

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313. 761-4700 800/521-0600



Order Number 9215001

**Structure and expression of a sugarcane gene encoding a
housekeeping phosphoenolpyruvate carboxylase**

Albert, Henrik Horst, Ph.D.

University of Hawaii, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106



**STRUCTURE AND EXPRESSION OF A SUGARCANE GENE
ENCODING A HOUSEKEEPING PHOSPHOENOLPYRUVATE
CARBOXYLASE**

**A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF
THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF**

DOCTOR OF PHILOSOPHY

IN

BOTANICAL SCIENCES (PLANT PHYSIOLOGY)

DECEMBER 1991

By

Henrik Horst Albert

Dissertation Committee:

S. S. M. Sun, Chairperson

Tom Humphreys

John A. Hunt

Andrew Maretzki

John I. Stiles

ACKNOWLEDGEMENTS

I want to thank the teachers, fellow students and other friends who have helped and supported me in my efforts to learn about the molecular biology of plants. In particular my advisor, Dr. Sam Sun, has been a friend in addition to an inspiring teacher. I look forward to maintaining a close relationship in the future.

The other members of my committee, Dr. Tom Humphreys, Dr. John Hunt, Dr. Andrew Maretzki and Dr. John Stiles, have also been very generous with their time and assistance. Dr. Hunt performed the "silent base change" sequence comparison analysis, and took the time to explain its significance to me. Dr. Maretzki and others at Hawaiian Sugar Planter's Association (HSPA) provided me with stimulating discussion and with plant tissue.

In addition, HSPA supported me financially with a fellowship for three years of graduate study.

My parents, Ingeborg and Fritz Albert, have supported me emotionally and materially at an age when they might reasonably expect to be supported themselves.

My wife, Kathleen Houtchens, has been my companion, teacher, and friendly critic. Kath was instrumental in first stimulating my interest in biology, but I love her anyway.

Mahalo to all of you!

ABSTRACT

Two genes encoding phosphoenolpyruvate carboxylase (PEPC) have been isolated from the C-4 plant, sugarcane (*Saccharum* hybrid var. H32-8560), SCPEPCD1 and SCPEPCD2. SCPEPCD1 is ca. 6800bp long, with 10 exons. Positions of the 9 introns are identical to those in a maize C-4 PEPC gene (Hudspeth & Grula 1989) and *M. crystallinum* CAM and housekeeping PEPC genes (Cushman et al 1989). The entire gene sequence from -1561 to ca. 300bp downstream of the putative poly-A addition signal is reported. DNA binding sites conserved in the upstream region of maize and sorghum C-4 specific PEPC genes could not be recognized in this gene. Deduced amino acid sequence of this gene is more similar to plant housekeeping PEPCs than to C-4 specific forms. A motif proposed to act as a substrate site in light mediated activation of photosynthetic PEPC enzymes is present in the SCPEPCD1 protein; evidence is presented for the presence of this site in other housekeeping PEPC proteins. SCPEPCD1 is expressed at low levels in all tissues, but is diminished in leaves during light induced greening. SCPEPCD2 is

expressed relatively abundantly in greening leaves but at much lower levels in stems, a pattern consistent with C-4 specific PEPC genes.

TABLE OF CONTENTS

| | |
|---|-----|
| ACKNOWLEDGEMENTS | iii |
| ABSTRACT | iv |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xii |
| LIST OF ABBREVIATIONS | xv |
| CHAPTER I. INTRODUCTION | 1 |
| LITERATURE REVIEW | 3 |
| C-4 photosynthesis and phosphoenolpyruvate carboxylase | 3 |
| Other PEPC functions | 8 |
| Characteristics of PEPC isozymes | 9 |

| | |
|--|-----------|
| PEPC multi-gene family in C-4 plants | 12 |
| Tissue-specific distribution of PEPC | 14 |
| Localization of C-4 leaf-specific PEPC | 17 |
| Environmental control of expression | 18 |
| Developmental control of expression | 21 |
| DNA methylation | 24 |
| PEPC cDNA and genomic clones | 27 |
| Sequence analysis of PEPC genes | 27 |
| C-4 PEPC genes | 27 |
| Housekeeping PEPC genes | 31 |
| | |
| CHAPTER II. DEVELOPMENT OF HYPOTHESIS | 33 |
| | |
| CHAPTER III. MATERIALS AND METHODS | 35 |
| CHEMICALS | 35 |
| GROWTH OF PLANTS | 35 |
| EXTRACTION OF DNA FROM SUGARCANE LEAVES . . | 36 |
| EXTRACTION OF DNA FROM SUGARCANE STALKS . . | 39 |

| | |
|--|----|
| PREPARATION OF DNA FOR CLONING IN | |
| BACTERIOPHAGE λ REPLACEMENT VECTOR | 40 |
| DEPHOSPHORYLATION OF INSERT DNA | 40 |
| SIZE FRACTIONATION OF INSERT DNA | 41 |
| PREPARATION OF λ VECTOR DNA | 42 |
| LIGATION OF PREPARED SUGARCANE INSERT DNA | |
| TO EMBL4 ARMS | 42 |
| IN VITRO PACKAGING OF RECOMBINANT DNA | 43 |
| SCREENING THE LIBRARY | 44 |
| SMALL-SCALE PREPARATION OF λ DNA | 45 |
| SOUTHERN BLOTS | 46 |
| COPY NUMBER CALCULATIONS | 46 |
| RNA EXTRACTION | 47 |
| SMALL-SCALE RNA ISOLATION | 50 |
| ISOLATION OF POLY-A PLUS RNA | 51 |
| NORTHERN BLOTS | 52 |
| PRIMER EXTENSION cDNA TO DETERMINE | |
| TRANSCRIPTION START SITE | 52 |
| REVERSE TRANSCRIPTION-PCR | 53 |

| | |
|---|----|
| PCR CLONING OF mRNA 5' END | 55 |
| SEQUENCE ANALYSIS | 58 |
| | |
| CHAPTER IV. RESULTS | 60 |
| CONSTRUCTION OF GENOMIC LIBRARY | 60 |
| ISOLATION OF PEPC GENES | 61 |
| Subcloning | 64 |
| Southern Analysis | 65 |
| EXPRESSION | 68 |
| Northern analysis | 68 |
| SCPEPCD1 | 68 |
| SCPEPCD2 | 71 |
| Detection of SCPEPCD1 mRNA by PCR amplification . | 71 |
| STRUCTURE OF SCPEPCD1 | 77 |
| Primer extension to identify transcription start site | 77 |
| Cloning 5' end of SCPEPCD1 mRNA | 78 |
| Sequencing of SCPEPCD1 | 80 |
| Intervening sequences | 82 |

| | |
|---|-----|
| CHAPTER V. DISCUSSION | 90 |
| MULTIPLE PEPC GENES IN SUGARCANE | 90 |
| EXPRESSION | 92 |
| Distribution and relative abundance of PEPC | |
| transcripts | 92 |
| STRUCTURE OF SCPEPCD1 | 96 |
| Is SCPEPCD1 unique in the sugarcane genome? | 96 |
| Introns and Exons | 97 |
| Mapping the 5' end of SCPEPCD1 | 97 |
| Promoter | 101 |
| SCPEPCD1 5' end located in CpG-rich island | 104 |
| Deduced SCPEPCD1 protein structure | 106 |
| EVOLUTIONARY RELATIONSHIP OF PLANT PEPC | |
| GENES | 116 |
| CONCLUSION | 119 |
| REFERENCES | 122 |

