Unraveling the function of bacterial-type phosphoenolpyruvate carboxylase in vascular plants

by

Michael Kien Yin Ting

A thesis submitted to the Department of Biology
in conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
September 2016

Copyright © Michael Kien Yin Ting, 2016
Abstract

Two distinct phosphoenolpyruvate carboxylase (PEPC) isozymes occur in vascular plants and green algae: plant-type PEPC (PTPC) and bacterial-type PEPC (BTPC). PTPC polypeptides typically form a tightly regulated cytosolic Class-1 PEPC homotetramer. BTPCs, however, appear to be less widely expressed and to exist only as catalytic and regulatory subunits that physically interact with co-expressed PTPC subunits to form hetero-octameric Class-2 PEPC complexes that are highly desensitized to Class-1 PEPC allosteric effectors. Yeast two-hybrid studies indicated that castor plant BTPC (RcPPC4) interacts with all three *Arabidopsis thaliana* PTPC isozymes, and that it forms stronger interactions with AtPPC2 and AtPPC3, suggesting that specific PTPCs are preferred for Class-2 PEPC formation. In contrast, Arabidopsis BTPC (AtPPC4) appeared to interact very weakly with AtPPC2 and AtPPC3, suggesting that BTPCs from different species may have different physical properties, hypothesized to be due to sequence dissimilarities within their ~10 kDa intrinsically disordered region. Recent RNA-seq and microarray data were analyzed to obtain a better understanding of *BTPC* expression patterns in different tissues of various monocot and dicot species. High levels of *BTPC* transcripts, polypeptides and Class-2 PEPC complexes were originally discovered in developing castor seeds, but the analysis revealed a broad range of diverse tissues where abundant *BTPC* transcripts are also expressed, such as the developing fruits of cucumber, grape, and tomato. Marked *BTPC* expression correlated well with the presence of ~116 kDa immunoreactive BTPC polypeptides, as well as Class-2 PEPC complexes in the immature fruit of cucumbers and tomatoes. It is therefore hypothesized that in vascular plants BTPC and thus Class-2 PEPC complexes maintain anaplerotic PEP flux in tissues with elevated malate levels that would potently inhibit ‘housekeeping’ Class-1 PEPCs. Elevated levels of malate can be used by
biosynthetically active sink tissues such as immature tomatoes and cucumbers for rapid cell expansion, drought or salt stressed roots for osmoregulation, and developing seeds and pollen as a precursor for storage lipid and protein biosynthesis.
Co-Authorship

All chapters were co-authored with Dr. W. C. Plaxton, and performed within the Plaxton lab. Chapter 3 was in collaboration with Dr. Wayne Snedden (Queen’s University Biology Dept.) who provided guidance for the molecular work and the reagents for the success of the yeast two-hybrid study.
Acknowledgements

After finishing my Biochemistry degree at the University of Calgary, I decided to travel the world for a few years. Being away from the lab setting, I began to realize how important science was for me, and knew I was meant to continue on with research. Unfortunately, being away from the lab environment had undoubtedly made my lab skills a bit rusty. I’m extremely grateful towards Dr. W. C. Plaxton who took a chance by accepting me as a graduate student. He has given me great guidance and has been an exceptional supervisor during my MSc. He encourages the best out of his students and helps them to thoroughly develop their critical thinking. I would also like to thank all the members of the Plaxton who have helped me grow as a researcher. Special recognition goes to Eric Fedosejevs and Howard Teresinski who have been wonderful colleagues for both support and insightful scientific conversations. Next, I would like to thank my committee members, Dr. Kenton Ko and Dr. Wayne Snedden for their inputs in moving my project in the proper direction. Wayne has particularly been a fantastic person to turn to for both guidance and moral support and has been like a second supervisor for me. Special acknowledgment goes to Dr. Tomas Babak who taught me proper bioinformatics analysis. Learning how to code from scratch was incredibly difficult but the long hours spent learning was well worth it, especially since the transcriptomic data really helped to bring together my thesis. Data analysis was made possible thanks to the computing power of the High Performance Computing Virtual Laboratory (HPCVL). Finally, I’d like to thank Dr. Jaqueline Monaghan for the use of her Bio-Rad imaging system, and all the little tips that have made numerous experiments so much simpler and easier. Conversations (and beers!) with Jacqueline are inspirational and constantly remind me why it is that I love science.
# Table of Contents

Abstract .......................................................................................................................... ii
Co-Authorship ................................................................................................................... iv
Acknowledgements ............................................................................................................ v
Table of Contents .............................................................................................................. vi
List of Figures .................................................................................................................. viii
List of Tables .................................................................................................................... x
List of Abbreviations ........................................................................................................ xi

Chapter 1 . Literature Review and Introduction ............................................................... 1
    PEPC function and regulation in vascular plants ......................................................... 1
    BTPC function and regulation in vascular plants ....................................................... 3
    Thesis objectives ........................................................................................................... 9

Chapter 2 . Analysis of bacterial-type PEP carboxylase expression in vascular plants indicates a widespread anaplerotic role in diverse sink tissues that accumulate high levels of malate ...... 10
    Abstract ..................................................................................................................... 10
    Introduction ............................................................................................................... 11
    Materials and Methods ............................................................................................. 15
    Results ....................................................................................................................... 19
        Identification and analysis of vascular plant BTPC genes ........................................ 19
        BTPC transcriptomic analysis: RNA-seq and microarray data ............................... 20
    Discussion .................................................................................................................. 23
Chapter 3 . Biochemical and molecular studies of AtPPC4, the enigmatic bacterial-type phosphoenolpyruvate carboxylase of *Arabidopsis thaliana* ................................................................. 56

Abstract .................................................................................................................................. 56

Introduction .................................................................................................................................. 57

Materials and methods ............................................................................................................... 60

Results and discussion ............................................................................................................... 66

AtPPC4 does not co-immunopurify with AtPPC3 from Arabidopsis flower bud extracts ... 66

Heterologous expression of *AtPPC4* in *E.coli* ........................................................................ 67

*AtPPC4* overexpression in Arabidopsis suspension cell cultures and *in planta* ............... 68

Yeast two-hybrid analysis of the ability of AtPPC4 and RcPPC4 to interact with the

Arabidopsis PTPC isozymes AtPPC1-AtPPC3 ......................................................................... 69

Chapter 4 . General Discussion ............................................................................................. 78

Diverse functions of BTPC in vascular plants ........................................................................ 78

Future directions ....................................................................................................................... 82

Literature Cited .......................................................................................................................... 86

Appendix A: List of Primers .................................................................................................... 100

Appendix B: RNA-seq/microarray experiments and RNA-seq accession numbers............. 102

Appendix C: Expression and purification of recombinant AtPPC4 ........................................ 106
List of Figures

Figure 1.1. Regulation of the PEP branch point in plant primary metabolism................................. 2

Figure 1.2. Recruitment of Class-2 PEPC to outer mitochondrial membrane (OMM). .................... 6

Figure 2.1. Model of alternate pathways supporting fatty acid synthesis in leucoplasts of
developing COS.................................................................................................................................. 28

Figure 2.2. Analysis of genomic structures of various BTPC genes. ................................................ 30

Figure 2.3. Partial alignment of deduced BTPC polypeptides............................................................ 31

Figure 2.4. Alignment of the intrinsically disordered region (IDR) of deduced BTPC
polypeptides. ..................................................................................................................................... 32

Figure 2.5. Ricinus communis (castor) BTPC (RcPPC4) expression. ............................................. 33

Figure 2.6. Arabidopsis thaliana BTPC (AtPPC4) expression............................................................. 35

Figure 2.7. Brassica rapa (field mustard) BTPC expression............................................................... 36

Figure 2.8. Fragaria vesca (strawberry) BTPC expression................................................................. 37

Figure 2.9. Zea mays (maize) BTPC expression.................................................................................. 38

Figure 2.10. Populus trichocarpa (poplar tree) BTPC1 and BTPC2 expression................................. 40

Figure 2.11. Oryza sativa (rice) BTPC expression............................................................................... 41

Figure 2.12. Persea americana (avocado) BTPC expression.............................................................. 42

Figure 2.13. Gossypium raimondii (cotton) BTPC expression............................................................ 43

Figure 2.14. Glycine max (soybean) BTPC1-BTPC3 expression......................................................... 44

Figure 2.15. Cucumis sativus (cucumber) BTPC expression.............................................................. 45

Figure 2.16. Vitis vinifera (grape) BTPC expression........................................................................... 47

Figure 2.17. Solanum lycopersicum (tomato) BTPC expression........................................................ 48

Figure 2.18. Citrus sinensis (orange) BTPC expression....................................................................... 49
Figure 2.19. *Manihot esculenta* (cassava) *BTPC1* and *BTPC2* expression. ........................................ 50

Figure 2.20. Immunodetection of BTPC from plant tissues with abundant *BTPC* transcripts. .... 51

Figure 2.21. Co-IP of BTPC from cucumber and tomato extracts using anti-RcPTPC. ............ 52

Figure 2.22. Detection of active Class-2 PEPC in tomato fruit extracts via non-denaturing PAGE. ........................................................................................................................................ 53

Figure 2.23. Model of the mechanisms that control malate levels in a plant cell.................... 55

Figure 3.1. Protein quantification of the four Arabidopsis PEPC isozymes from various tissues. ........................................................................................................................................ 72

Figure 3.2. Attempted co-IP of AtPPC4 from Arabidopsis flower buds. ................................. 73

Figure 3.3. Attempted co-purification of recombinant AtPPC4 using His-tagged AtPPC3 via metal ion affinity chromatography................................................................. 74

Figure 3.4. *AtPPC4* overexpression in Arabidopsis suspension cell cultures. .................... 75

Figure 3.5. *AtPPC4* overexpression in planta........................................................................ 76

Figure 3.6. Yeast two-hybrid analysis of Arabidopsis versus castor BTPC interactions with various Arabidopsis PTPC isozymes. ........................................................................ 77

Figure 4.1. Model of BTPC and thus Class-2 PEPC function in osmotically stressed plants..... 84

Figure 4.2. Model proposing diverse functions for Class-2 PEPC in tissues with elevated levels of malate................................................................................................................ 85
List of Tables

Table 2.1. List of all vascular plants with published $BTPC$ expression data............................... 29

Table 2.2. Estimated intracellular malate concentration of various plant tissues ......................... 54
List of Abbreviations

AAT   aspartate amino transferase
ABA   abscisic acid
ABRE  ABA responsive cis-acting element
ATPase Adenosine triphosphatase
ATP-PFK ATP-dependent phosphofructokinase
AtBTPC Arabidopsis thaliana bacterial-type PEPC (AtPPC4)
AtPPC1 Arabidopsis thaliana plant-type PEPC gene - At1g53310
AtPPC2 Arabidopsis thaliana plant-type PEPC gene - At2g42600
AtPPC3 Arabidopsis thaliana plant-type PEPC gene - At3g14940
AtPPC4 Arabidopsis thaliana bacterial-type PEPC gene - At1g68750
BAM   Binary Alignment Map format
BAR   Bio-Analytic Resource database (available online from University of Toronto)
BCIP  5-bromo-4-chloro-3-indolylphosphate
BTPC  Bacterial-type PEPC
BY-2  bright yellow-2 tobacco cells
CA    carbonic anhydrase
CAM   crassulacean acid metabolism
CDPK  calcium dependent protein kinase
Co-A  coenzyme A
Co-IP  co-immunoprecipitation
COS   castor oil seed
DAP   days after pollination
DO dropout supplement (lacking an essential amino acid)
DDO double dropout supplement (lacking leucine and tryptophan)
DPA days post-anthesis
DPDS 2,2′-dipyridyl disulphide
DTT dithiothreitol
ECL enhanced chemiluminescence
EDTA ethylenediaminetetraacetic acid
eFP electronic fluorescent pictograph
EGTA (ethylene-dioxy)diethylenedinitrolotetracetic acid
FPKM fragments per kilobase of transcript per million mapped reads
FW fresh weight
GABA γ-Aminobutyric acid
Glu-6-P glucose-6-phosphate
HPLC high performance liquid chromatography
$\textit{IC}_{50}$ inhibitor concentration yielding 50% inhibition of enzymatic activity
IDR intrinsically disordered region
IgG immunoglobulin G
IPTG isopropyl β-D-1-thiogalactopyranoside
kDa kilodalton
LiAc lithium acetate
MDH malate dehydrogenase
ME malic enzyme
MS Murashige Skoog media
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr</td>
<td>relative molecular mass</td>
</tr>
<tr>
<td>NADH</td>
<td>nicotinamide adenine dinucleotide (reduced)</td>
</tr>
<tr>
<td>NADPH</td>
<td>nicotinamide adenine dinucleotide phosphate (reduced)</td>
</tr>
<tr>
<td>NADP-ME</td>
<td>NADP-malic enzyme</td>
</tr>
<tr>
<td>NBT</td>
<td>nitro blue tetrazolium</td>
</tr>
<tr>
<td>OAA</td>
<td>oxaloacetate</td>
</tr>
<tr>
<td>OD#</td>
<td>absorbance of light at # nm wavelength</td>
</tr>
<tr>
<td>OMM</td>
<td>outer mitochondrial membrane</td>
</tr>
<tr>
<td>p118</td>
<td>118 kDa polypeptide</td>
</tr>
<tr>
<td>p116</td>
<td>116 kDa polypeptide</td>
</tr>
<tr>
<td>p110</td>
<td>110 kDa polypeptide</td>
</tr>
<tr>
<td>p107</td>
<td>107 kDa polypeptide</td>
</tr>
<tr>
<td>PA</td>
<td>phosphatidic acid</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PDC</td>
<td>pyruvate dehydrogenase complex</td>
</tr>
<tr>
<td>PEP</td>
<td>phosphoenolpyruvate</td>
</tr>
<tr>
<td>PEPC</td>
<td>phosphoenolpyruvate carboxylase</td>
</tr>
<tr>
<td>PEPCK</td>
<td>phosphoenolpyruvate carboxylase kinase</td>
</tr>
<tr>
<td>PK</td>
<td>pyruvate kinase</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>PPCK</td>
<td>Ca$^{2+}$ independent PEPC kinase</td>
</tr>
<tr>
<td>PPDK</td>
<td>pyruvate orthophosphate dikinase</td>
</tr>
<tr>
<td>PPI</td>
<td>pyrophosphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PPiase</td>
<td>pyrophosphatase</td>
</tr>
<tr>
<td>PPi-PFK</td>
<td>pyrophosphate-dependent phosphofructokinase</td>
</tr>
<tr>
<td>PTPC</td>
<td>plant-type PEPC</td>
</tr>
<tr>
<td>PVDF</td>
<td>polyvinylidene difluoride</td>
</tr>
<tr>
<td>PVPP</td>
<td>polyvinyl(polypyrrolidone)</td>
</tr>
<tr>
<td>QDO</td>
<td>quadruple drop out supplement (no leucine, tryptophan, adenine or histidine)</td>
</tr>
<tr>
<td>RcBTPC</td>
<td><em>Ricinus communis</em> bacterial-type PEPC</td>
</tr>
<tr>
<td>RcPPC4</td>
<td><em>Ricinus communis</em> bacterial-type PEPC gene or transcripts</td>
</tr>
<tr>
<td>RcPPC3</td>
<td><em>Ricinus communis</em> plant-type PEPC gene or transcripts</td>
</tr>
<tr>
<td>RMA</td>
<td>robust multi-array average</td>
</tr>
<tr>
<td>ROS</td>
<td>reactive oxygen species</td>
</tr>
<tr>
<td>RPKM</td>
<td>reads per kilo base transcript per million reads</td>
</tr>
<tr>
<td>SAM</td>
<td>sequence alignment map format</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SRA</td>
<td>sequence read archive</td>
</tr>
<tr>
<td>TCA</td>
<td>tricarboxylic acid</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N,N-tetramethyl ethylenediamine</td>
</tr>
<tr>
<td>UofT</td>
<td>University of Toronto</td>
</tr>
</tbody>
</table>
Chapter 1. Literature Review and Introduction

PEPC function and regulation in vascular plants

Phosphoenolpyruvate carboxylase (PEPC) is a tightly regulated enzyme of vascular plants and green algae, situated at a crucial branch point in primary metabolism that catalyzes the irreversible β-carboxylation of phosphoenolpyruvate (PEP) to form oxaloacetate (OAA) and inorganic phosphate (Pi) (O’Leary et al. 2011b). PEPC and pyruvate kinase (PK) play a prominent role in regulating plant glycolysis and respiration (O’Leary et al. 2011b). In C₄ and Crassulacean acid metabolism (CAM) plant leaves, PEPC has been well characterized due to its essential role in atmospheric CO₂ assimilation during photosynthesis (Izui et al. 2004). These PEPCs are classified as photosynthetic (C₄ PEPC) and exhibit higher kinetic efficiency than non-photosynthetic PEPC (C₃ PEPC), with reduced sensitivity to allosteric inhibition (Paulus et al. 2013). C₃ PEPCs are found in C₃ leaves and non-photosynthetic tissues, and are primarily responsible for the anaplerotic replenishment of tricarboxylic acid (TCA) cycle intermediates that have been consumed for biosynthesis and nitrogen assimilation (Fig. 1.1) (O’Leary et al. 2011b). In addition, PEPC has been implicated in a variety of other key processes such as carbon storage, energy production, abiotic stress acclimation, seed germination, seed development, and cell expansion (O’Leary et al. 2011b).
Figure 1.1. Regulation of the PEP branch point in plant primary metabolism. PEPC and PK play a prominent role in regulating plant glycolysis and respiration by consuming PEP to relieve feedback inhibition of ATP-PFK/PPi-PFK. Malate derived from PEPC can be transported into the mitochondria as respiratory substrate for energy production, but is typically used for a variety of other key functions. Red dashed lines represent feedback inhibition whereas green dashed lines represent allosteric activation. Abbreviations: ATP-PFK, ATP-dependent phosphofructokinase; PPI-PFK, pyrophosphate dependent phosphofructokinase; PEPC, phosphoenolpyruvate carboxylase; PKc, cytosolic pyruvate kinase; AAT, aspartate amino transferase; MDH, malate dehydrogenase; PDC, pyruvate dehydrogenase complex; Fru-6-P, fructose-6-phosphate; Fru-1,6-P2, fructose-1,6-bisphosphate. Figure adapted from O’Leary et al. (2011b).
Plant PEPCs belong to a small multigene family that contain several closely related
*plant-type* PEPC (*PTPC*) encoding ~105 kDa polypeptides that typically assemble to form a
homotetramer known as Class-1 PEPCs (O’Leary et al. 2011b). PEPCs are generally localized to
the cytosol, but there has been one report of a unique chloroplast targeted Class-1 PEPC in rice
(Masumoto et al. 2010). Class-1 PEPCs have been well characterized and are post-translationally
regulated in numerous ways. Glucose-6-phosphate (Glu-6-P) is an allosteric activator, while
malate, aspartate, and glutamate are allosteric inhibitors (O’Leary et al. 2011b). In addition, all
PTPC subunits contain a conserved serine residue near their N-terminus that activates Class-1
PEPCs upon phosphorylation by a dedicated, Ca\(^{2+}\)-independent, PTPC protein kinase (PPCK),
which has been shown to increase sensitivity to allosteric activators, while decreasing sensitivity
to inhibitors such as malate (O’Leary et al. 2011b). In addition, monoubiquitination of a
conserved lysine residue near the PTPC active site has been shown to interfere with substrate
binding while enhancing sensitivity to inhibitors such as malate and aspartate (Uhrig et al. 2008).
It appears that monoubiquitination serves to attenuate Class-1 PEPC activity until flux through
PEPC is required (Uhrig et al. 2008, Shane et al. 2013). Interestingly, anionic phospholipids such
as phosphatidic acid (PA) specifically inhibit C\(_{4}\) PEPCs (Monreal et al. 2010). This requires
further clarification but it offers an interesting new element to PEPC regulation.

