose-6-phosphate has also been observed in the C_4 *Flaveria* (Westhoff *et al.*, 1997; Engelmann *et al.*, 2003; Gowik and Westhoff, 2011).

The evolutionary drivers and molecular basis of the C_4 specific properties are still not well understood. Analysis of the evolution of the amino acid sequence of C_4 PEPC has

shown that at least 22 sites have been driven by positive selection in grasses and sedges. Of these sites, three are also observed in C_4 *Flaveria* (Christin *et al.*, 2007; Besnard *et al.*, 2009). Some of these mutations have been shown to be responsible for key C_4 specific kinetic properties. Of these, a mutation for alanine to serine at position 780 (*Zea mays* numbering, accession number NM_001161348.2) has been identified as an important determinant of the low specificity for PEP of the C_4 form of the enzyme (Bläsing, Westhoff and Svensson, 2000). Analyses of *Flaveria* mutants have further indicated that the region from position 296 to 437 (301 to 422 in *Z. mays*) is responsible for the rest of the change in specificity for PEP (Blasing, Westhoff and Svensson, 2000; Engelmann *et al.*, 2002). These two regions were however not involved controlling in the IC_{50} for malate (Jacobs *et al.*, 2008). However, the effect of these mutants on bicarbonate specificity remains unknown. As the effect of these mutations on bicarbonate specificity has not been investigated, the link between bicarbonate specificity and PEP specificity is unclear. Of the other amino acid replacements evaluated experimentally, the mutation of position 890 (*Z. mays* numbering) from arginine to glycine reduced sensitivity to uncompetitive inhibition (Paulus, Schlieper and Groth, 2013). Additionally, monocot specific mutation of position 100 from asparagine to serine increased PEPC sensitivity to activation by neutral amino acids, such as glycine (González-Segura *et al.*, 2018).

In this work, we use site-directed mutagenesis to investigate the effects of three C_4 specific amino acid mutations in genes encoding C_4 and non-C_4 from the grass genus *Panicum* and the eudicot genus *Flaveria*. The first of these three mutations is the serine/alanine replacement at position 780, which was changed in most C_4 lineages (Christin *et al.*, 2007; Besnard *et al.*, 2009) and shown to be a major determinant of PEP affinity in *Flaveria* (Blasing, Westhoff and Svensson, 2000). The second is a histidine/arginine mutation at position 665, which occurred in both C_4 *Flaveria* and *Panicum*, but has yet to be experimentally investigated. The third is alanine/serine transition at position 761, a mutation that occurred in most grasses, but not *Flaveria*, and lies in the active site. For each of these

sites, the C_4-specific amino acid residue is generated in the non-C_4 gene, and vice versa. The mutants generated are then evaluated biochemically, using the approaches described previously (Chapters 2 and 3). Our work thereby evaluates the effect of each of these mutations in distantly-related non-C_4 genes, and sheds new light onto the enzyme properties selected for the C_4 context.

RESULTS

Site directed mutagenesis and protein purification

Site specific mutations were introduced into the PEPC genes from *Flaveria trinervia* and *Flaveria pringlei* (Chapter 2), and *Panicum queenslandicum* and *Panicum pygmaeum* (Chapter 3). For the investigation of C_4 mutation H665N, the arginine amino acid was mutated to a histidine in *F. trinervia* and *P. queenslandicum* PEPC, and the histidine amino acid was mutated to an arginine in *F. pringlei* and *P. pygmaeum* PEPC. For the investigation of grass specific C_4 mutation S761A, the alanine amino acid was mutated to serine in *P. queenslandicum* PEPC, and the serine amino acid was mutated to alanine in *F. trinervia* PEPC and *P. pygmaeum* PEPC. For the investigation of the C_4 mutation A780S, the serine amino acid was mutated to alanine in *F. trinervia* PEPC and *P. queenslandicum* PEPC, and the alanine amino acid was mutated to serine in *F. pringlei* PEPC and *P. pygmaeum* PEPC.

Mutants were expressed in *E. coli* as described previously (Chapter 2). Expressed protein was purified to > 95% by SDS PAGE with a single immobilised metal column as described previously (Chapter 2; Supp. Figure S 1).

Kinetic analysis of mutant PEPCs varying bicarbonate and PEP

The specificity for bicarbonate of mutant enzymes was determined using a gas-tight assay system as described previously (Chapter 2) and compared with native PEPCs (Chapter 2 and 3). Assays were performed at a saturating PEP concentration, while varying the concentration of bicarbonate (Supp Figure S 2). A change in bicarbonate specificity was not observed in mutants at position 665 (Figure 2A), position 761 (Figure 3A), or position 780 (Figure 1A).

The specificity for PEP of the mutant enzymes was determined at saturating bicarbonate. A change in PEP specificity was not observed in mutants at position 665 (Figure 2B), or position 761 (Figure 3B). A decrease in PEP specificity was observed for the A780S mutants of *P. pygmaeum* and *F. pringlei,* and an increase in PEP specificity was

observed for the S780A mutants of *P. queenslandicum* and *F. trinervia* (Figure 1B). PEP and bicarbonate specificity constants are summarised in Supp. Table 1.

PEPC inhibition by malate

The effect of these mutations on the inhibition by malate was also assessed. All the mutants showed mixed inhibition in the presence of PEP at pH 8.0. No change in the inhibition parameters at limiting or saturating PEP were observed for mutants of positions 665 (Figure 2C and D), 761 (Figure 3C and D), or 780 (Figure 1C and D). Malate inhibition constants are summarised in Supp. Table 1.

Amino acid biosynthetic cost

The energetic cost of producing PEPC for each species was estimated using amino acid expression cost in terms of high-energy phosphate bonds using values previously determined (Heizer, Raymer and Krane, 2011). This biosynthetic cost was taken as the number of high-energy phosphate bonds required to synthesise the amino acid in *E. coli*. The biosynthetic costs obtained show *P. queenslandicum* PEPC requires hydrolysis of 0.08% fewer high-energy phosphates than the *P. pygmaeum PEPC*. Likewise, the *F. trinervia* PEPC also requires hydrolysis of 0.08% fewer high-energy phosphates than the *F. pringlei* PEPC (Supp. Table 2).

Differential Scanning Fluorimetry

The melting temperature of native PEPC and mutants was determined using differential scanning fluorimetry. Comparison of the melting temperature of native PEPCs showed that C_3 PEPCs *F. pringlei* (T_m = 53.5 ± 2.1°C) and *P. pygmaeum* (T_m = 47.4 ± 0.1°C), were slightly more stable than the C_4 PEPCs, *F. trinervia* (T_m = 40.0 ± 0.8°C, P = 0.003) and *P. queenslandicum* (T_m = 42.8 ± 0.4°C, P = 0.0007; Supp. Figure 7). Mutants at positions 665, 761 and 780, showed no change in melting temperature compared to the native enzyme (Supp. Figure 8B, C, and D; $P > 0.05$). Protein stability does not appear to be affected by the investigated mutations.

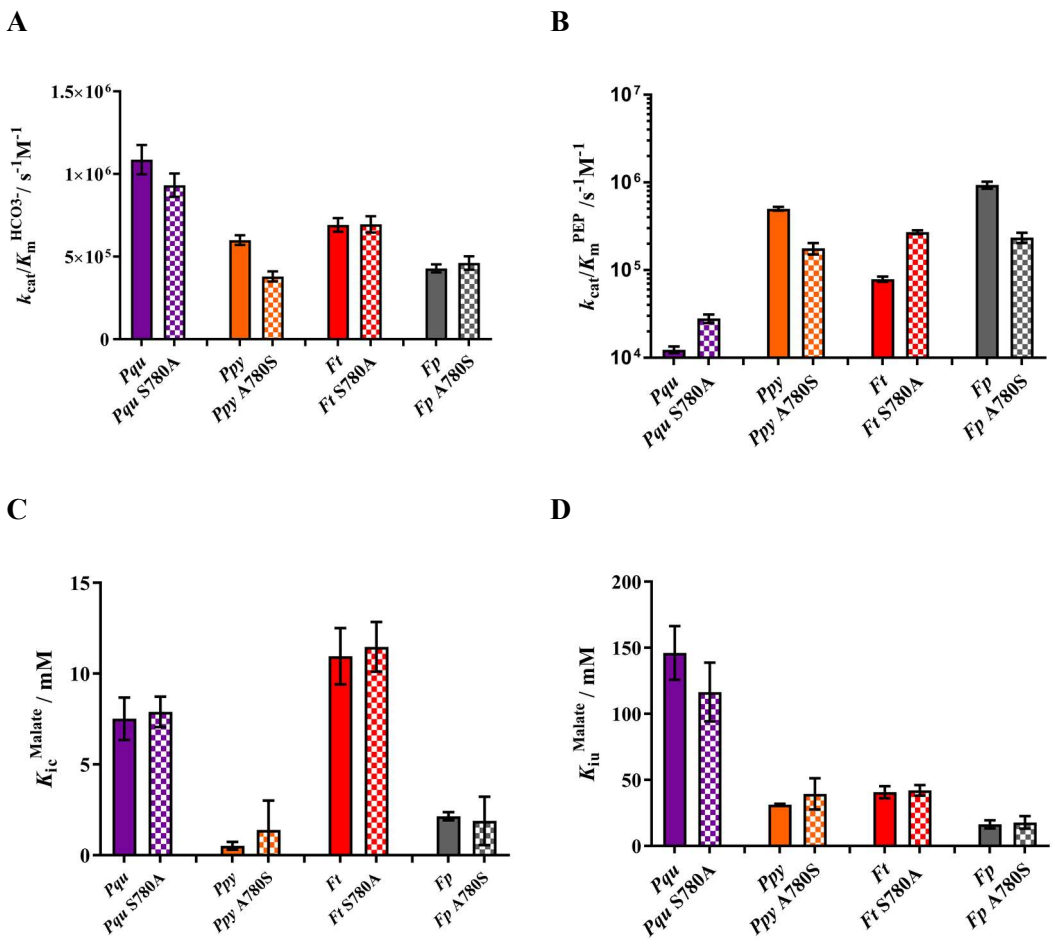


Figure 1: Kinetic parameters determined from the investigation of the C_4 mutation A780S. error bars represent standard errors based on fitted curves. Specificity constants for the native enzymes have been given previously (Chapter 2 and 3). **A** Summary of bicarbonate specificity parameter for 780 mutants, values derived from Supp. Figure 2. No effect of mutations detected ($P > 0.05$). **B** Summary of PEP specificity parameter for 780 mutants, values derived from Supp. Figure 3. An increase is specificity for PEP is observed for the S780A mutants ($P = 0.0001$), and a decrease in specificity is observed in A780S ($P = 0.0001$). **C** Summary of K_{ic}^{Malate} for 780 mutants values derived from Supp. Figure 4. No effect of mutations detected ($P > 0.05$). **D** Summary of K_{ic}^{Malate} for 780 mutants values derived from Supp. Figure 5. No effect of mutations detected ($P > 0.05$).

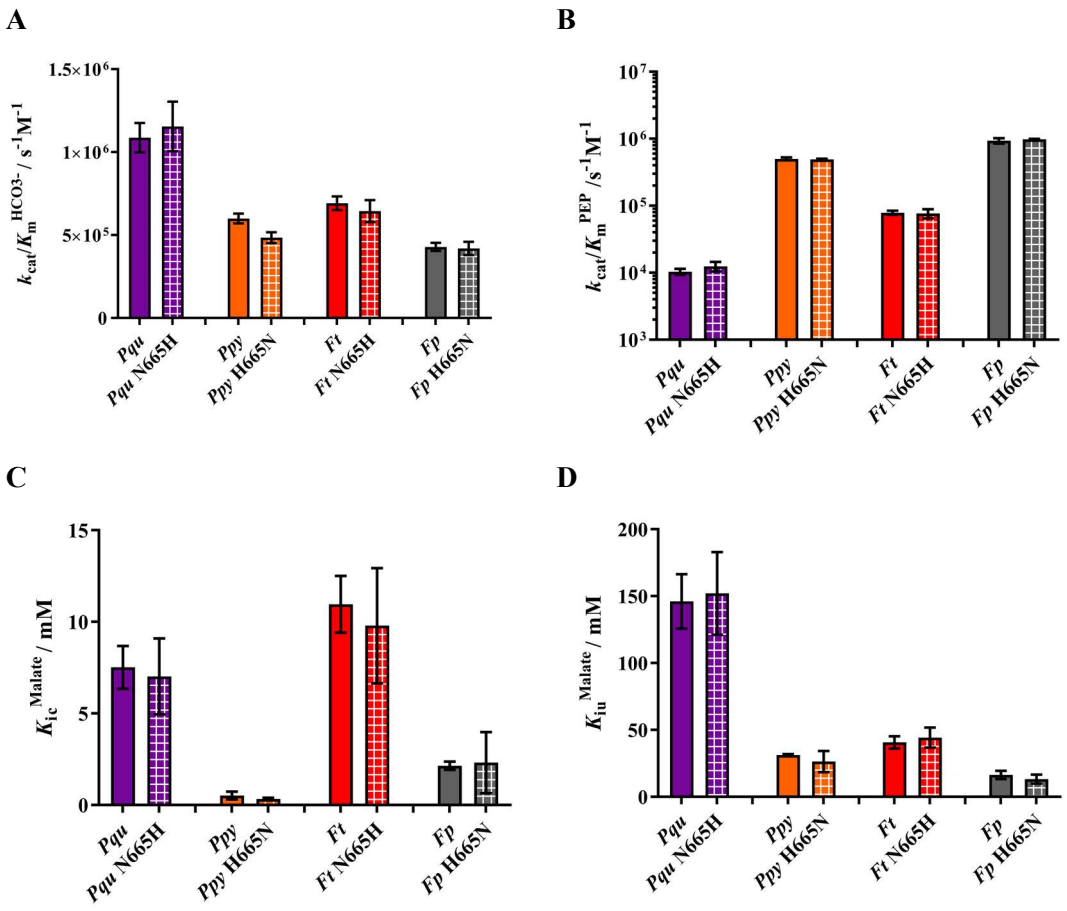


Figure 2: Kinetic parameters determined from the investigation of the C_4 mutation H665N. error bars represent standard errors based on fitted curves. Specificity constants for the native enzymes have been given previously (Chapter 2 and 3)**. A** Summary of bicarbonate specificity parameter for 665 mutants, values derived from Supp. Figure 2. No effect of mutations detected ($P > 0.05$). **B** Summary of PEP specificity parameter for 665 mutants, values derived from Supp. Figure 3. No effect of mutations detected ($P > 0.05$). **C** Summary of K_{ic}^{Malate} for 665 mutants' values derived from Supp. Figure 4. No effect of mutations detected ($P > 0.05$). **D** Summary of K_{ic}^{Malate} for 665 mutants' values derived from Supp. Figure 5. No effect of mutations detected ($P > 0.05$).

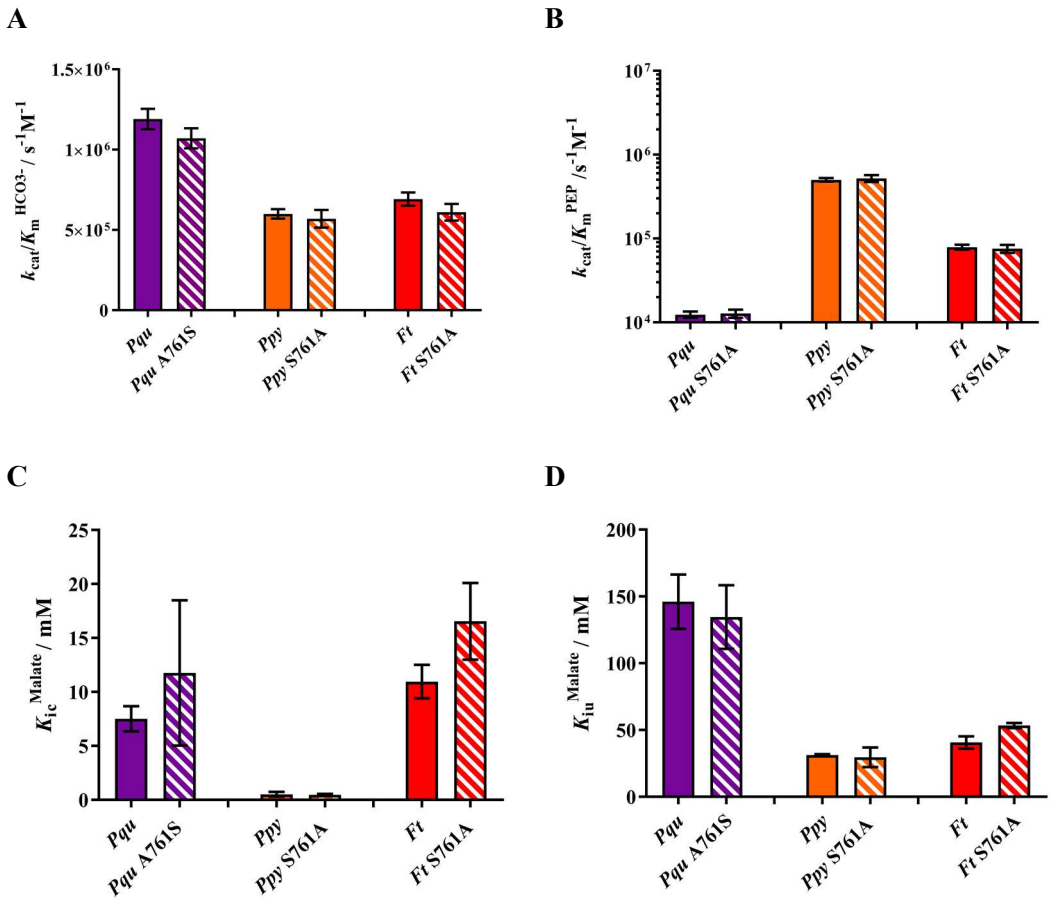


Figure 3: Kinetic parameters determined from the investigation of the grass specific C_4 mutation S761A. error bars represent standard errors based on fitted curves. Specificity constants for the native enzymes have been given previously (Chapter 2 and 3). **A** Summary of bicarbonate specificity parameter for 761 mutants, values derived from Supp. Figure 2. No effect of mutations detected ($P > 0.05$). **B** Summary of PEP specificity parameter for 761 mutants, values derived from Supp. Figure 3. No effect of mutations detected ($P > 0.05$). **C** Summary of K_{ic}^{Malate} for 761 mutants' values derived from Supp. Figure 4. No effect of mutations detected ($P > 0.05$). **D** Summary of K_{ic}^{Malate} for 761 mutants' values derived from Supp. Figure 5. No effect of mutations detected ($P > 0.05$).

DISCUSSION

C_4 PEPC has evolved from an ancestor with a non-carbon fixing role and adapted to meet the demands of the carbon fixing role by increasing its expression level (Bräutigam *et al.*, 2011; Moreno-Villena *et al.*, 2018), and by changes in key kinetic properties such as specificity for PEP and bicarbonate, and sensitivity to inhibitors (Chapter 2). These C_4 specific properties have been shown to be convergent in nature, being present in distantly related C_4 lineages (Chapter 3). The structural origin of these changes in kinetic properties has not been fully elucidated, even though C_4 specific amino acids have been selected for convergently across C_4 grasses, sedges and eudicots (Christin *et al.*, 2007; Besnard *et al.*, 2009). As a result of limited experimental investigation, the functional effect of these convergent C_4 amino acid changes, and therefore their selective advantages, are not fully understood.