LIST OF TABLES

| | |
|--|-----|
| Table 1. <i>Saccharum</i> hybrid var. H32-8560 genomic library | 60 |
| Table 2. Rate of synonymous base changes among plant PEPC genes | 117 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1. PEPC activity in sugarcane tissues | 15 |
| Figure 2. PEPC activity during sugarcane leaf greening | 16 |
| Figure 3. Restriction maps of SCPEPCD λ clones. | 62 |
| Figure 4. Genomic Southern blot probed with 2.5kb <i>Eco</i> RI fragment from λ 3-2-1. | 63 |
| Figure 5. Sequence of λ 3-2-1 and λ 2-3-1 in the region containing the point of divergence. | 64 |
| Figure 6. Plasmid subclones of SCPEPCD1 and 2. | 65 |
| Figure 7. Genomic Southern probed with 0.45kb fragment from SCPEPCD1 exon 9 | 66 |
| Figure 8. SCPEPCD1 copy number reconstruction | 67 |
| Figure 9. Northern blot hybridized to SCPEPCD1 riboprobe | 69 |
| Figure 10. Comparison of RNA pools homologous to SCPEPCD1 and SCPEPCD2 in leaf and stem | 70 |
| Figure 11. Oligonucleotide primers used for PCR amplification of cDNA and predicted products from mRNA and DNA. | 72 |
| Figure 12. RT-PCR time course | 73 |

| | |
|---|-----|
| Figure 13. RT-PCR amplification of SCPEPCD1 cDNA from various sugarcane tissues; SCEX1-19mer & SCEX2-20mer . . . | 74 |
| Figure 14. RT-PCR with SCUTL-18mer(s) and SCEX2-20mer primer pair | 75 |
| Figure 15. RT-PCR with SCCAP-19mer and SCEX2-20mer primer pair | 76 |
| Figure 16. Primer extension from SCUTL-18mer(a) | 77 |
| Figure 17. Amplification of SCPEPCD1 5' end by RACE PCR . . . | 78 |
| Figure 18. Sequence of cloned RACE PCR product aligned to SCPEPCD1. | 80 |
| Figure 19. Sequence fragments assembled to generate 5' sequence for SCPEPCD1 | 81 |
| Figure 20. Structure of SCPEPCD1 | 82 |
| Figure 21. Complete nucleotide and deduced amino acid sequence of SCPEPCD1. | 83 |
| Figure 22. PEPC transcription start sites vs. plant consensus | 98 |
| Figure 23. PEPC transcription start sites vs. eukaryote consensus . | 99 |
| Figure 24. Plant PEPC gene TATA boxes | 101 |
| Figure 25. Frequency of G or C in SCPEPCD1 5' region. | 104 |

| | |
|---|-----|
| Figure 26. Frequency of CpG in SCPEPCD1. | 106 |
| Figure 27. Pileup* alignment of higher plant PEPC proteins | 107 |
| Figure 28. Similarity of higher plant PEPC proteins. | 110 |
| Figure 29. Pairwise comparisons; SCPEPCD1 to other plant PEPC proteins | 111 |
| Figure 30. PEPC active site | 111 |
| Figure 31. Proposed PEPC light activation site | 113 |
| Figure 32. Deduced amino acid sequence of SCPEPCD1 amino terminus | 113 |
| Figure 33. Alternate translation start sites for <i>M. crystallinum</i> housekeeping PEPC gene | 114 |
| Figure 34. Higher plant PEPC amino termini | 115 |

LIST OF ABBREVIATIONS

| | |
|----------------|------------------------------------|
| AMP | adenosine monophosphate |
| ATP | adenosine triphosphate |
| β ME | 2-mercaptoethanol |
| bp | base pair |
| cDNA | complementary DNA |
| CAM | Crassulacean Acid Metabolism |
| CIP | calf intestinal phosphatase |
| DEPC | diethylpyrocarbonate |
| dNTP | deoxynucleotide triphosphate |
| DTT | dithiothreitol |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | ethylenedinitrilo-tetraacetic acid |
| EtBr | ethidium bromide |
| EtOH | ethanol |
| GuCN | guanidine thiocyanate |
| HTF | <i>Hpa</i> II tiny fragments |
| HOAc | acetic acid |
| kb | kilobase |

| | |
|------------------------|---|
| kDa | kiloDalton |
| K_M | Michaelis constant |
| mRNA | messenger RNA |
| MNF1 | maize nuclear factor 1 |
| MNF2a | maize nuclear factor 2a |
| MNF2b | maize nuclear factor 2b |
| <i>M. crystallinum</i> | <i>Mesembryanthemum crystallinum</i> |
| NaOAc | sodium acetate |
| ON | overnight |
| PCR | polymerase chain reaction |
| PEPC | phosphoenolpyruvate carboxylase |
| pfu | plaque forming units |
| poly-A | poly-adenylated |
| RACE | rapid amplification of cDNA ends |
| RNase | ribonuclease |
| RT | room temperature |
| RUBISCO | ribulose 1,5-bisphosphate carboxylase/oxygenase |
| SDS | sodium dodecyl sulfate |
| TdT | terminal deoxynucleotidyl transferase |

| | |
|-----------|---------------------|
| U | unit |
| UTL | untranslated leader |
| V_{MAX} | maximum velocity |

CHAPTER I. INTRODUCTION

Understanding the mechanisms of cell-type differentiation is a central problem in the molecular biology of eukaryotes. An organism made up of multiple tissues and cell-types, each with different structures and functions, requires complex systems for control of gene expression. Cells with identical genetic information require specific gene regulation systems if they are to differentiate structurally and functionally as they follow different developmental pathways.

Phosphoenolpyruvate carboxylase (PEPC) in higher plants is an attractive system in which to study cell-type or organ-specific gene expression. Numerous PEPC isozymes and allozymes exist, with different patterns of abundance and activity, and roles in different metabolic or synthetic pathways. These different PEPC forms are encoded by a multi-gene family, with very different patterns of expression for different family members. C-4 plants have one PEPC form which is specifically involved in C-4 photosynthesis; this isozyme is kinetically and chromatographically distinct from other PEPC isozymes,

and has a unique pattern of expression. This C-4 specific PEPC isozyme is variously referred to as leaf-specific PEPC, photosynthetic PEPC or green PEPC. Crassulacean Acid Metabolism (CAM) plants also have a unique PEPC isozyme, and this form is specifically involved in CAM photosynthesis. In addition to these specifically photosynthetic forms, all plants have other PEPC isozymes not involved in photosynthesis, which are sometimes grouped as "housekeeping" forms (also referred to as non-photosynthetic or non-autotrophic). Expression of the C-4 and CAM specific PEPC genes have been shown to be induced by specific stimuli; such inducing stimuli have not been identified for housekeeping PEPC genes, and their expression may be "constitutive": expressed in most or all tissues at most or all times. Genes encoding the C-4 and CAM specific isozymes have been cloned and sequenced, as have housekeeping PEPC genes from C-4 and CAM plants; to date there is no published sequence data for a housekeeping PEPC gene from a C-3 plant. Here we describe a preliminary expression comparison of two PEPC encoding genes from sugarcane, a C-4 plant, and a more detailed analysis of the structure and expression of one of these, putatively a housekeeping PEPC gene.

Pioneer work in describing the C-4 pathway was done by Kortschak et al (1965) in Hawaii, working with sugarcane. Sugarcane was then, and is still today, of major economic importance in Hawaii. Our choice of a sugarcane gene encoding phosphoenolpyruvate carboxylase (EC4.1.1.3.1.) as a model of organ- and cell-type specific gene expression was intended to continue and build on this tradition of C-4 plant research, to increase our understanding of the physiology of this agronomically important plant at the molecular level, and to begin compiling a collection of sugarcane promoters with defined patterns of expression for possible future crop improvement efforts.

LITERATURE REVIEW

C-4 photosynthesis and phosphoenolpyruvate carboxylase

The C-4 pathway is an adaptation which functions in many tropical plants, allowing efficient photosynthetic carbon fixation to be carried on under high temperature and light intensity conditions, with limiting amounts of water. CO_2 compensation points, the concentration of ambient CO_2 at which carbon gained from fixation equals that lost to

photorespiration, and a measure of photosynthetic efficiency, are much lower for C-4 plants than for C-3 plants.

The C-4 pathway was discovered independently by groups in Hawaii and Australia, both working with sugarcane. They found that malate, aspartate and oxaloacetate, all four-carbon molecules, are the major early products (intermediates) of carbon fixation in these plants (Kortschak et al, 1965; Hatch & Slack 1966; Hatch 1971). In *Chlorella pyrenoidosa*, the green algae used by Calvin and Benson (1949) to first describe photosynthetic carbon fixation, and in most other plants, 3-phosphoglycerate, a three carbon molecule is the first stable intermediate. Later work has shown that the C-4 pathway functions in many plant species, appearing in numerous, widely divergent taxa (Björkman 1976), thereby raising interesting questions about the evolution of this pathway. C-4 crop plants include corn, sorghum and millet in addition to sugarcane.

In C-4 plants, unlike C-3 and CAM plants, the process of photosynthetic carbon fixation is divided between two distinct cell-types. Mesophyll cells and bundle sheath cells, adjacent within the leaves, are

structurally and functionally distinct, with different patterns of gene regulation.

The C-4 pathway can be seen as a CO₂ concentrating mechanism added to the C-3 pathway found in all plants (Björkman 1976). Carbon as atmospheric CO₂ is first fixed by carboxylation of phosphoenolpyruvate to form oxaloacetate in leaf mesophyll cells, then transported to leaf bundle sheath cells. Here CO₂ is released, available for fixation and reduction to form sugars and starches in the C-3 pathway.