**BTPC function and regulation in vascular plants**

All plant and green algal genomes also contain a *bacterial-type* PEPC (*BTPC*) gene that
is more closely related to prokaryotic PEPC genes (O’Leary et al. 2011b). Vascular plant *BTPC*
genes encode polypeptides of ~116-118 kDa that are classified by three main criteria: 1) the C-
terminal tetrapeptide is (R/K)NTG, a feature that is typical of prokaryotic PEPCs; 2) the lack of
the distinct N-terminal serine phosphorylation motif (acid-base-XX\textsubscript{p}SIDAQLR) present in all PTPCs; and 3) an intrinsically disordered region (IDR) of approximately 10 kDa (O’Leary et al. 2011b). Significant BTPC expression appears to be limited to specific tissues and to vary among different plant species. In the model plant Arabidopsis thaliana, BTPC was reported to be exclusively expressed in developing pollen and salt stressed roots (Sanchez and Cejudo 2003, Sánchez et al. 2006, Igawa et al. 2010) whereas in castor oil plants (Ricinus communis), BTPC transcripts are abundant in the developing seeds and immature leaves (Gennidakis et al. 2007, O’Leary et al. 2011a). In addition, BTPC appears to exist only as a catalytic and regulatory subunit of novel hetero-octameric Class-2 PEPC complexes, which have thus far only been shown to occur in green algae, developing castor oil seeds (COS), expanding castor leaves, and developing lily pollen (Rivoal et al. 2001, Blonde and Plaxton 2003, O’Leary et al. 2009, 2011a, Igawa et al. 2010). Class-2 PEPC exhibits biphasic PEP saturation kinetics since both PTPC and BTPC subunits are catalytically active (O’Leary et al. 2009). Although BTPC subunits have a lower affinity for PEP, they exhibit higher maximal velocity ($V_{\text{max}}$) and are desensitized to allosteric effectors such as malate and aspartate, relative to PTPC (O’Leary et al. 2009). Class-2 PEPC has thus been hypothesized to act as a 'metabolic overflow mechanism' to allow continued production of malate (derived from PEP) in biosynthetically active tissues, which could otherwise reach concentrations that would inhibit Class-1 PEPC whose half maximal inhibitory concentration ($I_{50}$) for malate are typically very low (i.e. $<0.2$ mM) (O’Leary et al. 2011b). This is especially apparent in developing castor endosperm, where in vivo malate concentrations (i.e. ~5 mM) are up to 80 fold higher than the $I_{50}$ value of COS Class-1 PEPC (O’Leary et al. 2011b). The seed specific expression of a malate insensitive prokaryotic PEPC (i.e. corynebacterium PEPC) in Vicia narbonensis altered metabolic fluxes from sugars/starch into
organic and amino acids (Rolletschek et al. 2004), suggesting that the malate insensitive BTPC may function similarly by promoting anaplerotic carbon flow. Although COS Class-1 PEPC localizes diffusely throughout the cytosol, immunogold labelling of developing COS, as well as imaging of fluorescent protein-tagged fusions of castor PTPC and BTPC transiently expressed in tobacco Bright Yellow-2 (BY-2) cells, revealed that castor Class-2 PEPC dynamically associates with the outer mitochondrial membrane (OMM) (Park et al. 2012). It was concluded that: 1) BTPC largely localizes to the OMM while PTPC localizes to the cytosol; 2) upon their co-expression in tobacco BY-2 cells, castor BTPC can recruit castor PTPC onto the OMM to form a Class-2 PEPC; and 3) BTPC’s unique IDR mediates its association with PTPC. The unique kinetic and regulatory properties of Class-2 PEPC, and dynamic subcellular targeting to the mitochondrial surface support the hypothesis that it facilitates rapid refixation of respired CO₂ while sustaining a large anaplerotic flux to replenish TCA cycle C-skeletons. This is especially important in developing seeds and pollen because of high respiration rates that result in increased CO₂ production. If CO₂ concentrations were allowed to steadily rise, mitochondrial respiration and thus seed development would cease. A possible mechanism around this inhibition, especially in bulky tissues such as developing COS where gas exchange with the atmosphere is likely severely restricted, is the conversion of CO₂ into HCO₃⁻ by mitochondrial carbonic anhydrase (CA) (Braun and Zabaleta 2007, Lee et al. 2011), and its refixation into OAA by mitochondrial-associated Class-2 PEPC (as illustrated in Fig. 1.2). This synergy between CA and Class-2 PEPC to refix respired CO₂ would prevent respiratory inhibition and simultaneously provide a key co-substrate for Class-2 PEPC (HCO₃⁻) for generation of biosynthetic precursors needed for storage oil and protein production.
Figure 1.2. Recruitment of Class-2 PEPC to outer mitochondrial membrane (OMM).
Association of Class-2 PEPC to OMM may prevent respiratory inhibition from accumulated CO₂ in bulky tissues such as developing castor beans, where atmospheric gas exchange is likely severely restricted. $I_{50}$ (malate) values are derived from O’Leary et al (2011c) and Blonde and Plaxton (2003).

1) 4 BTPC subunits rapidly recruit 4 PTPC subunits to form Class-2 PEPC. The association of the malate insensitive BTPC subunit also desensitizes the associated PTPC subunits of Class-2 PEPC to malate inhibition.

2) Class-2 PEPC is recruited to the OMM, an association mediated by the BTPC subunits. 3) Respired CO₂ from the mitochondria is converted to HCO₃⁻ by mitochondrial carbonic anhydrase, and used as substrate by mitochondrial-associated Class-2 PEPC for OAA production.

4) OAA is converted to aspartate via aspartate amino transferase (AAT) or converted to malate via malate dehydrogenase (MDH) for storage protein and storage oil biosynthesis, respectively.
Interestingly, the addition of KCN (a respiratory inhibitor) resulted in a significant in vivo repartitioning of BTPC from the OMM into the cytosol of tobacco BY-2 cells (Park et al. 2012). This pattern follows a similar trend observed with mitochondria derived from Arabidopsis suspension cell cultures, wherein KCN addition resulted in the repartitioning of glycolytic enzymes from the OMM to the cytosol (Graham et al. 2007). Numerous cytosolic glycolytic isozymes were shown to interact with the OMM to form a metabolon in which intermediates can be channeled between active sites of sequential enzymes in the complex (Giegé et al. 2003, Graham et al. 2007). One potential mode of recruitment is PA synthesis (via phospholipase C and diacylglycerol kinase or phospholipase D), which is activated upon various stresses to act as a lipid second messenger, or to recruit proteins to membranes (Testerink et al. 2004). Membrane PA composition (such as in the OMM) is generally low but rapidly and transiently accumulates in lipid bilayers following biotic and abiotic stresses (Mcloughlin et al. 2013). Interestingly, both PEPC and glyceraldehyde 3-phosphate dehydrogenase, derived from tomato and Arabidopsis, were purified using PA-affinity chromatography, and the affinity of PEPC for PA appeared to increase following osmotic stress treatment (Monreal et al. 2010, Mcloughlin et al. 2013). In addition, analysis of the Arabidopsis PTPC isozyme, AtPPC3, revealed that it contains an endophilin-A1 BAR domain, which is known to bind anionic phospholipids such as PA (Monreal et al. 2010). These domains may allow recruitment of PEPC to membranes following PA synthesis, which has been documented for C₄ PEPC from sorghum leaves (Monreal et al. 2010). Since glyceraldehyde 3-phosphate dehydrogenase is one of the glycolytic enzymes that dynamically associates with the OMM (Graham et al. 2007), it will be interesting to determine the relation between PEPC and glyceraldehyde 3-phosphate dehydrogenase, since their co-elution from a PA-affinity column may signify that the two proteins are associated in some way
(Mcloughlin et al. 2013). Currently, more clarification is needed for the possible metabolon formation of BTPC with other glycolytic enzymes on the OMM.

Information regarding the regulation of BTPC is not as extensive as what's known for PTPC. Although BTPC lacks the activating N-terminal serine phosphorylation site present in all PTPCs, studies from green algae show that Class-2 PEPC is predominantly regulated by inhibitory phosphorylation at a different site (Rivoal et al. 2002b). In developing castor endosperm, Ser$^{425}$ and Ser$^{451}$ have been identified as novel in vivo BTPC phosphorylation sites, and that phosphorylation at these sites is inhibitory (O’Leary et al. 2011c, Dalziel et al. 2012, Hill et al. 2014). Although the kinase responsible for phosphorylating COS BTPC at Ser$^{425}$ remains unknown, Hill and co-workers (2014) purified and characterized the native COS kinase responsible for phosphorylating castor BTPC at Ser$^{451}$ and provided kinetic evidence that it belongs to the castor calcium-dependent protein kinase (CDPK) family. It was subsequently identified as RcCDPK1, one of 20 predicted castor CDPK isozymes (Ying et al. 2016). Ser$^{451}$ of castor BTPC and residues flanking this phosphosite are highly conserved in other BTPCs, including Arabidopsis BTPC (AtPPC4; At1G68750), suggesting this is a universal mechanism of BTPC post-translational control in vascular plants (O’Leary et al. 2011c, Dalziel et al. 2012). This is supported by the fact that recombinant AtCPK4 and AtCPK11 (Arabidopsis orthologs of RcCDPK1) effectively phosphorylated recombinant castor BTPC at Ser$^{451}$ (Ying et al. 2016), suggesting that Arabidopsis BTPC may be a target for one of these kinases in developing Arabidopsis pollen where AtBTPC, AtCPK4 and AtCPK11 are all highly expressed (Harper et al. 2004, Igawa et al. 2010).
Thesis objectives

Research done with castor bean BTPC and Class-2 PEPC complexes has provided an excellent working model for the function of Class-2 PEPC in vascular plants. However, the lack of characterized BTPCs (and Class-2 PEPC complexes) from other plant species has made it difficult to assign a universal function for this conserved protein. In addition, the limited BTPC expression data that’s currently available shows variable expression patterns in different plants, suggesting that there may be alternate tissue-specific functions in different plant species. The questions that drive this research are: In what tissues and developmental stages is BTPC commonly expressed, and what is the reason for the marked differences in AtBTPC expression compared to COS BTPC? This thesis addresses these two questions to expand our knowledge of the function of BTPC in vascular plants. Recent advances in next generation sequencing have made it possible to analyze BTPC expression from various plant tissues of varying developmental stages. My analysis has revealed novel plants and tissues where BTPC was not expected to be expressed. In addition, the previous problems with recombinant production of Arabidopsis BTPC (AtBTPC) in Escherichia coli (E. coli) have made it difficult to characterize (Gennidakis et al. 2007). Currently, there is no evidence that AtBTPC is a functional PEPC that is capable of forming a Class-2 PEPC complex. Overexpression of AtBTPC in planta could help to clarify its role and regulation in Arabidopsis, which has previously been implicated in pollen development and salinity stress tolerance (Sánchez et al. 2006, Wang et al. 2012).
Chapter 2. Analysis of bacterial-type PEP carboxylase expression in vascular plants indicates a widespread anaplerotic role in diverse sink tissues that accumulate high levels of malate


In preparation for submission to *New Phytologist*

**Abstract**

PEP carboxylase plays an essential role in all plant cells, particularly the anaplerotic replenishment of TCA cycle intermediates withdrawn for biosynthesis and N-assimilation. Most PEPCs studied to date exist as tightly regulated Class-1 PEPC homotetramers composed of identical PTPC subunits. The discovery in green algae and then developing castor beans of unusual, allosterically-desensitized Class-2 PEPC heteromeric complexes composed of tightly associated PTPC and BTPC subunits added a further layer of complexity to the physiological functions and control of this important enzyme. The abundance of BTPC/Class-2 PEPC complexes during the major oil accumulating stages of COS development suggests that BTPC plays an important role in maintaining PEP metabolism to support fatty acid synthesis in non-green oilseeds. Unfortunately, BTPC has not yet been documented in other oilseeds, and the limited data available for *BTPC* expression suggests alternative functions in different plant species. To clarify the function of BTPC in vascular plants, the current study analyzed publicly...
available RNA-seq and microarray datasets and established two distinct patterns of BTPC expression in vascular plants. Species such as Arabidopsis thaliana, field mustard, strawberry, rice, maize, and poplar mainly exhibit pollen-specific BTPC expression. By contrast, BTPC transcripts were particularly abundant in a variety of non-pollen heterotrophic sinks of several dicot species. Apart from immature castor seeds, these included developing cotton, soybean, and cucumber embryos, as well as the developing fruits of avocado, grape, and tomato. Immunoreactive ~118 kDa BTPC polypeptides were detected on immunoblots of clarified extracts of immature tomato fruit and cucumber ovary. Co-immunoprecipitation of BTPC using anti-RcPTPC antibodies demonstrated that tomato and cucumber BTPC physically interact with co-expressed PTPC subunits to form a Class-2 PEPC complex. BTPC and thus Class-2 PEPC complexes appear to be broadly expressed in diverse heterotrophic sinks that accumulate high levels of malate. I hypothesize that Class-2 PEPC has a widespread function in vascular plants to sustain rapid anaplerotic carboxylation of PEP to OAA in sink tissues whereby the corresponding Class-1 PEPC would be subject to potent allosteric inhibition by accumulated malate. In addition, the in vivo regulatory phosphorylation sites (Ser^{425} and Ser^{451}) of castor bean BTPC are conserved in all of the deduced BTPC sequences, suggesting universal post-translational regulatory elements for vascular plant BTPC.

**Introduction**

Phosphoenolpyruvate carboxylase (PEPC, EC 4.1.1.31) is a tightly controlled cytosolic enzyme that catalyses the irreversible β-carboxylation of PEP in the presence of HCO_{3}^{-} to yield OAA and Pi. C₄ and CAM photosynthetic PEPCs have been extensively characterized at the structural and regulatory levels (Izui et al. 2004). PEPC also plays essential roles in all plant
cells, particularly the anaplerotic replenishment of TCA cycle intermediates that have been withdrawn for biosynthesis and N-assimilation (O’Leary et al. 2011b). Most plant PEPCs studied to date exist as Class-1 PEPC homotetramers composed of identical PTPC subunits of ~100-110 kDa. The activity of Class-1 PEPCs is: 1) potently modulated by various allosteric effectors, particularly the feedback inhibitor malate; and 2) subject to reciprocal reversible phosphorylation (activating) and monoubiquitination (inhibiting) at highly conserved serine and lysine residues (Uhrig et al. 2008, O’Leary et al. 2011b, Shane et al. 2013). However, annotation of Arabidopsis, rice, and soybean genomes revealed that plant PEPC belongs to a small multigene family that includes several PTPC genes, along with an unexpected and enigmatic BTPC gene that had not been previously characterized (Sanchez and Cejudo 2003, Sullivan et al. 2004). Initial insights into plant BTPC function arose when novel high molecular weight Class-2 PEPC heteromeric complexes exhibiting unique kinetic and physical properties were isolated from unicellular green algae and then the oil-rich endosperm of developing COS (Rivoal et al. 1998, 2001, Blonde and Plaxton 2003, O’Leary et al. 2011b). Tryptic peptide sequencing by liquid chromatography tandem mass spectrometry identified RcPPC3 and RcPPC4 (i.e., castor PTPC and BTPC, respectively) as the subunits of COS Class-2 PEPC (Blonde and Plaxton 2003, Gennidakis et al. 2007). The unique kinetic properties of BTPC and largely restricted expression during the major oil accumulating stages of COS development, strongly suggests that BTPC and Class-2 PEPC play an important role in supporting fatty acid and storage lipid synthesis (Gennidakis et al. 2007, O’Leary et al. 2009, 2011a,b). BTPC transcripts have also been reported during the major oil accumulating stages of immature peanut (Arachis hypogaea) seeds (Yu et al. 2010, Pan et al. 2013), suggesting that BTPC is involved in the metabolism of a variety of developing oilseeds.

Plant fatty acid synthesis occurs in plastids and requires stoichiometric equivalents of
ATP, NADH, and NADPH for each acetyl addition (derived from acetyl-CoA) onto the growing fatty acid chain (Baud and Lepiniec 2010). Green photoheterotrophic oilseeds such as canola (Brassica napus) can generate the ATP and reducing power within their plastids via photosynthesis, whereas non-green oilseeds such as castor, maize, peanuts, and sunflower must generate reducing power and ATP through alternative mechanisms (Baud and Lepiniec 2010). One possibility is through the oxidation of PEPC-derived malate via the plastidial isozymes of NADP-malic enzyme (ME_p) and pyruvate dehydrogenase complex (PDC_p) which generates all the acetyl-CoA, NADPH and NADH needed for fatty acid synthesis (Fig. 2.1) (Smith et al. 1992). Support for this mechanism was provided with the discoveries that maximal ME_p activity coincided with the major storage lipid accumulating phase of developing COS (Shearer et al. 2004), and that a specific malate/Pi transporter exists within the COS leucoplast envelope (Eastmond et al. 1997). Exogenous malate promoted maximal rates of fatty acid synthesis by leucoplasts isolated from developing COS, in which malate was estimated to reach upwards of 5 mM (Smith et al. 1992). This concentration of malate is very inhibitory for COS Class-1 PEPC (I_{50}(malate) ~0.07 mM) but far less so for castor BTPC (I_{50}(malate) ~11 mM) (O’Leary et al. 2009, 2011c). Furthermore, COS Class-2 PEPC associates with the outer mitochondrial envelope, an association mediated by its BTPC subunits (Park et al. 2012). The unique kinetic properties and dynamic association of Class-2 PEPC with the outer mitochondrial envelope support the hypothesis that they facilitate rapid refixation of respiratory CO_2, while sustaining a large anaplerotic flux to replenish TCA cycle C-skeletons that have been withdrawn for storage oil and protein synthesis in developing COS (O’Leary et al. 2009, 2011a,b, Park et al. 2012). This could be particularly advantageous in bulky, non-photosynthetic oilseeds such as COS that likely have limited atmospheric gas exchange. Unfortunately, further support for this hypothesis
has not been available since neither BTPC polypeptides nor Class-2 PEPC complexes have been documented in any other seeds.