The alanine to serine mutation at position 780 reduces specificity for PEP without changing the bicarbonate specificity.

Position 780 lies on an α-helix above the active site of PEPC in the *Z. mays* crystal structure (Matsumura *et al.*, 2002). The mutation to serine at position 780 is selected for in the C_4 *Flaveria* and grasses and has been shown to be an important determinant for C_4 activity (Hermans and Westhoff, 1992; Blasing, Westhoff and Svensson, 2000). When this site was mutated to serine in the C_3 *Flaveria* enzyme, the mutant showed kinetic properties towards PEP similar to a C_4 enzyme with a decrease in k_{cat}/K_m^{PEP} (Blasing, Westhoff and Svensson, 2000; Engelmann *et al.*, 2002; Svensson, Bläsing and Westhoff, 2003). The structural basis of how this mutation changes the PEP specificity is not fully understood, the change to serine at this position may give rise to hydrogen bonding interactions with PEP or other parts of the enzymes (Blasing, Westhoff and Svensson, 2000). This change may also affect how the enzyme interacts with bicarbonate. Previous work has shown the alanine to serine mutation does not affect inhibition by malate at a single concentration of PEP (Jacobs *et al.*, 2008).

Our investigation shows that when this residue is mutated from alanine to serine in the non-C_4 enzymes from *Panicum pygmaeum* and *Flaveria pringlei* the resulting mutants have a comparable decrease in specificity towards PEP (Figure 3B). For the C_4 enzymes from *Panicum queensladicum* and *Flaveria trinervia,* the S780A mutants show an increase in specificity for PEP. As reported earlier, these point mutations do not account for the full difference in behaviour between the C_3 and C_4 forms of the enzyme (Blasing, Westhoff and Svensson, 2000).

None of these four mutants exhibited a change in bicarbonate specificity or malate inhibition (Figure 1). Our results conform to the previously described results in the *Flaveria* enzymes (Blasing, Westhoff and Svensson, 2000) and shows that similar behaviour is seen in the PEPC enzymes from the grasses, *P. pygmaeum* and *P. queensladicum*. Notably, despite the significance of this residue for the productive interaction with PEP, mutations at this position do not appear to change the sensitivity of the enzyme to the other substrate, bicarbonate, or to the inhibitor, malate. This mutation is seen in the majority of C_4 PEPCs (Christin *et al.*, 2007; Besnard *et al.*, 2009). A similar magnitude change in specificity for PEP is observed in both of these C_4 enzymes when mutated, which is not combined with any significant change in inhibitor sensitivity or bicarbonate specificity. This indicates that this C_4 specific mutation is responsible for the same change in C_4 specific properties in PEPC evolutionary distance C_4 species. This further supports the hypothesis that low PEP specificity is not sacrificed to improve another enzyme property, but is an essential C_4 property as discussed previously (Chapter 2).

The conserved C_4 specific change at position 665 is functionally silent with respect to investigated properties.

The H665N mutation is conserved in the C_4 grass and the C_4 *Flaveria* PEPC. In the *Zea mays* PEPC crystal structure (Kai, Matsumura and Izui, 2003), the C_α of residue 665 is approximately 22 Å away from the C_α of catalytically essential residue H177. The H665N mutation can be considered to be chemically conserved as the side-chain nitrogen in

asparagine can substitute for an imidazole nitrogen in the histidine side chain and retain similar electrostatic and hydrogen bonding interactions (Fersht, 2002). This mutation is shared among C_4 grasses, *Flaveria* and sedges, and was assigned to C_4-specific positive selection with a 0.995 posterior probability (Christin *et al.*, 2007; Besnard *et al.*, 2009).

Our work shows a mutation from histidine to asparagine at position 665 does not change the bicarbonate specificity, PEP specificity or malate inhibition (Figure 2). This suggests that this mutation might has been selected for a different reason. When the cost of amino acid biosynthesis is considered, histidine costs 29.0 in number of high energy bonds while asparagine costs 18.5 (Heizer, Raymer and Krane, 2011). Combined with the massive upregulation of C_4 PEPC, a single mutation away from an expensive amino acid may be important in reducing the overall cost of enzyme production. Comparison of the overall biosynthetic cost of PEPC production, however, suggests that there is no major difference in cost between the C_3 and C_4 forms of the enzyme. C_4 specific mutations can require an increase in biosynthetic cost. For example position 573, which is glutamate in non-C_4 PEPCs, and is lysine or glutamine in C_4 grasses, sedges and eudicots (Besnard *et al.*, 2009). The nature of these adaptation is not fully understood; however, this change involves an inversion of electrostatic interactions and increase in cost from 9.5 to 37.0/10.5. Some of the selected changes in amino acid side chain may well be epistatic, selected to enable other directly functional changes. Epistasy has been shown to be an important driving force in the selection of amino acids (Kimura, 1985; Halabi *et al.*, 2009; Tracewell and Arnold, 2009; Breen *et al.*, 2012).

The conserved change in the active site of PEPC at position 761 is also functionally silent.

Position 761 lies in the active site between amino acids thought to be essential for substrate interactions *i.e.* R^{759}P**A**KRR^{764}. In this motif, R^{759} is essential for PEP binding and the KRR^{764} region is thought to be essential for bicarbonate binding in *Z. mays* PEPC (Matsumura *et al.*, 1999; Izui *et al.*, 2004). The mutation from serine to alanine has not been

observed in *Flaveria* but occurred in most C_4 origins within grasses, where it was assigned to positive selection with a 0.999 posterior probability (Christin *et al.*, 2007).

Investigation of the S761A mutation has shown that the C_4 specific change does not appear to contribute to bicarbonate specificity, PEP specificity or malate inhibition (Figure 3). This position is predicted to be part of a casein kinase II phosphorylation site (Hulo, 2006; Christin *et al.*, 2007). Casein kinase II has important functions in regulation in higher plants (Kanekatsu *et al.*, 1998; Ogiso *et al.*, 2010). The much higher expression level and concentration of PEPC makes the protein more likely to be a target of non-specific phosphorylation. As this amino acid is part of a loop of the active site, any unnecessary modifications may affect protein activity, such as reducing the mobility of active site lid or the interactions substrates. This would be undesirable and create selection pressure for mutation. This mutation is observed in PEPCs from older C_4 species such as *P. queenslandicum, Z. mays and Alternanthera pugens*. However, it is not observed in the PEPC from one of the most recent C_4 origins, *Flaveria* (Christin *et al.*, 2011). This would suggest that this is a later adaption as phosphorylation would be more likely with increases in expression level.

The co-option of PEPC for C_4 carbon fixation involved the adaptation of the enzyme to meet the demands of a central metabolic role. As well as an increase in expression level, convergent amino acid changes have been observed across C_4 PEPCs, some of which have been investigated and were shown to be important for key kinetic properties. The mutation K890G (*Z. mays* numbering) has been shown to decrease the malate sensitivity (Paulus, Schlieper and Groth, 2013), A780S is partially responsible for a decrease in PEP specificity (Blasing, Westhoff and Svensson, 2000), and K100S is responsible for monocot-specific sensitivity to activation by glycine (González-Segura *et al.*, 2018). In this report, we have focused on three residues at positions 665, 761 and 780. The mutations at these positions are strongly selected in the C_4 forms of PEPC. We have examined the effects of these mutations

in both C_4 and C_4 forms of PEPC in both *Flaveria* and *Panicum*. At position 780, the same change is observed on mutation in the C_3 PEPC, in *Panicum* and *Flaveria*, but no change of bicarbonate or inhibition kinetics was observed. Remarkably, there is no change in the kinetic properties or stability of mutations at the other two of these sites; the core biochemical properties of substrate specificity and inhibitor sensitivity appear unconnected to these highly selected amino acids.

This observation is particularly surprising in the case of the active-site residue 761. The mutation at this site from serine to alanine removes a putative casein kinase II phosphorylation site. If the serine is phosphorylated, this would become disadvantageous due to the location in the active site. With the increase in C_4 expression level, the avoidance of incorrect phosphorylation could be a driving force for selection. Position 665 mutation is from histidine to arginine and results in a limited change in chemical functionality, suggesting this mutation could be driven by the biosynthetic cost optimisation. Selection for lower cost amino acids in highly expressed protein in prokaryotes has been observed (Akashi and Gojobori, 2002; Heizer *et al.*, 2006).

Position 780 has long been identified as an important determinant in C_4 activity (Hermans and Westhoff, 1992). We have shown that mutations at this position confer the change in PEPC's evolutionary distant origins of C_4 photosynthesis. Positions 665 and 761 have been shown to be functionally silent with respect to substrate specificity and inhibitor sensitivity. However, the changes in amino acid properties of these sites suggest that these positions are driven by enzyme epistasis and other factors with the cell not encountered at low expression levels.

EXPERIMENTAL PROCEDURES

Unless otherwise stated, reagents and components were from Sigma. For purification, unless otherwise stated the equipment used procured from GE Healthcare. Unless otherwise stated enzymes and *E. coli* strains were from NEB.

DNA Mutagenesis

Mutations were introduced to PEPC genes of *Flaveria trinervia* and *Flaveria pringlei* (Chapter 2), *Panicum queenslandicum* and *Panicum pygmaeum* (Chapter 3) using the appropriate primers (Table 3) and Q5 DNA polymerase and the KLD enzyme kit. DNA product was verified with agarose gel. High competency DH5α cells were transformed with the DNA product. Resultant plasmids were Sanger sequenced (GATC Biotech) using the appropriate primers.

Protein Expression

For protein expression, BL21λ(DE3) strain of *E. coli* (NEB) was used. Chemically competent *E. coli* cells were transformed with each of the plasmids. Eight litres of cultures were grown in LB medium at 37°C to OD_{600} 0.8. Cultures were cooled to 4°C for one hour prior to recombinant protein induction with 0.5 mM IPTG (Fischer). Cultures were then incubated at 18°C for 18 hours. Cells were harvested by centrifugation at 5,422 × *g* for 25 minutes and stored at -80°C.

Protein Purification

Cells were suspended with IMAC buffer (25 mM Tris, 0.5 M NaCl, 0.3 M glycerol, 20 mM imidazole (Acros Scientific)), 10 ml per 2 L of culture with 50 µl of 50mg ml^{-1} DNase I and 100 µl of 100 mg ml^{-1} Pefabloc. Cells were passed twice through a cell disruptor (Constant Systems) before centrifugation at 26,902 × *g* for 40 minutes. The supernatant was passed through a 0.45 µm pore filter (Elkay Labs.). PEPC was separated from soluble protein with a prepacked 1 ml nickel affinity column using an ÄKTA™ Pure 25 L Chromatography System. The loaded column was washed with 50 column volumes of IMAC buffer, then 50 column volumes of IMAC buffer containing 150 mM imidazole. Pure PEPC was eluted with 10 column volumes of IMAC buffer containing 400 mM imidazole.

Protein eluted from IMAC purification was loaded onto a Sephadex G50 desalting column (Amersham Biosciences) and rebuffered in storage buffer (20 mM Tris, 5% v/v glycerol, 150 mM KCl, 1 mM DTT (AnaSpec. Inc.)). Protein was aliquoted and frozen at -80°C until use. Mutant enzyme stability was verified with circular dichroism (Supp Figure 6).

Enzyme Quantification

PEPC enzyme concentration was quantified by absorption at 280 nm, the native enzyme extinction coefficient was used (Chapters 2 and 3), assuming no change in extinction coefficient on mutation. The enzyme extinction coefficient was calculated using the ExPASy protein parameter tool and corrected by determining the absorbance of the protein denatured in 6 M guanidine hydrochloride (Gill and von Hippel, 1989). The difference between the denatured and folded protein at 280 nm was used to adjust the extinction coefficient of the protein. The extinction coefficient for *Flaveria trinervia* PEPC was determined to be 120480 M^{-1} cm^{-1}, and the extinction coefficient for *Flaveria pringlei* PEPC was determined to be 117030 M^{-1} cm^{-1}. The extinction coefficient for *Panicum queenslandicum* PEPC was determined to be 105805 M^{-1} cm^{-1}, and the extinction coefficient for *Panicum pygmaeum* PEPC was determined to be 111514 M^{-1} cm^{-1}. It is assumed that all enzyme used to start the assay was active.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE) Analysis

Total protein concentration for purification efficiency was determined using the BCA Pierce quantification kit (Thermo Scientific). Concentration was determined using a standard curve performed with bovine serum albumin, over a concentration range 0 – 2.0 $mgml^{-1}$.

Protein samples were analysed for purity using SDS PAGE analysis. Protein samples were quantified using the BCA Pierce method, 25 µg of cell lysis and 5 µg of pure protein fractions were denatured in 2 × SDS PAGE loading dye (200 mM Tris.HCl pH 6.8, 2 % SDS, 20 % Glycerol, 0.01% Bromophenol blue (BDH Laboratory Supplies) and 7 % β-mercaptoethanol). Protein was loaded onto an 8% acrylamide SDS gel (Expedeon) with 2 µl of Blue Prestained Protein Standard Broad Range (11-190 kDa) (NEB). Gels were run for 50

minutes at 200 V with 1 × Tris/Glycine/SDS running buffer (Geneflow). Gels were stained with InstantBlue (Expedeon) and imaged with a ChemiDoc™ MP (BioRad).

Enzyme Assays

PEPC activity was measured spectroscopically at 340 nm by coupling to NADH-malate dehydrogenase. Assays with a high fixed concentration of bicarbonate were observed using a FLUOstar plate reader (BMG Labtech) using the 340 nm ± 5 nm absorbance filter (BMG Labtech). Plate reader assays were conducted in a reaction volume of 150 µl at 25°C. Typical reaction mixture contained 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$ (Fluka), 10 mM $KHCO_3$. 0.2 mM NADH (Fischer) and 0.1 Uμl^{-1} malate dehydrogenase. Assays were initiated with the addition of PEPC enzyme. Rates were calculated with a NADH calibration curve.

Assays at a range of bicarbonate concentrations were observed with a Cary Bio 300 spectrophotometer (Agilent Technologies) in the same reaction buffer, in a total reaction volume of 600 µl. In bicarbonate assays, the water and tricine buffer were sparged with nitrogen for 18 hours prior to use in assays. Bicarbonate assays were constructed under a nitrogen flow. Assays were performed in a sealed cuvette. The reaction was initiated with the addition of 50 nM PEPC, delivered with a gastight syringe (Hamilton). Bicarbonate concentrations were controlled with the addition of freshly prepared potassium bicarbonate. Background bicarbonate was determined using an endpoint assay with no potassium bicarbonate, run for 30 minutes. Rates were calculated using the Cary analysis software.

Circular Dichroism Analysis

Protein samples were made to 2 µM in 200 µl of PEPC storage buffer. Samples was analysed in a J-810 Spectropolarimeter (Jasco). Reading were taken four times from 200 to 280 nm, in a method described in (Greenfield, 2007).

Differential Scanning Fluorimetry

Protein at a concentration of 75 µgml^{-1} was suspended in 10 mM HEPES pH 7.5, 150 mM sodium chloride, and 1 × SYPRO orange dye (Invitrogen). Samples were incubated in a RT-PCR machine. Samples were run on a temperature scan from 35 to 95°C at, 1°C min^{-1}, in the method described in (Niesen, Berglund and Vedadi, 2007).

Data Analysis

Kinetic parameters were evaluated by non-linear regression analysis in Igor Pro (Version 7.0.8.1; Wavemetrics Inc., Lake Oswego Orgeon). The statistical significance of changes in kinetic parameters were then analysed with GraphPad Prism 7 for Windows (Version 7.04, GraphPad Software, Inc.) using the analysis of variance function (ANOVA). The following equations were used:

(Equation 1)

$$v_i/_{[E_T]} = \frac{k_{cat} \times [S]}{K_m + [S]}$$

Equation 1, where $v_i/[E_T]$ is the steady state rate divided by the total enzyme concentration, k_{cat} is the first order rate constant, K_m is the Michaelis constant, and [S] is the substrate concentration.

(Equation 2)

$$k^{app} = \frac{k}{1 + {[i]}/{K_i}}$$

Equation 2, where k^{app} is an apparent steady-state rate constant (i.e. apparent k_{cat}/K_m or k_{cat}), k is the uninhibited constant, i is the inhibitor concentration, and K_i is the inhibition constant.

(Equation 3)

$$y = LL + \frac{(UL - LL)}{1 + \exp(\frac{T_m - x}{a})}$$

Equation 3, where y is fluorescence at a given wavelength, x is the temperature, LL is the maximum intensity of fluorescence, UL is the minimum intensity of fluorescence, a denotes the slope of the curve, and T_m is the melting temperature of the protein.

REFERENCES

Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*, *Proceedings of the National Academy of Sciences*, 99(6), pp. 3695–3700.

Atkinson, R. R. L., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M. and Osborne, C. P. (2016) C_4 photosynthesis boosts growth by altering physiology, allocation and size, *Nature Plants*. Nature Publishing Group, 2(5), p. 16038.

Aubry, S., Brown, N. J. and Hibberd, J. M. (2011) The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway, *Journal of Experimental Botany*, 62(9), pp. 3049–3059.

Besnard, G., Muasya, a. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. (2009) Phylogenomics of C_4 Photosynthesis in Sedges (*Cyperaceae*): Multiple Appearances and Genetic Convergence, *Molecular Biology and Evolution*, 26(8), pp. 1909–1919.

Blasing, O. E., Westhoff, P. and Svensson, P. (2000) Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria-a conserved serine residue in the carboxyterminal part of the enzyme is a major determinant for C_4-specific characteristics, *Journal of Biological Chemistry*, 275(36), pp. 27917–23.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., Carr, K. M., Gowik, U., Mass, J., Lercher, M. J., Westhoff, P., Hibberd, J. M. and Weber, A. P. M. (2011) An mRNA Blueprint for C_4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C_3 and C_4 Species, *Plant Physiology*, 155(1), pp. 142–156.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P. and Weber, A. P. M. (2014) Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species, *Journal of experimental botany*, 65(13), pp. 3579–3593.

Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. and Kondrashov, F. A. (2012) Epistasis as the primary factor in molecular evolution, *Nature*. Nature Publishing Group, 490(7421), pp. 535–538.

von Caemmerer, S. and Furbank, R. T. (2003) The C_4 pathway: an efficient CO_2 pump, *Photosynthesis Research*, 77(2/3), pp. 191–207.

Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J. and Osborne, C. P. (2013) Parallel Recruitment of Multiple Genes into C_4 Photosynthesis, *Genome Biology and Evolution*, 5(11), pp. 2174–2187.

Christin, P.-A., Osborne, C. P., Sage, R. F., Arakaki, M. and Edwards, E. J. (2011) C_4 eudicots are not younger than C_4 monocots, *Journal of experimental botany*, 62(9), pp. 3171–3181.

Christin, P., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. (2007) C_4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes, *Current Biology*, 17(14), pp. 1241–1247.

Ehleringer, J. and Björkman, O. (1977) Quantum yields for CO_2 uptake in C_3 and C_4 plants, *Plant Physiology*, 59(577), pp. 86–90.

Engelmann, S., Bläsing, O. E., Gowik, U., Svensson, P. and Westhoff, P. (2003) Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria*- a gradual increase from C_3 to C_4 characteristics., *Planta*, 217(5), pp. 717–25.

Engelmann, S., Bläsing, O. E., Westhoff, P. and Svensson, P. (2002) Serine 774 and amino acids 296 to 437 comprise the major C_4 determinants of the C_4 phosphoenolpyruvate carboxylase of *Flaveria trinervia*, *FEBS Letters*, 524, pp. 11–14.