If the C-4 pathway can be described as a spatial separation of carbon fixation and carbon reduction, then the CAM pathway can be described as a temporal separation of these same functions. At night, when water losses are minimized, CAM plants open their stomata, harvesting CO₂, which is fixed as four carbon organic acids (primarily malic acid) in mesophyll cells, and stored in the vacuoles of these same cells. During the day, when water losses are greatest, CAM plant stomata are closed. Organic acids accumulated in the vacuole move to the cytoplasm and CO₂ is released, available for reduction in the C-3 pathway. In this way, the plant is able to utilize abundant daytime light

energy to reduce carbon, without incurring the water losses that occur during times of high temperature and light intensities (Salisbury & Ross 1978).

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RUBISCO) catalyzes the first carbon-fixation step in the C-3 pathway: the carboxylation of ribulose 1,5-bisphosphate with atmospheric CO₂. Oxygen (O₂) is an alternate substrate for RUBISCO, and as the ambient concentration of CO₂ drops, the reaction with oxygen is favored increasingly. This oxidation of ribulose 1,5-bisphosphate, or "photorespiration", represents a net loss of fixed carbon and energy for the plant. In tropical conditions, with leaf stomata closed to reduce water loss, CO₂ concentration in the mesophyll air space drops, and photorespiration increases, greatly reducing the photosynthetic efficiency of the plant. C-4 plants "scavenge" CO₂ from the mesophyll air space and effectively concentrate it in the bundle sheath cells, where photosynthetic carbon fixation can proceed efficiently, with very low levels of photorespiration, despite low mesophyll concentrations of CO₂.

This spatial separation of photosynthetic functions requires structural specialization. C-4 plant leaves have evolved "kranz" anatomy:

mesophyll cells surround the bundle sheath cells, which are arranged as a sheath around the vascular bundles. In cross section these structures resemble a wreath, "kranz" in German.

Functional specialization of C-4 leaf cells involves cell-type specific gene expression. Enzymes involved in the initial assimilation of atmospheric CO₂ into C-4 dicarboxylic acids are localized in mesophyll cells, while enzymes involved in decarboxylation and refixation of this carbon in the C-3 pathway are localized in the bundle sheath cells. Enzymes found predominantly or exclusively in the bundle sheath cells include NADP-malic enzyme, RUBISCO and other C-3 cycle enzymes. Pyruvate, orthophosphate dikinase, NADP-malate dehydrogenase and phosphoenolpyruvate carboxylase (PEPC) are located primarily in the mesophyll cells (Brogie et al, 1984). PEPC catalyzes the carboxylation of phosphoenolpyruvate, forming oxaloacetate. In mesophyll cells of C-4 plants this reaction constitutes the initial carbon fixation reaction in photosynthesis (for reviews see Stiborova 1988; Andreo et al, 1987).

Other PEPC functions

While the role of PEPC in carbon fixation in leaves of C-4 and CAM plants is clearly established, PEPC function in C-3 plants, non-photosynthetic tissues of C-4 or CAM plants, or in bacteria is not as clearly understood. In non-photosynthetic plant tissues, anaplerotic functions: providing carbon skeletons for the Krebs cycle and biosynthetic pathways, refixing CO₂ generated by high levels of respiration (Hudspeth & Grula 1989), production of malate for ionic balance (eg in stomata guard cells) and generation of NADPH (Ting & Osmond 1973b) have been proposed as the role for PEPC. Some workers also propose a role for PEPC in C-3 photosynthesis (Matsuoka & Hata 1987), producing malate as a product of photosynthesis, rather than as a photosynthetic intermediate as in C-4 photosynthesis (Ting & Osmond 1973b). The addition of glutamate allows *E coli* defective for PEPC to grow on media containing sugar or glycerol as the sole carbon source (Izui et al, 1986), presumably because the glutamate is a source of four carbon skeletons to replenish the Krebs cycle. This would be consistent with an anaplerotic role for PEPC in *E coli*.

Characteristics of PEPC isozymes

Phosphoenolpyruvate carboxylase (PEPC) in its native form is a homotetramer of approximately 400 kilodaltons (kDa), each monomer having a molecular weight of approximately 100 kDa (Iglesias & Andreo 1989).

Ting and Osmond (1973b) divide higher plant PEPC enzymes into four groups associated with different metabolic pathways:

C-4 photosynthetic PEPC with high K_M for PEP and Mg and high

V_{MAX} ;

C-3 photosynthetic (leaf?) PEPC with low K_M for PEP and Mg and

low V_{MAX} ;

CAM photosynthetic PEPC with low K_M for PEP and high V_{MAX} ;

Non-autotrophic PEPC with low K_M for PEP and Mg and low

V_{MAX} .

While kinetic properties of the C-3 leaf forms and the non-autotrophic forms are similar, these workers group them separately based on chromatographic differences. Etiolated leaves of C-4 plants contain a PEPC enzyme which is similar kinetically and chromatographically to C-3 leaf enzymes (Ting & Osmond 1973b; Goatley & Smith 1974).

Reflecting the relative importance of these functions, PEPC is an abundant protein in the leaves of C-4 plants, while in C-3 plants it is present at far lower levels. PEPC activity per mg protein in the leaves of *Flaveria trinervia*, a C-4 species, is 15 times the level found in leaves of *F. cronquistii*, a C-3 member of the same genus (Adams et al, 1986). Peptide maps of the predominant PEPC species in these *Flaveria* species are very different, further evidence that PEPC exists in more than one form, and in fact electrofocusing gel electrophoresis reveals seven distinct PEPC forms among three species of *Flaveria* (Adams et al, 1986). Different species may have different PEPC allozymes, and at least some species contain several different isozymes.

Recently an active site peptide has been isolated and sequenced from maize leaf PEPC using pyridoxal 5'-phosphate inactivation and reduction with sodium borohydride (Jiao et al 1990). This peptide is highly conserved in all known PEPC amino acid sequences, and contains an invariant lysine residue (Lys 606 in the maize leaf PEPC) suggesting a critical role for this residue in PEPC activity.

In addition to a light mediated increase in the amount of PEPC protein in leaves of C-4 plants, the specific activity of C-4 leaf PEPC is

also light regulated (Jiao & Chollet 1988). PEPC from light-grown maize leaves is "two- to threefold more active and less sensitive to feedback inhibition by L-malate than the corresponding dark form enzyme" and these activity changes correlate to the degree of seryl phosphorylation, which is approximately 50% greater in the light-form enzyme (Jiao & Chollet 1990). *In vitro* phosphorylation of serine residue 15 of maize leaf PEPC by a maize green leaf protein-serine kinase (Jiao & Chollet 1990) or mammalian cyclic AMP-dependent protein kinase (A-kinase)(Terada et al 1990) decreases the half-saturation concentration of PEP, and decreases the inhibitory effect of malate. These enzymatic and regulatory changes are similar to light induced activation of PEPC *in vivo*, and indicates that phosphorylation of this serine residue near the amino terminus of leaf specific PEPC enzymes is the mechanism of light regulation of PEPC activity. Jiao and Chollet (1990) propose the sequence K/R-X-X-S as the structural motif required for the leaf kinase substrate. Deduced amino acid sequences from the published data show that this motif is present in the maize and sorghum leaf PEPC isozymes, and is not present in the housekeeping PEPC isozyme from *Mesembryanthemum crystallinum* (the common ice plant). The motif is also present in the CAM specific PEPC

isozyme from *M. crystallinum*; CAM PEPC activity is also modulated by light, but in this case higher enzyme activity occurs in darkness, not light.