The occurrence of BTPC polypeptides and Class-2 PEPC complexes has thus far only been described for two vascular plant species; i.e., developing seeds and immature leaves of castor, and developing lily pollen (Blonde and Plaxton 2003, Gennidakis et al. 2007, Igawa et al. 2010, O'Leary et al. 2011a). There are numerous reasons why the detection of BTPC polypeptides and Class-2 PEPC complexes has been restricted to so few species. Although they share conserved PEPC catalytic domains, PTPCs and BTPCs exhibit less than 40% sequence identity (Sanchez and Cejudo 2003, Gennidakis et al. 2007). Since PEPC research has been largely focused on PTPC-containing Class-1 PEPCs, most PEPC antibodies have been raised against PTPC polypeptides that don’t recognize BTPC (Gennidakis et al. 2007, Uhrig et al. 2008). Proper detection and analysis of BTPC polypeptides requires the production and use of BTPC-specific antibodies. Second, BTPC studies have generally not profiled a wide range of tissues from monocot and dicot species. Of the studies that profiled BTPC transcriptional expression, as few as four tissue sources were selected for analysis (Table 2.1). In addition, the selected tissues may not have been at the proper developmental stage during which BTPC transcripts and polypeptides accumulate. This is especially crucial for developing Arabidopsis pollen in which BTPC polypeptides were shown to exist for a couple days in the immature flower buds before disappearing (Igawa et al. 2010). Third, BTPC may also be induced under specific abiotic stress conditions, which the plants may not have been exposed to prior to analysis. This is evident from the induction of BTPC transcripts by osmotic stress in Arabidopsis roots (Sánchez et al. 2006). The recent advancements made with next-generation sequencing have made it possible to overcome the aforementioned issues by searching for BTPC transcripts.
in RNA-seq databases created from various plant tissues at varying developmental stages.

The aim of the current study was to assess BTPC expression patterns via bioinformatics analysis of publicly available RNA-seq and microarray datasets. In addition, immunoblots of soluble protein extracts from several tissues exhibiting abundant BTPC transcripts were probed using highly specific BTPC antibodies generated against RcBTPC (RcPPC4) (O’Leary et al. 2009). The results provide unexpected and novel insights into the expression patterns and potential functions of BTPC and Class-2 PEPC complexes in the plant kingdom.

Materials and Methods

Plant materials

All plants were cultivated in a greenhouse at 22-26 °C and 30-40% humidity under natural light supplemented with 16 h of artificial light. Tomatoes (Beefsteak variety) and cucumber (Marketmore variety) were supplemented every 7 d with a commercial 20-20-20 fertilizer following formation of the first fruit. Developing COS endosperm (cv. Baker 296) were staged as previously described (Greenwood and Bewley 1982) and harvested between stages V-VII (i.e. mid- to full cotyledon). Whole tomato fruits were harvested at the immature green stage (7-14 d post-anthesis (DPA)). Cucumber flowers were manually pollinated with a paint brush, and ovaries harvested within 7 d. All tissues were snap frozen in liquid N₂ and stored at -80 °C.

Identification of BTPC genes and analysis of corresponding RNA-seq datasets for BTPC expression

Although numerous plant genomes have been sequenced, our analysis was restricted to the current set of publicly available plant genome sequences having robust gene annotations. The
full length coding sequence of \textit{RcPPC4} (Genbank ID: EF634318.1) was subjected to BLAST analysis in the Joint Genome Institute (JGI) Phytozome v11.0 database to search for putative \textit{BTPC} genes. A search was then done on the Sequence Read Archive (SRA) library from NCBI (http://trace.ncbi.nlm.nih.gov/Traces/sra/) to locate all RNA-seq studies available for each respective species. Plants with abundant data for relevant tissues were further analyzed. If the authors of the RNA-seq study included annotated transcript data in the supplementary files, the reads per kilobase of gene per million reads (RPKM) or fragments per kilobase of transcript per million mapped reads (FPKM) values for \textit{BTPC} were directly taken from these files and used for analysis. If no annotated data was available, the raw reads were downloaded from the SRA and the data sets were aligned to the reference genomes (obtained from the "JGI Phytozome v11.0") using Bowtie v1.1.2 (Langmead et al. 2009). Alignments were done using restricted parameters that only reported unique single alignments. The Sequence Alignment Map (SAM) format files were subsequently converted into Binary Alignment Map (BAM) format using ‘SAMtools’ (Li et al. 2009). Alignment data was analyzed using bedtools (Quinlan and Hall 2010). The number of mapped reads to corresponding \textit{BTPC} genes were then normalized to RPKM values (formula in Appendix B).

\textit{Microarray analysis of \textit{BTPC} expression}

The Arabidopsis \textit{BTPC} gene identifier (At1g68750; \textit{AtPPC4}) was used as query sequence in the ‘expressolog tree’ at the University of Toronto (UofT) Bio-Analytic Resource (BAR) database to find all data associated with \textit{AtBTPC} and corresponding orthologs. Microarray data were visualized using available electronic fluorescent pictograph (eFP) browsers from the BAR website (Winter et al. 2007).
Protein extraction

Proteins were extracted from COS endosperm as previously described (O’Leary et al. 2011a). Quick-frozen COS was ground to a fine powder in liquid N\textsubscript{2} using a mortar and pestle, followed by the addition of 2 volumes (w/v) of extraction buffer which contained 50 mM HEPES-KOH (pH 7.5), 1 mM EDTA, 1 mM EGTA, 25 mM NaF, 1 mM Na\textsubscript{3}VO\textsubscript{4}, 1 mM Na\textsubscript{2}MoO\textsubscript{4}, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol, 10 mM MgCl\textsubscript{2}, 1% (w/v) poly(vinylpolypyrrolidone) (PVPP), 2 mM phenylmethylsulphonyl fluoride (PMSF), 2 mM 2,2'-dipyridyl disulphide (DPDS), and 5 µl ml\textsuperscript{-1} ProteCEASE-100 (G-Biosciences). Homogenates were centrifuged at 4 °C and 15,000 g for 10 min, and the supernatant filtered through 2 layers of Miracloth (EMD Millipore). Extracts were frozen in liquid N\textsubscript{2} for future use or immediately boiled in 2x Laemmli sample buffer for 5 min. Protein extractions of all tomato and cucumber tissues were done with a modified protocol as previously described (Guillet et al. 2002). The extraction buffer contained 500 mM Tris-HCl (pH 7.5), 5 mM EDTA, 5 mM EGTA, 25 mM NaF, 1 mM Na\textsubscript{3}VO\textsubscript{4}, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol, 5 mM MgSO\textsubscript{4}, 5 mM NaHCO\textsubscript{3}, 10 mM L-ascorbic acid, 1 mM DTT, 1% (w/v) PVPP, 2 mM PMSF, 2 mM DPDS and 5 µl ml\textsuperscript{-1} ProteCEASE-100 (G-Biosciences). Tissues were homogenized and clarified extracts prepared as described above.

Electrophoresis, immunoblotting, and in-gel PEPC activity staining

Non-denaturing and SDS-PAGE, and immunoblotting were performed using a Mini-PROTEAN II system (Bio-Rad) according to manufacturer’s instructions and as previously described (Rivoal et al. 2002a, Blonde and Plaxton 2003). Anti-RcPTPC (i.e., anti-RcPPC3) and anti-RcBTPC (i.e. anti-RcPPC4) immune sera were raised in rabbits as previously described
Antigenic polypeptides were visualized using a peroxidase-conjugated α-rabbit secondary antibody (Sigma cat. A0545) with Clarity™ Western ECL Blotting Substrate (Bio-Rad) and imaged using a ChemiDoc Touch Imaging System (Bio-Rad). For non-denaturing PAGE, clarified extracts were mixed with 5x native-PAGE loading buffer consisting of 300 mM Tris-HCl (pH 6.8) and 50% (v/v) glycerol, with the addition of a few bromophenol blue crystals to create a dye front. Proteins were resolved in a 7% acrylamide gel run at 180 V for 1 h at 4 °C. The gel was either electroblotted onto polyvinylidene difluoride (PVDF) for immunoblotting or incubated for 10 min in a PEPC activity stain consisting of 100 mM Tris-HCl (pH 8.0), 1 mM EDTA, 20 mM MgCl₂, 30 mM NaHCO₃, 5 mM PEP, and 3 mg ml⁻¹ Fast Violet B. Gels were then incubated in 5% (v/v) acetic acid for stain fixation.

**Co-immunoprecipitation**

Anti-RcPTPC-IgG (Gennidakis et al. 2007) was purified from corresponding immune serum using Pierce Protein A Chromatography Cartridges (Thermo Fisher Scientific). The IgG was then coupled onto Surebeads™ Protein-G Magnetic Beads (Bio-Rad) followed by a 1 h incubation at 24 °C with protein extracts. After thorough washing, bound proteins were eluted with 20 mM glycine-HCl (pH 2.0), followed by neutralization of the eluate with phosphate buffer as per manufacturer’s protocol. Eluates were boiled in 2x-Laemmli sample buffer for 5 min before subjecting to SDS-PAGE.
Results

Identification and analysis of vascular plant BTPC genes

Putative BTPC genes were identified via BLAST analysis of publicly available monocot and dicot genome sequences using the cDNA-derived gene sequence of castor BTPC (RcBTPC/RcPPC4) (Gennidakis et al. 2007). PTPC genes have a highly conserved structure composed of approximately 10 exons interrupted by 9 introns, irrespective of whether they originate from C₃, C₄, or CAM plants (Chollet et al. 1996). However, BTPC genes are characterized by ~20 exons (Sanchez and Cejudo 2003) (Fig. 2.2B). This difference is consistent with the hypothesis that both types of PEPC genes evolved independently from a common ancestral gene very early during the evolution of green algae (Sanchez and Cejudo 2003, O’Leary et al. 2011b). As illustrated in the sequence alignment of Fig. 2.3, all of the deduced BTPC polypeptides exhibit 77-92% sequence identity with RcBTPC. By contrast, the sequence identity within their corresponding IDR tended to be much lower (Fig. 2.4). The putative BTPCs were further classified according to three criteria, namely the: 1) presence of the C-terminal tetrapeptide (R/K)NTG characteristic of non-archaeal, prokaryotic-like PEPCs; 2) absence of the N-terminal serine phosphorylation motif (acid-base-XX₉SIDAQLR) present in all PTPCs; and 3) presence of an approximate 10 kDa IDR (corresponding to residues 325-467 of RcBTPC) (O’Leary et al. 2011b). Furthermore, all deduced BTPC polypeptides ranged from ~116-120 kDa and contained the characteristic domains and residues believed to be important for PEPC structure and catalysis (Chollet et al. 1996, Izui et al. 2004) (Fig. 2.3). In addition, the genomes of several species (i.e. cassava, field mustard, poplar, and soybean) appear to contain two or three BTPC genes, whereas the others (i.e. Arabidopsis, avocado, castor, cotton, cucumber, grape, maize, orange, rice, and tomato) only appear to contain a single BTPC gene (Fig. 2.2).
Interestingly, analysis of the promoter region revealed that many BTPC genes contain an abscisic acid (ABA) responsive cis-acting element (ABRE) (Nakashima et al. 2006) as illustrated in Fig. 2.2A. ABA is a phytohormone that plays an important role in plant adaptation to abiotic stresses such as drought and salt (Nakashima et al. 2006).

**BTPC transcriptomic analysis: RNA-seq and microarray data**

Fifteen vascular plant species were analyzed for tissue-specific BTPC expression (summarized in Appendix B). The various species were selected based upon availability of their annotated genome sequences, as well as corresponding RNA-seq and/or microarray transcriptomic datasets for their developing fruits or seeds, leaves, pollen, and roots. To corroborate the reliability of my comparative transcriptomic analysis, recent castor plant RNA-seq data were correlated with the corresponding COS BTPC expression profile previously obtained via RT-PCR using RcBTPC-gene specific primers (O’Leary et al. 2011a, Brown et al. 2012). Both RNA-seq and RT-PCR analyses consistently demonstrated that RcBTPC transcripts are particularly abundant during early stages of COS development, occurring at lower levels in male flowers and immature leaves, but essentially absent in the germinating seeds (Fig. 2.5). Comparing the abundance of mRNA transcripts is relatively subjective when analyzing different RNA-seq datasets. However, for the current comparison of BTPC expression an RPKM value of 10 or higher will be considered 'abundant' since immature castor leaves that exhibit a RPKM of about 15 (Fig. 2.5) contain immunologically detectable 118-kDa BTPC polypeptides, as well as a Class-2 PEPC complex composed of tightly associated BTPC and PTPC subunits (O’Leary et al. 2011a). As discussed below, BTPC expression appears to follow two distinct patterns in the plants that were analyzed; i.e., those in which BTPC shows mainly pollen-specific expression,
relative to plants like castor where $BTPC$ is highly expressed in heterotrophic, non-pollen sink tissues such as developing seeds, tubers, or fruits.

Species that exhibit mainly pollen-specific $BTPC$ expression

Field mustard ($Brassica rapa$) $BTPC1$ shares a similar expression profile to Arabidopsis $BTPC$, being generally low except for the flowers and roots (Figs. 2.6 and 2.7). This is not surprising considering that both species belong to the crucifer family. However, annotation of the field mustard genome identified a second $BTPC$ gene ($BTPC2$) which exhibited a different expression profile from $BTPC1$. Overall, pollen-specific $BTPC$ expression appears to be quite common among monocots and dicots. Strawberry, maize, rice, and poplar fall under this category, as they express abundant $BTPC$ transcripts in their developing anthers/pollen (Figs. 2.8-2.11). Interestingly, all monocots analyzed to date (i.e., rice, lily, and maize) demonstrate significant $BTPC$ expression in pollen, suggesting an evolutionary role in pollen development prior to the monocot/dicot divergence.

Species that exhibit marked $BTPC$ expression in heterotrophic sinks other than pollen

Similar to castor, abundant $BTPC$ transcripts are present in developing seeds of avocado, cotton, and soybean (Figs. 2.12 - 2.14). Although developing soybeans generally exhibit low levels of $BTPC$ transcripts, the early stages of soybean embryo development are indeed characterized by higher levels of $BTPC1$ transcripts (Fig. 2.14). In addition to its developing ovary, cucumber exhibits marked $BTPC$ expression in leaves, roots, and stems (Fig. 2.15). Interestingly, striking $BTPC$ expression also occurs in developing fruits of avocado, grape and tomato (Figs. 2.12, 2.16, and 2.17). Although RNA-seq data are available for orange fruit, attempts at aligning the data to the $Citrus sinensis$ genome resulted in poor alignment scores.
However, excellent alignments were obtained for immature and mature orange exocarp (peels) and leaves, in which \textit{BTPC} transcripts also appear to be fairly abundant (Fig. 2.18). Finally, developing cassava tubers also express considerable levels of \textit{BTPC1} transcripts (Figs. 2.19).

\textit{Significant BTPC expression shows close correlation with presence of immunoreactive BTPC polypeptides and Class-2 \textit{PEPC} complexes in tomato and cucumber plants}

Previous research with the castor oil plant established that its unique pattern of tissue-specific \textit{BTPC} transcriptional expression is closely correlated with the occurrence of 118 kDa \textit{BTPC} polypeptides and Class-2 \textit{PEPC} complexes (Blonde and Plaxton 2003, Gennidakis et al. 2007, O’Leary et al. 2011a). In order to establish the degree to which significant \textit{BTPC} expression is paralleled by the presence of \textit{BTPC} polypeptides in other species, clarified extracts from immature tomato fruit and cucumber ovaries were subjected to SDS-PAGE and immunoblotting using anti-Rc\textit{BTPC}. As seen in Fig. 2.20, ~118 kDa anti-Rc\textit{BTPC} immunoreactive polypeptides were detected on immunoblots of clarified extracts from both tissues. Interestingly, immunoblots of cucumber tendril extracts also revealed apparent ~118 kDa \textit{BTPC} polypeptides (Fig. 2.20), which was unexpected owing to the relatively low abundance of \textit{BTPC} transcripts in this tissue (Fig. 2.15).

Next, co-IP using anti-Rc\textit{PTPC} was performed on protein extracts prepared from developing cucumber ovaries and tomato fruit. Developing COS endosperm served as a positive control; \textit{i.e.}, castor \textit{BTPC} (Rc\textit{PPC4}; p118) co-IP’d with the endogenous \textit{PTPC} (Rc\textit{PPC3}; p107) in a near 1:1 ratio (Fig. 2.21), confirming the results of Uhrig and co-workers (2008). Analysis of tomato and cucumber co-IP eluates revealed that \textit{BTPC} from both species co-IP’d with endogenous \textit{PTPC} polypeptides (Fig. 2.21). In addition, non-denaturing \textit{PAGE} of clarified
extracts from immature tomato fruit followed by in-gel PEPC activity staining and parallel immunoblotting using subunit specific antibodies confirmed the presence of an active, high molecular weight Class-2 PEPC complex composed of associated BTPC and PTPC polypeptides that co-migrated with the ~900 kDa Class-2 PEPC complex from developing COS (Fig. 2.22).