Fersht, A. (2002) *Structure and Mechanism in Protein Science*. World Scientifc (Series in Structural Biology).

Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Analytical Biochemistry*, 182(2), pp. 319–326.

González-Segura, L., Mújica-Jiménez, C., Juárez-Díaz, J. A., Güémez-Toro, R., Martinez-Castilla, L. P. and Muñoz-Clares, R. A. (2018) Identification of the allosteric site for neutral amino acids in the maize C_4 isozyme of phosphoenolpyruvate carboxylase: The critical role of Ser-100, *Journal of Biological Chemistry*, 293(26), pp. 9945–9957.

Gowik, U. and Westhoff, P. (2011) The path from C_3 to C_4 photosynthesis., *Plant Physiology*, 155(1), pp. 56–63.

Greenfield, N. J. (2007) Using circular dichroism spectra to estimate protein secondary structure, *Nature Protocols*, 1(6), pp. 2876–2890.

Halabi, N., Rivoire, O., Leibler, S. and Ranganathan, R. (2009) Protein Sectors: Evolutionary Units of Three-Dimensional Structure, *Cell*, 138(4), pp. 774–786.

Hatch, M. D. (1987) C_4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure, *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 895(2), pp. 81–106.

Heizer, E. M., Raiford, D. W., Raymer, M. L., Doom, T. E., Miller, R. V. and Krane, D. E. (2006) Amino Acid Cost and Codon-Usage Biases in 6 Prokaryotic Genomes: A Whole-Genome Analysis, *Molecular Biology and Evolution*, 23(9), pp. 1670–1680.

Heizer, E. M., Raymer, M. L. and Krane, D. E. (2011) Amino Acid Biosynthetic Cost and Protein Conservation, *Journal of Molecular Evolution*, 72(5–6), pp. 466–473.

Hermans, J. and Westhoff, P. (1992) Homologous genes for the C_4 isoform of phosphoenolpyruvate carboxylase in a C_3 and a C_4 *Flaveria* species, *MGG Molecular & General Genetics*, 234(2), pp. 275–284.

Hulo, N. (2006) The PROSITE database, *Nucleic Acids Research*, 34(90001), pp. D227–D230.

Izui, K., Matsumura, H., Furumoto, T. and Kai, Y. (2004) Phosphoenolpyruvate Carboxylase: A New Era of Structural Biology, *Annual review of plant biology*, 55, pp. 69–84.

Jacobs, B., Engelmann, S., Westhoff, P. and Gowik, U. (2008) Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*: Determinants for high tolerance towards the inhibitor L-malate, *Plant, Cell and Environment*, 31, pp. 793–803.

Kai, Y., Matsumura, H. and Izui, K. (2003) Phosphoenolpyruvate carboxylase: three-dimensional structure and molecular mechanisms, *Archives of Biochemistry and Biophysics*, 414(2), pp. 170–179.

Kanekatsu, M., Saito, H., Motohashi, K. and Hisabori, T. (1998) The β subunit of chloroplast ATP synthase (CF0CF1-ATPase) is phosphorylated by casein kinase II, *IUBMB Life*, 46(1), pp. 99–105.

Kimura, M. (1985) The role of compensatory neutral mutations in molecular evolution, *Journal of Genetics*, 64(1), p. 7.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. and Gowik, U. (2017) C_3 cotyledons are followed by C_4 leaves: intra-individual transcriptome analysis of *Salsola soda* (*Chenopodiaceae*), *Journal of Experimental Botany*, 68(2), pp. 161–176.

Matsumura, H., Terada, M., Shirakata, S., Inoue, T., Yoshinaga, T., Izui, K. and Kai, Y. (1999) Plausible phosphoenolpyruvate binding site revealed by 2.6 Å structure of Mn^{2+} - bound phosphoenolpyruvate carboxylase from *Escherichia coli*, *FEBS Letters*, 458(2), pp. 93–96.

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*, 10(12), pp. 1721–1730.

McKown, A. D., Moncalvo, J.-M. and Dengler, N. G. (2005) Phylogeny of *Flaveria* (*Asteraceae*) and inference of C_4 photosynthesis evolution, *American Journal of Botany*, 92(11), pp. 1911–1928.

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. and Christin, P. A. (2018) Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis, *Molecular Biology and Evolution*, 35(1), pp. 94–106.

Niesen, F. H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nature Protocols*, 2(9), pp. 2212–2221.

Ogiso, E., Takahashi, Y., Sasaki, T., Yano, M. and Izawa, T. (2010) The Role of Casein Kinase II in Flowering Time Regulation Has Diversified during Evolution, *Plant Physiology*, 152(2), pp. 808–820.

Paulus, J. K., Schlieper, D. and Groth, G. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution, *Nature communications*, 4, p. 1518.

Sage, R. F. (2004) The evolution of C_4 photosynthesis, *New Phytologist*, 161(2), pp. 341–370.

Sage, R. F., Christin, P.-A. and Edwards, E. J. (2011) The C_4 plant lineages of planet Earth, *Journal of experimental botany*, 62(9), pp. 3155–3169.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) Photorespiration and the Evolution of C_4 Photosynthesis, *Annual Review of Plant Biology*, 63(1), pp. 19–47.

Skillman, J. B. (2007) Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark, *Journal of Experimental Botany*, 59(7), pp. 1647–1661.

Svensson, P., Bläsing, O. E. and Westhoff, P. (1997) Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase- a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3)., *European journal of biochemistry*, 246(2), pp. 452–60.

Svensson, P., Bläsing, O. E. and Westhoff, P. (2003) Evolution of C_4 phosphoenolpyruvate carboxylase, *Archives of Biochemistry and Biophysics*, 414(2), pp. 180–188.

Tausta, S. L., Miller Coyle, H., Rothermel, B., Stiefel, V. and Nelson, T. (2002) Maize C_4 and non-C_4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor, *Plant Molecular Biology*, 50(4/5), pp. 635–652.

Tracewell, C. A. and Arnold, F. H. (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time, *Current Opinion in Chemical Biology*, 13(1), pp. 3–9.

Westhoff, P., Svensson, P., Ernst, K., Bläsing, O., Burscheidt, J. and Stockhaus, J. (1997) Molecular Evolution of C_4 Phosphoenolpyruvate Carboxylase in the Genus *Flaveria*, *Australian Journal of Plant Physiology*, 24(4), p. 429.

SUPPLEMENTARY FIGURES

SDS PAGE Analysis

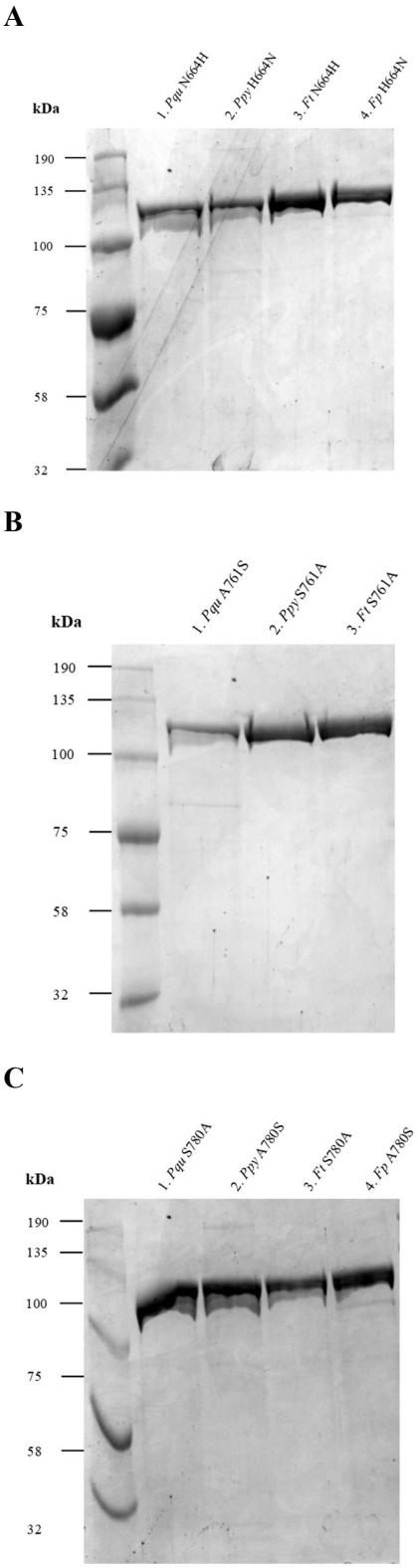


Figure S 1: 8 % acrylamide SDS PAGE analysis of mutants. *ca.* 6 µg of PEPC protein was loaded into each well. **A** Analysis of position 665 mutants. **B** Analysis of position 761 mutants. **C** Analysis of position 780 mutants.

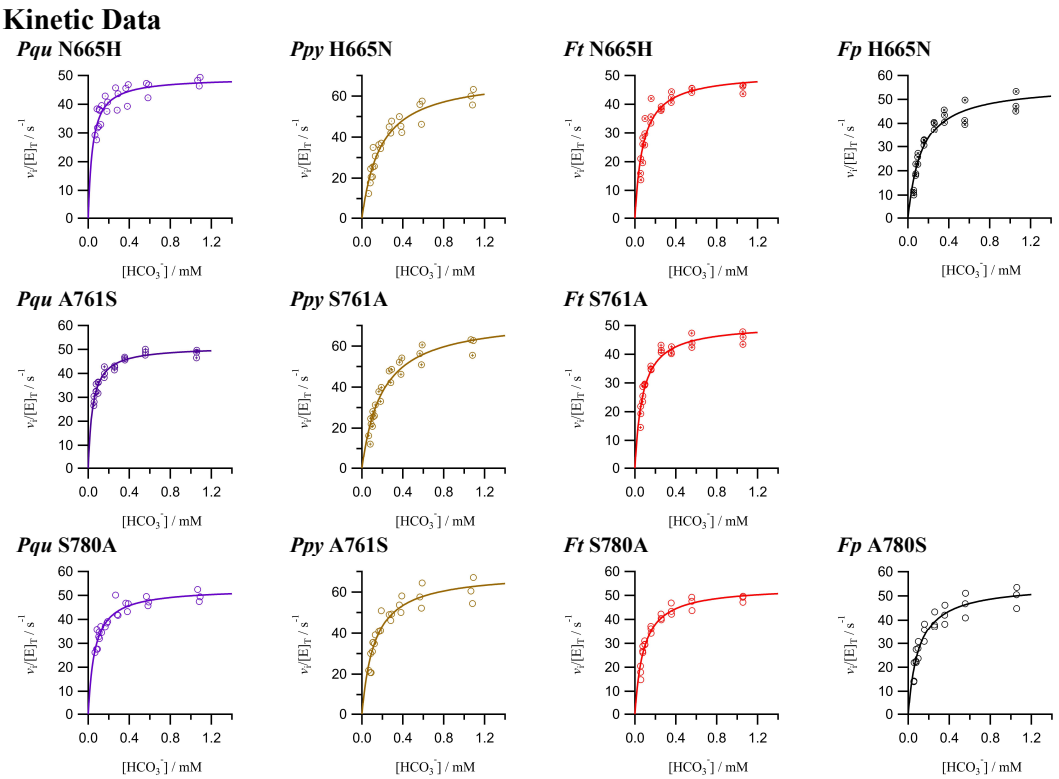


Figure S 2: Rate of oxaloacetic acid formation, catalysed by PEPC mutants, varying the concentration of bicarbonate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM magnesium chloride, 0.2 mM NADH, 0.01 Uµl^{-1} malate dehydrogenase and 50 nM PEPC. Performed at saturating PEP, concentration dependent on the K_m^{PEP} of the mutant (50 mM PEP for Pqu mutants, 10 mM PEP for Ft mutants, and 5 mM for Fp and Ppy mutants). Lines are described by equation 1, k_{cat}/K_m values from the fit of the line are summarised in Table 1.

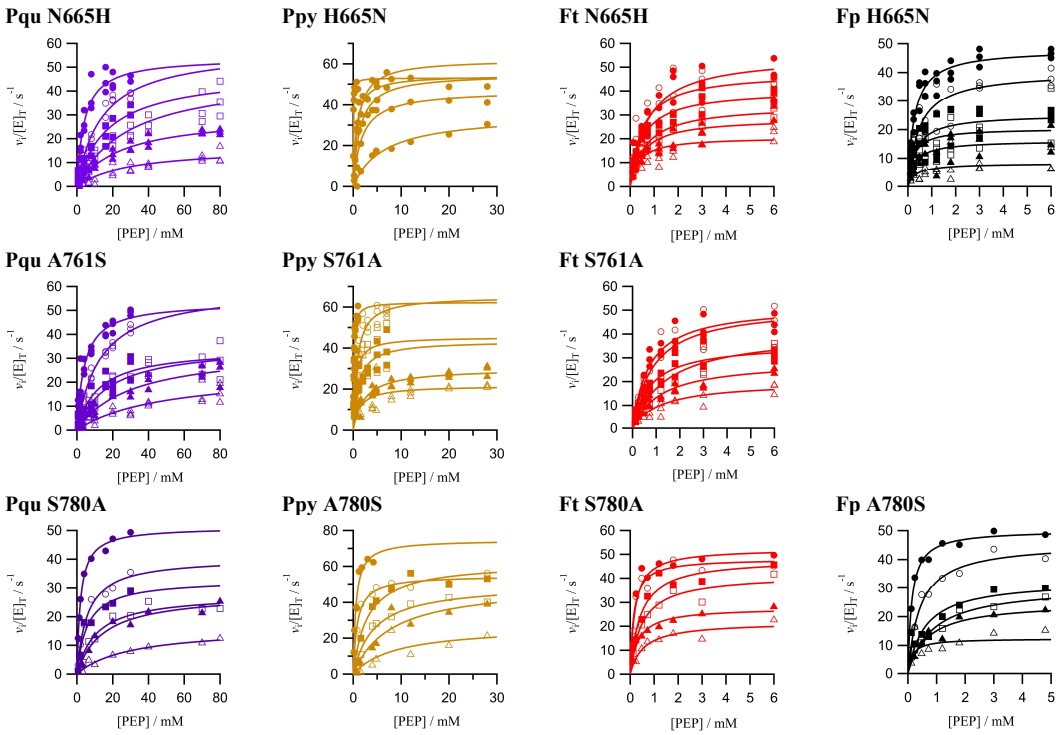


Figure S 3: Primary plot of PEPC mutants inhibited by malate. Assays conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 10 mM $KHCO_3$, 0.2 mM NADH and 0.01 Uμl^{-1} malate dehydrogenase initiated with 5 or 10 nM of mutant PEPC. The lines are described by equation 1.

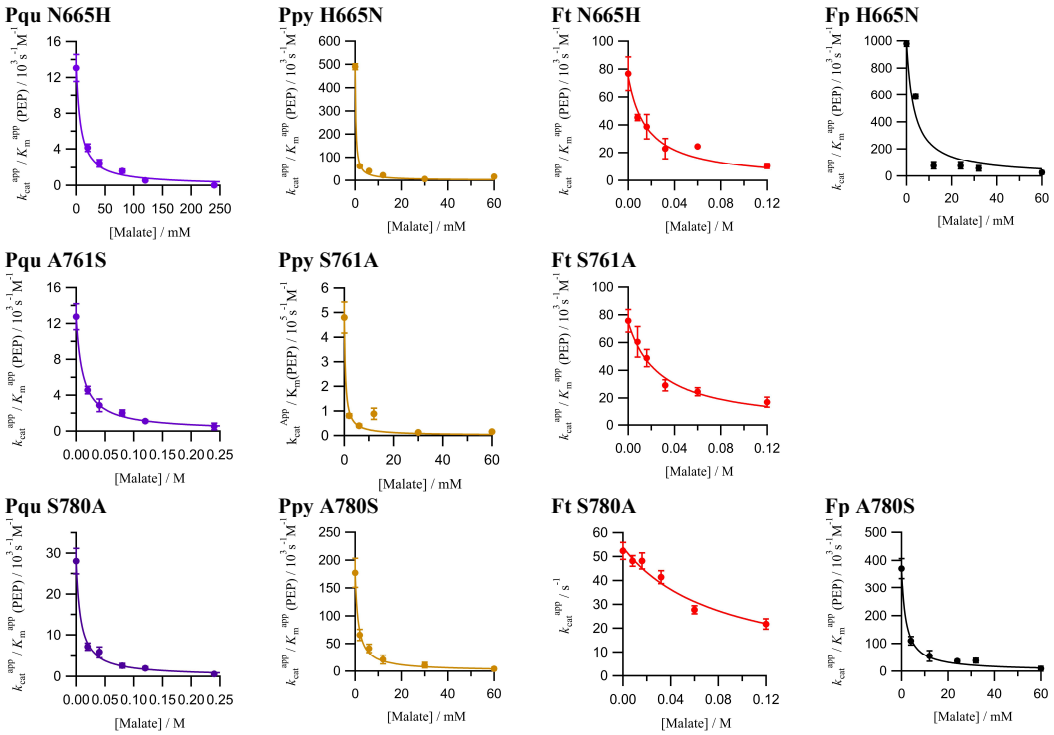


Figure S 4: Plots of competitive inhibition behaviour by malate for mutants of the PEPC enzymes. Secondary plot of $k_{cat}^{app}/K_m^{appPEP}$ against malate concentration, described by equation 2. Error bars represent standard errors based on fitted curves.

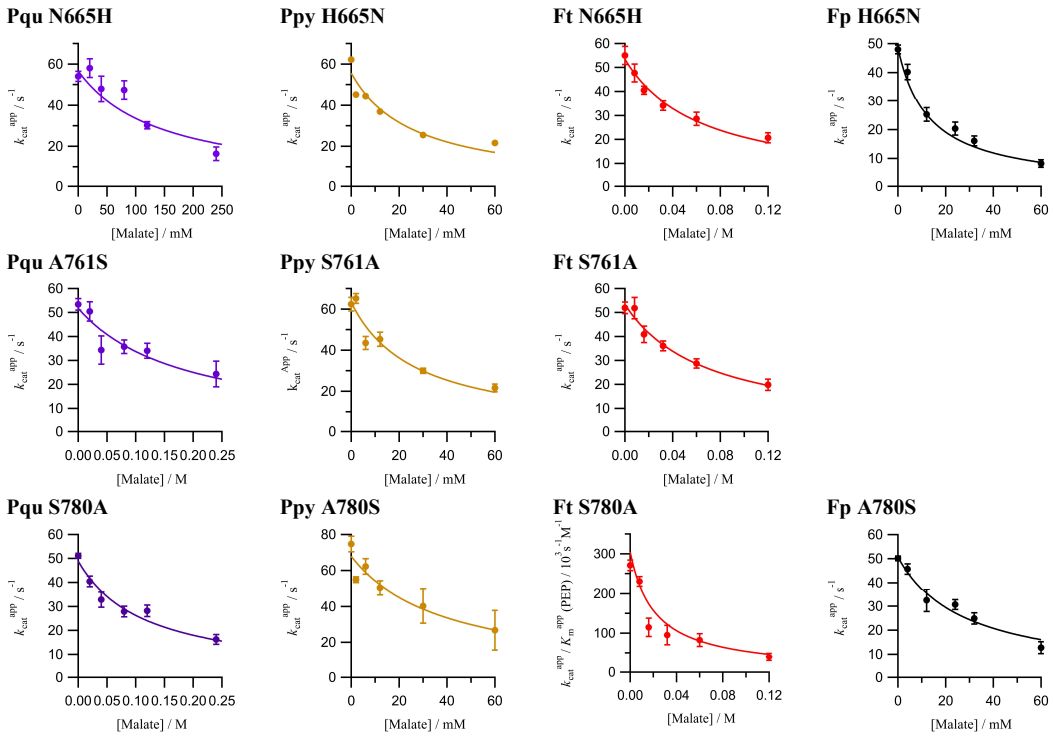


Figure S 5: Plots of uncompetitive inhibition behaviour by malate for mutants of the PEPC enzymes. Secondary plot of k_{cat}^{app} against malate concentration, described by equation 2. Error bars represent standard errors based on fitted curves.