PEPC multi-gene family in C-4 plants

In higher plants the multiple PEPC isozymes are encoded by a multi-gene family. Hybridization of cloned PEPC sequences to genomic Southern blots has been used to characterize the multi-gene family encoding PEPC in maize, with one group estimating three to six genes in an inbred line, Pioneer B73 (Harpster & Taylor 1986). Another group (Hudspeth et al, 1986) using the same line showed that a partial PEPC cDNA (approximately 400 base pairs (bp)) made to green leaf RNA hybridizes to four distinct fragments in a *Hind* III digest of maize genomic DNA, under moderately stringent conditions. The same cDNA clone was used to identify two types of PEPC clones in a maize genomic library. These clones differ in their restriction maps and in the intensity with which they hybridize to this probe, indicating different degrees of sequence homology to the probe, with only the highly homologous clone representing a "photosynthetic" member of the gene family. Using these

different genomic clones (or subclones thereof) to probe northern blots, these workers report abundant RNA in green leaves hybridizing to the photosynthetic gene under moderate stringency, but barely detectable signals in etiolated leaves and roots. Probing with the "non-photosynthetic" gene gives lesser, approximately equal signals in all three tissue types. In each case the major hybridization band is at approximately 3.5 kb. Later work indicates that in maize there is only one photosynthetic member of the gene family, and that under "standard blot conditions" there is very little cross hybridization between this gene and other members of the family. Three other members of the family, including the non-photosynthetic gene mentioned above do "exhibit high levels of cross-hybridization" (Grula & Hudspeth 1987; Hudspeth & Grula 1989). Higher stringency northern blot analysis of RNA from other maize tissues indicates that the photosynthetic PEPC gene is transcribed at high levels not only in mature green leaves, but generally in photosynthetic or potentially photosynthetic tissue such as young green leaves, inner leaf sheaths, tassels and husks, with very low levels of expression in non-photosynthetic tissue such as roots and seeds. Non-photosynthetic PEPC genes are expressed most highly in seedling

stems, with relatively high levels in roots, husks, seeds, inner leaf sheaths and mature green leaves (Hudspeth & Grula 1989).

Tissue-specific distribution of PEPC

As discussed above, physical and kinetic studies of PEPC isozymes extracted from different organs of C-3 or C-4 plants indicate an organ specific pattern of distribution of these isozymes (Ting & Osmond 1973a; Ting & Osmond 1973b, Goatly & Smith 1974). Root PEPC is different from leaf PEPC, and PEPC from etiolated C-4 leaves is different from that extracted from green C-4 leaves. Some of these studies do not effectively separate similar forms present in the same organ at the same time, and so they may in some cases represent "averaged" results from several isoforms.

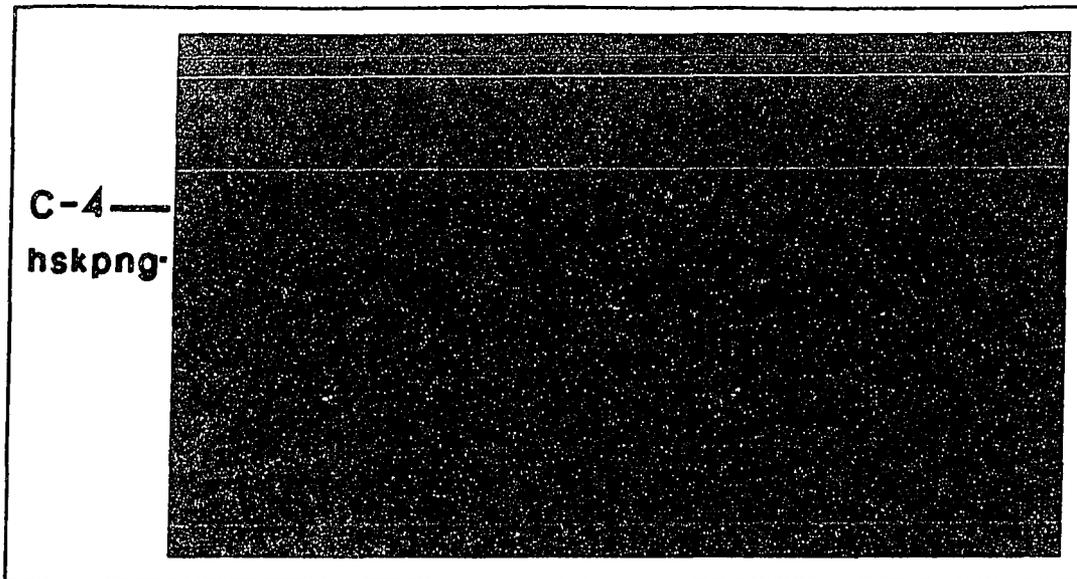


Figure 1. PEPC activity in sugarcane tissues (S. Sun unpublished results).

A zymogram of native proteins extracted from different sugarcane tissues shows a "slow" migrating band of PEPC activity present only in green leaves, with another faster migrating band present in all tissues (fig. 1).

In etiolated leaves, only the fast band is detectable, but as the leaves are exposed to light, the slow band appears, and becomes predominant (fig. 2). Presumably the slow band is one or more C-4 specific PEPC isozymes, and the fast band is made up of the housekeeping PEPC isozymes.

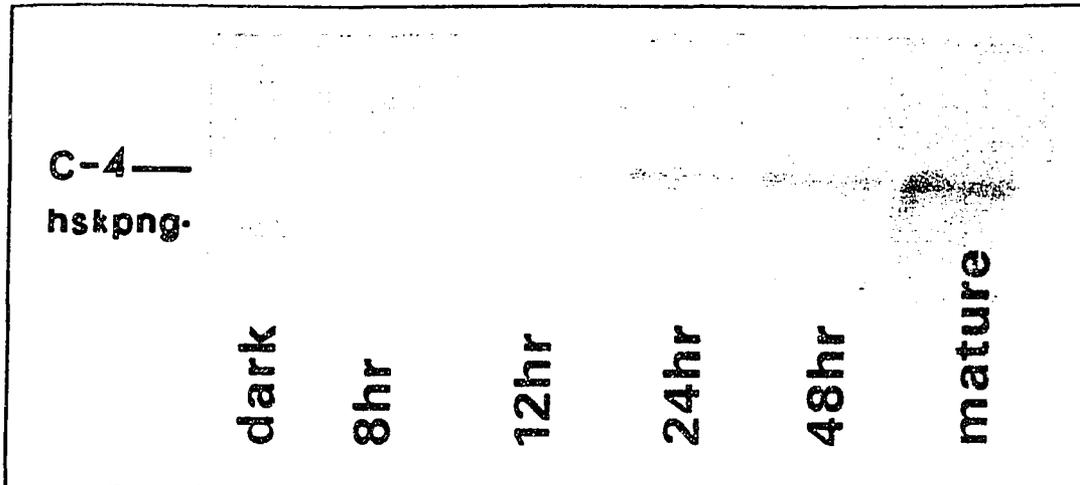


Figure 2. PEPC activity during sugarcane leaf greening (S. Sun unpublished results).

Studies following the distribution or induction of specific housekeeping isoforms have not been reported in C-4 plants.

In *M. crystallinum*, a facultative CAM plant (the C-3 pathway operates during unstressed conditions; the CAM pathway is induced by drought stress) transcription from a CAM-specific PEPC gene (PPC1) and a housekeeping PEPC gene (PPC2) have been studied (Cushman et al 1989). Under unstressed conditions, the CAM-specific transcript is more abundant in roots than in leaves; during salt stress this relationship is reversed as the transcript pool increases in leaves while in roots it gradually decreases. Transcripts from PPC2, the housekeeping PEPC gene, are also more abundant in roots than in leaves during unstressed

conditions, but under salt stress this transcript decreases in both organs. Southern blot analysis indicates that PPC1 exists as a single copy, but there may be several genes similar to PPC2; therefore it is uncertain whether the transcription pattern observed for this housekeeping PEPC gene represents RNA from this specific gene, or may be a pool from several similar genes.

Localization of C-4 leaf-specific PEPC

Enzymatic and mechanical treatment of maize (a C-4 plant) leaves allows the separation of cell-types, yielding relatively pure preparations of mesophyll cells and vascular strands, including bundle sheath cells (Brogie et al, 1984). Comparison of proteins from the two cell-types by gel electrophoresis and identification of specific proteins with monospecific antibodies shows major differences in protein composition, with PEPC and several light harvesting and electron transport proteins present at high levels in the mesophyll cells but not detectable in bundle sheath. Both subunits of RUBISCO are abundant in bundle sheath cells, but not detectable in mesophyll. *In vitro* translation of poly A+ RNA from these two cell-types yields a similar pattern of protein distribution,

indicating that cell-type specific regulation of these genes is determined, at least in part, by the level of translatable mRNA (Broglie et al, 1984). Northern blot analysis shows that the levels of translatable mRNA reflect the steady state pool size of these transcripts, levels of PEPC mRNA are high in mesophyll cells and low or undetectable in bundle sheath (Sheen & Bogorad 1987) with the opposite pattern for the small subunit of RUBISCO (Broglie et al, 1984; Sheen & Bogorad 1987). *In Situ* hybridization of maize leaf sections with radioactive probes specific for PEPC and RUBISCO large subunit confirms the cell-type specific distribution of transcripts for these genes in a C-4 plant (Martineau & Taylor 1986).