**Discussion**

The current results indicate that significant *BTPC* expression is generally restricted to a diverse variety of biosynthetically active, heterotrophic sink tissues that tend to accumulate relatively high concentrations of malate (Table 2.2). Although it is difficult to determine the cytosolic malate concentration in plant cells, it has previously been estimated that orange vacuoles store ~70% of the total malate, indicating that up to 30% of the total malate may be cytosolic (Echeverria and Valich 1988). Since the typical I_{50}(malate) value of Class-1 PEPCs are in the sub-mM range (Law and Plaxton 1995, Blonde and Plaxton 2003, O’Leary et al. 2009, 2011c), the presence of a Class-2 PEPC could be crucial for maintaining anaplerotic PEP carboxylation in biosynthetically active sinks where malate accumulates in the high-mM range (*e.g.* orange fruit where malate concentrations were determined to range from 18-67 mM (Albertini et al. 2006)). Interestingly, a recombinant PTPC derived from orange fruit endocarp was highly desensitized to malate inhibition, with an unusually high I_{50}(malate) value of ~5 mM (Perotti et al. 2010). However, native PEPC from clarified orange endocarp extracts co-migrated during non-denaturing PAGE with the ~900 kDa castor Class-2 PEPC hetero-octamer (Perotti et al. 2010). Immunoblotting orange fruit extracts with BTPC antibodies would help to clarify whether BTPC polypeptides are present in this high molecular mass PEPC activity staining band. Similar to citrus fruits, grape berries also accumulate high levels of malate, where concentrations
of 50-75 mM peak about a week before ripening (Wada et al. 2008). Interestingly, PEPC from clarified extracts of unripe grape berries exhibited an \( I_{50}(\text{malate}) \) value of \( \sim 1.5 \) mM (Diakou et al. 2000), which is very comparable to the \( I_{50}(\text{malate}) \) value of 1.6 mM obtained for total PEPC activity present in soluble protein extracts prepared from developing COS endosperm (Blonde and Plaxton 2003). The correlation of malate desensitized PEPC activity with high \( BTPC \) transcript levels in developing grape berries (Fig. 2.16) is consistent with the presence of a Class-2 PEPC complex; this will be an important area to investigate for future grape berry metabolism research.

\( BTPC \) may also play an important role in developing fiber (\textit{i.e.} trichome) cells of developing cotton seeds, which are amongst the fastest growing cells in the plant kingdom (Li et al. 2010). High levels of malate, along with hexoses and \( K^+ \), are used by the rapidly growing fiber cells to maintain turgor (Smart et al. 1998). Malate was also suggested to support leucoplast fatty acid synthesis for the production of membrane lipids by the fiber-producing cells, which are incorporated into the rapidly expanding plasma membrane and tonoplast (Li et al. 2010). Indeed, both PEPC activity and malate levels of elongating cotton fiber cells simultaneously peak between 5-10 DPA, during which the intracellular malate concentration is in excess of 50 mM (Li et al. 2010). Since cotton \( BTPC \) transcript levels also maximize this stage (Fig. 2.13), it will be interesting to determine whether \( BTPC \)-containing Class-2 PEPC complexes are present in this early phase of cotton seeds development.

Although cucumber fruit are not known for their acidic properties, they do accumulate fairly high levels of malate which was reported to be \( \sim 15 \) mM (Handley et al. 1983). Cucumbers are very fast growing and likely require high levels of malate to maintain turgor, which is supported by the discovery of \( BTPC \) and Class-2 PEPC complexes in developing cucumber
ovaries (Fig. 2.21). Interestingly, unfertilized cucumber ovaries are not rapidly growing but still contain relatively abundant BTPC polypeptides (Fig. 2.20). It will be interesting to determine the phosphorylation status of BTPC from unfertilized cucumber ovaries to determine whether inhibitory phosphorylation attenuates BTPC activity until fertilization (i.e., when greatly enhanced anaplerotic flux of PEP to OAA is likely required). This form of regulation may be more dominant in tissues where transcripts do not correlate well with protein level, such as cucumber tendrils, which appear to have abundant BTPC polypeptides (Fig. 2.20) but low levels of BTPC transcripts (Fig. 2.15).

Interestingly, PEPC in clarified extracts prepared from immature tomato fruit exhibited an $I_{50}(\text{malate})$ value of 1.1 mM (Guillet et al. 2002), which is within the range of a Class-2 PEPC (Table 2.2). The confirmation of an active Class-2 PEPC complex in developing tomato fruit (Fig. 2.22) suggests that it may be possible to predict its presence simply by calculating the $I_{50}(\text{malate})$ of PEPC. Tomato fruits accumulate upwards of 75 mM organic acids (primarily malate and citrate) which are used to maintain turgor for fruit expansion (Guillet et al. 2002). In immature tomato fruit, abundant BTPC transcripts (Fig. 2.17) coincided with the presence of the BTPC-containing Class-2 PEPC complex (Figs. 2.20 – 2.22), as well as rapid growth and marked accumulation of malate and citrate (Guillet et al. 2002). In addition, tomatoes contain a PTPC gene that is specifically expressed in fruit (Guillet et al. 2002), which indicates that it may encode a dedicated PTPC isozyme for Class-2 PEPC formation. Tomato fruit are similar to COS in that they undergo high rates of respiration, but are likely much less permeable to atmospheric gas exchange due to their bulky size (Sweetman et al. 2009). The presence of BTPC in tomatoes may also reflect its potential CO$_2$-refixing role by possibly associating with the OMM (see Park et al. 2012) to prevent respiratory inhibition by CO$_2$. 
The level of malate accumulation in a fruit cell is determined by the combination of its rates of synthesis, storage, and metabolism (Fig. 2.23). Cytosolic malate synthesis begins with the formation of OAA via PEPC, followed by the reversible reduction of OAA into malate by cytosolic MDH (Etienne et al. 2013). Cytosolic, mitochondrial, and plastidial-specific MDH isozymes exist in plant cells (Fernie and Martinoia 2009). The current dogma is that cytosolic MDH is involved (with PEPC) in net malate synthesis during fruit development, whereas net malate degradation associated with fruit ripening is due to mitochondrial MDH (Sweetman et al. 2009). Due to the approximate neutrality of cytosolic pH, cytosolic malate exists in its dianionic form, a form which can be specifically bound by tonoplast malate transporters and shunted into the vacuole, in conjunction with the influx of equivalent positive charges regulated by cation channels (e.g., potassium channels) and/or proton pumps such as the H⁺-ATPase and H⁺-PPiases (Fig. 2.23) (Etienne et al. 2013). Upon entering the acidic vacuole malate becomes protonated and trapped as malic acid until efflux (controlled by transporter proteins) is required (Etienne et al. 2013). Accumulation of malate in fruit cells was hypothesized to be mainly controlled by vacuolar storage, and that metabolism responds accordingly to regulate its cytosolic concentration (Etienne et al. 2013). Indeed, differences in tonoplast proton pumps has been linked to variation in fruit malate content (Etienne et al. 2013). Cellular malate levels of developing fruit are also determined by the rate conversion from malate into sugars which can occur through cytosolic NADP-ME and pyruvate orthophosphate dikinase (PPDK) (Etienne et al. 2013) or by the combined activities of cytosolic NAD-MDH and phosphoenolpyruvate carboxylase kinase (PEPCK), and subsequent gluconeogenesis from the PEP that is thereby produced (see Fig. 2.23 for further details) (Sweetman et al. 2009).
It is important to emphasize that the presence of BTPC appears specific to malate ‘accumulating’ tissues. Proteoid roots of Pi-deficient harsh hakea produce copious levels of malate that is excreted to solubilize Pi chelated to metal cations in the soil, rather than stored intracellularly (Shane et al. 2013). Malate excretion may prevent cytosolic malate from accumulating to levels that would inhibit Class-1 PEPC. As such, phosphorylation of PTPC subunits of Class-1 PEPC may be sufficient to maintain anaplerotic PEP flux, which is supported by abundant phosphorylated PTPC subunits, and undetectable BTPC polypeptides in proteoid roots of Pi-deficient harsh hakea (Shane et al. 2013). In addition, certain fruits such as kiwi, mango, and strawberry appear to rely less on malate and more on the hydrolysis of accumulated starch as their carbon source during ripening (Sweetman et al. 2009), and should thus not require BTPC nor a Class-2 PEPC complex. Indeed, BTPC transcripts were essentially absent in strawberry fruit (Fig. 2.8), providing further evidence in support for a role of BTPC in tissues that specifically accumulate high malate concentrations. This idea is also supported by flux studies and metabolite profiling of developing soybean embryos (a green oilseed), which showed that soybeans may rely on malate early in development, but switch to citrate as the primary carbon source for fatty acid synthesis in the latter stages of development (Allen and Young 2013, Wheeler et al. 2016). Soybean BTPC1 transcripts follow this trend, showing higher expression very early in development followed by the general low abundance of BTPC transcripts during mid- to later stages of soybean development (Fig. 2.14). It will be interesting to analyze BTPC transcripts in starch accumulating fruits such as mango and kiwi (once the relevant RNAseq data becomes available) to determine if they also exhibit low or non-existent BTPC expression.
Figure 2.1. Model of alternate pathways supporting fatty acid synthesis in leucoplasts of developing COS.

The model (adapted from O’Leary et al. 2011b) illustrates that the oxidation of malate produced via cytosolic PEPC is capable of generating all the acetyl-CoA and reducing power in the form of NAD(P)H, needed for fatty acid synthesis in the leucoplast. Abbreviations: MDH, malate dehydrogenase; ME_p, plastidal malic enzyme; PK_p, plastidal pyruvate kinase; PDC_p, plastidal pyruvate dehydrogenase complex.
### Vascular plant BTPC expression profile

<table>
<thead>
<tr>
<th>Species</th>
<th>Tissue</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>1) Developing pollen</td>
<td>(Igawa et al. 2010) and (Sanchez and Cejudo 2003)</td>
</tr>
<tr>
<td></td>
<td>2) Roots (induced under salt stress)</td>
<td>(Sánchez et al. 2006)</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>1) Roots (3-4 leaf stage)</td>
<td>(Yu et al. 2010)</td>
</tr>
<tr>
<td></td>
<td>2) Stem (3-4 leaf stage)</td>
<td>(Yu et al. 2010)</td>
</tr>
<tr>
<td></td>
<td>3) Seedling (3-4 leaf stage)</td>
<td>(Yu et al. 2010)</td>
</tr>
<tr>
<td></td>
<td>4) Developing seeds</td>
<td>(Pan et al. 2013)</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>1) Leaves (immature, young and mature)</td>
<td>(Sullivan et al. 2004)</td>
</tr>
<tr>
<td></td>
<td>2) Stem</td>
<td>(Sullivan et al. 2004)</td>
</tr>
<tr>
<td></td>
<td>3) Flower</td>
<td>(Sullivan et al. 2004)</td>
</tr>
<tr>
<td></td>
<td>4) Roots</td>
<td>(Sullivan et al. 2004)</td>
</tr>
<tr>
<td><em>Lilium longiflorum</em></td>
<td>1) Developing pollen</td>
<td>(Igawa et al. 2010)</td>
</tr>
<tr>
<td><em>Ricinus communis</em></td>
<td>1) Immature leaves</td>
<td>(O’Leary et al. 2011a)</td>
</tr>
<tr>
<td></td>
<td>2) Developing endosperm</td>
<td>(O’Leary et al. 2011a) and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gennidakis et al. 2007)</td>
</tr>
<tr>
<td></td>
<td>3) Developing cotyledon</td>
<td>(O’Leary et al. 2011a)</td>
</tr>
</tbody>
</table>

**Table 2.1. List of all vascular plants with published BTPC expression data.**

The expression data were derived from qRT-PCR or semi-qRT-PCR by the corresponding authors. The list highlights the scarcity and variability of current BTPC expression data within vascular plants. Although every sequenced plant genome appears to have at least one BTPC gene, little is known about its specific role among different plant species.
Figure 2.2. Analysis of genomic structures of various \textit{BTPC} genes.  
(A) Promoter analysis of \textit{BTPC} genes. Genomic sequences for the 5'-UTR (1000–1500 bp upstream) for each \textit{BTPC} gene were obtained from the JGI Phytozome v11.0 and analyzed on PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw/index.html). Analysis revealed that many \textit{BTPC} genes contain an ABA responsive cis-acting element (ABRE) with consensus sequence ACGTG (Nakashima et al. 2006). Purple marks (above the line) indicate location of the ABRE on the plus strand while the orange mark (below the line) indicates location on minus strand. (B) Genomic structures of putative \textit{BTPC} genes were visualized on the JGI Phytozome v11.0 genome browser. The avocado genome was not available on JGI and was excluded. Exons are indicated by tan colored boxes and the grey boxes represent UTR regions.
T-coffee alignment of putative BTPC polypeptides was done using Jalview (Waterhouse et al. 2009). Amino acid sequences for BTPC from *Ricinus communis* (Accession number EF634318) and *Arabidopsis thaliana* (Accession number NM 105548) were derived from corresponding cDNA sequences. The genomic sequence of avocado BTPC (contig ID: PA10006483) was annotated using GENSCAN (Burge and Karlin 1997). The remaining annotated BTPC sequences were derived from JGI Phytozome v11.0. Sequence identity is in reference to castor (*R. communis*) BTPC. Boxes I-III indicate conserved PEPC subdomains believed to be essential for PEPC activity (Izui et al. 2004). The red line highlights the absence of the N-terminal phosphorylation domain that is characteristic of all PTPCs. The box at the C-terminal end highlights the conserved prokaryotic-like tetrapeptide of (R/K)NTG that is characteristic of BTPCs.
Figure 2.4. Alignment of the intrinsically disordered region (IDR) of deduced BTPC polypeptides.

T-coffee alignment of putative BTPC IDRs (reference to castor BTPC (RcPPC4) amino acids 325-467) was done using Jalview. The percent identity is in reference to the IDR of RcPPC4, which is known to mediate binding with its PTPC partner, RcPPC3 (Park et al. 2012). The arrows indicate the inhibitory in vivo phosphorylation sites of developing COS BTPC (O’Leary et al. 2011c, Dalziel et al. 2012). The rectangle indicates a fragment of the IDR that is absent from Arabidopsis BTPC (as well F. vesca, O. sativa, Z. mays, and one from B. rapa and G. max), but is present other BTPCs, including castor.
Figure 2.5. *Ricinus communis* (castor) *BTPC* (*RcPPC4*) expression.
RNA-seq data was derived from Brown et al. (2012) using the castor variety 99N89I. Leaf tissue corresponded to expanding true leaves 10-15 cm in size. Germinating seed was obtained after 3 d of germination in moist vermiculite. Developing male flowers contain anthers and pollen (with sepals removed).
A)

At1g68750 268832_at  ATPC3

Arabidopsis eFP Browser at bar.utoronto.ca

Seed/ Silique 5
Shoot Apex Infl orescence
Seed/ Silique 4
Shoot Apex Transition
Cauline Leaf
Shoot Apex Vegetative
2nd Internode
Hypocotyl
Root
Vegetative Rosette
1st Node
24 h Imbied Seed
Dry seed
Entire Rosette after transition to flowering
Leaf 1 2 3 4 5 6 7 (dist/prox/petiole) 8 10 12 Caulline Senescent

Absolute
629.79
566.81
503.83
440.85
377.87
314.89
251.91
188.93
125.95
62.97
0.0

Maedox

Flowers 9 10 11 12 (sep/pet/stam/carp) 15 (flower/pedicel) 15 (sep/pet/stam/carp) mature pollen

eFP Browser by B. Vinecar, drawn by J. Alls and N. Provar. Data from Gene Expression Map of Arabidopsis Development: Schmid et al., 2005, Nat. Gen. 37:501, and the Nambara lab for the imbibed and dry seed stages. Data are normalized by the GCOS method, TGT value of 100. Most tissues were sampled in triplicate.
Figure 2.6. Arabidopsis thaliana BTPC (AtPPC4) expression.
Arabidopsis BTPC eFP (A) and expression histogram (B) are visualized using the Arabidopsis eFP browser available at the UofT BAR database using the gene identifier At1g68750.
Figure 2.7. *Brassica rapa* (field mustard) *BTPC* expression.

The data for the first 6 tissue sources (from left to right) were obtained from Tong et al. (2013). Callus tissue were obtained from tissue culture. Fully bloomed flowers, roots, stems and leaves were collected from 7 week old plants. Siliques were harvested 15 DAP. The remaining data (seedling to petiole) were obtained from Devisetty et al. (2014) from *B. rapa* genotypes R500 and IMB211. Seedlings and roots were harvested 10 d after germination. Vegetative meristems were collected from the plants when the third leaf reached 1 mm in size. Internode tissue was obtained from the stem between the fourth and fifth leaves, and inflorescence meristems were collected when these meristems became fully formed. Leaves were collected 28 d after germination. Siliques were obtained at full maturity. *BTPC1* corresponds to *Bra024730* and *BTPC2* corresponds to *Bra004016*. 
Figure 2.8. *Fragaria vesca* (strawberry) BTPC expression.
RNA-seq data for the first 30 tissues (anthers to the receptacle) were obtained from Hollender et al. (2014) using a 7th generation inbred line of wild strawberry, Yellow Wonder 5AF7. The numbers represent developmental stages of the tissues analyzed and are described by Hollender et al. (2012). The remaining data (ovule to leaf) and the developmental illustration were obtained from Kang et al. (2013) using the same strain of woodland strawberry.
Figure 2.9. *Zea mays* (maize) *BTPC* expression.
RNA-seq data and the developmental illustration was taken from Chen et al. (2014) using the inbred line B73. RPKM values were directly taken from the supplementary files using the gene identifier *GRMZM2G082780*. Abbreviations are: Em, embryo; En, endosperm; S, whole seed. For Em, En, and S samples, the numbers correspond to DAP. The remaining samples were staged as described by the authors.
Figure 2.10. *Populus trichocarpa* (poplar tree) *BTPC1* and *BTPC2* expression. Poplar *BTPC1* and *BTPC2* eFPs (A) and expression histograms (B) were derived from microarray data from Wilkins et al. (2009) and visualized using the poplar eFP browser at the UofT BAR. In panels A and B *POTRI.008G114200.1* (left) and *POTRI.010G131800.1* (right) represent expression profiles for poplar *BTPC1* and *BTPC2*, respectively.
Figure 2.11. *Oryza sativa* (rice) *BTPC* expression.
Rice *BTPC* eFP (A) and expression histogram (B) were derived from microarray data from Li et al. (2007) and visualized using the rice eFP browser at the UofT BAR using the *BTPC* gene identifier LOC_Os01g02050. All tissue were derived from the Nipponbare variety. Germinated seeds were grown for 2 weeks before harvesting roots and shoots. The remaining tissues were harvested from natural grown rice plants. Anthers, were collected 1-2 d before floret flowering. Unpollinated stigma and ovary were collected 0-1 d before floret flowering. Pistils were collected 5 DPA. Embryo and endosperm were collected 10 DPA.
Figure 2.12. *Persea americana* (avocado) BTPC expression.
RNA-seq data and the illustration for the various stages of fruit development were derived from Kilaru et al. (2015) using the Hass cultivar with the indicated ‘date of harvest’. Specific age of the developing fruit in terms of days after pollination (DAP) were not provided by the authors. The rest of the data and the illustration for fruit ripening was derived from Ibarra-Laclette et al. (2015) using the *Drymifolia* variety. Leaves consisted of mixed developmental stages of expanded leaves. Stems were derived from segments of young branches. Aerial buds were derived from developing buds emerging from the shoot apical meristem. Flowers were derived from whole inflorescences from a mixture of young, immature, and mature stages. Seeds are derived from fruit at the mature-green stage (approximately 8 months old). Roots were derived from *in-vitro* propagated seedlings.
Figure 2.13. *Gossypium raimondii* (cotton) BTPC expression. RNA-seq data for developing ovules was obtained from Patterson et al. (2012). Seedling data was obtained from Renny-Byfield et al. (2014). Petal tissue were harvested at full expansion just prior to pollination, as described by Rambani et al. (2014). Mature leaf data was obtained from Li et al. (2014). The image for developing cotton stages was obtained from Lee et al. 2007.
RNA-seq data was obtained from Collakova et al. (2013) using the cultivar Williams 82. Seeds that were measured as 3 mm long were tagged as ‘0 d’ and subsequently harvested every 5 d. BTPC1 corresponds to JGI identifier Glyma.02G130700, BTPC2 to Glyma.01G091000, and BTPC3 to Glyma.10G205500/Glyma10g34970.
Figure 2.15. *Cucumis sativus* (cucumber) *BTPC* expression.