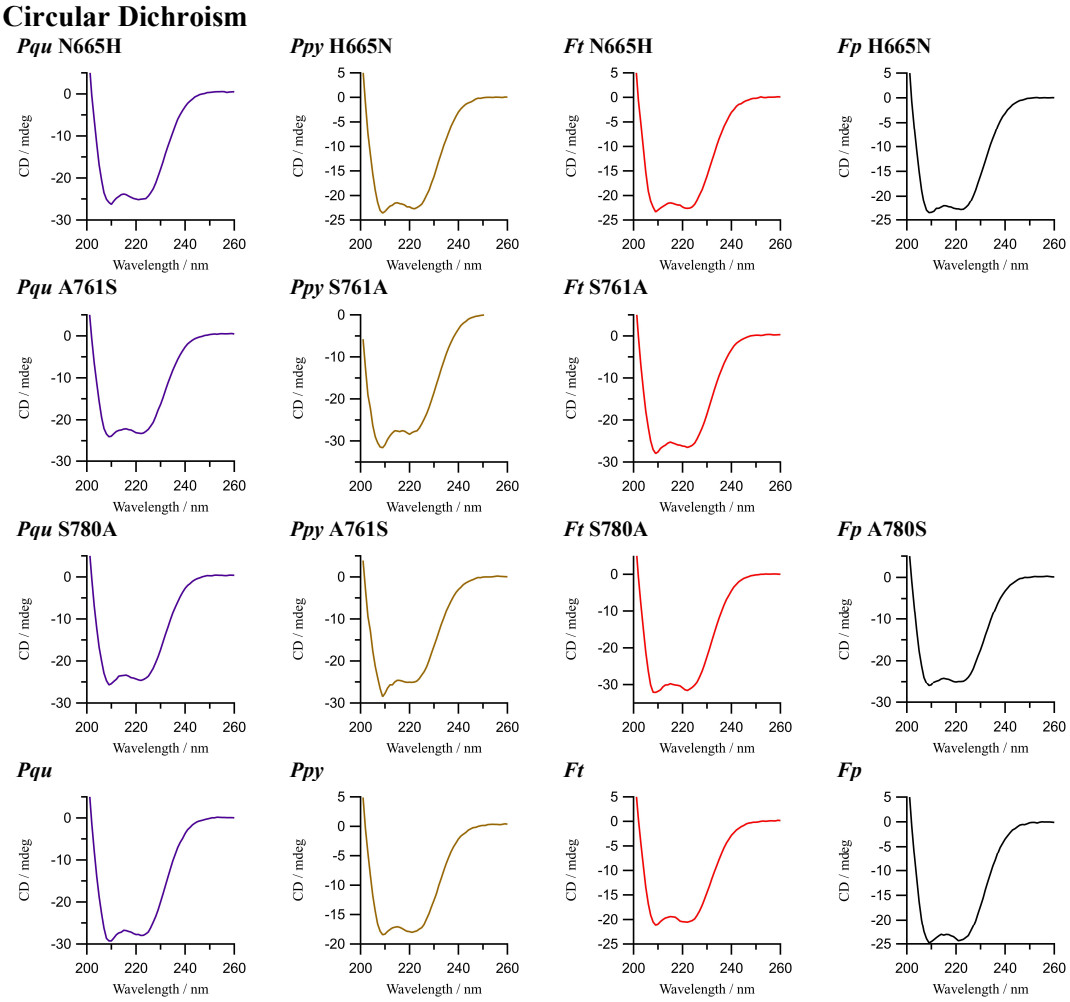


Figure S 6: Plot of circular dichroism result against wavelength. Distinctive peaks at 210 nm and 230 nm indicated the protein is correctly folded.

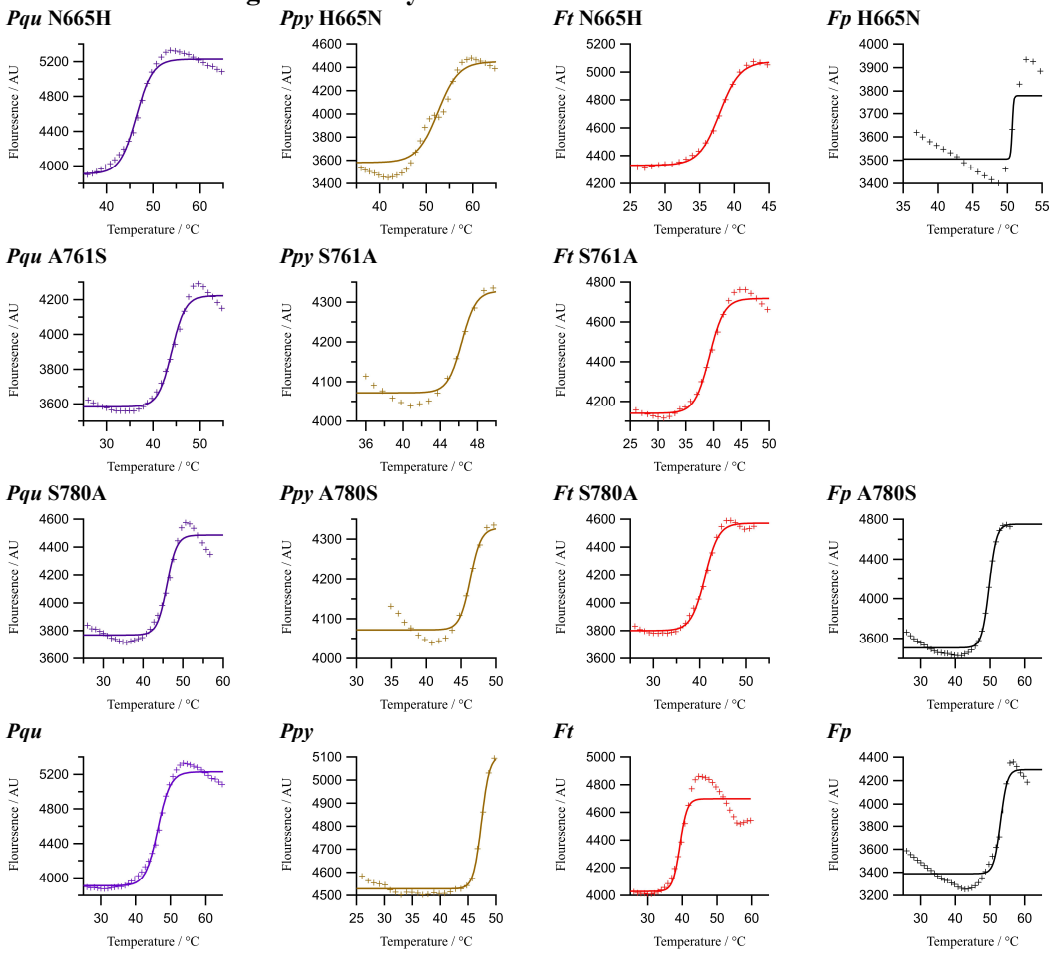


Figure S 7: Plots of fluorescence at 492 nm against temperature of the melting temperature for PEPC. Assays run in 10 mM HEPES pH 7.5, 150 mM NaCl, and 1 × SYPRO orange dye. The lines are described by equation 3, values derived are the melting temperatures of the PEPC enzymes. T_m values are summarised in Supp. Figure 8.

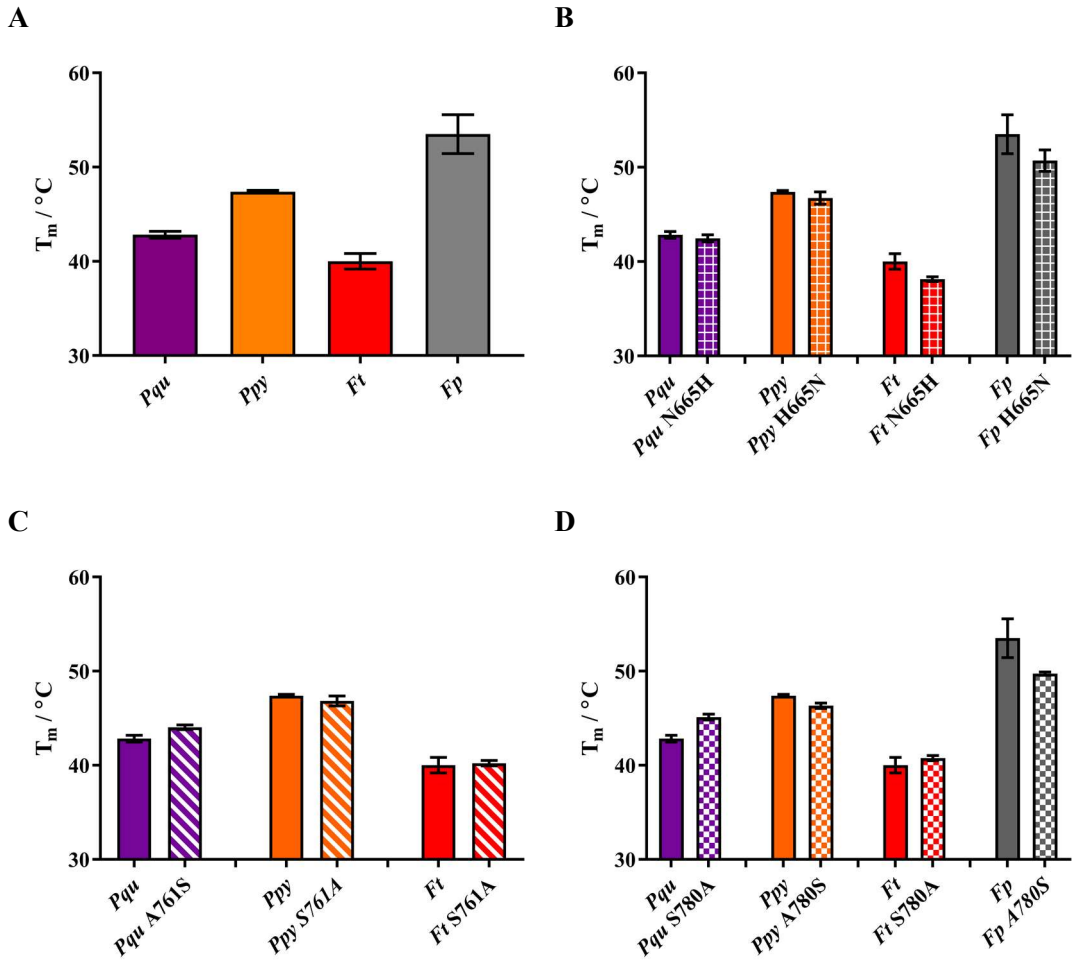


Figure S 8: Summary of Melting Temperatures from Differential Scanning Fluorimetry. Values are derived from Supp. Figure 7, error bars represent standard errors based on fitted curves. **A** Comparison of melting temperatures of native PEPC. **B** Comparison of melting temperatures of position 665 mutants with native PEPCs, no effect of mutations detected ($P > 0.05$). **C** Comparison of melting temperatures of position 761 mutants with native PEPCs No effect of mutations detected ($P > 0.05$). **D** Comparison of melting temperatures of position 780 mutants with native PEPCs No effect of mutations detected ($P > 0.05$).

Table 1: Summary of kinetic parameters found in this study. Standard errors are given, based on fitted theoretical curves.

PEPC Mutant	k_{cat}/K_m^{HCO3-} / $s^{-1}M^{-1}$	k_{cat}/K_m^{PEP} / $s^{-1}M^{-1}$	K_{ic}^{Malate} / mM	K_{iu}^{Malate} / mM
Pqu N665H	$1.15 \times 10^6 \pm 1.49 \times 10^4$	$1.25 \times 10^4 \pm 2.01 \times 10^3$	7.02 ± 2.07	152.02 ± 30.90
Ppy H665N	$3.84 \times 10^5 \pm 3.24 \times 10^4$	$5.01 \times 10^5 \pm 1.28 \times 10^4$	0.33 ± 0.06	26.38 ± 7.89
Ft H665N	$6.44 \times 10^5 \pm 6.66 \times 10^4$	$7.67 \times 10^4 \pm 1.20 \times 10^4$	9.79 ± 3.14	44.31 ± 7.48
Fp N665H	$4.20 \times 10^5 \pm 3.97 \times 10^4$	$9.79 \times 10^5 \pm 1.61 \times 10^4$	2.32 ± 1.67	13.24 ± 3.32
Pqu A761S	$1.16 \times 10^6 \pm 6.63 \times 10^4$	$1.27 \times 10^4 \pm 1.44 \times 10^3$	11.76 ± 6.73	134.61 ± 23.8
Ft S761A	$6.57 \times 10^5 \pm 4.95 \times 10^4$	$7.56 \times 10^4 \pm 8.13 \times 10^4$	16.51 ± 3.55	53.37 ± 1.91
Ppy A761S	$5.70 \times 10^5 \pm 5.46 \times 10^4$	$5.21 \times 10^5 \pm 4.90 \times 10^4$	0.46 ± 0.10	29.56 ± 7.89
Pqu S780A	$9.52 \times 10^5 \pm 7.08 \times 10^4$	$2.80 \times 10^4 \pm 3.13 \times 10^3$	7.90 ± 0.84	116.48 ± 22.30
Ppy A780S	$4.80 \times 10^5 \pm 3.05 \times 10^4$	$5.01 \times 10^5 \pm 1.28 \times 10^4$	1.40 ± 1.61	39.41 ± 11.90
Ft S780A	$6.41 \times 10^5 \pm 4.32 \times 10^4$	$2.71 \times 10^5 \pm 1.35 \times 10^4$	11.47 ± 1.37	42.01 ± 4.00
Fp A780S	$5.38 \times 10^5 \pm 5.14 \times 10^4$	$2.35 \times 10^5 \pm 3.22 \times 10^4$	1.89 ± 1.33	17.79 ± 4.77

Table 2: Table summarising the calculated cost of biosynthesis of PEPC. Cost is the biosynthetic cost, in number of high-energy phosphate bonds, for each PEPC in *E. coli*.

PEPC Species	Cost
Panicum queenslandicum (C_4)	24530.5
Panicum pygmaeum (C_3)	24550.5
Flaveria trinervia (C_4)	25016.0
Flaveria pringlei (C_3)	25037.0

Primers

Table 3:Summary of the primers used in this study for cloning and sequencing.

Primer	Sequence, 5' to 3'
Ft S780A For	TTTGCATGGACTCAGACC
Flav 780 Rev	GATCCATGGAATGGCTCT
Fp A780S For	TTTTCATGGACTCAGACC
Ppy A780S For	TTTTCGTGGACCCAGACA
Ppy A780S Rev	AATCCACGGAATTGCACG
Pqu S780A For	CTGGATCTTCGCATGGACGCAGACGAGG
Pqu S780A Rev	GGGATCGCGCGGAGCGAC
Flav S761A For	CCAGCGAAAAGAAAACCT
Flav S761A Rev	GCGACTTCCAATATTCAT
Pqu A761S For	CCGTCAAAGAGGAAGCCC
Pqu A761S Rev	CCGGCTGCCGATGTTCAT
Ppy S761A For	TAGTCGTCCGGCGAAACGTAAACCG
Ppy S761A Rev	CCAATATTCATACGACCATATTC
Ft N665H For	ATTCATGGGTCTTTAAGAGTGACAGTT
Flav 665 Rev	GGTGTCCGGTGGTTGAGA
Fp H665N For	ATTAACGGGTCGTTAAGA
Pqu N665H For	CCGGACACCATTCATGGGTCCATCC
Pqu N665H Rev	CGGCTGCGACAGGATGGC
Ppy H665N For	ATTAATGGTTCTCTGCGT
Ppy H665N Rev	GGTATCCGGTGGCTGGCT
Ppy Mutant Seq For	AGGTTGCCAAACGTTAT
Ppy Mutant Seq Rev	TGAATGTTGCGGATATC
Flav Mutant Seq Rev	AGATTCTTGCTGTCTT
Flav Mutant Seq For	TTGCAAAAGAGTTTGG
Pqu Mutant Seq Rev	GGATGTTCTTGCTGTC
Pqu Mutant Seq For	CGGTGTGAAGCTGA
Pqu_1323_Seq_For	CGTGAAGCTGGACAT
Pqu_1752_Seq_Rev	ATGACCTGCTGCTTG
Ppy_1291_Seq_For	GATGGTAGTCTGCTGG
Ppy_1791_Seq_Rev	GCTATCGCTATAACCA
Flav_1303_Seq_For	AGACAAGTGTCGACTT
Flav_1832_Seq_Rev	TTGTAGAGCTGCCATG
T7 Promotor	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG

General Discussion

This thesis has explored the adaptations of the enzyme PEPC with a change in function from a general metabolic role to that of a core enzyme in carbon fixation. With the establishment of a C_4 cycle in a plant species, the gene encoding the enzyme is massively upregulated (Bräutigam *et al.*, 2011, 2014; Külahoglu *et al.*, 2014; Lauterbach *et al.*, 2017; Moreno-Villena *et al.*, 2018). My work focused on the comparison of the kinetic changes between PEPC enzymes encoded by orthologous genes from closely-related C_4 and non-C_4 species. PEPC is encoded by a small multigene family, with different gene lineages issued from recurrent gene duplications, some of which happened before the emergence of land plants (Christin *et al.*, 2007; Gowik and Westhoff, 2011). During this long history, the expression patterns of the different genes diverged, as did their coding sequences (Gehrig, Heute and Kluge, 1998, 2001; Monson, 2003; Christin *et al.*, 2007; Moreno-Villena *et al.*, 2018). However, the rate of modification of both expression patterns and amino acid sequences was markedly higher during the transition from non-C_4 to C_4-specific genes (Christin *et al.*, 2007, 2014; Moreno-Villena *et al.*, 2018). The non-C_4 enzymes of extant taxa have consequently been considered as a proxy for the pre-C_4 ancestral sequences, an assumption that is corroborated by selection analyses that show most changes concentrated in C_4 branches (Christin *et al.*, 2007, 2014; Besnard *et al.*, 2009), but should be validated in the future via the comparison of multiple non-C_4 PEPC. In my work, the genes were expressed in *E. coli* and purified, which facilitated analysis of homogenous protein. Enzymes isolated directly from plant leaves represent a mixture of those encoded by different alleles, as well as paralogs. In addition, the post-translational modification state of PEPC enzymes extracted from leaves depend on the time of day the plant leaves are harvested, adding variation among species (Doncaster and Leegood, 1987; Nimmo *et al.*, 1987; McNaughton *et al.*, 1989). Purified PEPC was assayed at controlled concentrations of phosphoenolpyruvate and bicarbonate. The PEPC enzymes were also assayed in the presence of the inhibitors malate and aspartate at limiting and saturating phosphoenolpyruvate. In addition, site-directed mutagenesis was used to introduce amino

acid residues usually associated with C_4-specific PEPC into non-C_4 PEPCs, and vice versa. The properties of these mutant proteins were then used to shed light on the nature of the adaptations PEPC underwent during C_4 evolution.