Environmental control of expression

In addition to cell-type specific regulation, photosynthetic PEPC genes are also under the influence of environmental factors. Like numerous other genes encoding proteins involved in photosynthesis, some PEPC genes have been shown to be induced by light. In maize (Sheen & Bogorad 1987, Hudspeth et al, 1986, Nelson & Langdale 1989 and references therein) and sorghum (Thomas et al, 1987) mRNA encoding

PEPC increases significantly when etiolated leaves are exposed to light. In sorghum this increase can be induced by a short exposure to red light; if red exposure is followed by short exposure to far-red light, the induction of PEPC mRNA is not observed (Thomas et al, 1987). This red/far-red reversability is considered classical evidence that the response is mediated by the photoreceptor phytochrome.

In sorghum, monoclonal antibodies have been used to show that the photosynthetic PEPC isozyme is induced by light in greening leaves, while at least one antigenically distinct PEPC isozyme, present in etiolated leaves, is unaffected or reduced (Thomas et al, 1987). In sugarcane, PEPC extracted from green leaves shows physical and kinetic properties typical of the C-4 photosynthetic enzyme, while PEPC extracted from etiolated leaves shows characteristics more typical of C-3 plants (Goatly & Smith 1974). It appears from this data that photosynthetic PEPC genes are light regulated, while those encoding isozymes with non-photosynthetic functions are not.

The adaptive advantage of photosynthetic genes evolving to be light regulated seems clear. Interestingly, in both rice and wheat, two gramineous species which lack the C-4 pathway, at least one PEPC

isozyme has been reported which is also regulated by light (Matsuoka & Hata 1987). In wheat this isozyme has been shown to be structurally related to the main isozyme in maize leaves by peptide mapping and immunoblotting (Matsuoka & Hata 1987). This seems to support a role for PEPC in C-3 photosynthesis, long postulated by some workers.

Some of the earliest work tracing the path of carbon in C-3 photosynthetic carbon fixation showed a small but significant fraction of ^{14}C from labeled CO_2 incorporated in phosphoenolpyruvate (10%) and malate (3%) after five seconds of photosynthesis. Pulse-chase experiments show that these products accumulate during the chase, indicating that they are final products, and not intermediates of carbon fixation as in the C-4 pathway (Bassham et al, 1950). Other workers have seen these results as evidence that malic acid, not being an intermediate, is an end product of C-3 photosynthesis, and therefore PEPC has a role in C-3 photosynthesis (Ting & Osmond 1973A&B). Details of this aspect of C-3 carbon fixation have not been demonstrated, but if true, the genes encoding these isozymes may constitute a "missing link" in the evolution of the C-4 pathway.

In plants with Crassulacean acid metabolism (CAM), another carbon fixing strategy well-adapted to arid environments, PEPC also plays a central role. As in C-4 plants, it catalyzes the fixation of CO₂ into oxaloacetate. In the common ice plant (*M. crystallinum*), a facultative CAM plant, PEPC activity in leaves increases up to 50% in response to osmotic stress. After five days of salt stress, transcripts of the CAM-specific PEPC gene (PPC1) increase about 30-fold in leaves, while those of a housekeeping PEPC gene (PPC2) decrease slightly. Nuclear run-off experiments show that these differences in steady state levels of these two transcripts are due at least in part to differences in transcription (Cushman et al 1989). In CAM plants as in C-4 plants, PEPC is encoded by a multi-gene family, and those members of the family encoding photosynthetic isozymes are regulated in part by environmental signals.

Developmental control of expression

Environmental and cell-type signals regulating expression of photosynthetic PEPC genes do not operate independently of each other or of developmental factors. In light-grown maize leaves, significant PEPC mRNA and protein accumulation does not occur until four to six cm from

the leaf base meristem, which is the region where mesophyll and bundle sheath cells display fully differentiated morphologies. In contrast, mRNA encoding both large and small subunits of RUBISCO starts to accumulate significantly nearer the base of the leaf, before the leaf cell-types are fully differentiated, and before the RUBISCO polypeptides are detectable (Martineau & Taylor 1985).

Morphological differentiation and illumination are necessary for expression of the photosynthetic PEPC gene in maize mesophyll cells, but not sufficient. In foliar leaves, veins are closely spaced, so that there are only two mesophyll cell files between veins. In husk leaves veins are more widely spaced, with up to 20 files of mesophyll cells between veins. In foliar leaves all of the mature mesophyll cells express PEPC at significant levels, while RUBISCO expression is not detectable. In husk leaves this pattern of gene expression is present in mesophyll cells adjacent to veins, but those mesophyll cells more than several cells away express RUBISCO and lack significant levels of PEPC, suggesting a gradient of some regulatory trans acting factor(s) centered on veins (Nelson & Langdale 1989 and references therein).

It appears from these studies that C-3 carbon fixation may be the "default" state, and that the genes encoding the C-4 enzymes are expressed only in cells which have differentiated and exist in a specific spatial anatomical relationship. In parenchyma cells isolated from sugarcane stalks, labeled carbon from $^{14}\text{CO}_2$ appears predominantly in 3-phosphoglycerate after three seconds, indicating that the C-3 pathway is operative (Kortschak & Nickel 1970). In apparent contradiction to this need for cell-type differentiation and trans-acting regulatory factors from neighboring cell-types is a report of C-4 photosynthesis in stable callus culture derived from stem explants of *Froelichia gracilis* (Laetsch & Kortschak 1972). These workers reported four carbon molecules (aspartate and malate) as the first labeled compounds eight seconds after exposure to $^{14}\text{CO}_2$ in this undifferentiated tissue. Eight seconds is a rather long time over which to determine first carbon fixation intermediates; perhaps these labeled four carbon acids are really final products. However, if these results actually represent C-4 carbon fixation, serious questions are raised about our understanding of the separation of photosynthetic functions in different cell-types in C-4 leaves.

DNA methylation

In higher plants and vertebrates a high percentage of cytosine residues in the pattern CpG or CpXpG are methylated at the five position on the cytosine ring. Because eukaryotic DNA repair enzymes do not recognize deamination of 5-methylcytosine to give thymine, the frequency of CpG and CpXpG in higher plant genomes is far lower than expected for a random distribution. Unmethylated CpG and CpXpG sites tend to be clustered, and these clusters are often centered around the 5' end of genes. These hypomethylated clusters are often referred to as "CpG islands" or "HTF (HpaII Tiny Fragments) islands" (Bird 1986; Antequera & Bird 1988). HpaII is a methylation-sensitive restriction enzyme with the recognition sequence CCGG.

Antequera & Bird (1988) suggest that CpG islands are non-methylated in all or most tissues, so hypomethylation correlates to potential transcription, not actual transcription. Also, that genes expressed only in specific cell types, such as zein, tend not to be located in CpG islands. For zein, a maize seed storage protein expressed only in endosperm cells during seed development methylation of specific DNA

sites has been shown correlated to tissue specific and possibly developmentally regulated expression (Bianchi & Viotti 1988)

Methyl-sensitive and -insensitive isoschizomer restriction enzymes have been used with Southern blot analysis to investigate the methylation status of several C-4 photosynthetic genes in greening maize leaves. Ngerprasisiri et al (1989) report methylation of PEPC and several other maize nuclear C-4 photosynthetic genes inversely correlated to expression of these genes in mesophyll and bundle sheath cells of etiolated, greening and mature green leaves. For these nuclear genes, this group observed the same differential methylation patterns in fully green leaves as in greening and etiolated leaves, indicating that reduced levels of methylation precede the time of major transcriptional activation. In fully green leaves this pattern of methylation is also reported in a chloroplast gene, the gene for the large subunit of RUBISCO. The gene is less highly methylated in bundle sheath cells, where it is highly expressed, than in mesophyll cells, where its expression is much lower. In this case, however, this difference in methylation is not observed in etiolated leaves, and only becomes apparent in greening leaves (Ngerprasisiri et al, 1989). The Ngerprasisiri group used the

isoschizimers *Bst*NI and *Eco*RII (recognition site = CCA/TGG) in their work. Langdale et al (1991) carried out similar experiments on maize, also using *Bst*NI and *Eco*RII, and 25 other methylation sensitive restriction enzymes. They report no methylation patterns which correlate with gene expression for either the small or large subunit RUBISCO genes. For the C-4 PEPC gene, they do report a *Pvu*II (recognition site = CAGCTG) site greater than 3 kb upstream of the transcription start site which is hypomethylated in mesophyll cells. This hypomethylated site is not present in etiolated leaves, and only becomes apparent after 24 hours of leaf greening, thus correlates very well to the accumulation of PEPC transcripts during leaf greening.

The experiments by Langdale et al (1991) also confirm that for the maize C-4 PEPC gene, the "immediate 5' region...and some of the coding region, represents an unmethylated CpG island". As the C-4 PEPC gene is expressed in a highly cell-type specific manner, this runs counter to the proposed pattern of CpG islands being associated with constitutively expressed genes.