RNA-seq data was obtained from Li et al. (2011) using the sativus line 9930. Ovary tissues were harvested 7 DPA. Details on the exact developmental stages of the remaining tissues were not included by the authors (likely the same as when ovary were harvested).
Grape eFP Browser by R. Patel. Images adapted and data derived from The Grapevine Expression Atlas Reveals a Deep Transcriptome Shift Driving the Entire Plant into a Maturation Program. Mananna Ascoli et al., (2012) Plant Cell; 24(5): 3489–3505. Data were NimbleGen derived and were normalized using RMA and were provided as linearized data. All tissues were sampled in triplicate.
Figure 2.16. *Vitis vinifera* (grape) *BTPC* expression.
Grape *BTPC* eFP (A) and expression histogram (B) were derived from microarray data from Fasoli et al. (2012) and visualized using the grape eFP browser at the UofT BAR with the gene identifier *VIT_01S0011G02740.*
RNA-seq data for developing tomato fruit was obtained from Osorio et al. (2011) using the Ailsa Craig cultivar. The remaining tissues were derived from the Money Maker cultivar according to Sato et al. (2012). Leaves were derived from 2-3 week old fully expanded leaves. Stems were derived from the upper stem, below apical meristem. The exact age of the stem and root tissue was not provided by the authors, although speculated to be collected at the same time as the leaves. Flowers were derived from whole flowers (excluding sepals) at full anthesis. Pollen samples were derived from the cultivar Hazera 3042 according to Pressman et al. (2002) consisting of a mixture of developmental stages.
Figure 2.18. *Citrus sinensis* (orange) *BTPC* expression.
RNA-seq data was obtained from Martinelli et al. (2012) using the Valencia variety. Leaves were collected from randomly selected areas in the canopy. Fruit peel (exocarp) were obtained from healthy plants of unspecified age.
Figure 2.19. *Manihot esculenta* (cassava) *BTPC1* and *BTPC2* expression.
RNA-seq data was derived from Wang et al. (2014) using the cassava cultivar Arg7. Leaves were harvested from ‘fresh young leaves’. The specific age of the roots (cassava tubers) and stem were not provided by the authors. *BTPC1* corresponds to the JGI identifier *Manes.14G095900* and *BTPC2* corresponds to *Manes.06G074600*. 

![Bar chart showing expression levels of BTPC1 and BTPC2 in different tissues.](image)
Figure 2.20. Immunodetection of BTPC from plant tissues with abundant BTPC transcripts.
Protein extracts (50 µg lane⁻¹) were taken from stage V COS endosperm (2), immature tomato fruit 7 DAP (3), fertilized cucumber ovary 7 DAP (4), unfertilized cucumber ovary (5), and cucumber tendril (6). Samples were subjected to SDS-PAGE and immunoblotting using anti-RcBTPC. Antigenic polypeptides were visualized using a peroxidase-conjugated secondary antibody and ECL detection. Purified recombinant chimeric Class-2 PEPC (50 ng) (O’Leary et al. 2009) consisting of 118 kDa RcPPC4 and 107 kDa AtPPC3 was used as a positive control (1).
Figure 2.21. Co-IP of BTPC from cucumber and tomato extracts using anti-RcPTPC. Co-IP was performed using anti-RcPTPC-IgG coupled onto protein-G magnetic beads (Bio-Rad). Protein extracts were taken from immature green tomatoes (10 DPA) and fertilized cucumber ovaries. Stage V developing COS endosperm was used as a positive control. The beads were incubated for 1 h at 24 °C in clarified protein extracts. After thorough washing and elution of bound proteins, the co-IP eluates were subjected to SDS-PAGE followed by Coomassie Blue R-250 staining and immunoblotting using anti-RcPTPC or anti-RcBTPC. Antigenic polypeptides were visualized using a peroxidase-conjugated secondary antibody and ECL detection.
Figure 2.22. Detection of active Class-2 PEPC in tomato fruit extracts via non-denaturing PAGE.
Protein extracts were taken from immature tomato fruit (7 DAP - 260 µg protein). Extract from developing COS (stage V - 120 µg protein) was used as a positive control. Antigenic polypeptides were visualized using a peroxidase-conjugated secondary antibody and ECL detection.
### Table 2.2. Estimated intracellular malate concentration of various plant tissues.

For reference, the $I_{50}(\text{malate})$ values from purified COS Class-1 PEPC (Blonde and Plaxton 2003) and recombinant COS Class-2 PEPC (O’Leary et al. 2011c) are listed at the top. Early estimates of COS endosperm malate concentrations were around 5 mM (Smith et al. 1992). More recent metabolite profiling revealed far higher total malate during COS development. *Asterisk indicates that the malate concentrations were calculated relative to tissue fresh weight, assuming 1 gFW·mL$^{-1}$ cell volume. Stages of castor seeds are described according to Greenwood and Bewley (1982). Soybean seeds are staged accordingly to Wheeler et al. (2016).
Figure 2.23. Model of the mechanisms that control malate levels in a plant cell.
Vacuolar storage has been proposed as the main regulator of citrus fruit malate content (Etienne et al. 2013). For reversible reactions, the larger arrows indicate the thermodynamically favored direction under physiological conditions. NAD-MDH can be involved in net synthesis or net conversion of malate into OAA depending on the isozymes and specific conditions. Cytosolic MDH generally favors the forward reaction (ie. OAA to malate) whereas mitochondrial MDH has been implicated in malate oxidation during fruit ripening (Sweetman et al. 2009).
Chapter 3 . Biochemical and molecular studies of AtPPC4, the enigmatic bacterial-type phosphoenolpyruvate carboxylase of Arabidopsis thaliana

In preparation for Journal of Experimental Botany

Abstract

Characterization of PEPC from green algae and then developing castor beans led to the discovery of novel, allosterically-desensitized Class-2 PEPC hetero-octameric complexes composed of tightly interacting PTPC and BTPC subunits. The main objective of the current study was to assess the ability of the poorly characterized BTPC from the model plant Arabidopsis thaliana (AtPPC4) to form a Class-2 PEPC complex with endogenous Arabidopsis PTPC isozymes, particularly AtPPC3. After optimizing heterologous expression of non-tagged AtPPC4 within the soluble fraction of Escherichia coli, co-purification of AtPPC4 with recombinant His-tagged AtPPC3 was attempted using metal-ion affinity chromatography. Although bound proteins eluting with 150 mM imidazole from this column were enriched with active AtPPC3, the recombinant AtPPC4 was catalytically inactive and eluted in the non-bound fractions, indicating minimal interaction between AtPPC4 and AtPPC3. Similarly, yeast two-hybrid experiments demonstrated that AtPPC4 seemed to interact weakly with AtPPC3 (as well as its homolog AtPPC2). By contrast, parallel yeast two-hybrid studies showed that RcPPC4 (i.e., castor BTPC) showed a far stronger interaction with all three Arabidopsis PTPC isozymes, particularly AtPPC2 and AtPPC3. It is hypothesized that the apparent difference in the PTPC-
binding ability of AtPPC4 versus RcPPC4 arises from differences within their respective intrinsically disordered regions (IDR) which are common to many proteins and function to provide a docking site to promote protein-protein interactions. In particular, the IDR of RcPPC4 contains a 15 amino acid insertion (residues 404-418) not found in the IDR of AtPPC4. It is notable that plants such as Arabidopsis and maize that mainly exhibit pollen-specific BTPC expression are missing this approximate 2 kDa insertion within their IDR, whereas it is largely conserved in BTPC IDRs of plants such as castor and tomato that exhibit marked BTPC expression and Class-2 PEPC formation in non-pollen heterotrophic sink tissues. In addition, attempts at overexpression of AtPPC4 in Arabidopsis suspension cells and in planta failed to produce significant levels of AtPPC4 polypeptides.

**Introduction**

PEP carboxylase (EC 4.1.1.31) is a tightly regulated enzyme that has been intensively studied owing to its pivotal role in assimilating atmospheric CO$_2$ during C$_4$ and crassulacean acid metabolism photosynthesis (O’Leary et al. 2011b). Much attention has also been devoted to the essential non-photosynthetic functions of PEPC, particularly the anaplerotic replenishment of tricarboxylic acid cycle intermediates withdrawn during biosynthesis and nitrogen assimilation. To fulfill its diverse roles and complex regulation, plant PEPC belongs to a small multi-gene family encoding PTPCs, along with a distantly-related BTPC (O’Leary et al. 2011b). *PTPC* genes encode closely related 100-110-kDa polypeptides containing conserved serine phosphorylation and lysine monoubiquitination sites and that typically oligomerize as tetrameric Class-1 PEPCs (Tripodi et al. 2005, Uhrig et al. 2008, O’Leary et al. 2011b). By contrast, plant *BTPC* genes encode distantly related 116-118-kDa polypeptides that are more similar to
prokaryotic PEPCs. BTPCs from green algae, developing castor beans, and lily pollen tightly associate with co-expressed PTPC isozymes to form unusual Class-2 PEPC complexes (Rivoal et al. 1998, 2001, Gennidakis et al. 2007, Igawa et al. 2010). However, despite a number of published studies (Sanchez and Cejudo 2003, Sánchez et al. 2006, Igawa et al. 2010, Wang et al. 2012), very little is known about the biochemical or functional properties of AtPPC4, the single BTPC encoded by the genome of Arabidopsis thaliana. Extensive studies of RcPPC4 (i.e., castor BTPC) and Class-2 PEPC complex formation in the oil-rich endosperm of developing COS (Tripodi et al. 2005, Gennidakis et al. 2007, O’Leary et al. 2009, 2011a) indicated that AtPPC4 might have a similar function during Arabidopsis seed development; i.e., to support malate production as a precursor for long chain fatty acid synthesis. However, this idea is challenged by the facts that: 1) AtPPC4 transcripts and polypeptides cannot be detected in developing Arabidopsis seeds (Sanchez and Cejudo 2003); and 2) suppression of AtPPC4 by artificial microRNA did not influence Arabidopsis seed development, nor storage lipid content or composition (Wang et al. 2012). On the other hand, AtPPC4 may have an important function during Arabidopsis pollen development, where the gene is highly expressed (Fig. 2.6) (Igawa et al. 2010). Much like developing seeds, pollen is a biosynthetically active sink that accumulates significant amounts of storage carbohydrates and lipids (Paupière et al. 2014). However, an AtPPC4-null mutant exhibited normal pollen morphology, plant growth, and fertility, indicating that BTPC is not essential for pollen development or plant survival (at least under the experimental growth conditions that were employed) (Igawa et al. 2010). Regardless, the high expression of AtPPC4 in pollen indicates it may have an important role in pollen development. AtPPC4 may also be important for abiotic stress acclimation since it was also highly upregulated in roots of salt or drought stressed Arabidopsis (Sánchez et al. 2006). Wang et al. (2012) used
microRNA to down regulate \textit{AtPPC4} and showed that root growth was impaired in the transgenic lines. This inhibition was partially relieved by salt treatment, which was attributed to the compensatory \textit{PTPC} induction (\textit{i.e.}, \textit{AtPPC1-AtPPC3}), thus showing a link between \textit{BTPC} and \textit{PTPC} expression (Wang et al. 2012).

Although BTPCs from green algae, castor seeds and leaves, and lily pollen tightly interact with co-expressed PTPCs to form functional Class-2 PEPC complexes (Rivoal et al. 2001, Gennidakis et al. 2007, Igawa et al. 2010, O’Leary et al. 2011a), there is no evidence for this occurring with \textit{AtPPC4}. If \textit{AtPPC4} does indeed form a Class-2 PEPC \textit{in vivo}, it is hypothesized that \textit{AtPPC3} is its preferential/native PTPC binding partner. This is supported by the fact that \textit{RcPPC4} is co-expressed with and tightly interacts with the \textit{AtPPC3} ortholog \textit{RcPPC3} in developing castor beans, and that \textit{RcPPC4} and \textit{RcPPC3} are the \textit{p118} and \textit{p107} subunits, respectively, of native Class-2 PEPC from developing COS, whereas polypeptides corresponding to the other two PTPC isozymes encoded by the castor genome (\textit{i.e.}, \textit{RcPPC1} and \textit{RcPPC2}) were never detected (Blonde and Plaxton 2003, Gennidakis et al. 2007, Uhrig et al. 2008). Like castor, the Arabidopsis genome encodes three PTPC isozymes, \textit{AtPPC1}, \textit{AtPPC2}, and \textit{AtPPC3} whose deduced polypeptides share >85\% sequence identity (Sánchez et al. 2006). \textit{AtPPC1} and \textit{AtPPC2} transcripts are constitutively expressed as housekeeping PEPCs in most organs (Sanchez and Cejudo 2003, Winter et al. 2007). Single knockout mutants of \textit{atppc1} or \textit{atppc2} did not generate noticeable phenotypes, suggesting that the two genes can compensate for the lost function of the other, whereas the double \textit{atppc1/atppc2} mutant exhibited a severe growth arrest, highlighting the importance of having at least one of the two PTPC isozymes for normal Arabidopsis development (Shi et al. 2015). \textit{AtPPC3} appears to be a more specialized PTPC isozyme that shares a similar pollen and root specific expression pattern with \textit{AtPPC4},
suggesting that these two PEPC isozymes are present in the same tissue (Sanchez and Cejudo 2003, Sánchez et al. 2006). This is further supported by the ‘pep2pro’ proteomics database (http://fgcz-pep2pro.uzh.ch/) that provides protein quantification from mass spectroscopy data (Baerenfaller et al. 2011) and confirmed that AtPPC4 and AtPPC3 are the most abundant PEPCs present in Arabidopsis pollen (Fig. 3.1).

The goal of the present study was to expand our knowledge on AtPPC4 properties and function by determining whether it: 1) is capable of interacting with a PTPC to form a Class-2 PEPC complex; and 2) exhibits preferential association with the PTPC isozyme, AtPPC3. To answer these questions production of recombinant AtPPC4 was attempted in E. coli and in planta, for biochemical characterization. Co-IP studies of developing Arabidopsis flower bud extracts were attempted using anti-RcPPC3 antibodies. The yeast two-hybrid system was also used to compare in vivo protein interactions between AtPPC4 and AtPPC1-AtPPC3.

**Materials and methods**

**Plant materials**

*Arabidopsis thaliana* (cv. Col-0) were cultivated at 20 °C in growth chambers (16/8 h photoperiod at 100-125 µmol m⁻²s⁻¹). Arabidopsis seedlings were also cultivated in Murashige Skoog (MS) liquid media, with gentle agitation (100 r.p.m.) at 20 °C and continuous illumination (40-60 µmol m⁻²s⁻¹). Heterotrophic Arabidopsis (cv. Ler-0) suspension cells were cultured in liquid MS media, supplemented with 0.5 mg L⁻¹ α-naphthalene acetic acid (NAA) and 0.05 mg L⁻¹ kinetin as previously described (Veljanovski et al. 2006). Cells were maintained in the dark at 24 °C on a rotational shaker (125 r.p.m.).
Protein extraction, electrophoresis and immunoblotting

Proteins were extracted from liquid grown seedlings and 7-d old suspension cell cultures as previously described (O’Leary et al. 2011a). Quick-frozen cells were ground to a fine powder in liquid N$_2$ using a mortar and pestle. Extraction buffer contained 50 mM HEPES-KOH (pH 7.5), 1 mM EDTA, 1 mM EGTA, 25 mM NaF, 1 mM Na$_3$VO$_4$, 1 mM Na$_3$MoO$_4$, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol, 10 mM MgCl$_2$, 1% (w/v) polyvinyl(poly pyrrolidone), 2 mM phenylmethylsulfonyl fluoride, 2 mM 2,2'-dipyridyl disulphide, 5 µl ml$^{-1}$ ProteCEASE-100 (G-Biosciences). Homogenates were centrifuged at 4 °C and 15,000 g for 10 min, and filtered through 2 layers of Miracloth (EMD Millipore). Extracts were quick frozen in liquid N$_2$ and stored at -80 °C for future use or immediately boiled in 2x-Laemmli sample buffer for 5 min.

Electrophoresis and Immunoblotting

Electrophoresis and immunoblotting were performed as described in Chapter 2. Anti-AtPPC4 was raised in rabbits against a heterologously expressed, C-terminal portion of AtPPC4 as previously described (Gennidakis et al. 2007). The Precision Plus Protein Ladder (Bio-Rad) was used as the molecular weight marker for SDS-PAGE.

PEPC activity and protein concentration assays

PEPC activity was determined as previously described (O’Leary et al. 2009). Activity was assayed at 24 °C by monitoring NADH oxidation at 340 nm in a final volume of 200 µl using a Spectromax Plus microplate reader (Molecular Devices). The assay mixture contained 25 mM Hepes-KOH (pH 8.0), 10% (v/v) glycerol, 2 or 10 mM PEP (for Class-1 PEPC and Class-2 PEPC respectively), 5 mM NaHCO$_3$, 10 mM MgCl$_2$, 1 mM DTT, 0.15 mM NADH and 5 units
ml\(^{-1}\) malate dehydrogenase from pig heart (Sigma Aldrich). One unit of PEPC is defined as the amount of enzyme needed for the production of 1 \(\mu\)mol of OAA min\(^{-1}\). Protein concentrations were determined using the Bradford assay with bovine \(\gamma\)-globulin as the protein standard as previously described (Blonde and Plaxton 2003).