My investigations of the genus *Flaveria* involved comparison of PEPCs from the C_4 species *Flaveria trinervia* and the C_3 *Flaveria pringlei* (Chapter 2), two species that diverged in the last 3 million years (Christin *et al.*, 2011). It was shown the C_4 *Flaveria* PEPC had a bicarbonate specificity one third higher than the non-C_4 PEPC. Both PEPCs were shown to be inhibited by malate at limiting and saturating phosphoenolpyruvate, and aspartate only inhibited at limiting PEP. The C_4 *Flaveria* PEPC was shown to be less sensitive to both inhibitors when compared to the non-C_4 PEPC. While the *Flaveria* genus represents a good system to study C_4 evolution because of the diversity of photosynthetic types and close relatedness among the species (Engelmann *et al.*, 2003; Westhoff, 2004; McKown, Moncalvo and Dengler, 2005), some of the most successful and earliest origins of C_4 photosynthesis are found in the grass family (Giussani *et al.*, 2001; Christin *et al.*, 2011; Sage, Christin and Edwards, 2011). I consequently decided to compare the C_4 PEPC from the grass *Panicum queensladicum* and the non-C_4 PEPC encoded by the orthologous gene from the C_3 grass *Panicum pygmaeum* that belongs to the same tribe (Chapter 3). Comparison of the two sets of species indicates convergence in C_4 PEPCs of *Flaveria* and *Panicum* with respect to inhibition, PEP specificity and bicarbonate despite their evolutionary distance. Further, the increased divergence time between the grasses compared to the *Flaveria* has resulted in quantitatively larger C_4-specific kinetic properties in the carbon fixing PEPC of *P. queenslandicum*.

Amino acid differences between C_4 and non-C_4 PEPC of grasses have been investigated in the past independently of the kinetic properties (Christin *et al.*, 2007; Besnard *et al.*, 2009). I consequently decided to investigate the biochemical significance of key C_4-specific amino acid changes on the kinetic properties of the encoded enzyme (Chapter 4). The C_4 amino acid mutations H665N and A780S (*Zea mays* numbering), which are shared

by *Flaveria* and *Panicum*, and S761A, which is only observed in *Panicum*, were experimentally evaluated. Position 780 was shown to be an important contributing mutation to the C_4-characteristic low specificity for PEP and did not contribute to inhibition sensitivity or bicarbonate specificity (Chapter 4). Surprisingly, investigation of position 665 and 761 did not detect functional effects with respect to substrate specificity or inhibition. Further investigation of 665 indicated that this mutation is chemically conservative but less biosynthetically costly. This position will be under selection in all species, however this pressure increases with the massive increase in expression level. Analysis of position 761 indicated that this position is in the active site and the serine in the non-C_4 form of the enzyme and is predicted to contribute to a casein II kinase site. When mutated to alanine, this phosphorylation site is removed. The combination of high expression levels of PEPC and the presence of casein II kinase could result in incorrect phosphorylation; preventing this could act as a driving force for selection on non-synonymous mutations.

Overall this work sheds light on the evolution of the C_4 carbon fixing enzyme PEPC, and on the properties that are selected for the C_4 role. It has shown the importance of specific kinetic and structural properties across the spectrum of C_4 species, as well as the selective forces driving of the evolution of these properties.

Insight into the Metabolic Changes of the C_4 Cell

My investigations of C_4 PEPC have shown that the enzyme has a higher K_m^{PEP} and a lower specificity for PEP, when compared to non-C_4 PEPC. While this conclusion was already reached with investigations in the eudicot genera *Flaveria* and *Alternanthera* (Chapter 2; Svensson, Bläsing and Westhoff, 1997; Gowik *et al.*, 2006), I have shown that the same pattern is observed in the distantly-related C_4 grasses (Chapter 3). Differences in K_m^{PEP} have previously been associated with two regions of the PEPC amino acid sequence, including the position 780 (*Z. mays* numbering) and the region from 301 to 422. None of these regions affects sensitivity to inhibition (Chapter 4; Jacobs *et al.*, 2008) and the mutation at position 780 does not change the bicarbonate specificity (Chapter 4). This

demonstrates that specificity for PEP and bicarbonate are not linked, so that the lower specificity for PEP in C_4 PEPC cannot be seen as a side-effect of other protein adjustments but is a property that is directly selected for in the C_4 context. The concentration of PEP has been estimated to be 0.38 mM in the mesophyll cells of *Z. mays* (Arrivault *et al.*, 2017). The concentration of *Z. mays* PEPC can be estimated at 0.14 mM in the whole leaf (McNaughton *et al.*, 1989; Jiao and Chollet, 1991) , but the enzyme concentration is almost certainly higher when considering solely the mesophyll cells. At these concentrations, a non-C_4 type K_m^{PEP} would result in a substantial fraction of the intracellular PEP being bound to the enzyme. The order-of-magnitude larger C_4 K_m^{PEP} results in much more of the PEP pool being unbound and thus available to other enzymes and metabolic pathways (Gowik and Westhoff, 2011).

The major role of C_4 PEPC is fixing atmospheric carbon dioxide in the form of bicarbonate. Investigation of C_4 PEPCs in *Flaveria* and *Panicum* has shown an increase in specificity for bicarbonate compared to the C_3 PEPCs (Chapters 2 and 3). C_4 species typically grow in environments where carbon dioxide availability is reduced (Hatch, 1987), and in tropical and subtropical environments they are some of the most productive species (Cerling *et al.*, 1997; Sage, 2004; Osborne and Beerling, 2006; Osborne and Freckleton, 2009). Carbonic anhydrase (CA) converts carbon dioxide to bicarbonate and is necessary for the C_4 cycle in low carbon dioxide conditions (Osborn *et al.*, 2017). In hotter climates, carbon dioxide becomes limiting in plant cells due to the solubility of carbon dioxide decreasing with increasing temperature (Ku and Edwards, 1977). The increased bicarbonate affinity of C_4-specific PEPC is likely necessary to ensure high fluxes when bicarbonate concentrations are limiting. Carbonic anhydrase is upregulated in the mesophyll of C_4 *Flaveria* (Ludwig, 2016), contributing to an increase in the pool of bicarbonate. However, carbonic anhydrase can only accelerate the approach to equilibrium, so in the presence of a high flux of bicarbonate into the C_4 pathway, bicarbonate is likely to remain limiting in some conditions. Therefore, being able to maintain an adequate reaction rate in the presence of

low bicarbonate concentrations is likely to be crucial for C_4-specific PEPC, explaining the observed changes. The differences between non-C_4 and C_4 PEPCs are markedly more important in grasses than in *Flaveria* (Chapter 3). The high affinity for bicarbonate in grasses might reduce their dependence on carbonic anhydrase, explaining that knock-downs of the enzyme are not fatal in C_4 grasses (Studer *et al.*, 2014).

When compared to non-C_4 forms, C_4 PEPCs have lower sensitivity to both malate and aspartate (Chapters 2 and 3). This is likely to be a response to a higher concentration of these metabolites in the mesophyll cells of C_4 plants compared to C_3 plants. *Zea mays* mesophyll cell concentrations of aspartate and malate have been estimated to be 1.17 mM and 11.18 mM, respectively (Arrivault *et al.*, 2017). This represents an increase in inhibitor concentration compared to the concentrations that the non-C_4 ancestral PEPC experienced. A high sensitivity to inhibition by the products is likely necessary to tightly control the activity of PEPC in its ancestral anaplerotic role. It however becomes highly handicapping for the C_4-specific enzyme that is readily exposed to large pools of these metabolites (Stitt and Zhu, 2014). A dramatic decrease in sensitivity to both inhibitors likely becomes strongly selected for once a rudimentary C_4 cycle is established, as might be necessary to improve the efficiency of the C_4 pathway.

The changes in kinetic properties selected for in C_4-specific PEPC shed light on the changes in the chemistry of the C_4 cell. The decrease in specificity for PEP and sensitivity to aspartate and malate suggest a greater pool of metabolites within the cell, confirming other lines of evidence. The increase in specificity for bicarbonate reflects an increased demand for this substrate the C_4 cells.

Convergent Evolution of the C_4 PEPC

The exact role of specific non-C_4 PEPCs are not known, although some forms have been connected to specific functions, such as generating C_4 acid pools for the glyoxylate cycle in seed germination or a fulfilling a housekeeping role for various metabolic pathways (Sangwan, Singh and Plaxton, 1992; O'Leary, Park and Plaxton, 2011). The non-C_4 PEPCs

investigated so far have a high specificity for PEP and a high sensitivity to inhibitors (Chapter 2 and 3), and it is likely that all non-C_4 isoforms have similar properties. Indeed, the non-C_4 PEPC from *P. pygmaeum* and *F. pringlei*, which diverged *ca.* 150 million years ago (Christin *et al.*, 2011), have broadly similar kinetic properties and have a 91.2 % amino acid similarity (Chapter 3). The non-C_4 enzymes are highly sensitive to inhibitors, indicating they are active when the downstream products are at low concentration and need regenerating to fulfil a role in the very different chemical environment of the C_4 cell. A high specificity for PEP was also observed in other plant non-C_4 PEPCs (Dong *et al.*, 1998; Gowik *et al.*, 2006), suggesting this is a general pattern. There is some variation in specificity to bicarbonate between *P. pygmaeum* and *F. pringlei* PEPCs, but their values remain lower than those of the C_4-specific PEPC (Chapters 2 and 3). This variation suggests that bicarbonate specificity is not under strong stabilizing selection for the housekeeping role.

The role of the C_4-specific PEPC is comparatively well understood. The C_4 PEPCs from *F. trinervia* and *P. queenslandicum* have an 88.1 % amino acid similarity. These genes have a much higher similarity with their respective non-C_4 counterparts (similarity > 90 %), as expected due to shared evolutionary history. However, the C_4-specific PEPCs of *F. trinervia* and *P. queenslandicum* have much more in common in terms of kinetic behaviours. Both have a high specificity for bicarbonate, a low specificity for PEP and a low sensitivity for inhibitors, at both a limiting PEP and saturating PEP (Chapters 2 and 3). Comparison of sequences also indicate that both PEPCs share some C_4-specific amino acids. It has been shown that glycine at position 884 (*Flaveria* numbering) is important for a decreased sensitivity toward inhibition at saturating PEP (Paulus, Niehus and Groth, 2013; Paulus, Schlieper and Groth, 2013), a position shared by both the C_4 PEPCs from *F. trinervia* and *P. queenslandicum.* My investigation of the C_4 specific mutation A780S has shown that this mutation results in the same decrease in PEP specificity in *Flaveria* and *Panicum* (Chapter 4). This C_4 mutation lies near the access point of the active site and may interact with PEP or

other enzyme substrates as they enter (Blasing, Westhoff and Svensson, 2000). However, it does not contribute to C_4 bicarbonate specificity or competitive inhibition (Chapter 4).

My work shows that the kinetic properties that are selected for in C_4 PEPC, and by extension the driving forces of selection, are the same in the evolution of C_4 PEPC in *Flaveria* and *Panicum*. This results in the same mutations being observed in distantly related enzymes.

Greater adaptation of PEPC after full establishment of C_4

The differences between C_4 and non-C_4 PEPC are quantitatively greater in *Panicum* than in *Flaveria* (Chapter 3). While the starting points of evolution represented by their non-C_4 relatives are similar, the C_4 PEPC of *P. queenslandicum* has a higher specificity for bicarbonate, a lower specificity for PEP and a lower sensitivity to inhibition at saturating PEP than its counterpart in *Flaveria*. These differences may be linked to the time spent as C_4. Indeed, the C_4 pathway of *P. queenslandicum* established *ca.* 16 million years ago compared to *ca.* 3 million in the case of *F. trinervia* (Christin *et al.*, 2008, 2011).

The C_4 PEPCs have a similar inhibition sensitivity at limiting PEP (Chapter 2 and 3), which would suggest that this value is already optimised in *Flaveria*. *P. queenslandicum* PEPC has a lower sensitivity to inhibition at saturating PEP than the C_4 *Flaveria*. Analysis of crystal structures of PEPC bound to aspartate has indicated that the binding site is composed of the amino acids R641, Q673, K829, R884 and R888 in *F. pringlei* (Matsumura *et al.*, 2002; Paulus, Schlieper and Groth, 2013). Apart from the 884 mutation, position 673 is the only amino acid site not shared with *Panicum*. In both *P. pygmaeum* and *P. queenslandicum* PEPC, this site is a histidine. Both *Panicum* enzymes have a lower sensitivity to inhibition at saturating PEP compared to their *Flaveria* counterparts (Chapter 4). This suggests that the decreased sensitivity to malate might be related to the specific gene recruited for the C_4 role. This site is a phenylalanine in *Z. mays* and does not interact with the aspartate inhibitor in the crystal structure, which suggests that the site plays no part in inhibition in grasses (Matsumura *et al.*, 2002). The PEPC of the C_4 *Alternanthera* presents

the same amino acids as that of the C_3 *Alternanthera* at all positions that are involved in the allosteric inhibition site including the arginine at position 884 *(Flaveria* numbering). This suggests that C_4 *Alternanthera* PEPC has a high sensitivity to inhibition by malate, although this prediction remains to be tested.

Panicum queenslandicum PEPC has a 50 times lower specificity for PEP than the C_3 *P. pygmaeum*, compared to the 10 times decrease observed in *Flaveria* (Chapters 2 and 3). The PEP specificity of the C_4 *Panicum* is similar to that observed in *Z. mays* (Janc, O'Leary and Cleland, 1992). This may reflect further optimisation of the C_4 cycle after full establishment, which is dependent on the time spent as C_4. After the establishment of an initial C_4 cycle, the maintenance of large pools of metabolites such as PEP might have been further increased through secondary decreases of this enzyme specificity for this substrate.

Bicarbonate specificity is quantitatively greater in the C_4 PEPC of *P. queenslandicum* PEPC than in any of the other enzymes investigated in this work (Chapters 2 and 3). This is likely to further reflect time-dependent optimisation of C_4 PEPC. A C_4 PEPC with lower affinity for bicarbonate, as observed in *Flaveria*, might be sufficient to sustain a C_4 cycle in conditions where CO_2 availability is not excessively low given high enough activity of carbonic anhydrase. Subsequent increases of bicarbonate affinity might improve the efficiency of the C_4 pathway and allow its maintenance even in conditions of extreme CO_2 depletion. This hypothesis is supported by the greater photosynthetic rates achieved in C_4 grasses compared to *Flaveria* (Ubierna *et al.*, 2013), and implies that adaptation of bicarbonate affinity is not necessary for the development of a C_4 cycle, but is involved in the follow-up period of adaptation of the existing trait.

Analysis of one mutation often observed in C_4-specific PEPC suggests that the enzyme undergoes selection for the C_4 function that is not directly linked to its kinetic properties. Indeed, the investigation presented here of the mutation S761A (*Z. mays* numbering) showed the change was functionally silent with respect to substrate specificity

and inhibitor sensitivity (Chapter 4). This mutation is not shared with the C_4 *Flaveria* but is observed in the C_4 *Alternanthera*, *Z. mays*, and *Panicum*. Its lack of kinetic effect is particularly surprising as this site lies on an active site functional loop. When it is occupied by a serine, this site forms part of a casein kinase II site, and the S761A mutation removes the possibility of phosphorylation (Hulo, 2006). With the increase in expression of the gene in C_4 *Panicum* (Moreno-Villena *et al.*, 2018), the likelihood of this amino acid being erroneously phosphorylated is increased. The location of this residue in the active site makes this phosphorylation undesirable and the driving force for this amino acid replacement might have been the prevention of erroneous phosphorylation. Similarly, the decrease of bicarbonate affinity, this mutation is not present in *Flaveria*, showing that the change is not essential for the C_4 function of PEPC. It can be associated to the period of adaptation that follows the initial emergence of a C_4 pathway.

Acquisition Rates of C_4 Specific Properties in C_4 PEPC

Given the different amounts of time spent in a C_4 state, *Flaveria* PEPC is likely to present mostly those changes that are extremely important for the C_4 function and therefore selected for early after the emergence of a C_4 physiology, while the older C_4 PEPC from *Panicum* might present more changes linked to the adaptation of existing C_4 enzymes. By comparing the properties of PEPC from *Panicum* and *Flaveria*, it is therefore possible to infer which of the properties are most important for the C_4 function of PEPC.

The change of sensitivity to inhibitors is similar in *Panicum* and *Flaveria* (Chapters 2 and 3). The photorespiratory pump (C_2 cycle) establishes important components of the C_4 cycle and might therefore represent an evolutionary intermediary state for C_4 evolution (Heckmann *et al.*, 2013; Williams *et al.*, 2013; Mallmann *et al.*, 2014). The C_2 pathway shuttles carbon dioxide using glycine, and the resulting nitrogen imbalance is corrected with amino acids shuttles such as the malate/aspartate shuttle, which is facilitated by PEPC (Dal'Molin *et al.*, 2010; Bräutigam and Gowik, 2016). When the C_4 cycle first emerges, PEPC is not adapted for the C_4 context (Dunning *et al.*, 2017). PEPC has a high degree of

control over photosynthetic flux in C_4 species at high light and ambient carbon dioxide, and the importance of PEPC increases when CO_2 decreases (Bailey *et al.*, 2000). The early versions of C_4-specific PEPC not adapted for the C_4 cycle likely represented a limiting step for the cycle due to their high sensitivity for inhibitors (Chapter 2 and 3), leading to the rapid selection for forms with decreased sensitivity to inhibitors as observed in both *Flaveria* and *Panicum*. The presence of an aspartate/malate shuttle in species that have not fully developed the C_4 cycle suggests that early development of C_4 PEPC involved presence of higher concentrations of inhibitors. The magnitude of increase in tolerance to malate and aspartate, and the comparable change between *Flaveria* and *Panicum*, suggests that this property is extremely important for the C_4 context, and was one of the earliest changes during C4 PEPC evolution.

The non-C_4 PEPC of *Flaveria* and *Panicum* have a high specificity for PEP, which is decreased by a factor 10 in C_4 *Flaveria* and 50 in C_4 *Panicum* (Chapters 2 and 3). The expression levels of pyruvate phosphate dikinase (PPDK), the enzyme responsible for PEP regeneration, increase with the expression level of PEPC across C_3-C_4 intermediates, C_4-like and C_4 species in *Flaveria* as the species become more C_4-like (Mallmann *et al.*, 2014). PPDK has a degree of control over photosynthetic flux, although this remains less important than PEPC (Furbank *et al.*, 1997; Matsuoka *et al.*, 2001). The relative changes during C_4 evolution of PEP specificity compared to inhibitor sensitivity suggest that the adaptation of PEP specificity is less important for C_4 PEPC evolution than inhibitor tolerance. PEP specificity adaptation might therefore happen slightly later during C_4 evolution. The increase in PPDK expression levels observed early during C_4 evolution might be required to provided high amounts of PEP to compensate for the high PEP specificity of the early versions of C_4 PEPC.

Changes observed in bicarbonate specificity between non-C_4 and C_4 PEPCs are modest when compared with other changes investigated. The specificity for bicarbonate slightly increased in C_4 PEPCs of both *Flaveria* and *Panicum*, and the change was larger in

Panicum (Chapters 2 and 3). Specificity for bicarbonate of non-C_4 PEPC is already expected to be high to facilitate the enzyme role, potentially explaining that relatively fewer modifications are required for the C_4 role. The low magnitude of change in bicarbonate specificity compared to other properties suggests that bicarbonate affinity was not a high priority target for C_4 adaptation of PEPC when compared to PEP specificity and inhibitor sensitivity.