PEPC cDNA and genomic clones

The study of PEPC in higher plants has been facilitated by the use of genomic clones encoding PEPC in maize (Hudspeth et al, 1986; Matsuoka & Minami 1989), *M. crystallinum* (Cushman & Bohnert 1989a,b) and sorghum (Cretin et al, 1991), and cDNA clones from maize (Harpster & Taylor 1986; Hudspeth et al, 1986; Izui et al, 1986), sorghum (Thomas et al, 1987; Cretin et al, 1990, 1991) several species of *Flavaria* (S. Sun, unpublished results) and several others, including genes from bacteria and cyanobacteria (Izui et al, 1986).

Sequence analysis of PEPC genes

C-4 PEPC genes

The first published higher plant PEPC sequence was of a maize leaf PEPC cDNA clone synthesized by Izui et al (1986). This clone has been shown to be nearly full length. The sequence of this clone includes an open reading frame 2805 nucleotides, a 3' untranslated region of 222 nucleotides and a poly-A tail of 64 residues. The 5' untranslated leader and part of the amino terminus coding sequence are missing.

Comparison of deduced amino acid sequences from this clone shows 43%

homology to the endogenous *E. coli* PEPC, and 33% homology to the PEPC of *Anacystis nidulans*, a cyanobacteria.

Full-length genomic clones for photosynthetic PEPC genes from two maize cultivars, B73 (Hudspeth & Grula 1989) and Golden Cross Bantam (Matsuoka & Minami 1989) have been completely sequenced. The two genes are very similar; both encode polypeptides of 970 amino acid residues, and contain ten exons with nine introns. Normal TATA consensus sequences are absent in the upstream regulatory region, but both have a TATA-like sequence, TATTTGAA, approximately 30 base pairs upstream of the transcription start site. Upstream regions of both genes contain numerous sequence elements thought to be potentially involved in gene regulation either because they are repeated or because they have homology to known regulatory elements in other genes. These include an eight base pair sequence (CCTTATCCT) at approximately -650 which fits a consensus sequence (CCTTATCAT) found in almost all phytochrome mediated light regulated genes (Matsuoka & Minami 1989; Grob & Stüber 1987). Matsuoka and Minami (1989) report a six bp sequence (CCGCCC) present twice in the upstream region and twice in the first intron as being identical to the "core sequence" of the Sp1

binding site present in numerous mammalian and mammalian virus genes (Kadonaga et al, 1986 and references therein). A similar sequence (CCGCC/G) is also present in the B73 gene, with two copies upstream and one in the first intron. Recent reports of intron sequences significantly effecting gene expression levels in some plant genes makes this element particularly intriguing (Callis et al, 1987; Dean et al, 1989). Also present are four (B73) or six (Golden Cross Bantam) copies of a G/C rich imperfect direct repeat (C/aC/tCT/cgC/tXXCCACATCC/t) which is also present in RUBISCO small subunit (SSU) genes in maize and wheat (Hudspeth & Grula 1989). The B73 cultivar additionally contains two 29 bp imperfect direct repeats (AACAGCAC/GCG/AAGCCAAGCCAAAAGGAGC).

A sorghum C-4 PEPC gene has been cloned as a full-length cDNA (Cretin et al, 1990) and a genomic clone containing the 5' region (Cretin et al, 1991). This gene is very similar to the maize C-4 PEPC genes; it encodes a protein of 952 amino acid residues, its putative TATA box is identical to the maize C-4 genes, and it shares several of the upstream repeats.

At present no *in vivo* evidence for functional significance of any of these sequence elements has been reported, however, as discussed in the following, recent work has shown that several of these elements act as sequence specific binding sites for nuclear proteins *in vitro*.

Kano-Murakami et al (1991) have identified a maize nuclear protein which binds to 4 copies of the G/C rich repeat (mentioned above) in a fragment from -570 to -272 of the maize C-4 PEPC gene. Competition and mutagenesis experiments show that binding by this protein, which they call PEP-1, is sequence specific for the G/C repeat. In gel retardation assays, the PEP-1 binding activity was present in nuclear extracts from green or etiolated leaves, but not in extracts from stems or roots. Yanagisawa & Izui (1990) identified what is presumably the same protein (they refer to it as MNF2a: maize nuclear factor 2a) and a related factor MNF2b, which shows similar affinity for the G/C repeats, but is chromatographically distinct. They also identify MNF1, which binds a repeat sequence (RS1) present at -886 to -849 and -846 to -807. Binding activities corresponding to all three of these factors are present in nuclear extracts from green or etiolated leaves, but not in stems or roots. MNF2a activity was much higher in extracts from etiolated leaves than

from green leaves, while MNF1 and MNF2b activities were approximately equal in green and etiolated leaves. This led the authors to speculate MNF1 and MNF2b may be "positive regulators involved in leaf-specific expression, and that [MNF2a] is a negative regulator related to the light-dependent expression in leaves."

In the putative binding region for MNF1, there is also "one copy of a sequence with some homology to the "G-box""; a consensus element shown to act as a binding site for a nuclear factor (GBF) and involved in light-regulated and/or leaf specific gene expression in dicots (Yanagisawa & Izui 1990).

Housekeeping PEPC genes

A full length housekeeping PEPC genomic clone from *M. crystallinum* (Cushman & Bohnert 1989a) has been sequenced, and in sorghum, a full length cDNA and a genomic clone of the 5' region (Cretin et al 1991) of a housekeeping PEPC gene. In maize, the 5' region of two PEPC genes which are not the (single) C-4 specific PEPC, and so are presumably housekeeping PEPC genes (but may be pseudogenes) have been sequenced (Yanagisawa & Izui, 1989).

In many respects these genes are similar to the C-4 PEPC genes, but notably they lack the repeats present in the upstream regions of the maize and sorghum C-4 genes.

CHAPTER II. DEVELOPMENT OF HYPOTHESIS

We have isolated two different PEPC encoding genes from sugarcane. Transcripts homologous to one of these genes (SCPEPCD2) can be easily detected by northern blot analysis in greening leaf poly-A RNA, but only at a much lower level in poly-A RNA from stems. Transcripts homologous to the other gene (SCPEPCD1) are apparently far less abundant, but can be detected at low levels in the non-photosynthetic organs stems, shoots, roots and in white callus; in etiolated, greening or mature leaves transcripts homologous to SCPEPCD1 are still less abundant.

Overall, and with respect to known structural features of PEPC, the deduced amino acid sequence of SCPEPCD1 is more similar to the sequence of housekeeping PEPC genes from *M. crystallinum* or sorghum than to the sequence of C-4 specific PEPC genes from maize or sorghum.

Here we report the structure and expression of a PEPC encoding gene from sugarcane (SCPEPCD1), and evidence for the hypothesis: SCPEPCD1 encodes a housekeeping PEPC isozyme, and this gene, alone or in concert with other very similar genes which make up a

"housekeeping" subgroup of the sugarcane PEPC gene family are transcribed at very low levels in many or all organs of the sugarcane plant.

CHAPTER III. MATERIALS AND METHODS

CHEMICALS

All chemicals used were reagent grade or better, and purchased from Fisher Scientific unless otherwise noted.

GROWTH OF PLANTS

Field grown plants (*Saccharum* hybrid var. H32-8560) were harvested from Hawaii Sugar Planters' Association breeding station, Maunawili, Oahu.

Shoots were germinated by cutting ca. 8 cm. sections containing nodes from mature stems. These stem sections were soaked in H₂O at 52° for 20', then for 1' in a RT solution of Benlate fungicide in H₂O at 600 mg/L. The treated stem sections were then placed in plastic bags and incubated in darkness at 33° until shoots and/or roots developed to the desired size. Shoots germinated in this way could be used to generate dark grown leaves. To do so, germinating shoots were removed from plastic bags after six days and transplanted to moist vermiculite trays in

complete darkness. These trays were then incubated in a growth chamber in complete darkness at 27° for an additional six days before harvest of etiolated leaf tissue, or transfer to light for 18 or 48 hours before harvest of greening leaf tissue.

Alternatively, etiolated or greening light tissue was obtained by cutting several tillers on a young plant back to ca. 2 cm, and placing the plant in a dark growth chamber at 27° for 7 days before harvest of new etiolated leaf tissue, or transfer to light.

EXTRACTION OF DNA FROM SUGARCANE LEAVES

Young leaves from variety H32-8560 were cut, frozen in liquid nitrogen and stored at -80°.