*Cloning of constructs*

All restriction enzymes used were obtained from New England BioLabs (NEB). PCR was performed using Q5 High-Fidelity Polymerase (NEB) and purified using Purelink Quick Gel Extraction kit (Invitrogen) as per the manufacturer’s protocol. All primers and restriction sites used are listed in Appendix A. All full length PEPC amplifications were done using touchdown PCR using a MyCycler™ (Bio-Rad). Cycle parameters were 98 °C for 1 min followed by 10 cycles of 98 °C for 10 s, 60 °C for 30 s (dropping 1 °C for each cycle), 72 °C for 2 min, followed by 35 cycles of 98 °C for 10 s, 50 °C for 30 s, 72 °C for 1 min 45 s, and a final extension of 72 °C for 2 min. Constructs used for producing recombinant N-terminal-His-tagged-AtPPC3 and full length AtPPC4 were previously described (Gennidakis et al. 2007). For constitutive expression in planta under the 2x35S cauliflower mosaic virus promoter, AtPPC4 was cloned into pMDC32 (Curtis and Grossniklaus 2003) using BamHI and SalI restriction sites. For the yeast-two hybrid constructs, gene templates (RcPPC4, AtPPC4, and AtPPC3) were obtained from vectors previously described (Gennidakis et al. 2007, O’Leary et al. 2009) or from Arabidopsis seedling cDNA (AtPPC1 and AtPPC2). AtPPC4 and RcPPC4 were cloned into pGBK7 (Clontech) while the Arabidopsis PTPC isozymes (AtPPC1, AtPPC2 and AtPPC3) were cloned into pGADT7-AD (Clontech). Directional cloning was done for AtPPC4, RcPPC4, and AtPPC3 with ligations performed overnight at 16 °C using T4 DNA ligase (NEB). Since
AtPPC1 and AtPPC2 had no convenient restriction sites for the pGADT7-AD vector, the Gibson assembly cloning kit (NEB) was used. pGADT7-AD was digested overnight with NdeI and XmaI, and gel purified using NucleoSpin Gel and PCR-clean up kit (Machery Nagel) prior to Gibson assembly. Vector and genes were assembled in a 1:4 molar ratio as per the manufacturer’s recommendations. All gene constructs were confirmed via Sanger sequencing.

**Heterologous expression of recombinant PEPCs**

Plasmid constructs were transformed into *E. coli* (BL21-CodonPlus (DE3)-RIL) (Stratagene) using chemical transformation. Recombinant protein expression was as previously described (O’Leary et al. 2009). *E. coli* were cultured at 37 °C to an *OD*$_{600}$ of 0.6 before induction. AtPPC4-overexpressing cells were induced with 100 µM IPTG at 24 °C for 24 h (unless otherwise stated), whereas AtPPC3-overexpressing cells were induced with 2% (w/v) lactose for 12 h at 24 °C. Cultures were centrifuged at 3,000 g for 10 min and the cell pellet resuspended (1:5; w/v) in Buffer A which contained 50 mM NaH$_2$PO$_4$ (pH 8.0), 300 mM NaCl, and 15% (v/v) glycerol. The cells were either immediately lysed or quick frozen in liquid N$_2$ and stored at -80 °C. Chromatography was conducted using an AKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare) at 24 °C. Recombinant PEPCs were purified as previously described (O’Leary et al. 2009) with slight modifications. *E. coli* cells containing recombinant AtPPC4 (approx. 16 gFW) or AtPPC3 (approx. 8 gFW) were combined in a 2:1 ratio and lysed with 3 passes through a French press at 20,000 p.s.i. After centrifugation, the supernatant was loaded at 1 ml min$^{-1}$ onto a column (1.6 x 10 cm) of PrepEase$^\text{TM}$ His-tagged High Yield Purification Ni$^{2+}$ affinity resin (U.S. Biochemical Corp.). The column was washed with buffer A until *OD*$_{280}$ approached baseline followed by elution with buffer A containing 150
mM imidazole (pH 8.0). Eluted proteins were concentrated to 2 ml with an Amicon Ultra-15 centrifugal filter unit (100-kDa cut-off).

**Nucleic acid extraction and RT-PCR**

RNA and gDNA were extracted using Trizol reagent (Thermo Fisher Scientific) as per the manufacturer’s protocol. For amplification of full length *AtPPC1* and *AtPPC2*, cDNA was synthesized from RNA using Superscript III reverse transcriptase (Thermo Fisher Scientific) as per the manufacturer’s protocol. For semi-quantitative RT-PCR, cDNA was synthesized with QuantiTect-RT kit (Qiagen) or using M-MuLV-reverse transcriptase (NEB) as per the manufacturers’ protocol. Arabidopsis *ACTIN2* (reference gene) and *AtPPC4* were amplified with gene-specific primer pairs listed in Appendix A, using cDNA (~600 ng) as template and GoTaq Hot Start Polymerase (Promega). Typical cycle parameters were 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s, 49 °C for 15 s, 72 °C for 34 s, and a final extension of 64 °C for 1 min. PCR products were resolved on 1% (w/v) agarose gels and visualized with Red Safe (iNtRON Biotechnology).

**Yeast culturing and transformation**

Baker’s yeast (PJ69-4A) was cultured according to the Yeast Protocols Handbook (Clontech) and transformation was done with slight modifications to the lithium acetate (LiAc)-mediated procedure. Dropout (DO) supplements were purchased from Clontech. Double dropout (DDO) supplement lacking leucine and tryptophan was used to screen uptake of both bait and prey vectors. Quadruple dropout (QDO) supplement lacking leucine, tryptophan, adenine and histidine was used to screen for interaction between bait and prey proteins. Yeast were cultured
in YPAD media in a 30 °C incubator at 220 r.p.m. When an \(OD_{600}\) of 0.4-0.6 was attained, 50 ml of yeast culture was centrifuged at 3000 r.p.m. for 5 min, and the pellet was washed with sterile water before resuspending the cells in 1 ml of 100 mM LiAc. Cells were aliquoted into 100 µl portions with the supernatant removed before adding transformation mix consisting of 50% (w/v) PEG-3350 (240 µl), 1 M LiAc (36 µl), 2 mg ml\(^{-1}\) sheared salmon sperm DNA (25 µl) and 0.5-1 µg of plasmid DNA (1-3 µl). Yeast were incubated in the transformation mix for 30 min at 30 °C with shaking. Dimethylsulfoxide (35 µl) was added to increase transformation efficiency, followed by a 15 min heat shock at 42 °C. Cells were harvested, resuspended in 0.5 ml sterile water, and plated (200 µl) onto auxotrophic media plates. \(NCK-1\) (wormbase.org ID=ZK470.5) in \(pGADT7-AD\) and \(Vab-1\) (wormbase.org ID=M03A1.1) in \(pGBKT7\) were obtained from Prof. Ian Chin-Sang (Queens University Biology Department); \(NCK-1\) and \(Vab-1\) were used as a positive control as they are \textit{Caenorhabditis elegans} proteins that strongly interact in yeast two hybrid tests (Mohamed et al. 2012). Semi-quantitative serial dilutions were done by growing yeast to an \(OD_{600}\) of 0.4-0.8 and then diluting with water. An aliquot (25 µl) of each dilution was plated on DDO or QDO plates and incubated at 30 °C for 86 h.

\textit{Arabidopsis thaliana transformation via Agrobacterium tumefaciens}

The floral dip method was used to stably transform Arabidopsis plants (Zhang et al. 2006) using Agrobacteria (GV3101) with binary vector \(pMDC32-AtPPC4\). Transformation of Arabidopsis suspension cells was done using the same construct according to the protocol obtained from the 'Arabidopsis Biological Resource Center' of Ohio State University (https://abrc.osu.edu/sites/abrc.osu.edu/files/Transformationprotocol_VIB.pdf). Agrobacteria (GV3101) containing the binary vector were cultured for 2 d at 28 °C, centrifuged at 3,000 r.p.m.
and washed in MS media before being adjusted to an \( \text{OD}_{600} \) of 1.0. Transformation was done by combining 3 ml of 2-d old Arabidopsis suspension cells, 200 \( \mu \text{l} \) of MS washed Agrobacterium (\( \text{OD}_{600} \) of 1.0) and 6 \( \mu \text{l} \) of 100 mM acetoseringone in a sterile culture plate. The plate was sealed with Micropore surgical tape and co-cultured at 130 r.p.m. for 3 d. On the third day, the cell mixture was washed 2x with MS media containing 200 \( \mu \text{g ml}^{-1} \) timentin and 50 \( \mu \text{g ml}^{-1} \) carbenicillin to kill off any remaining agrobacteria, and plated on MS agar with carbenicillin, timentin, and hygromycin (20 \( \mu \text{g ml}^{-1} \) each) for selection of transformed cells. The plate was sealed with micropore surgical tape and incubated for 2 weeks until sufficient cells were present on the plate to inoculate into fresh liquid MS media.

**Immunoprecipitation**

Attempted co-IP experiments were performed as described in Chapter 2. Proteins were extracted as described above from Arabidopsis floral buds of varying inflorescent stages.

**Results and discussion**

AtPPC4 does not co-immunopurify with AtPPC3 from Arabidopsis flower bud extracts

A relatively straightforward way to test potential interaction between AtPPC4 and AtPPC3 is by co-IP using anti-RcPPC3 (i.e., anti-RcPTPC) antibodies, which was very effective for co-IP of RcPPC4 with RcPPC3 from developing COS extracts (Uhrig et al. 2008) and which is immuno-reactive with Arabidopsis PTPC. Since AtPPC3 and AtPPC4 are both highly expressed in developing Arabidopsis pollen (Sanchez and Cejudo 2003, Igawa et al. 2010), this is the ideal tissue for the attempted co-IP. However, harvesting enough Arabidopsis pollen or anthers for the experiment was not feasible, so developing flower buds were instead used in an
attempt to enrich native AtPPC4 via its co-IP with AtPPC3. SDS-PAGE followed by protein staining of the IP eluate resolved a p110:p107 doublet, but immunoblotting with anti-RcPPC3 demonstrated that both polypeptides are PTPCs (Fig. 3.2) (with the p110 likely corresponding to a monoubiquitinated version of the p107, as previously established with lily pollen PTPC (Igawa et al. 2010)) . However, ~116 kDa anti-AtPPC4 immunoreactive polypeptides were never detected in the IP eluate. The lack of detection may be due to the low protein abundance of AtPPC4; since AtPPC4 expression is pollen-specific AtPPC4 polypeptides may have been diluted by other floral bud proteins during the extraction. Indeed, anti-AtPPC4 immunoblotting of the clarified bud extracts prior to IP failed to detect immunoreactive ~116 kDa AtPPC4 polypeptides (Fig. 3.2). Alternatively, this result could signify that AtPPC4 doesn't strongly interact with AtPPC3, the dominant PTPC isozyme of Arabidopsis pollen. Since AtPPC4 is confirmed to be present in pollen via mass spectrometry (Fig. 3.1) this could also mean that in vivo, AtPPC4 might function without being in a Class-2 PEPC complex, or that it might associate with a non-PEPC protein.

**Heterologous expression of AtPPC4 in *E.coli***

Recombinant AtPPC4 expressed in *E. coli* was documented as being insoluble and inactive (Sánchez et al. 2006, Gennidakis et al. 2007). Thus, different *E. coli* expression strains (BL21, SoluBL21, and Arctic express), and induction temperatures (4, 24, and 37 °C) and times (24, and 48 h) were tested in an attempt to increase the solubility of AtPPC4. BL21 induced for 24 h at 24 °C yielded maximal expression of full-length (*i.e.*., 116 kDa) soluble AtPPC4 polypeptides in clarified *E. coli* lysates as judged via anti-AtPPC4 immunoblotting (Appendix C). However, in contrast to heterologously expressed RcPPC4 (O’Leary et al. 2009),
recombinant AtPPC4 was catalytically inactive as PEPC activity in the soluble fraction was not increased relative to non-induced controls. Recombinant RcPPC4 was stabilized when co-purified in the presence of heterologously-expressed AtPPC3 (O’Leary et al. 2009). This principle was applied in the purification of recombinant AtPPC4. Thus, *E. coli* cells heterologously expressing non-tagged AtPPC4 and His-tagged AtPPC3 were combined in a 2:1 ratio (gFW) prior to lysis using the French press. As seen in Fig. 3.3, the AtPPC3 was highly enriched after being eluted from the subsequent nickel-affinity column with 150 mM imidazole. However, AtPPC4 failed to bind to the resin as it eluted with other proteins in the unbound fractions. The overall results indicate that either: 1) the recombinant, soluble AtPPC4 was misfolded and thus catalytically inactive, while missing proper PTPC-interaction domains, or 2) AtPPC4 is a non-functional PEPC that lacks the PTPC-interaction domains characteristic of RcPPC4.

**AtPPC4 overexpression in Arabidopsis suspension cell cultures and in planta**

Recombinant RcPPC4 readily formed a Class-2 PEPC complex when combined *in vitro* with native AtPPC1 purified from the Arabidopsis cell cultures (O’Leary et al. 2009). Thus, stable overexpression of AtPPC4 was attempted in the Arabidopsis suspension cells, to overcome the potential problem of protein misfolding in *E. coli*. Under the 2x35S promoter, two independently transformed cell lines designated as Tr1 and Tr2 were expressing more *AtPPC4* transcripts than corresponding wild-type cells (Fig. 3.4A), with the Tr2 line being magnitudes higher. Nevertheless, immunoblots of soluble protein extracts from both overexpression lines did not yield detectable levels of immunoreactive ~116 kDa AtPPC4 polypeptides (Fig. 3.4B). A possible explanation as to why no AtPPC4 polypeptides were detected is that AtPPC4 was
immediately degraded due to instability without the presence of sufficient PTPC, which was otherwise dedicated towards Class-1 PEPC formation. Alternatively, since AtPPC1 is the predominant PTPC isozyme expressed in Arabidopsis suspension cell cultures (Gregory et al. 2009) it may be the ‘incorrect’ PTPC for Class-2 PEPC formation (as discussed below in the yeast two hybrid section). A pair of in planta constitutive overexpression lines were also produced, along with an empty vector control. Both AtPPC4 overexpression lines produced more abundant AtPPC4 transcripts than wild type or empty vector lines (Fig. 3.5B). However, similar to AtPPC4 overexpression in suspension cell cultures, full length (i.e., ~116 kDa) immunoreactive AtPPC4 polypeptides were never detected following anti-AtPPC4 immunoblotting of clarified seedling extracts using chromogenic detection methods. Although ECL detection revealed faint immunoreactive anti-AtPPC4 polypeptides (Fig. 3.5D), there was no significant difference in protein abundance between the different plant lines. In the future, it will be necessary to obtain the AtPPC4 knockout mutant to confirm whether the faint immunoreactive band is indeed AtPPC4. In addition, the dual over-expression of AtPPC3 with AtPPC4, might stabilize AtPPC4 as subunits within a Class-2 PEPC complex (i.e. with AtPPC3).

**Yeast two-hybrid analysis of the ability of AtPPC4 and RcPPC4 to interact with the Arabidopsis PTPC isozymes AtPPC1-AtPPC3**

Next the yeast two-hybrid system was employed to assess the ability of AtPPC4 to interact with AtPPC1-AtPPC3. All three Arabidopsis PTPC isozymes were tested to determine if AtPPC4 has a preferential binding partner. As recombinant RcPPC4 strongly interacts with AtPPC3 in vitro (O’Leary et al. 2009), an RcPPC4 construct was also made as a positive control, as well as to determine if RcPPC4 exhibits a preferential affinity for AtPPC3 relative to AtPPC1.
and AtPPC2. In addition, a pair of *C. elegans* proteins, Vab1 (Ephrin receptor-1, used as bait) and NCK (non-catalytic region of a tyrosine kinase, used as prey), were selected as an additional positive control as they yield strong interaction signals in the yeast two hybrid system (Mohamed et al. 2012). Indeed, the Vab1 and NCK positive control yielded such a strong interaction that yeast growth on QDO media was not disrupted (Fig. 3.6B). Auto-activation of the two BTPC proteins was tested against NCK; as these negative controls showed no growth on QDO plates (Fig. 3.6B), no auto-activation of RcPPC4 and AtPPC4 occurred. The yeast two hybrid results indicated that RcPPC4 interacts with all 3 Arabidopsis PTPC isozymes (Fig. 3.6C), corroborating the PTPC-binding promiscuity of this castor BTPC as previously reported (O’Leary et al. 2009). Interestingly, it appears that RcPPC4 might interact better with AtPPC2 and AtPPC3 than with AtPPC1, suggesting that BTPCs may have a preferential PTPC isozyme as their interacting partner within a Class-2 PEPC complex. The apparent stronger interaction of RcPPC4 with AtPPC3 comes as no surprise since AtPPC3 is the ortholog of RcPPC3, the native binding partner of RcPPC4 in the Class-2 PEPC complex of developing castor beans (O’Leary et al. 2011a). In contrast, AtPPC4 did not appear to interact with any of the PTPCs (Fig. 3.6D). However, the original yeast colonies used to prepare the yeast dilutions shown in Fig. 3.6A did in fact grow (although minimally) when AtPPC2 or AtPPC3 were the prey proteins. It is feasible that the Gal4-DNA-BD (16.9 kDa) domain fused to the N-terminus of AtPPC4 hinders its ability to recruit a PTPC as binding partner, resulting its apparent ‘weak’ interaction with AtPPC2 and AtPPC3. This is particularly relevant with RcPPC4 which was expected to exhibit a strong interaction with the various PTPC isozymes, especially AtPPC3, but only a medium interaction was observed (*i.e.*, relative to Vab1 and NCK) (Fig. 3.6C). Fusion of a ~27 kDa green fluorescent protein to RcPPC4’s N-terminus abolished its *in vivo* interaction with RcPPC3.
following their transient co-expression in tobacco suspension cells (Park et al. 2012). The lack of growth of AtPPC4 expressing yeast during the serial dilution test could be due to the fact that AtPPC4 has a lower endogenous affinity for AtPPC3 (relative to RcPPC4), and thus even minor steric hindrance at its N-terminal may block its ability to interact with AtPPC3. It would be ideal to repeat the experiment with the Gal4-domain fused to the C- rather than N-terminal end of the two BTPCs to minimize potential steric hindrance. In addition, one of the drawbacks to the yeast two hybrid system is that it only provides a semi-quantitative determination of protein-protein interaction strengths while being largely dependent on how well bait and prey proteins are respectively expressed. The differences could also be due to the fact that RcPPC4 is expressed better than AtPPC4 in yeast. Future confirmation of bait and prey polypeptide abundance (via immunoblotting) are required before conclusions can be drawn.