Overall my work has shown that the adaptation of PEPC for the C_4 context continues after the initial establishment of a C_4 cycle, and concerns both kinetic and structural properties. This indicates that the development of a highly efficient C_4 cycle requires long evolutionary times. The sequence of the changes might be dictated by their importance, with larger effect mutations most likely to be selected for earlier in the development of C_4 PEPC. However, epistasy might also be involved, with some modifications providing an advantage only once others have been fixed. Differentiating these two scenarios would require establishing the adaptive landscape of C_4-specific mutations. This could be performed by expanding the site-specific mutations conducted in Chapter 4 to cover multiple combinations, as has been done in other systems (e.g. Weinreich *et al.*, 2006).

Conclusion

C_4 photosynthesis is a complex assemblage of genetic and biochemical changes in plants that provides a significant increase in photosynthetic efficiency. The C_4 process evolved convergently in many species by co-opting enzymes from other biochemical pathways to create a carbon concentrating mechanism (Sage, Christin and Edwards, 2011). The identity of the co-opted enzymes varies among C_4 species, however the enzyme phosphoenolpyruvate carboxylase (PEPC) is always utilised for primary carbon fixation. The enzyme is massively upregulated in C_4 species (Moreno-Villena *et al.*, 2018), however the full extent of adaptations undertaken by the enzyme were not fully understood before this dissertation.

A comparison of PEPCs in the genus *Flaveria*, used as a model for probing the evolution of C_4 cycle, was made (Chapter 2). It was shown that an increase in specificity for bicarbonate was selected for in C_4 PEPC. This is likely a response to the metabolic demand of the bicarbonate substrate in C_4 species. The C_4 PEPC was also shown to have a lower specificity for the PEP, but the reason for this change had remained elusive before my work. Investigation of an amino acid change that is partially responsible for the change in PEP affinity, namely A780S (*Z. mays* numbering), suggests that this decrease in substrate specificity does not result from a sacrifice to increase bicarbonate specificity or decrease sensitivity to competitive inhibition by substrates (Chapter 3). These differences in chemical properties with respect to bicarbonate, PEP and inhibitors between the C_4 and non-C_4 PEPC shed light on a changing metabolite composition of the C_4 cell. In the non-C_4 role, the enzyme has a high specificity for substrates and is highly regulated by inhibition. However, it would seem both key properties are modified to sustain high flux metabolites in the C_4 context (Stitt and Zhu, 2014; Arrivault *et al.*, 2017).

The evolution of C_4 PEPC was subsequently probed in the grass genus *Panicum* (Chapter 2). An increase in specificity for bicarbonate, a decrease in specificity for PEP, and a decrease in sensitivity to inhibition by aspartate and malate were observed in the C_4 PEPC

compared to its non-C_4 counterpart. The changes seen between non-C_4 and C_4 PEPC of *Panicum* are quantitatively greater than the changes seen in *Flaveria*. Some of the differences in inhibitor sensitivity between *Panicum* and *Flaveria* C_4 PEPC predate the evolution of C_4. However, some changes may reflect the time the species have respectively spent as C_4. This is evident with respect to the larger decrease in PEP specificity and increased specificity for bicarbonate in C_4 *Panicum* compared to the C_4 *Flaveria*. Some of the amino acid changes observed in the C_4 enzymes were shown to be functionally silent with respect to bicarbonate PEP and inhibition (Chapter 4). This indicates that there are more driving forces acting on C_4 PEPC than the kinetic behaviour of the system.

My work indicates that a key adaptation of C_4 PEPC is an increased specificity for bicarbonate, which is a response to the increased demand for this substrate in the C_4 context. In addition, the evolution of C_4 PEPC is accompanied by decreases in specificity to PEP and decreased sensitivity to inhibition, which reflect the need to maintain metabolite flux in the presence of high concentrations of feedback inhibitors, both through the C_4 pathway and through other PEP requiring reactions. Further, secondary non-kinetic driving forces have been observed. The kinetic properties of several other C_4 amino acid mutations have not been investigated. Some of these changes may not have kinetic functions, and selection at these positions may be driven by some yet to be identified selective pressures.

Overall, my work indicates that the convergent origins of the C_4 phenotype are caused by convergent enzymatic modifications, which suggests a limited number of possible responses to selective pressures created by metabolic innovations. The enzymatic properties found in the most efficient C_4 plants constitute excellent targets for bioengineering attempts to improve both C_3 and C_4 crops. Indeed, while the engineering of C_4 photosynthesis in crops lacking this trait would boost productivity, existing C_4 crops might be improved by human-mediated incorporation of properties observed in some wild C_4 species. This could represent a use of insights gained from comparative studies for agronomical purposes.

References

Arrivault, S., Obata, T., Szecówka, M., Mengin, V., Guenther, M., Hoehne, M., Fernie, A. R. and Stitt, M. (2017) Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation., *Journal of experimental botany*, 68(2), pp. 283–298.

Bailey, K. J., Battistelli, A., Dever, L. V, Lea, P. J. and Leegood, R. C. (2000) Control of C_4 photosynthesis: effects of reduced activities of phosphoenolpyruvate carboxylase on CO_2 assimilation in *Amaranthus edulis L.*, *Journal of experimental botany*, 51, pp. 339–46.

Besnard, G., Muasya, a. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P.-A. (2009) Phylogenomics of C_4 Photosynthesis in *Sedges* (*Cyperaceae*): Multiple Appearances and Genetic Convergence, *Molecular Biology and Evolution*, 26(8), pp. 1909–1919.

Blasing, O. E., Westhoff, P. and Svensson, P. (2000) Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria conserved serine residue in the carboxyterminal part of the enzyme is a major determinant for C_4-specific characteristics, *Journal of Biological Chemistry*, 275(36), pp. 27917–23.

Bräutigam, A. and Gowik, U. (2016) Photorespiration connects C_3 and C_4 photosynthesis, *Journal of experimental botany*, 67(10), pp. 2953–2962.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., Carr, K. M., Gowik, U., Mass, J., Lercher, M. J., Westhoff, P., Hibberd, J. M. and Weber, A. P. M. (2011) An mRNA Blueprint for C_4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C_3 and C_4 Species, *Plant Physiology*, 155(1), pp. 142–156.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P. and Weber, A. P. M. (2014) Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species, *Journal of experimental botany*, 65(13), pp. 3579–3593.

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V. and Ehleringer, J. R. (1997) 'Global vegetation change through the Miocene/Pliocene boundary', *Nature*, 389(6647), pp. 153–158.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V. and Salamin, N. (2008) Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses, *Current Biology*, 18(1), pp. 37–43.

Christin, P.-A., Osborne, C. P., Sage, R. F., Arakaki, M. and Edwards, E. J. (2011) C_4 eudicots are not younger than C_4 monocots, *Journal of experimental botany*, 62(9), pp. 3171–3181.

Christin, P., Arakaki, M., Osborne, C. P., Bräutigam, A., Sage, R. F., Hibberd, J. M., Kelly, S., Covshoff, S., Wong, G. K.-S., Hancock, L. and Edwards, E. J. (2014) Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism, *Journal of Experimental Botany*, 65(13), pp. 3609–3621.

Christin, P., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. (2007) C_4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes, *Current Biology*, 17(14), pp. 1241–1247.

Dal'Molin, C. G. de O., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M. and Nielsen, L. K. (2010) C_4GEM, a Genome-Scale Metabolic Model to Study C_4 Plant Metabolism, *Plant Physiology*, 154(4), pp. 1871–1885.

Doncaster, H. D. and Leegood, R. C. (1987) Regulation of phosphoenolpyruvate carboxylase activity in maize leaves., *Plant physiology*, 84(1987), pp. 82–87.

Dong, L.-Y., Masuda, T., Kawamura, T., Hata, S. and Izui, K. (1998) Cloning, Expression, and Characterization of a Root-Form Phosphoenolpyruvate Carboxylase from *Zea mays*: Comparison with the C_4-Form Enzyme, *Plant and Cell Physiology*, 39(8), pp. 865–873.

Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P. and Christin, P.-A. (2017) Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives, *Evolution*, 71(6), pp. 1541–1555.

Engelmann, S., Bläsing, O. E., Gowik, U., Svensson, P. and Westhoff, P. (2003) Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria*- a gradual increase from C_3 to C_4 characteristics., *Planta*, 217(5), pp. 717–25.

Furbank, R. T., Chitty, J. A., Jenkins, C. L. D., Taylor, W. C., Trevanion, S. J., von Caemmerer, S. and Ashton, A. R. (1997) Genetic Manipulation of Key Photosynthetic Enzymes in the C_4 Plant *Flaveria bidentis*, *Australian Journal of Plant Physiology*, 24(4), p. 477.

Gehrig, H. H., 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(1), pp. 107–114.

Gehrig, H., Heute, V. and Kluge, M. (2001) New Partial Sequences of Phosphoenolpyruvate Carboxylase as Molecular Phylogenetic Markers, *Molecular Phylogenetics and Evolution*, 20(2), pp. 262–274.

Giussani, L. M., Cota-Sánchez, J. H., Zuloaga, F. O. and Kellogg, E. A. (2001) A molecular phylogeny of the grass subfamily *Panicoideae* (*Poaceae*) shows multiple origins of C_4 photosynthesis, *American Journal of Botany*, 88(11), pp. 1993–2012.

Gowik, U., Engelmann, S., Bläsing, O. E., Raghavendra, a. S. and Westhoff, P. (2006) Evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Alternanthera*: Gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 *Alternantheras*, *Planta*, 223, pp. 359–368.

Gowik, U. and Westhoff, P. (2011) C_4 Phosphoenolpyruvate Carboxylase, in Raghavendra, A. S. and Sage, R. F. (eds) *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*. 2011 ed. Dordrecht, Netherlands: Springer, pp. 257–275.

Hatch, M. D. (1987) C_4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure, *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 895(2), pp. 81–106.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M. and Lercher, M. J. (2013) Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape, *Cell*, 153(7), pp. 1579–1588.

Hulo, N. (2006) The PROSITE database, *Nucleic Acids Research*, 34(90001), pp. D227–D230.

Jacobs, B., Engelmann, S., Westhoff, P. and Gowik, U. (2008) Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*: Determinants for high tolerance towards the inhibitor L-malate, *Plant, Cell and Environment*, 31, pp. 793–803.

Janc, J. W., O'Leary, M. H. and Cleland, W. W. (1992) A kinetic investigation of phosphoenolpyruvate carboxylase from *Zea mays*, *Biochemistry*, 31(28), pp. 6421–6426.

Jiao, J. -a. and Chollet, R. (1991) Posttranslational Regulation of Phosphoenolpyruvate Carboxylase in C_4 and Crassulacean Acid Metabolism Plants, *Plant Physiology*, 95(4), pp. 981–985.

Ku, S. B. and Edwards, G. E. (1977) Oxygen Inhibition of Photosynthesis, *Plant Physiology*, 59(5), pp. 991–999.

Külahoglu, C., Denton, A. K., Sommer, M., Mass, J., Schliesky, S., Wrobel, T. J., Berckmans, B., Gongora-Castillo, E., Buell, C. R., Simon, R., De Veylder, L., Brautigam, A. and Weber, A. P. M. (2014) Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species, *The Plant Cell*, 26(8), pp. 3243–3260.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. and Gowik, U. (2017) C_3 cotyledons are followed by C_4 leaves: intra-individual transcriptome analysis of *Salsola soda* (*Chenopodiaceae*), *Journal of Experimental Botany*, 68(2), pp. 161–176.

Ludwig, M. (2016) Evolution of carbonic anhydrase in C_4 plants, *Current Opinion in Plant Biology*. Elsevier Ltd, 31, pp. 16–22.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P. and Gowik, U. (2014) The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*, *eLife*, 3(3), p. e02478.

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*, 10(12), pp. 1721–1730.

Matsuoka, M., Furbank, R. T., Fukayama, H. and Miyao, M. (2001) Molecular Engineering of C_4 Photosynthesis, *Annual Review of Plant Physiology and Plant Molecular Biology*, 52(1), pp. 297–314.

McKown, A. D., Moncalvo, J.-M. and Dengler, N. G. (2005) Phylogeny of *Flaveria* (*Asteraceae*) and inference of C_4 photosynthesis evolution, *American Journal of Botany*, 92(11), pp. 1911–1928.

McNaughton, G. A. L., Fewson, C. A., Wilkins, M. B. and Nimmo, H. G. (1989) Purification, oligomerization state and malate sensitivity of maize leaf phosphoenolpyruvate carboxylase, *The biochemical journal*, 261(2), pp. 349–55.

Monson, R. K. (2003) Gene Duplication, Neofunctionalization, and the Evolution of C_4 Photosynthesis, *International Journal of Plant Sciences*, 164(S3), pp. S43–S54.

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. and Christin, P. A. (2018) Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis, *Molecular Biology and Evolution*, 35(1), pp. 94–106.

Nimmo, G. A., McNaughton, G. A. L., Fewson, C. A., Wilkins, M. B. and Nimmo, H. G. (1987) Changes in the kinetic properties and phosphorylation state of phosphoenol pyruvate carboxylase in *Zea mays* leaves in reponse to light and dark, *FEBS Letters*, 213(1), pp. 18–22.

O'Leary, B., Park, J. and Plaxton, W. C. (2011) The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs, *Biochemical Journal*, 436(1), pp. 15–34.

Osborn, H. L., Alonso-Cantabrana, H., Sharwood, R. E., Covshoff, S., Evans, J. R., Furbank, R. T. and von Caemmerer, S. (2017) Effects of reduced carbonic anhydrase activity on CO_2 assimilation rates in *Setaria viridis*: a transgenic analysis, *Journal of experimental botany*, 68(2), pp. 299–310.

Osborne, C. P. and Beerling, D. J. (2006) Nature's green revolution: The remarkable evolutionary rise of C_4 plants, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361(1465), pp. 173–194.

Osborne, C. P. and Freckleton, R. P. (2009) Ecological selection pressures for C_4 photosynthesis in the grasses, *Proceedings of the Royal Society: Biological Sciences*, 276(1663), pp. 1753–1760.

Paulus, J. K., Niehus, C. and Groth, G. (2013) Evolution of C_4 Phosphoenolpyruvate Carboxylase: Enhanced Feedback Inhibitor Tolerance Is Determined by a Single Residue, *Molecular Plant*, 6(6), pp. 1996–1999.

Paulus, J. K., Schlieper, D. and Groth, G. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution, *Nature communications*, 4, p. 1518.

Sage, R. F. (2004) The evolution of C_4 photosynthesis, *New Phytologist*, 161(2), pp. 341–370.

Sage, R. F., Christin, P.-A. and Edwards, E. J. (2011) The C_4 plant lineages of planet Earth, *Journal of experimental botany*, 62(9), pp. 3155–3169.

Sangwan, R. S., Singh, N. and Plaxton, W. C. (1992) Phosphoenolpyruvate Carboxylase Activity and Concentration in the Endosperm of Developing and Germinating Castor Oil Seeds, *Plant Physiology*, 99(2), pp. 445–449.

Stitt, M. and Zhu, X.-G. (2014) The large pools of metabolites involved in intercellular metabolite shuttles in C_4 photosynthesis provide enormous flexibility and robustness in a fluctuating light environment, *Plant, Cell & Environment*, 37(9), pp. 1985–1988.

Studer, A. J., Gandin, A., Kolbe, A. R., Wang, L., Cousins, A. B. and Brutnell, T. P. (2014) A Limited Role for Carbonic Anhydrase in C_4 Photosynthesis as Revealed by a ca1ca2 Double Mutant in Maize, *Plant Physiology*, 165(2), pp. 608–617.

Svensson, P., Bläsing, O. E. and Westhoff, P. (1997) 'Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase- a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3).', *European journal of biochemistry*, 246(2), pp. 452–60.

Ubierna, N., Sun, W., Kramer, D. M. and Cousins, A. B. (2013) The efficiency of C_4 photosynthesis under low light conditions in *Zea mays*, *Miscanthus x giganteus* and *Flaveria bidentis*, *Plant, Cell & Environment*, 36(2), pp. 365–381.

Weinreich, D. M., Delaney, N. F., DePristo, M. A. and Hartl, D. L. (2006) Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins, *Science*, 312(5770), pp. 111–114.

Westhoff, P. (2004) 'Evolution of C_4 Phosphoenolpyruvate Carboxylase. Genes and Proteins: a Case Study with the Genus *Flaveria*, *Annals of Botany*, 93(1), pp. 13–23.

Williams, B. P., Johnston, I. G., Covshoff, S. and Hibberd, J. M. (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis, *eLife*, 2(2), p. e00961.

Appendix: PEPC Gene Cloning Design, PEPC *E. coli* Expression and Purification Optimisation, and Bicarbonate Assay Design

Molecular Cloning of PEPC

A high throughput method was desired for molecular cloning to expedite the cloning process for several genes, for which the technique Ligation Independent Cloning (LIC) was chosen. LIC uses the 3'-5' exonuclease activity of T4 polymerase to create single stranded overhangs that are complementary between the vector and the insert (Aslanidis and de Jong, 1990). Overhang generation is halted by the presence of a nucleotide in the reaction mixture, which is not present in the complementary DNA sequence of the desired overhang, but immediately after the sequence i.e. GTP is added to the T4 polymerase digestion mix where the complementary strand to overhang is composed of A, C and T, and in this instance, CTP is added to the complementary strand (Figure 1). When combined the vectors and inserts with overhangs anneal, and the strands form a single plasmid with nicks in the DNA either side of the insert. *E. coli* is transformed with the DNA and the nicks in the cloned plasmid are repaired by the replication mechanisms of the bacteria. It has been demonstrated that LIC is a high throughput technique suitable for the cloning of genes as well as multiple gene assembly (Stols *et al.*, 2002; Berrow *et al.*, 2006; Schmid-Burgk *et al.*, 2013).

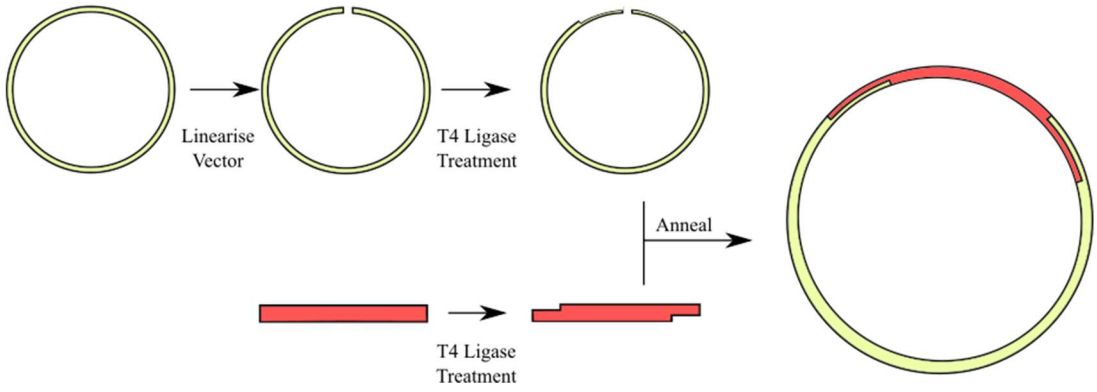


Figure 1: Diagram of ligation independent cloning technique. Vector DNA represented in green, insert DNA represented in red.