Frozen leaves (approx. 75 grams) were ground under a small amount of liquid nitrogen in a Waring commercial blender at high speed for 60 seconds. Powdered leaves were transferred to mortar and pestle prechilled at -20°, and ground further for about 15 minutes, with periodic additions of small amounts of liquid nitrogen to keep tissue frozen. Five to seven volumes of grinding buffer were added to the mortar, and the

powder was allowed to thaw. This slurry was ground extensively (about 30 minutes).

This slurry was filtered through four layers of sterile cheesecloth two times, with moderate wringing. The filtered solution, in 250 ml Nalgene bottles, was centrifuged in a GSA (Sorvall) rotor at 5K for five minutes at 4°. The pellets were resuspended in 50 ml cold chloroplast lysis buffer, and incubated on ice for ten minutes. Nuclei were pelleted for five minutes at 5K, 4°, in a GSA rotor, and resuspended in 18 ml nuclei lysis buffer. Proteinase K was added to 0.2 mg/ml, and the solution incubated at 37° for 1.5 hours. Membrane debris was pelleted for two minutes at 5K, 4° in a GSA rotor, and the supernatant transferred to a sterile 50 ml Falcon tube. TE was added to 20 ml, and 20 g CsCl were dissolved in the solution. 0.75 ml 10 mg/ml EtBr were added, and the solution transferred to a 40 ml Nalgene tube and centrifuged at 12K, 4° for 30 minutes in an SS34 (Sorvall) rotor to remove insoluble material. The supernatant was transferred to a 39 ml polyallomer tube, topped up with CsCl solution (1g CsCl/ml TE), and centrifuged in a VTi 50 rotor for 18 hours at 42K, 20°.

The resulting DNA band was removed with a 16 gauge hypodermic needle, and extracted repeatedly with isoamyl alcohol to remove EtBr. The DNA solution was then dialyzed against 2 l of TE at 4° for two days, with four changes of buffer. The dialysed DNA solution was transferred to a 30 ml Corex tube, and 0.1 volumes of 3M NaOAc, pH5.2 were added and mixed. 2.5 volumes of -20° EtOH were gently layered on top of this. A small glass rod was used to spool high molecular weight DNA from the interface. The spooled DNA was washed three times with 70% EtOH, and briefly air-dried. The dried DNA was then resuspended in 0.8 ml TE. DNA quantity and purity was assayed spectrophotometrically, and approximate size was estimated by electrophoresis of an aliquot on a 0.2% agarose/TBE gel at 4°, 23V, and comparing to uncut lambda DNA. "Digestibility" of the DNA was tested by digesting aliquots with *Eco* RI and *Mbo* I restriction enzymes.

Grinding buffer

0.3 M sucrose
50 mM Tris-HCl, pH 8.0
5 mM MgCl₂

Chloroplast lysis buffer

15% sucrose
50 mM Tris-HCl, pH 8.0
50 mM EDTA
2% Triton X100

Nuclei lysis buffer

30 mM Tris-HCl, pH 8.0
10 mM EDTA
10% Sarkosyl

EXTRACTION OF DNA FROM SUGARCANE STALKS

DNA was extracted from ca. 110 g. of soft, etiolated tissue from two mature field grown stalks (H32-8560) essentially as described above for extraction of DNA from leaves. Small pieces were cut from soft stem regions directly into liquid N₂. More extensive grinding with mortar and pestle was required to reduce frozen stem tissue to a fine powder. This fine powder was suspended in 350 ml grinding buffer, and processed as above. Stem tissue yielded more DNA and of higher molecular weight than was obtained from leaves.

PREPARATION OF DNA FOR CLONING IN BACTERIOPHAGE

REPLACEMENT VECTOR

Genomic DNA was partially digested with *Mbo* I to obtain 20 kb insert fragments. Trial reactions varying both restriction enzyme concentration and time of digestion were carried out to determine conditions which produced the optimum yield of 20 kb fragments. Large scale reactions often failed to reproduce the results obtained in small scale pilot reactions. Very exacting reproduction of pilot reaction conditions (tube size, concentration of DNA, etc) was required to reproduce pilot results in large scale reactions.

DEPHOSPHORYLATION OF INSERT DNA

Some libraries were constructed without size fractionating the insert DNA; in these cases insert DNA was dephosphorylated. Genomic DNA which had been partially digested with *Mbo* I to optimize for 14 to 20 kb sizes was ethanol precipitated, washed twice with 70% ethanol, dried briefly under vacuum, and resuspended in TE at 1 $\mu\text{g}/\mu\text{l}$. Ten μg of this DNA was treated with one unit Calf Intestinal Phosphatase (CIP) (Promega Biotech) in a reaction volume of 50 μl which contained 1X

reaction salts and buffer as provided by Promega. After a 30 minute incubation at 37°, 5 µl 100mM trinitriloacetic acid were added, and the reaction was incubated at 70° for ten minutes to denature the enzyme. This was followed by precipitation with NaOAc and ethanol, two washes in 70% ethanol, brief drying, and resuspension in TE at 0.5 µg/µl.

10X CIP reaction buffer (Promega)

500 mM Tris-HCl, ph 9.0
10 mM MgCl₂
1 mM ZnCl₂
10 mM Spermidine

SIZE FRACTIONATION OF INSERT DNA

To optimize the number of 20 kb insert fragments available for ligation, partially *Mbo*I digested sugarcane was size fractionated by zonal centrifugation through a 5 to 20% NaCl gradient. 100 µg of digested DNA in a volume of 1.8 mls was layered onto a 15 ml 5 to 20% NaCl continuous gradient and centrifuged at 16K rpm in an SW28.1 rotor (Beckman) at 21° for 22 hrs (acceleration and decelation 7). Ten drop fractions (ca. 0.5 ml) were collected, and a 10 µl aliquot from each fraction was analyzed on an 0.4% agarose gel. The four optimum

fractions (18-21) were pooled, EtOH precipitated, washed twice with 70% EtOH, and resuspended in TE at 0.75 $\mu\text{g}/\mu\text{l}$.

PREPARATION OF λ VECTOR DNA

Bacteriophage λ EMBL4 DNA was purified by CsCl step gradient centrifugation, essentially as described by Sambrook et al (1989). Purified vector DNA was restricted with *Bam* HI and *Sal* I, then precipitated and resuspended in 10 mM MgCl_2 . This solution was incubated at RT for several hours to allow λ cohesive termini to anneal. The annealed arms were then separated from stuffer fragments by centrifugation through a continuous 5 to 20% NaCl gradient, as described above. Peak fractions containing arms were pooled, mixed with an equal volume H_2O , precipitated with an equal volume isopropanol, washed twice with 70% EtOH and resuspended in TE at 0.5 $\mu\text{g}/\mu\text{l}$.

LIGATION OF PREPARED SUGARCANE INSERT DNA TO EMBL4 ARMS

1.3 μg insert DNA was ligated to 3.0 μg arms overnight at 14°. The reaction was carried out in a final volume of 10 μl which was

66 mM tris-Cl (pH 7.6), 5 mM MgCl₂, 5 mM DTT, 1 mM ATP and contained 2 U T4 DNA ligase.

IN VITRO PACKAGING OF RECOMBINANT DNA

Ligation products were packaged using "Gigapack Gold" (Stratagene) packaging extracts per the suppliers instructions. The library was titered on *E. coli* host strains LE 392 (*supE44 supF58 hsdR514 galK2 galT22 metB1 trpR55 lacY1*), Q 358 (*supE hsdR φ80'*), Q 359 (*supE hsdR φ80' P2*) and K 802 (F- *lacY1* or $\Delta(lac)6$ *supE44 galK2 galT22 mcrA rfbD1 metB1 mcrB1 hsdR2 (r_K-m_K+)*). The titer on K 802 (6.2×10^6) was ca. five fold higher than on any other strain, presumably because highly methylated sugarcane DNA is better tolerated by this *McrA*-, *McrB*- strain (Woodcock et al, 1989). Unfortunately, this strain is *RecA*+, so the risk of incurring some types of cloning artifacts is increased.

This primary library was amplified on K 802, and the amplified library was titered on permissive (Q 358) and non-permissive (Q 359) strains to determine what percent of the library was made up of

recombinants. The titer on these two strains was essentially equal, indicating nearly 100% recombinants.