If future experiments confirm that RcPPC4 does indeed have a stronger affinity for PTPCs relative to AtPPC4, it is tempting to hypothesize that this is due to significant sequence variation within their respective IDRs (Fig. 2.4), a region that is essential for the ability of RcPPC4 to associate with the surface of mitochondria and recruit RcPPC3 as binding partner in vivo (Park et al. 2012). It is particularly notable that RcPPC4 (as well as cucumber and tomato BTPC) contains a 15 amino acid insertion that makes them ~2 kDa larger than AtPPC4 (Fig. 2.4). Interestingly, plants that mainly exhibit pollen-specific BTPC expression generally appear to be missing this 2 kDa insertion within their IDR (Fig. 2.4) as well. It would thus be interesting to assess the impact of mutating AtPPC4 on its ability to recruit AtPPC3 as a binding partner by: 1) substituting the IDR of AtPPC4 with the IDR of RcPPC4; and 2) inserting the extra 2 kDa portion that is present in the IDR of RcPPC4 (i.e., residues 404-418) into the corresponding position of the IDR of AtPPC4.
**Figure 3.1. Protein quantification of the four Arabidopsis PEPC isozymes from various tissues.**

Quantification was visualized using the pep2pro “TAIR10 wos database” (http://fgcz-pep2pro.uzh.ch/) (Baerenfaller et al. 2011). Tandem mass spectrometry spectral peptide counts are displayed as log2-transformed values (darker colors with higher values are indicative of more abundant peptide counts; AtPPC1, At1g53310; AtPPC2, At2g42600; AtPPC3, At3g14940; AtPPC4, At1g68750).
Figure 3.2. Attempted co-IP of AtPPC4 from Arabidopsis flower buds.
Protein A purified anti-RcPPC3-IgG was coupled to protein-G magnetic beads. The beads were incubated with a clarified Arabidopsis flower bud extract. The clarified extract (25 µg) and IP eluate (500 ng) was subject to SDS-PAGE followed by protein staining with Coomassie Blue R-250 and parallel immunoblotting using anti-RcPPC3 and anti-AtPPC4. Detection of antigenic polypeptides were visualized using a peroxidase conjugated secondary antibody and ECL detection.
Figure 3.3. Attempted co-purification of recombinant AtPPC4 using His-tagged AtPPC3 via metal ion affinity chromatography. BL21(DE3) E. coli cells expressing AtPPC4 or AtPPC3-His were combined in a 2:1 ratio (gFW), before lysing in a French press. The clarified/soluble extract was loaded onto a nickel affinity column. Proteins that flowed through the column were collected and designated as the ‘unbound proteins’. Bound proteins were eluted with 150 mM imidazole and concentrated. Various fractions were subjected to SDS-PAGE followed by immunoblotting using anti-RcPPC3 (A) or anti-AtPPC4 (B). Recombinant Class-2 PEPC (50 ng) containing an equivalent ratio of p118 RcPPC4 subunits and p107 AtPPC3 subunits (O’Leary et al. 2009) was used as a positive control for the PTPC immunoblot. Detection of antigenic polypeptides were visualized using an alkaline phosphatase conjugated secondary antibody and chromogenic detection.
Figure 3.4. *AtPPC4* overexpression in Arabidopsis suspension cell cultures.
Two separate transformations were performed to generate independent transgenic lines (Tr1 and Tr2) for *AtPPC4* overexpression under the 2x35S promoter (*pMDC32* vector). (A) semi-quantitative RT-PCR to determine relative *AtPPC4* expression. Arabidopsis *ACTIN2* was used as the reference gene. (B) Immunoblot for detection of *AtPPC4* and *AtPPC1* (Gregory et al. 2009) in clarified extracts prepared from Tr1, Tr2, and WT cells (20 µg protein lane⁻¹). Recombinant Class-2 PEPC (50 ng protein) (O’Leary et al. 2009) was used as a *AtPPC1* positive control, whereas total cell lysate from *E. coli* expressing *AtPPC4* was used as an *AtPPC4* positive control. Detection of antigenic polypeptides were visualized using an alkaline phosphatase conjugated secondary antibody and chromogenic detection.
Figure 3.5. *AtPPC4* overexpression *in-planta*.
The floral dip method was used to generate the transgenic lines. Two independent transformations were done as replicates (Tr1 and Tr2) for constitutive *AtPPC4* expression under the 2x35S promoter. In addition, an empty vector transformation (E) was used as a negative control. DNA and RNA samples were taken from 7-d old liquid grown seedlings (T2 generation). (A) PCR of gDNA for presence of T-DNA insert to confirm transformation. (B) semi-quantitative RT-PCR to detect expression levels of *AtPPC4*, relative to *ACTIN2*. (C) Immunoblot detection of PTPC and AtPPC4 in clarified extracts prepared from 7-d old liquid grown seedlings (T2 generation) (50 µg protein lane−1). Detection of antigenic polypeptides were visualized using an alkaline phosphatase conjugated secondary antibody and chromogenic detection. (D) Same samples as ‘C’ but detection was done using enhanced chemiluminescence (ECL). Recombinant Class-2 PEPC (50 ng) (O’Leary et al. 2009) was used as a PTPC positive control. Total cell lysate from *E. coli* expressing AtPPC4 was used as an AtPPC4 positive control.
Figure 3.6. Yeast two-hybrid analysis of Arabidopsis versus castor BTPC interactions with various Arabidopsis PTPC isozymes.

DDO, double dropout media lacking leucine and tryptophan was used to screen for presence of both bait and prey plasmids, whereas QDO, quadruple drop out media lacking leucine, tryptophan, alanine and histidine was used to screen for presence of both plasmids and for protein-protein interactions. (A) Eight transformed colonies were randomly selected and incubated on DDO or QDO plates for 7 d at 30 °C to test bait-prey interaction. (B, C and D) Determination of semi-quantitative bait-prey interaction strength via serial dilutions of select yeast colonies. (B) Positive and negative controls for the yeast two hybrid experiment using the Vab1 and NCK proteins from C. elegans. (C and D) Interaction of RcBTPC (C) and AtPPC4 (D) with AtPPC1-AtPPC3, the three PTPC isozymes encoded by the Arabidopsis genome. All dilutions were incubated for 86 h at 30 °C.
Chapter 4 . General Discussion

Diverse functions of BTPC in vascular plants

Results of the preceding chapters indicate that BTPCs from different plant species exhibit remarkably different expression patterns and biochemical properties. For example, the marked dissimilarities in tissue-specific BTPC expression patterns outlined in Chapter 2 indicated that BTPC serves different roles in castor and Arabidopsis plants. It is notable that field mustard, maize, rice, and strawberry BTPCs are all missing an approximate 2 kDa insertion within their respective IDRs, and like Arabidopsis mainly exhibit pollen-specific BTPC expression (Fig. 2.4). This suggests that the unique IDR sequence and expression patterns of different BTPCs may give important clues into their specific functions. It is possible that RcPPC4 requires a higher affinity for PTPC because of the need to form and then recruit a Class-2 PEPC complex to the OMM, due to substantial atmospheric gas exchange restrictions that are likely prevalent in the ‘bulky’ endosperm of developing COS. Interestingly, BTPCs from species such as tomato and cucumber that also develop bulky sink tissues with similar atmospheric gas-exchange restrictions, also contain an analogous ~15 amino acid insert within their IDR (i.e., at the same position as RcPPC4), while forming a Class-2 PEPC complex with co-expressed PTPC subunits (Fig. 2.21). It will be interesting to confirm whether the tomato and cucumber Class-2 PEPCs also localize to the OMM. Immunogold labeling of immature tomato fruit and cucumber embryos using BTPC- and PTPC-specific antibodies coupled with transmission electron microscopy would clarify the subcellular localization of their respective BTPCs and PTPCs, as previously documented for developing COS (Park et al. 2012).
In contrast, Arabidopsis BTPC may play a completely different role than castor BTPC, as suggested by previously established induction of \textit{AtPPC4} transcripts in salt or drought stressed Arabidopsis (Sánchez et al. 2006). The osmotic imbalance imposed by drought or salinity stress leads to loss of turgor, membrane disorganization, protein instability, and production of reactive oxygen species (ROS) (Krasensky and Jonak 2012). Maintenance of cell turgor is essential for sustaining key physiological processes such as stomatal opening, photosynthesis, and cell expansion, which is particularly important for root elongation and water uptake (Zivcak et al. 2016). Various osmolytes and biocompatible solutes accumulate in osmotically stressed plants. In particular, proline rises to very high levels to function as a biocompatible solute that helps maintain turgor, scavenge ROS, and stabilize proteins (Krasensky and Jonak 2012). The accumulation of proline during stress is also universal among eubacteria and protozoa (Zivcak et al. 2016). The anaplerotic production of OAA by PEPC directly supports the biosynthesis of several amino acids, including glutamate, the precursor for proline biosynthesis (Krasensky and Jonak 2012). Elevated glutamate levels have also been documented to stimulate γ-aminobutyric acid (GABA) synthesis via glutamate decarboxylase; GABA also rapidly accumulates during salt stress as a biocompatible osmolyte and a ROS scavenger (Shelp et al. 1999). Most importantly, organic anions downstream of the PEPC reaction (\textit{i.e.}, fumarate and malate) accumulate to high levels in Arabidopsis following salt or drought stress, and contribute up to 25% of the osmotic potential needed for maintaining turgor (Hummel et al. 2010). Once again, \textit{BTPC} expression correlates with the accumulation of malate and provides a compelling rationale for the upregulation of \textit{AtPPC4} under drought and salt stress.

It is therefore hypothesized that the expression of \textit{AtPPC4} under osmotic stress may be a response to the higher demand for organic anions and biocompatible solutes, in which increased
levels of malate and glutamate (PEPC inhibitors) are required to produce the wide range of protective osmolytes (Fig. 4.1). Furthermore, all the biocompatible osmolytes produced downstream of the PEPC reaction function as available carbon reserves that facilitate plant recovery from the osmotic stress (Zivcak et al. 2016). The presence of an ABRE in the promoter of many vascular plant BTPC genes (Fig. 2.2A) may indicate that abiotic stress acclimation may be a common role for BTPC across the plant kingdom.

Interestingly, BTPC may play a similar role in developing pollen. Pollen of numerous species accumulate massive levels of proline which is believed to act as an energy source for the high demands of pollen germination and tube elongation, and for the synthesis of proline rich cell wall proteins (Mattioli et al. 2012). In addition, proline is also postulated to act as a compatible osmolyte in preparation for the dehydration of mature pollen grains (Mattioli et al. 2012). Indeed, proline, aspartate, and glutamate have been documented to comprise upwards to 80% of the free amino acids in developing pollen (Sangwan 1978), creating an environment rich in potent allosteric inhibitors of Class-1 PEPC (i.e. malate, aspartate and glutamate) (O’Leary et al. 2011b). The natural desiccation of pollen grains imposes osmotic stress similar to drought stressed roots, and Class-2 PEPC may help to sustain elevated levels of malate for the biosynthesis of biocompatible osmolytes such as proline. Proline deficient transgenic Arabidopsis plants exhibited decreased pollen viability, highlighting the importance of accumulating this amino acid for proper pollen development (Mattioli et al. 2012).

To summarize, there are multiple reasons why a plant cell would need to accumulate high levels of malate (Fernie and Martinoia 2009). First, malate serves as a key osmolyte for the maintenance of turgor pressure, which is important for stomatal opening in guard cells as well as rapidly expanding tissues such as developing fruit and immature leaves. In ripening fruit, the
accumulated malate eventually serves as an important carbon source for fueling cellular respiration, as well as gluconeogenesis (Etienne et al. 2013, Osorio et al. 2013). Next, malate can be utilized as a precursor for fatty acid synthesis within the plastid. This is particularly advantageous for non-green oilseed, such as castor bean, in which plastidic malate oxidation into pyruvate and hence acetyl-CoA by the combined reactions of NADP-ME and PDC generates all the reducing power (i.e., NAD(P)H) needed for fatty acid elongation (Fig. 2.1). In addition, rapidly expanding cells require fatty acids for production of membrane lipids that are incorporated into the growing plasma membrane and tonoplast. High levels of malate can also be used for environmental stress acclimation. Under limited water supply as would occur during drought or salinity stress, plants will decrease expansion of aerial tissue and allocate carbon to non-photosynthetic organs such as the roots (Hummel et al. 2010). Indeed, Arabidopsis increases its root mass under drought and salinity stress while accumulating high levels of malate and proline to maintain osmotic potential (Hummel et al. 2010). Malate serves the dual purpose of being an osmolyte, as well as being an available carbon source and respiratory substrate during stress recovery (Hummel et al. 2010). In addition, the massive accumulation of proline (and other biocompatible solutes) that accommodates salt stress acclimation and pollen development may result in elevated glutamate levels (another allosteric PEPC inhibitor) since proline is primarily synthesized through glutamate. Osmotically stressed tissues that accumulate such Class-1 PEPC inhibitors would clearly benefit by the presence of a Class-2 PEPC for maintaining anaplerotic C-flux. Regardless of the role, the overall conclusion from results presented in Chapter 2 is that significant BTPC expression occurs in vascular plant tissues characterized by elevated intracellular levels of malate, levels that would otherwise potently inhibit allosterically-sensitive Class-1 PEPCs.
Future directions

My discovery of BTPC and Class-2 PEPC in developing fruit has opened up various doors for future research. One of the drawbacks of castor is that there is currently no efficient castor transformation protocol, making it difficult to test the role of BTPC and Class-2 PEPC in developing COS via RcPPC4 silencing. By contrast, tomato transformation protocols are well established (Sharma et al. 2009, Arshad et al. 2014), so it would be most interesting to assess the impact of RNAi-mediated silencing of tomato BTPC on fruit development and organic acid metabolism. In addition, it will be essential to further clarify the role of BTPC during salinity stress acclimation in plants other than Arabidopsis. Past work with castor BTPC has primarily focused on developing COS, but it is currently unknown if BTPCs that appear dedicated towards malate production for seed development and/or tissue expansion are also important during salinity or drought stress acclimation. As mentioned above, the fact that some plants have more than one BTPC gene may signify that certain BTPC isozymes are dedicated for specific roles. In addition, as discussed in Chapter 3, detailed analysis of the IDR sequence of different BTPCs may allow for prediction of which isozymes are dedicated towards certain purposes.

One of the original goals of my thesis research was to further characterize regulatory BTPC phosphorylation at the highly conserved Ser^{425} and Ser^{451} (relative to RcPPC4 of developing COS) in other plant BTPCs. The original strategy was to obtain Arabidopsis BTPC and check for substrate compatibility with AtCPK4/AtCPK11, RcCDPK1 orthologs which both effectively phosphorylate recombinant RcPPC4 at Ser^{451} (Ying et al. 2016). This goal was not accomplished mostly because of the difficulty in obtaining native AtPPC4, as well as the considerable instability of heterologously expressed AtPPC4 (Chapter 3). However, the high abundance of BTPC polypeptides and thus Class-2 PEPC in immature tomato fruit and cucumber embryos
(Chapter 2) revealed a new source of stable BTPC that can be used for such studies. Analysis of the deduced amino acid sequences of tomato and cucumber BTPC revealed the presence of both phosphosite originally discovered in COS BTPC (Figure 2.3), making them ideal candidates for future BTPC phosphorylation studies. The ease of obtaining tomatoes and cucumbers makes it potentially possible to purify and identify the as yet unknown protein kinase responsible for phosphorylating BTPC at Ser$^{425}$, which has been predicted to be a proline-directed kinase (O’Leary et al. 2011c). Such a discovery would expand our knowledge of BTPC regulation, and ultimately reveal a major control in plant metabolism.

Finally, past attempts at determining the 3-D structure of recombinant Class-2 PEPC using X-ray crystallography have been hindered due to the difficulty of forming suitable crystals (M. Connell, W. Plaxton, and Z. Jia, unpublished research). With the discovery of BTPC polypeptides in a variety of new plants, other Class-2 PEPC orthologs may be more suitable for crystal formation. The comparison of the 3D-crystal structures of a Class-1 PEPC versus Class-2 PEPC would reveal the structural and conformational changes that occur when Class-1 PEPC binds co-expressed BTPC subunits to form the novel Class-2 PEPC hetero-oligomeric complex. In particular, it would be a significant breakthrough to eventually establish a structural model for the architecture of tightly associated BTPC and PTPC subunits within plant and algal Class-2 PEPCs (including the role of the unique IDR of BTPC) and the location of conserved binding domains between the BTPC and PTPC subunits.
Figure 4.1. Model of BTPC and thus Class-2 PEPC function in osmotically stressed plants. Osmotic stress requires elevated levels of organic anions and other biocompatible solutes for maintenance of osmotic potential. Free amino acids (particularly proline and GABA) along with malate and fumarate have all been documented to accumulate in various plants subjected to osmotic stress. Elevated malate levels would likely result in potent inhibition of Class-1 PEPC activity (O’Leary et al. 2011b). However, the allosterically desensitized Class-2 PEPC complex has been hypothesized to sustain significant anaplerotic OAA production under high malate conditions (O’Leary et al. 2011a, b); the OAA can then be used for biosynthesis of various osmolytes/biocompatible solutes such as proline and fumarate that are downstream of the PEPC reaction. Red dotted lines indicate feedback allosteric inhibition, whereas green dotted lines indicate allosteric activation. Abbreviations: MDH, malate dehydrogenase; AAT, aspartate amino transferase; P5CS, pyrroline-5-carboxylate synthetase; P5CR, pyrroline-5-carboxylate reductase; GDC, glutamate decarboxylase.
Figure 4.2. Model proposing diverse functions for Class-2 PEPC in tissues with elevated levels of malate.

PEPC has been implicated in numerous key processes in plant cells, including storage protein synthesis and N-assimilation (O’Leary et al. 2011b). However, certain tissues require elevated intracellular levels of malate (i.e. >50 mM in some cases), which would essentially inactivate allosterically-sensitive Class-1 PEPCs (see Table 2.3). Under these conditions, maintenance of anaplerotic PEP carboxylation to OAA may require BTPC expression, and hence Class-2 PEPC complex formation. The increased malate supply can serve a variety of purposes.
Literature Cited


Phosphoenolpyruvate carboxylase from C4 leaves is selectively targeted for inhibition by anionic phospholipids. Plant Physiology 152:634–638.