Traditional molecular cloning using restriction enzymes is disadvantageous in comparison as it requires specific restriction sites that may result in the introduction of non-native amino acids into the expressed protein. Restriction enzyme choices can be limited by the sequence of the insert. LIC does not need T4 ligase to anneal the insert and vector fragment, removing a step which may cause non-specific ligation. LIC has the advantage

that it does not rely on sequence specific DNA cleavage and potentially may be used with any vector. However, LIC does require PCR with large primers that may be difficult to optimise.

The pET His6 TEV LIC cloning vector (pET-1B; Addgene Plasmid #29653) was chosen as a suitable vector for cloning PEPC which is optimised for LIC. The vector encodes an N terminal poly-histidine fusion tag and a tobacco etch virus protease cleavage site (TEV) to remove the tag, as TEV does not cleave PEPC. The pET-1B vector was gifted from the Scott Gardia lab.

Isolation of PEPC Gene from Leaf RNA

Eight C_3 and C_4 grass species were selected for RNA extraction. Species that were selected were raised from seeds from a seed bank and all had partial or complete genome or transcriptome libraries (Atkinson *et al.*, 2016), allowing the design of primer for PCR of the *ppc-1P3* gene (Christin *et al.*, 2015). Leaf samples were harvested at midday in full light and flash frozen in liquid nitrogen during peak photosynthetic period. RNA was extracted from frozen tissue using RNA plant extraction kits.

RNA samples were treated with reverse transcriptase to produce cDNA libraries. Primers were designed from alignments of the *ppc-1P3* from transcriptome data and genome data of the species (Atkinson *et al.*, 2016). Highly conserved regions were chosen to maximise the likelihood of PCR success. Primers were designed to screen cDNA libraries for the presence of *ppc-1P3* gene. PCR with screening primers produced a *ca.* 2000 base pair fragment in the presence of the PEPC encoding gene ('pcc_1072_for' and 'ppc_3037_rev'). A *ca.* 2000 base pair band was seen in the PCR product from the cDNA of *Panicum queenslandicum* and *Panicum pygmaeum*. The PCR product was then sequenced which confirmed this band corresponded to the *ppc-1P3* gene fragment. Primers were designed to encapsulate *ca.* 200 base pairs either side of the start and the stop codon of the gene. The PCR products from reactions with these primers were sequenced and the start and stop codon regions were determined.

PEPC Gene Cloning

Primers were designed for cloning, composed of a 15 base pair section complementary to the plasmid and a 10 base pair section specific to the terminal region of the PEPC gene. These primers were used with cDNA and a high-fidelity polymerase (Q5 high fidelity polymerase). The full gene for *P. queenslandicum* PEPC was sequenced and then cloned into the 1B plasmid using LIC. The full gene for *P. pygmaeum* was sequenced, however the yield of DNA was too low to clone. The gene sequence was synthesised by GeneArt (Thermo Fischer Scientific) in the pMA-T vector, optimised for expression in *E. coli*. The gene was sub cloned into the 1B plasmid using the LIC technique.

The *ppcA* genes for PEPC (*ppc-1E2* in Christin *et al.*, 2015) from *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3) as described in (Svensson, Bläsing and Westhoff, 1997), were gifted in the pTrc-99A vector. Primers were designed for LIC using the Genbank sequence. The genes were sub cloned into the 1B plasmid and the resulting clones were fully sequenced.

PEPC Expression Optimization

The optimal conditions for expression in *E. coli* were determined. Several conditions were tested to determine maximal protein expression in *E. coli*. The Rosetta, and BL21λ(DE3) strains of *E. coli* were transformed with each PEPC expression plasmid. Test samples were grown to a density of OD_{600} = 0.6-0.8, in either LB or 2YT medium. Samples were induced with either 0.5 mM or 1 mM IPTG and cooled for an hour at 4°C. Cells were then incubated overnight at either 18°C or 16°C. Protein production was then analysed by SDS PAGE (Figure 2 and 3).

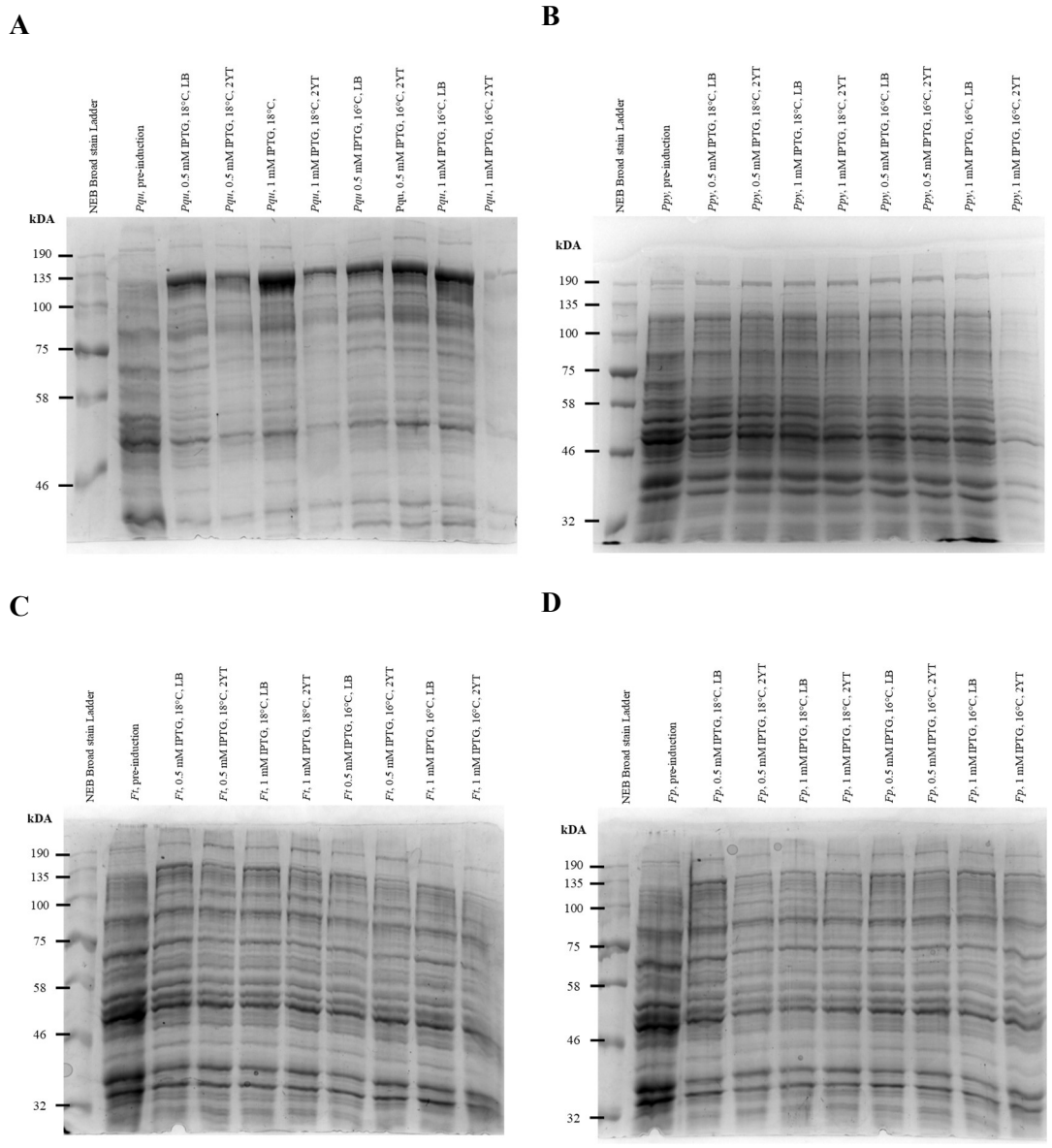


Figure 2: SDS PAGE analysis of PEPC induction in Rosetta strain *E.coli*. A *Panicum queenslandicum* PEPC induction trial. **B** *Panicum pygmaeum* PEPC induction trial. **C** *Flaveria trinervia* PEPC induction trial. **D** *Flaveria pringlei* PEPC induction trial.

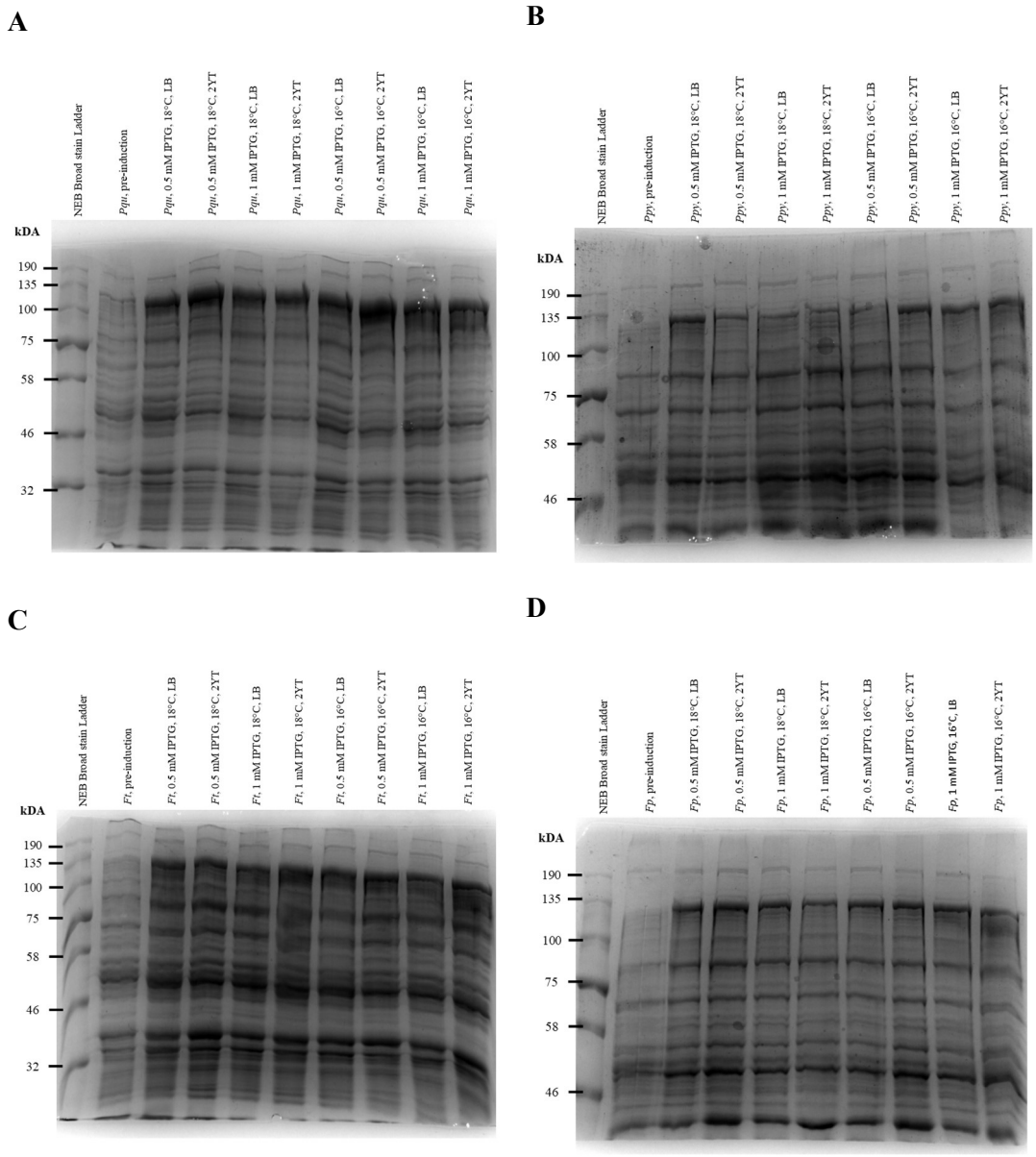


Figure 3: SDS PAGE analysis of PEPC induction in BL21λ(DE3) strain *E. coli*. A *Panicum queenslandicum* PEPC induction trial. **B** *Panicum pygmaeum* PEPC induction trial. **C** *Flaveria trinervia* PEPC induction trial. **D** *Flaveria pringlei* PEPC induction trial.

From analysis of SDS PAGE (Figure 2 and 3), PEPC is induced in most conditions with both strains, however the most consistent condition for protein induction was in BL21λ(DE3) strain *E. coli*, grown in LB medium, induced with 0.5 mM IPTG at a density of OD_{600} *ca.* 0.6-0.8, then cooled for one hour at 4°C followed by incubation for 18 hours at 18°C.

PEPC Purification Optimisation

The 1B plasmid facilitated the purification of PEPC using a metal ion affinity column with the poly histidine tag. If the tag adversely affected the protein, a TEV protease site was included so the tag could be cleaved to produce near native protein.

Eight litres of growth pellet were purified at a time to saturate the 1 ml nickel ion column and optimise yield of purification. Pellets were resuspended in binding buffer containing 20 mM imidazole, to reduce binding of contaminating proteins to the column, and 0.4 $mgml^{-1}$ Pefabloc protease inhibitor. The suspended bacteria solution was lysed with a cell disruptor and centrifuged. The soluble fraction of lysate was loaded onto the column and washed with 50 column volumes of binding buffer. It was found that by washing the column with 50 column volumes of buffer containing 150 mM imidazole, all contaminating protein was removed. PEPC was eluted with 400 mM imidazole buffer. The protein was then exchanged into a buffer suitable for assays with a G50 column. Analysis by SDS PAGE indicated that purified PEPC proteins were purified to > 95% purity (Figure 4).

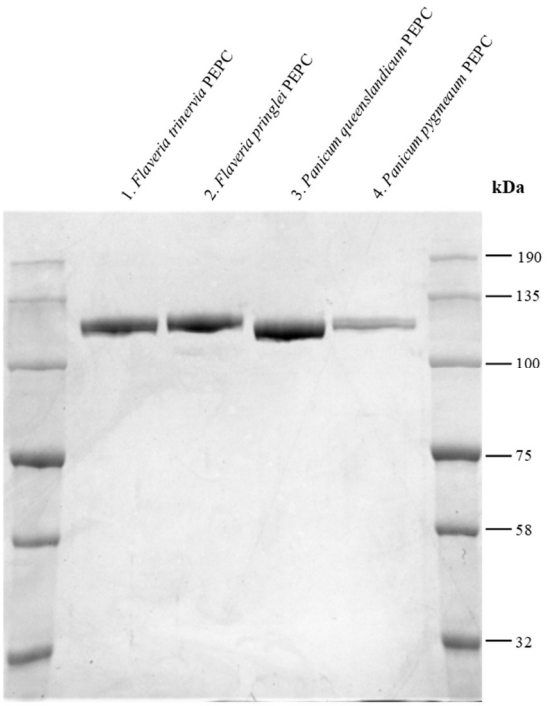


Figure 4: 8 % acrylamide SDS PAGE analysis of PEPC proteins. Lane 1 contains 5 µg of *Flaveria trinervia* PEPC, lane 2 contain 5 µg of *Flaveria pringlei* PEPC, lane 3 contain 5 µg of *Panicum queenslandicum* PEPC, lane 4 contains 5 µg *Panicum pygmaeum* PEPC.

PEPC Quantification

PEPC enzyme concentration was quantified by absorption at 280 nm. The enzyme extinction coefficient was calculated using the ExPASy protein parameter tool and corrected by determining the absorbance of the protein denatured in 6 M guanidine hydrochloride (Gill and von Hippel, 1989). The difference between the denatured and folded protein at 280 nm was used to adjust the extinction coefficient of the protein. The extinction coefficient for *F. trinervia* PEPC was determined to be 120480 M^{-1} cm^{-1} (Table 4), the extinction coefficient for *F. pringlei* PEPC was determined to be 117030 M^{-1} cm^{-1} (Table 5). Differences of -0.5% and -6.7% between predicted absorbance coefficients for *F. trinervia* and *F. pringlei* PEPC were observed respectively. The extinction coefficient for *P. queenslandicum* PEPC was determined to be 105805 M^{-1} cm^{-1} (Table 6), the extinction coefficient for *P. pygmaeum* PEPC was determined to be 111514 M^{-1} cm^{-1} (Table 7). Differences of 8.3 % and 4.6% between predicted absorbance coefficients for *P. queenslandicum* and *P. pygmaeum* PEPC were observed respectively. It is assumed that all the enzyme used to initiate the assay was active. For assays on single point mutants, the native enzyme extinction coefficient was used as it was assumed there would be no change in extinction coefficient with mutation.

PEPC Assay Design

PEPC has been assayed by coupling the enzyme to a second enzyme, malate dehydrogenase (MDH). PEPC produced oxaloacetate (OAA) from phosphoenolpyruvate (PEP) and bicarbonate, OAA was converted into malate by malate dehydrogenase (MDH) which also consumed NADH. The rate of NADH consumption was observed at 340nm. It has been observed that PEPC is most active at pH 8.0 (Chollet, Vidal and O'Leary, 1996). Assays were initiated by the addition of PEPC.

Assays performed with the presence of the inhibitors malate and aspartate were performed at limiting and saturating PEP, and were initiated by the addition of PEPC. Aspartate and malate are fast acting inhibitors and PEPC did not need to be pre-incubated prior to assay initiation (Wedding, Black and Meyer, 1990).

Assays of *F. trinervia* PEPC and *F. pringeli* PEPC, purified from expression from the 1B plasmid behaved similarly to the native form of the enzyme (Chapter 1; Svensson, Bläsing and Westhoff, 1997). This suggested that the N terminal fusion tag did not adversely affect the activity of PEPC and does not need removing.

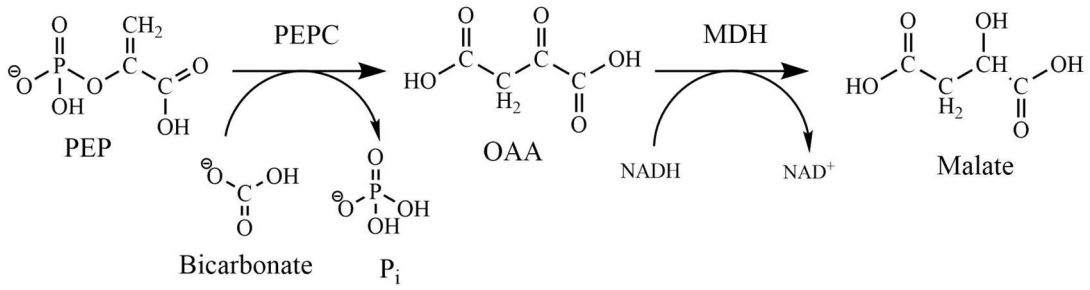


Figure 5: Diagram of coupled assay system of PEPC. Enzymes of the reaction listed above the reaction arrow.

Bicarbonate Assay Design

It has been shown that it is possible to control bicarbonate concentration in a PEPC assay in order to determine a K_m^{HCO3-} (Bauwe, 1986; Janc, O'Leary and Cleland, 1992; Dong *et al.*, 1998). Gas-tight cuvettes with tight-fitting septa were used to prevent atmospheric CO_2 contamination in bicarbonate-controlled assays. To reduce the background bicarbonate concentration, assay components were sparged with nitrogen gas. The water and tricine assay components were sparged with nitrogen for *ca.* 18 hours. These components made up at least 75% by volume of the assay solution. Tricine was buffered to pH 8.0 with solid potassium hydroxide to prevent bicarbonate contamination from potassium hydroxide solution. Other components such as NADH and PEP stock solutions were stored frozen which reduced bicarbonate in these components due to the freeze thaw cycle. Assays were composed and sealed under a nitrogen flow to prevent atmospheric CO_2 contamination during assay assembly. Assays were initiated by delivering the enzyme with a gastight syringe through the septa.