SCREENING THE LIBRARY

1 X 10⁶ pfu from the amplified library were screened on the host MB 406 (*supE*, *recB21*, *recC22*, *sbcB15*, *hflA*, *hflB*, *hsdR*), using six 22 X 33 cm baking dishes. After overnight incubation, these "megaplates" were placed at 4° for 1 hr. Phage were transferred to nitrocellulose sheets wet with H₂O and soaked in 6X SSC. After 4', the blots (phage side up) were denatured on 3MM (Whatman) paper saturated with 1.5M NaCl, 0.2N NaOH. After 4', the blots were neutralized on 3MM paper saturated with 3M NaCl, 0.5M tris-Cl, pH 7.2. After 4', the blots were laid on 3MM paper and allowed to air dry, then baked under vacuum at 80° for one hr. Blots were prehybridized in 100 mls of 6X SSD, two blots per heat seal bag, at 60° for at least 4 hrs. Prehybridization solution was removed and replaced with 100 mls fresh 6X SSD containing 30 µg/ml poly-A RNA, 10 µg/ml sheared and denatured salmon sperm DNA, and 4X10⁵ cpm/ml ³²P-labeled probe. Hybridization was carried out ca 16 hrs at 60°.

Because the probe was heterologous, initial washing was at reduced stringency. After hybridization, (6) blots were washed twice in 1L 3X SSC, 0.5% SDS, at RT for 1 hr, then three times in 1L of the same solution at 65°. Autoradiography of these blots showed ca. 1000 strongly hybridizing plaques per megaplate. In response to this high number, blots were rewashed at higher stringency; three times in 1L 0.5X SSC, 0.5% SDS at 65° for 1 hr. After this wash, autoradiography showed ca. 10 strongly hybridizing plaques/megaplate. 60 hybridizing plaques were picked using the large end of a pasteur pipette, and half of these were subject to two additional rounds of plaque purification. Of this half, five ultimately proved to contain PEPC genes.

6X SSC: 1X equals 150mM NaCl, 15mM Na citrate, pH7.2

6X SSD: equals 6X SSC plus 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% BSA and 0.5% SDS

SMALL-SCALE PREPARATION OF λ DNA

Phage DNA was purified from plate lysates by DEAE-cellulose chromatography by the method of Helms et al (1987).

SOUTHERN BLOTS

DNA from agarose gels was transferred to nylon membranes such as "Hybond +" (Amersham) by the alkaline transfer method of Reed & Mann (1985). Nucleic acids were fixed to the membrane by UV cross-linking using a "Stratolinker" (Stratagene) in "auto-link" mode. Hybridization was by the method of Church & Gilbert (1984).

COPY NUMBER CALCULATIONS

Our best estimate for the size of the sugarcane genome comes from Bennett and Smith (1976), who list 6.4 pg as the mass of DNA in the nucleus of a cell with the "unreplicated haploid chromosome complement" of a sugarcane variety with a chromosome number (116-117) similar to that of H32-8560, 115-116 (Sreenivasan et al 1987).

6.4 pg of DNA equals ca. 5.8×10^9 bp. For copy number experiments a 7.5kb (total) plasmid clone was used; this means a $1:1.3 \times 10^6$ mass ratio (genomic:cloned) would be an equimolar ratio. The southern blot shown in figure 8 contains $4 \mu\text{g}$ genomic DNA and 2.5, 10 and 40 pg plasmid DNA.

RNA EXTRACTION

(Sam Sun personal communication)

Numerous RNA extraction protocols were used, most with satisfactory results, as judged by the integrity of ribosomal RNA bands on denaturing gels. Disruption and denaturing of plant tissue in GuCN, followed by purification through a CsCl pad was found to be particularly effective, and relatively easy:

Using 5 to 100g of fresh tissue, ca. 2cm sections were cut directly into liquid N₂. These were ground thoroughly under N₂ in pre-chilled mortar and pestle, then transferred to a bottle pre-chilled to -20°. After allowing the powder to equilibrate at -20° for 30', 2 volumes (or a minimum of 30 ml) grinding buffer was added to the powder. Tissue was homogenized with ten bursts at top speed for 20" each (Brinkman PT2000 Polytron), keeping bottle on ice. The resulting slurry was filtered through miracloth into centrifuge bottles, wringing the miracloth to maximize recovery of the solution. This solution was centrifuged at 27,000 g, 15', 4°, then the supernatant filtered through fresh miracloth. 30ml filtrate was layered onto a 5ml CsCl pad, and centrifuged 26k, 16hr, 15° in an SW28 rotor. The supernatant was removed by aspiration, then the tube inverted and

allowed to drain for 5'. The tube bottom was cut off (at seam line) and 0.5ml 6M Gu-HCl buffer added to each pellet, which were allowed to resuspend on ice for 1hr, then pipetted up and down to resuspend, and transferred to 40ml Nalgene centrifuge tubes. Each tube bottom was rinsed with 0.5ml 6M Gu-HCl buffer 2X, and the rinse added to 40ml tube. Tubes were swirled at RT for 5' to thoroughly resuspend all RNA, then centrifuged 15k, 15' at 4° in SS34 rotor to pellet fines. The supernatant was removed and mixed with 2.5 volumes EtOH, then centrifuged 15k, 15', 4° in SS34. The supernatant was removed and the pellet dispersed in 5ml RT EtOH, then centrifuged 15k, 15', 4° in SS34. The supernatant was removed, and the pellet resuspended in 3ml RNase-free H₂O overnight at 4°. The resuspended RNA was mixed gently with 9ml 4M NaOAc, pH6.0, then incubated on ice 5', and centrifuged 15k, 30', 4° in SS34. The resulting pellet was resuspended in 3ml RNase-free H₂O, then mixed with 0.1 volume 1M NaOAc, pH5.0 and 2.5 v EtOH, then centrifuged 15k, 30', 4° in SS34. This pellet was washed 2X with 70% EtOH, then allowed to air dry briefly and resuspended in RNase-free H₂O.

This protocol resulted in a 2.3mg RNA yield from 5g tissue.

Grinding buffer, for 200ml

100g guanidine thiocyanate
1g Na lauroyl sarcosine
5ml 1M Na citrate, pH7.0
2ml β ME

dissolve all but β ME in sterile H₂O, adjust pH to 7.0 with NaOH. Filter through 0.4 μ filter, store at RT in dark bottle. Add β ME immediately before use.

CsCl pad

47.985g CsCl
1.861g EDTA (disodium salt)
dissolve in H₂O to 50ml final v, adjust to pH7.0 with NaOH, treat with 0.1% DEPC and autoclave

1M Na citrate

19.21g citric acid
dissolve in 60ml H₂O
adjust to pH7.0 with NaOH
adjust final volume to 100 ml, treat with 0.1% DEPC and autoclave

6M Gu-HCl buffer

0.32g EDTA (disodium salt)
0.07g NaOAc
dissolve in 25ml H₂O
adjust to pH6.5 with HOAc/NaOH
treat with 0.1% DEPC, autoclave
add 28.66g Gu-HCL
readjust pH to 6.5, adjust volume to 50ml
filter through 0.4 μ filter

4M NaOAc, pH6.0

16.41g NaOAc
dissolve in 40ml H₂O
adjust pH to 6 with HOAc
adjust volume to 50ml
treat with 0.1% DEPC and autoclave

SMALL-SCALE RNA ISOLATION

(Sunitha Midha personal communication)

Ca. 100-200 mg leaf (or other) tissue is cut into liquid N₂ in "Kontes" 1.5 ml microfuge tubes (Kontes), using flamed scissors to cut very small pieces.

With a Kontes pestle precooled in N₂, the tissue is ground (we use a 1/4" power drill at low speed) 1-2', until finely powdered.

250 μ l extraction buffer & 250 μ l phenol are added, both prewarmed to 80°. The sample is vortexed 30", mixed with 250 μ l Sevags, then vortexed again.

The slurry is spun 5' RT in microcentrifuge, then 250-300 μ l of the aqueous phase is pipetted to a fresh tube, and mixed with an equal volume 4 M LiCl. The RNA is allowed to precipitate at 4° ON.

The precipitate is pelleted by centrifugation at top speed in a microcentrifuge for 30' at 4°, washed with 100 μ l 100% EtOH, and centrifuged again 5'.

The pellet is decanted and dried in a speed-vac 1', then resuspended in 250 μ l DEPC water.

RNA is precipitated again with 0.1 vol 3M NaOAc, pH 5.5, 2 vol EtOH, then centrifuged 30'. After washing with 100% EtOH and brief drying, the pellet is resuspended in DEPC H₂O.

Typical yields are 25-50 μ g.

Extraction buffer:

0.1 M LiCl
0.1 M trisHCl, pH 8.0
0.1 M EDTA
1% SDS

ISOLATION OF POLY-A PLUS RNA

Poly-A plus RNA was selected by two rounds of oligo-dT chromatography, essentially as described by Sambrook et al (1989).