Appendix A: List of Primers

1) Primers used for sequencing full length constructs. *NOTE* Two of the primers used for sequencing *AtPPC2* have 1 bp that don’t match the gene sequence (indicated in red). These mismatches were designed in the middle and to not affect sequencing results. As such, the same sequencing primer could be used for sequencing all three Arabidopsis PTPC isozymes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>AtPPC1</em></td>
<td>1) Forward T7 promoter: 5'-TAATACGACCTCACTATAGGG-3'</td>
</tr>
<tr>
<td>GenBank: BT000647.1</td>
<td>2) Forward primer: 5'-ATTCCAGTTCTCTTCTTGATGGG-3'</td>
</tr>
<tr>
<td></td>
<td>3) Forward primer: 5'-GATCTTTCTAAAAACGAAGA-3'</td>
</tr>
<tr>
<td></td>
<td>4) Forward primer: 5'-ATTGATCTAATCGAAATGGT-3'</td>
</tr>
<tr>
<td></td>
<td>5) Reverse primer: pGADT7 multiple cloning site 5'-AAATTGAGATGTTGCAGATG-3'</td>
</tr>
<tr>
<td><em>AtPPC2</em></td>
<td>1) atppc4-NdeI-CT-Forward: 5'-TATATACATATGACGGAACACACAGACGA-3'</td>
</tr>
<tr>
<td>GenBank: AJ532902.1</td>
<td>2) atppc4-promoter-Reverse: 5'-CCTCAATACCCAGCCATACG-3'</td>
</tr>
<tr>
<td></td>
<td>3) atppc4-forward-middle: 5'-GGTTCTTGGATGGGAGGTG-3'</td>
</tr>
<tr>
<td></td>
<td>4) atppc4-middle-Reverse: 5'-GGGACTGAATAGCGAGATAG-3'</td>
</tr>
<tr>
<td></td>
<td>5) atppc4-forward-starts at 1944: 5'-ACCATGTCCCTGGTGGAGCGCTA-3'</td>
</tr>
<tr>
<td></td>
<td>6) atppc4-forward-starts at 2562: 5'-CCTCAATATAGGAAGCCGACCA-3'</td>
</tr>
<tr>
<td></td>
<td>7) atppc4-reverse-Sall: 5'-TATAGTCACTTTAACCAGGTATTTCTCATTCCTGC-3'</td>
</tr>
<tr>
<td><em>AtPPC3</em></td>
<td>Construct obtained previously described</td>
</tr>
<tr>
<td>NCBi: NM_112356.4</td>
<td>1) Forward T7 promoter: 5'-TAATACGACCTCACTATAGGG-3'</td>
</tr>
<tr>
<td></td>
<td>2) Forward primer: 5'-ATTCCAGTTCTCTTCTTGATGGG-3'</td>
</tr>
<tr>
<td></td>
<td>3) Forward primer: 5'-GATCTTTCTAAAAACGAAGA-3'</td>
</tr>
<tr>
<td></td>
<td>4) Forward primer: 5'-ATTGATCTAATCGAAATGGT-3'</td>
</tr>
<tr>
<td></td>
<td>5) Reverse primer: pGADT7 multiple cloning site 5'-AAATTGAGATGTTGCAGATG-3'</td>
</tr>
</tbody>
</table>

*RcBTPC/RcPPC4*  GenBank: EF634318

Construct obtained previously described
2) Primer pairs used for cloning and RT-PCR. Restriction sites are underlined.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Primer pair</th>
</tr>
</thead>
</table>
| AtPPC1-pGADT7      | Forward primer: contains 19 base pairs found upstream the NdeI site in pGADT7  
| (Gibson Assembly)  | 5’-CGACGTACCGATTACGCTCATATGGCGAATCGGAGTTGAG-3’  
|                    | Reverse primer: contains 17 base pairs found downstream the BamHI site in pGADT7  
|                    | 5’-TGCACTGCTGTCGATGGATCCTTAACCGGTGTGTTTGATGCACC-3’  |
| AtPPC2-pGADT7      | Forward primer: contains 19 base pairs found upstream the NdeI site in pGADT7  
| (Gibson Assembly)  | 5’-CGACGTACCGATTACGCTCATATGGCGAATCGGAGTTGAG-3’  
|                    | Reverse primer: contains 17 base pairs downstream the BamHI site in pGADT7  
|                    | 5’-TGCACTGCTGTCGATGGATCCTTAACCGGTGTGTTTGATGCACC-3’  |
| AtPPC3-pGADT7      | Forward primer: contains NdeI site  
| (Directional Cloning) | 5’-TATAGTACAGGCACGACAAGACGA-3’  
|                    | Reverse primer: contains XhoI site  
|                    | 5’-TATAGTACAGGCACGACAAGACGA-3’  |
| AtBTPC-pGBK7       | Forward primer: contains NdeI site  
| (Directional Cloning) | 5’-TATAGTACAGGCACGACAAGACGA-3’  
|                    | Reverse primer: contains SalI site  
|                    | 5’-TATAGTACAGGCACGACAAGACGA-3’  |
| RcBTPC-pGBK7       | Forward primer: contains NdeI site  
| (Directional Cloning) | 5’-TATAGTACAGGCACGACAAGACGA-3’  
|                    | Reverse primer: contains SalI site  
|                    | 5’-TATAGTACAGGCACGACAAGACGA-3’  |
| AtBTPC-pET30a      | Forward primer: contains NdeI site  
| C-terminal-his-tag | 5’-TATAGTACAGGCACGACAAGACGA-3’  
| (Directional Cloning) | Reverse primer: contains BamHI  
|                    | 5’-TATAGTACAGGCACGACAAGACGA-3’  |
| AtBTPC-pMDC32      | Forward primer: contains BamHI  
| (Directional Cloning) | 5’-TATAGTACAGGCACGACAAGACGA-3’  
|                    | Reverse primer: contains SalI  
|                    | 5’-TATAGTACAGGCACGACAAGACGA-3’  |
| Arabidopsis Actin2:| Forward primer: contains  
| Product size 108bp | Product size 108bp  
| (RT-PCR)           | 5’-GCCAACACTTGTGCTAGTGGG-3’  
|                    | Reverse primer: contains  
|                    | 5’-GCCAACACTTGTGCTAGTGGG-3’  |
| Arabidopsis AtBTPC:| Forward primer: contains  
| Product size 565bp | Product size 565bp  
| (RT-PCR)           | 5’-GCCAACACTTGTGCTAGTGGG-3’  
|                    | Reverse primer: contains  
|                    | 5’-GCCAACACTTGTGCTAGTGGG-3’  |
| pMDC32 t-DNA insert| Forward primer: contains  
| Product size 309 bp| Product size 309 bp  
|                   | 5’-GCCAACACTTGTGCTAGTGGG-3’  
|                   | Reverse primer: contains  
|                   | 5’-GCCAACACTTGTGCTAGTGGG-3’  |
Appendix B: RNA-seq/microarray experiments and RNA-seq accession numbers.

1) The type of data (microarray or RNA-seq) and the studies that generated the data are included in the table.

<table>
<thead>
<tr>
<th>Species</th>
<th>Data Type(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>RNA-seq and microarray</td>
<td>(Winter et al. 2007)</td>
</tr>
<tr>
<td>Avocado (Persea americana)</td>
<td>RNA-seq</td>
<td>(Ibarra-Laclette et al. 2015) (Kilaru et al. 2015)</td>
</tr>
<tr>
<td>Cassava (Manihot esculenta)</td>
<td>RNA-seq</td>
<td>(Wang et al. 2014)</td>
</tr>
<tr>
<td>Castor (Ricinus communis)</td>
<td>RNA-seq</td>
<td>(Brown et al. 2012)</td>
</tr>
<tr>
<td>Cotton (Gossypium raimondii)</td>
<td>RNA-seq</td>
<td>(Hovav et al. 2015) (Paterson et al. 2012) (Rambani et al. 2014) (Renny-Byfield et al. 2015) (Li et al. 2014)</td>
</tr>
<tr>
<td>Cucumber (Cucumis sativus)</td>
<td>RNA-seq</td>
<td>(Li et al. 2011)</td>
</tr>
<tr>
<td>Field mustard (Brassica rapa)</td>
<td>RNA-seq</td>
<td>(Tong et al. 2013) (Devisetty et al. 2014)</td>
</tr>
<tr>
<td>Grape (Vitis vinifera)</td>
<td>Microarray</td>
<td>(Fasoli et al. 2012)</td>
</tr>
<tr>
<td>Maize (Zea mays)</td>
<td>RNA-seq</td>
<td>(Chen et al. 2014)</td>
</tr>
<tr>
<td>Orange (Citrus sinensis)</td>
<td>RNA-seq</td>
<td>(Martinelli et al. 2012)</td>
</tr>
<tr>
<td>Poplar (Populus trichocarpa)</td>
<td>Microarray</td>
<td>(Wilkins et al. 2009)</td>
</tr>
<tr>
<td>Rice (Oryza sativa)</td>
<td>Microarray</td>
<td>(Li et al. 2007)</td>
</tr>
<tr>
<td>Soybean (Glycine max)</td>
<td>RNA-seq</td>
<td>(Collakova et al. 2013)</td>
</tr>
<tr>
<td>Strawberry (Fragaria vesca)</td>
<td>RNA-seq</td>
<td>(Hollender et al. 2014) (Kang et al. 2013)</td>
</tr>
<tr>
<td>Tomato (Solanum lycopersicum)</td>
<td>RNA-seq and microarray</td>
<td>(Sato et al. 2012) (Osorio et al. 2011) (Pressman et al. 2002)</td>
</tr>
</tbody>
</table>
2) List of RNA-seq experiments and accession numbers used for bowtie alignments. The “n/a” indicates that the data were directly taken from the supplementary files of the corresponding publication, or visualized directly on the UofT BAR using eFP browsers. In general, the gene annotation identifiers were obtained from the JGI Phytozome V11.0 database or from the UofT BAR database. For *Brassica rapa*, the alignment was done using the gene annotations from the plant genome database (http://www.plantgdb.org/), which was derived from an older version of the genome (Phytozome v8.0/Brassica v1.2). For the protein alignments, the updated gene models were obtained from the JGI Phytozome v11.0. For avocado, the genome and gene annotations are not publicly available, so the *BTPC* identifier was derived from the supplementary file provided by the corresponding RNA-seq study. RPKM values were calculated using the following formula: \( \text{RPKM} = \frac{10^9 \times C}{N \times L} \), where ‘C’ is the number of reads mapped to a gene, ‘N’ is the total mapped reads that aligned to the reference genome, and ‘L’ is the exon length for a gene.

<table>
<thead>
<tr>
<th>Species</th>
<th>BTPC gene identifier(s)</th>
<th>SRA experiment #</th>
<th>RNA-seq run #</th>
<th>Tissue source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>At1g68750(BAR)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td><em>Avocado (Persea americana)</em></td>
<td>UN00351 (Barra-Laclette et al. 2015) For full gene sequence, look in supplementary file from Kilaru et al. (2015) for contig ID:PA10006483</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td><em>Cassava (Manihot esculenta)</em></td>
<td>Manes.14G095900(JGI) Manes.06G074600(JGI)</td>
<td>SRP042139</td>
<td>SRR1299005</td>
<td>root-early</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR1299006</td>
<td>root-mid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR1299007</td>
<td>root-late</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR1299008</td>
<td>stem</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR1299009</td>
<td>leaf</td>
</tr>
<tr>
<td><em>Castor (Ricinus communis)</em></td>
<td>29738.m001022(JGI) For full mRNA sequence obtained from Genbank (EF634318.1)</td>
<td>ERP000877</td>
<td>ERR044267</td>
<td>stage 3 COS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ERR044268</td>
<td>stage 5 COS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ERR044269</td>
<td>germinating seed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ERR044270</td>
<td>leaf</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ERR044271</td>
<td>male dev. flowers (pollen)</td>
</tr>
<tr>
<td><em>Cotton (Gossypium raimondii)</em></td>
<td>Gorai.008G187900(JGI)</td>
<td>SRP017168</td>
<td>SRR617008</td>
<td>ovary 10dpa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRP017133</td>
<td>SRR617010</td>
<td>ovary 20dpa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRP028270</td>
<td>SRR617012</td>
<td>ovary 30dpa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRP009820</td>
<td>SRR616257</td>
<td>ovary 40dpa</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR943773</td>
<td>3 cm seedlings</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR389183</td>
<td>petals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mature leaves</td>
</tr>
<tr>
<td><em>Cucumber (Cucumis sativus)</em></td>
<td>Cucsa.280800(JGI)</td>
<td>SRP008779</td>
<td>SRR351489</td>
<td>ovary-fertilized</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351495</td>
<td>ovary-unfertilized</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351499</td>
<td>root</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351905</td>
<td>stem</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351906</td>
<td>leaf</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351908</td>
<td>male flower</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351910</td>
<td>tendril</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351911</td>
<td>tendril base</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR351912</td>
<td>female flower</td>
</tr>
<tr>
<td><em>Field mustard (Brassica rapa)</em></td>
<td>Brara.I03103(JGI) also equivalent to Brat024730 on PlantGDB</td>
<td>SRP017757</td>
<td>SRR643621</td>
<td>callus tissue</td>
</tr>
<tr>
<td></td>
<td>Brara.G02500(JGI) also equivalent to</td>
<td>SRP021896</td>
<td>SRR643622</td>
<td>flower tissue</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR643623</td>
<td>leaf(7week)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR643625</td>
<td>root(7week)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR643627</td>
<td>silique(15DAP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRR643628</td>
<td>stem tissue</td>
</tr>
</tbody>
</table>

103
<table>
<thead>
<tr>
<th>Plant</th>
<th>Gene ID/Accession</th>
<th>SRR Numbers</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grape (Vitis vinifera)</td>
<td>VIT_01S0011G02740(BAR) Gene sequence obtained from Uniprot</td>
<td>SRR837574, SRR837575, SRR837578, SRR837581, SRR837582, SRR837587, SRR837605, SRR837634</td>
<td>seedling, silique, root, internode, leaf, floral meristem, apical meristem, petiole</td>
</tr>
<tr>
<td>Maize (Zea mays)</td>
<td>GRMZM2G082780(JGI)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Orange (Citrus sinensis)</td>
<td>orange1.1g0015377m(JGI)</td>
<td>SRR867426, SRR867442, SRR867394, SRR867398</td>
<td>leaf-young, leaf-mature, fruit peel-immature, fruit peel-mature</td>
</tr>
<tr>
<td>Poplar (Populus trichocarpa)</td>
<td>POTRL010G131800.1(BAR) POTRL008G114200.1(BAR) Gene sequences obtained from JGI using same ID#</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Rice (Oryza sativa)</td>
<td>LOC_Os01g02050(JGI)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Soybean (Glycine max)</td>
<td>Glyma.02G130700(JGI) Glyma.01G091000(JGI) Glyma.10G205500(JGI)</td>
<td>SRR830182, SRR830185, SRR830188, SRR830191, SRR830194, SRR830197, SRR830200, SRR830203, SRR830206, SRR830209</td>
<td>developing seeds 5 DPA, developing seeds 10 DPA, developing seeds 15 DPA, developing seeds 20 DPA, developing seeds 25 DPA, developing seeds 30 DPA, developing seeds 35 DPA, developing seeds 40 DPA, developing seeds 45 DPA, developing seeds 55 DPA</td>
</tr>
<tr>
<td>Strawberry (Fragaria vesca)</td>
<td>gene04466-v1.0-hybrid(JGI)</td>
<td>SRR674026, SRR674028, SRR674030, SRR674032, SRR674034, SRR674036, SRR674038, SRR674040, SRR674042, SRR674044, SRR674046, SRR674048, SRR674050, SRR674052, SRR674054, SRR674056, SRR674058, SRR674060, SRR674062, SRR674064, SRR674066, SRR674068, SRR674070, SRR674072, SRR674074, SRR1106125, SRR1106894, SRR1106896, SRR1106902, SRR1107179, SRR1107189</td>
<td>ovule1, ghost3, ghost4, ghost5, embryo3, embryo4, embryo5, wall1, wall2, wall3, wall4, wall5, pith1, pith2, pith3, pith4, pith5, cortex1, cortex2, cortex3, cortex4, cortex5, seedling1, seedling2, leaf, Anther 6-7A, Anthers 6-7B, Anther 7-8A, Anther 7-8B, Anther 9 A, Anther 9B</td>
</tr>
<tr>
<td>Sample ID</td>
<td>Sample Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107611</td>
<td>Anther 10A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107612</td>
<td>Anther 10B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107613</td>
<td>Anther 11A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107639</td>
<td>Anther 11B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107952</td>
<td>Anther 12A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107953</td>
<td>Anther 12B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107955</td>
<td>Carpel 7-8A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107956</td>
<td>Carpel 7-8B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107977</td>
<td>Carpel 9A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107978</td>
<td>Carpel 9B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107979</td>
<td>Carpel 10A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107980</td>
<td>Carpel 10B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107981</td>
<td>Carpel 12A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107982</td>
<td>Carpel 12B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107983</td>
<td>Pollen A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107984</td>
<td>Pollen B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107991</td>
<td>Microspore 10A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107992</td>
<td>Microspore 10B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107993</td>
<td>Flower 1-4A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107994</td>
<td>Flower 1-4B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107995</td>
<td>Perianther 5-6A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107996</td>
<td>Perianther 5-6B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107997</td>
<td>Receptacle 6-7A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRR1107998</td>
<td>Receptacle 6-7B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tomato (*Solanum lycopersicum*)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Sample Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solyc04g006970.2(JGI)</td>
<td>root upper stem, expanded leaf (2-3 w old) flower- full anthesis fruit - mature green stage fruit -breaker stage fruit- ripe stage fruit (I) fruit (M) fruit (B) fruit (R)</td>
</tr>
<tr>
<td>SRP009216</td>
<td>mature pollen</td>
</tr>
<tr>
<td>SRP008367</td>
<td>mature pollen</td>
</tr>
<tr>
<td>SRP055068</td>
<td>mature pollen</td>
</tr>
</tbody>
</table>
Appendix C: Expression and purification of recombinant AtPPC4

*E. coli* cells were chemically transformed with *pET29a-AtPPC4*. Total cell lysate for IPTG induced BL21 cells (+) and non-induced cells (-) confirm production of recombinant AtPPC4 (116 kDa). Transformed BL21 and SoluBL21 cells were induced and cultured at 24 °C for 24 or 48 h before cell lyses (Cells grown at 4 °C and the Arctic express cells did not produce noticeable levels of recombinant AtPPC4). The clarified extracts were subjected to SDS-PAGE followed by (A) Coomassie Blue R-250 staining and (B)immunoblotting with anti-AtPPC4. Detection of antigenic polypeptides were visualized using an alkaline phosphatase conjugated secondary antibody and chromogenic detection.