End-point assays of PEPC were then used to determine the background concentration of bicarbonate of the sparged assay components. This was performed by composing the assays without adding bicarbonate. Assays were monitored for 30 minutes until a stable endpoint was reached. A high concentration (50 nM) of PEPC was used to ensure an endpoint was reached quickly. The difference in absorbance at 340 nm was used to calculate the bicarbonate concentration, due to the proportional relationship between the consumption of NADH and bicarbonate. Background bicarbonate was reduced to *ca.* 50 μM using this assay assembly system.

Experimental Procedure

Unless otherwise stated, reagents and components were from Sigma. For purification, unless otherwise stated the protein purification equipment was from GE Healthcare. For DNA treatment, unless otherwise stated enzymes and *E. coli* strains were from NEB.

RNA Extraction

Fresh leaves of species were selected from live organisms. Leaves that were chosen for extraction were young full-grown leaves, non-senescing. Leaf samples were flash frozen in liquid nitrogen and kept in liquid nitrogen until extraction.

For extraction, *ca.* 100 mg of leaf samples were then individually ground for *ca.* 30 minutes with a pestle and mortar under liquid nitrogen until the sample resembled a fine green powder. Powdered leaf sample was then transferred to a sample tube. RNA was then extracted with the RNeasy® Plant Mini Kit (Qiagen). Samples were treated with DNase and RNase inhibitor (Qiagen). Samples were then stored at -80°C.

cDNA synthesis

Samples of cDNA were generated from SuperScript™ II Reverse Transcriptase (Invitrogen). Synthesis was initiated with Oligo (dT) 12-18 primer (Invitrogen).

Polymerase Chain Reaction (PCR)

Screening PCR

PCR with G2Go Taq polymerase (Promega) in Go Taq green buffer with appropriate primers and ca. 100 ng of cDNA for cDNA screening and 5' and 3' PEPC gene sequence determination. A PCR reaction mix was assembled in the following concentrations:

Reagent	PCR Con[c]
Primer 1	1.0 µM
Primer 2	1.0 µM
DMSO	5 % (v/v)
dNTPs	0.2 µM
$MgCl_2$	3.0 mM
GoTAQ Green Buffer	1 ×
DNA Polymerase	1 Unit

The PCR mixture was incubated in a thermocycler on the routine, with primer appropriate annealing temperate:

	Step	Temperature /°C	Time / H:MM:SS
1	Initial Denaturation	94°C	0:00:30
2	Denaturation	94°C	0:00:10
3	Annealing	*	0:00:30
4	Extension	72°C	0:01:00
		Repeat 2-4 20 times	
36	Final Extension	72°C	0:10:00
37	Hold	4°C	0:10:00

Colony Screening

For colony screening, a bacterial colony was sampled with a pipette tip and added to the PCR mix. A PCR reaction mix was assembled with the following concentrations:

Reagent	PCR Con[c]
Primer 1	1.0 µM
Primer 2	1.0 µM
DMSO	5 % (v/v)
dNTPs	0.2 mM
GoTAQ Green Buffer	1 ×
DNA Polymerase	1 Unit

The PCR mixture was incubated in a thermocycler using the appropriate primers and the parameters shown:

	Step	Temperature /°C	Time / H:MM:SS
1	Initial Denaturation	94°C	0:00:30
2	Denaturation	94°C	0:00:10
3	Annealing	*	0:00:30
4	Extension	72°C	0:03:30
		Repeat 2-4 35 times	
36	Final Extension	72°C	0:10:00
37	Hold	4°C	0:10:00

Q5 High Fidelity PCR
PCR generation of inserts was performed with Q5 High Fidelity polymerase with *ca.* 100 ng of cDNA or *ca.* 50 ng of plasmid. The PCR reaction mix was assembled as follows:

Reagent	PCR Con[c]
Q5 Master Mix	1 ×
Primer 1	0.5 µM
Primer 2	0.5 µM

The PCR mixture was incubated in a thermocycler using the appropriate primers and the parameters shown:

	Step	Temperature /°C	Time / H:MM:SS
1	Initial Denaturation	98°C	0:00:30
2	Denaturation	98°C	0:00:10
3	Annealing	*	0:00:30
4	Extension	72°C	0:01:30
		Repeat 2-4 25 times	
26	Final Extension	72°C	0:02:00
27	Hold	4°C	0:10:00

PCR products and plasmid were Sanger sequenced (GATC Biotech) using the appropriate primers. PCR products from cDNA screening were prepared using ExoSAP clean-up. PCR products from Q5 PCR were prepared using a PCR clean-up kit (Qiagen).

Agarose Gel Analysis

PCR products were analysed with 1 % TBE agarose gel, stained with SYBR™ Safe (Invitrogen). Gels were run at 110 V for 30 minutes and imaged with a ChemiDoc™ MP (BioRad).

PCR Clean-up and Sanger Sequencing

Samples for DNA screening, colony screening and terminal sequence determination were prepared for sequencing using ExoSAP (Invitrogen). 5 µl of sample was treated with 2 µl of ExoSAP enzyme for 15 minutes at 37°C, then heat inactivated for 15 minutes at 80°C. Samples were then made up to 20 µl and then Sanger sequenced (GATC Biotech) with the appropriate primers.

Plasmid Preparations

Single colonies of DH5α strain *E. coli* were selected and used to inoculate LB culture with 30 µg ml^{-1} kanamycin. Cultures were incubated in a shaker at 37°C for 18 hours, 250 rpm. Plasmid DNA was then extracted from 5 ml (150 ml for midiprep) of overnight culture with a miniprep kit or midiprep kits (Qiagen). Plasmids were eluted with water, warmed to 65°C.

LIC Preparation of Vector and Insert

1B Vector Cloning Preparation

The pET His6 TEV LIC cloning vector (1B) was a gift from Scott Gradia (Addgene plasmid # 29653). 1 µg of DNA was linearized with 10 units of SspI-HF restriction enzyme in CutSmart Buffer, incubated at 2 hours at 37°C, then 65°C for 20 minutes to heat inactivate the enzyme. DNA was purified with agarose gel extraction. DNA was loaded onto a 0.8% TAE gel, stained with SYBR™ Safe. Gels were run at 90 V for 90 minutes. The appropriate band was excised with a scalpel and purified with a gel extraction kit (Qiagen) and a ChemiDoc™ MP (BioRad). DNA was eluted with water, warmed to 65°C.

Insert Cloning Preparation

Inserts were amplified with Q5 polymerase. PCR was optimised to produce one DNA band. PCR product was purified using a PCR clean-up kit (Qiagen). DNA was eluted with water, warmed to 65°C.

T4 Polymerase Treatment of Insert and Vector

Vector and insert were treated for 30 minutes at 22°C with T4 polymerase to generate complementary overhangs. With the insert, 2.5 mM of dCTP was added; with the Vector, 2.5 mM dGTP was added. Samples were then incubated for 20 minutes at 75°C to inactivate the enzymes. Vector and insert were combined in a 1:3 ratio, with 1 mM EDTA. DNA was incubated for 30 minutes prior to transformation.

Transformation of Competent Escherichia coli (E. coli)

Table 1: Summary of *E. coli* used, with genotype listed.

***E. coli* Strain**	**Genotype**
DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
BL21λ(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*
Rosetta™(DE3)	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R)

For cloning transformation, 50 µl of high competence DH5α strain *E. coli* cells were combined with 1 µl of DNA mix and incubated for 30 minutes on ice. The mix was heat-shocked at 42°C for 30 s. in a water bath at 42°C. Cells were then rested on ice for 15 minutes. 950 µl of SOC medium were added then incubated in a shaker for 1 hour at 37°C, 250 rpm. 50 µl of the cells were then plated on LB agar, 30 µg ml^{-1} kanamycin. Plates were then incubated for 18 hours at 37°C.

Transformations with BL21λ(DE3) and Rosetta™ strain (Novagen) *E. coli* were transformed were done using a similar method to DH5α protocol, with the following differences: after DNA and cells were combined, the cells were incubated for 10 minutes, and 450 µl of SOC medium were added. Rosetta™ cells transformation mixes were plated onto LB agar containing 30 µg ml^{-1} kanamycin and 50 µg ml^{-1} chloramphenicol.

Protein Expression

For protein expression, the BL2*1λ*(DE3) strain of *E. coli* was used. Eight litres of cultures were grown in LB medium at 37°C to OD_{600} = 0.6-0.8. Cultures were cooled to 4°C for one hour prior to recombinant protein induction with 0.5 mM IPTG (Thermo Fisher Scientific). Cultures were then incubated at 18°C for 18 hours. Cells were harvested by centrifugation at 5,422 × *g* for 25 minutes and stored at -80°C.

Protein Purification

Cells were suspended in IMAC buffer (25 mM Tris, 0.5 M NaCl, 0.3 M glycerol, 20 mM imidazole (Acros Scientific)), 10 ml per 2 litres of culture with 50 µl of 50mg ml^{-1} DNase I and 100 µl of 100 mg ml^{-1} Pefabloc. Cells were passed twice through a cell disruptor (Constant Systems) before centrifugation at 26,902 × *g* for 40 minutes. The supernatant was passed through a 0.45 µm filter (Elkay Labs.). PEPC was separated from soluble protein with a prepacked 1 ml nickel affinity column using an ÄKTA™ Pure 25 L Chromatography System. The loaded column was washed with 50 column volumes of IMAC buffer, then 50 column volumes of IMAC buffer containing 150 mM imidazole. Pure PEPC was eluted with 10 column volumes of IMAC buffer containing 400 mM imidazole.

Protein eluted from IMAC purification was loaded onto a Sephadex G50 desalting column (Amersham Biosciences) and rebuffered in a storage buffer (20 mM Tris, 5% v/v glycerol, 150 mM KCl, 1 mM DTT (AnaSpec. Inc.)). Protein was concentrated to *ca.* 12-15 µM with a Vivaspin 20 MWCO 3000 (Sartorius), aliquoted and frozen at -80°C until use.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE) Analysis

Total protein concentration was determined using the BCA Pierce quantification kit (Thermo Fischer Scientific). Concentration was determined using a standard curve of bovine serum albumin, over a concentration range 0 – 2.0 mg ml^{-1}.

Protein samples were analysed for purity using SDS PAGE analysis. Protein samples were quantified using the BCA Pierce method, 25 µg of cell lysate and 5 µg of pure protein elutions were denatured in 2 × SDS PAGE loading dye (200 mM Tris.HCl pH 6.8, 2 % SDS, 20 % Glycerol, 0.01% Bromophenol blue (BDH Laboratory Supplies) and 7 % β-

mercaptoethanol). Protein was loaded onto an 8% acrylamide SDS gel (6% stacking) with 2 µl of Blue Prestained Protein Standard Broad Range (11-190 kDa) (NEB). Gels were run for 50 minutes at 200 V with 1 × Tris/Glycine/SDS running buffer (Geneflow). Gels were stained with InstantBlue (Expedeon) and imaged with a ChemiDoc™ MP (BioRad).

Enzyme Assays

PEPC activity was measured spectroscopically at 340 nm by coupling to NADH-malate dehydrogenase. Assays with a high fixed concentration of bicarbonate were observed using a FLUOstar plate reader (BMG Labtech) using the 340 nm ± 5 nm absorbance filter (BMG Labtech). Plate reader assays were conducted in a reaction volume of 150 µl at 25°C. A typical reaction mixture contained 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$ (Fluka), 5 mM $KHCO_3$. 0.2 mM NADH (Thermo Fisher Scientific) and 0.1 U μl^{-1} malate dehydrogenase. Assays were initiated with the addition of PEPC enzyme. Rates were calculated with a NADH calibration curve.

Assays at a range of bicarbonate concentrations were observed with a Cary spectrophotometer (Agilent Technologies) in the same reaction buffer, in a total reaction volume of 600 µl. In bicarbonate assays, the water and tricine buffer were sparged with nitrogen for 18 hours prior to use in assays. Bicarbonate assays were constructed under a nitrogen flow. Assays were performed in a sealed cuvette. The reaction was initiated with the addition of 50 nM PEPC, delivered with a gastight syringe (Hamilton). Bicarbonate concentrations were controlled with the addition of freshly prepared potassium bicarbonate. Background bicarbonate was determined using endpoint assays with no potassium bicarbonate, run for 30 minutes. Rates were calculated using the Cary analysis software.

Primers

Table 2: Summary of primers used for PCR. Annealing temperature of PCR protocol with primers also listed.

Primer	Sequence, 5' to 3'	Annealing Temperature / °C
ppc_1072_for	TGCTTGYTKGCNAGAATGATGGC	54°C
ppc_3037_rev	TCATGGTSAGGATGAGSGTGTC	
Pqu_For 5'	GCGCCCCCCTCTCCAGCCACCTAGC	54°C
Pqu_Rev 5'	CCTGGAGGATGTCGAGGAAGCGCTC	
Pqu_For 3'	GAGCTTCAAGGTCGAGAAGCAGCCG	54°C
Pqu_Rev 3'	GCCGACACGTACATCAAGCGTG	
Ppy_Rev 3'	CAGTCCAAGCGTGAATACTG	54°C
Ppy_For 3'	AACAAGCCCGCCGGACTG	
Ppy_Rev 5'	GCGTCGTACTCGACGAGCTTGT	54°C
Ppy_For 5'	TTGAAGCCATCCGCGTCTCCCTCGC	
FlvFor1B	TACTTCCAATCCAATGCAATGGCTAACCGGAAT	72°C
FlvRev1B	TTATCCACTTCCAATGTTATTACTAACCGGTGTTCTGC	
PquFor1B	GACGACGACAAGATGGCGTCCTCCGAGCGCCACC	57°C
PquRev1B	GAGGAGAAGCCCGGTTAGCCCGTGTTCTGCATGCC	
PpyFor1B	TACTTCCAATCCAATGCAATGGCAAGCAG	67°C
PpyRev1B	TTATCCACTTCCAATGTTATTATTAACCGGTATTC	
T7 Promotor	TAATACGACTCACTATAGGG	58°C
T7 Terminator	GCTAGTTATTGCTCAGCGG	

Table 3: Summary of primers used for sequencing.

Primer	Sequence, 5' to 3'
Pqu_1323_Seq_For	CGTGAAGCTGGACAT
Pqu_1752_Seq_Rev	ATGACCTGCTGCTTG
Ppy_1291_Seq_For	GATGGTAGTCTGCTGG
Ppy_1791_Seq_Rev	GCTATCGCTATAACCA
Flav_1303_Seq_For	AGACAAGTGTCGACTT
Flav_1832_Seq_Rev	TTGTAGAGCTGCCATG

Protein Absorbance Coefficient Calculation

Table 4: Absorbance Coefficient Calculation for *Flaveria trinervia* PEPC. Absorbance determined by nanodrop. ε_{GdnHCl} calculated by ExPASy protein parameter tool. $\varepsilon_{Natural}$ calculated using the method described in Gill & von Hippel 1989.

Abs_{Nat}	Abs_{GdnHCl}	ε_{GdnHCl} / M^{-1} cm^{-1}	$\varepsilon_{Natural}$ / M^{-1} cm^{-1}	% Difference
0.354	0.353	119930	120480	0.46

Table 5: Absorbance Coefficient Calculation for *Flaveria pringlei* PEPC. Absorbance determined by nanodrop. ε_{GdnHCl} calculated by ExPASy protein parameter tool. $\varepsilon_{Natural}$ calculated using the method described in Gill & von Hippel 1989.

Abs_{Nat}	Abs_{GdnHCl}	ε_{GdnHCl} / M^{-1} cm^{-1}	$\varepsilon_{Natural}$ / M^{-1} cm^{-1}	% Difference
0.136	0.146	125430	117030	-6.70

Table 6: Absorbance Coefficient Calculation for *Panicum queenslandicum* PEPC. Absorbance determined by nanodrop. ε_{GdnHCl} calculated by ExPASy protein parameter tool. $\varepsilon_{Natural}$ calculated using the method described in Gill & von Hippel 1989.

Abs_{Nat}	Abs_{GdnHCl}	ε_{GdnHCl} / M^{-1} cm^{-1}	$\varepsilon_{Natural}$ / M^{-1} cm^{-1}	% Difference
0.69	0.75	115335	105805	8.3

Table 7: Absorbance Coefficient Calculation for *Panicum pygmaeum* PEPC. Absorbance determined by nanodrop. ε_{GdnHCl} calculated by ExPASy protein parameter tool. $\varepsilon_{Natural}$ calculated using the method described in Gill & von Hippel 1989.

Abs_{Nat}	Abs_{GdnHCl}	ε_{GdnHCl} / M^{-1} cm^{-1}	$\varepsilon_{Natural}$ / M^{-1} cm^{-1}	% Difference
2.85	2.98	116825	111514	4.6

References

Aslanidis, C. and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR)., *Nucleic acids research*, 18(20), pp. 6069–74.

Atkinson, R. R. L., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M. and Osborne, C. P. (2016) C_4 photosynthesis boosts growth by altering physiology, allocation and size, *Nature Plants*. Nature Publishing Group, 2(5), p. 16038.

Bauwe, H. (1986) An efficient method for the determination of K_m values for HCO_3^- of phosphoenolpyruvate carboxylase, *Planta*, 169(3), pp. 356–360.

Berrow, N. S., Büssow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G. E., Levy, N., Lieu, V., Owens, R. J., Peleg, Y., Pinaglia, C., Quevillon-Cheruel, S., Salim, L., Scheich, C., Vincentelli, R. and Busso, D. (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: A comparative study, *Acta Crystallographica Section D: Biological Crystallography*, 62(10), pp. 1218–1226.

Chollet, R., Vidal, J. and O'Leary, M. H. M. (1996) Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants, *Annual review of plant biology*, 47(1), pp. 273–298.

Christin, P.-A., Arakaki, M., Osborne, C. P. and Edwards, E. J. (2015) Genetic Enablers Underlying the Clustered Evolutionary Origins of C_4 Photosynthesis in Angiosperms, *Molecular Biology and Evolution*, 32(4), pp. 846–858.

Dong, L.-Y., Masuda, T., Kawamura, T., Hata, S. and Izui, K. (1998) Cloning, Expression, and Characterization of a Root-Form Phosphoenolpyruvate Carboxylase from *Zea mays*: Comparison with the C_4-Form Enzyme, *Plant and Cell Physiology*, 39(8), pp. 865–873.

Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Analytical Biochemistry*, 182(2), pp. 319–326.

Janc, J. W., O'Leary, M. H. and Cleland, W. W. (1992) A kinetic investigation of phosphoenolpyruvate carboxylase from *Zea mays*, *Biochemistry*, 31(28), pp. 6421–6426.

Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Höning, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes, *Nature Biotechnology*, 31(1), pp. 76–81.

Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R. and Donnelly, M. I. (2002) A New Vector for High-Throughput, Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site, *Protein Expression and Purification*, 25(1), pp. 8–15.

Svensson, P., Bläsing, O. E. and Westhoff, P. (1997) Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase- a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3)., *European journal of biochemistry*, 246(2), pp. 452–60.

Wedding, R. T., Black, M. K. and Meyer, C. R. (1990) Inhibition of Phosphoenolpyruvate Carboxylase by Malate, *Plant Physiology*, 92(2), pp. 456–461.