Role of the plastid envelope membrane in integrating the plastid into cellular metabolic networks

By

Andrea Bräutigam

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Genetics

2008

Abstract

Role of the plastid envelope membrane in integrating the plastid into cellular metabolic networks

by Andrea Bräutigam

Plastids are the defining organelle of the Archaeplastida which include all land plants. They can differentiate into several subtypes in a tissue dependent manner, for example, brightly colored chromoplasts in bell pepper fruits, green chloroplasts in leaves, or colorless proplastids in meristems, and each subtype is adapted to the cell it resides in. All plastid types of land plants are separated from the cytosol by two membranes, the inner envelope and the outer envelope membrane, and metabolites and signals have to cross these envelope membranes to connect to the remainder of the cell. This work is focused on integrating the plastids into the metabolic network of their respective cells; the adaptations of envelope membranes of different plastid subtypes were analyzed by comparative proteomics.

Initially, envelope proteomics of plants without a sequenced genome or transcriptome was established with the garden pea *Pisum sativum* as the new model. A novel sequencing technology, pyrosequencing, was used to create a transcriptome database which is suitable for proteomics applications. Data generated with pea leaf chloroplasts served as the template to which other differentiated plastid types were compared both qualitatively and semi-quantitatively.

The qualitative comparison of chloroplast and proplastid envelopes revealed specific adaptations of the proplastid envelope to its role in meristematic cells: Proplastids serve as cellular factories for amino acids, fatty acids and nucleotides for the proliferating cells. The transport protein complement is geared to import precursor metabolites and export products from the heterotrophic plastid. Chloroplast envelopes of pea were also compared to those of maize mesophyll cells in a semi-quantitative manner. Maize plants employ a specific subtype of photosynthesis called C4 photosynthesis which causes immense metabolite fluxes across the chloroplast envelope. The comparison revealed quantitative changes in the envelope protein composition indicating that the flux is accommodated by increased amounts of transport protein. It also revealed new candidate transport proteins for metabolite fluxes of C4 chloroplasts. One of these candidate transport proteins was characterized in more detail in the model plant *Arabidopsis thaliana* and it likely is a monocarboxylate transporter employed in photorespiration in C3 plants. Finally, proteome samples of the plastid associated membranes, the inner and the outer envelope of chloroplast and endomembrane system were investigated. GFP fusion protein analyses of candidate outer envelope residents demonstrated that the outer envelope is a dynamic system capable of producing extensions of the envelope membrane as well as vesicles. Plastid associated membranes are hypothesized to be part of the autophagy system for chloroplasts since they contain stromal proteins in addition to envelope and endomembrane system residents. The hypothesis is supported by GFP fusion protein analysis.

DEDICATED TO MY FAMILY

ACKNOWLEDGEMENTS

I would like to thank Dr. Andreas Weber for allowing me to work on exiting projects in his lab and for sharing his passion for original research.

I would also like to thank my guidance committee members, Dr. Kenneth Keegstra, Dr. Christoph Benning and Dr. Yair Shachar-Hill for advice and encouragement.

I would also like to thank my colleagues in the 'Weber Lab', past and present, my fellow students who became colleagues and then friends and my family for their continuing support.

Special thanks to the directors and administrator of the PhD program in genetics at Michigan State University.

Table of contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and literature review

Plants are the primary source of food for humankind and are increasingly considered for biofuel production. Since the arable land area is decreasing due to destruction of arable land for example by salt accumulation, climate change, or urbanization, supplying the human population with food is one of the challenges of the 21st century. Understanding the fundamental processes in plant cells will aid the development of crops with higher yields or altered nutrient composition as well as engineering of more efficient crops both for biofuel and food production. I focus my attention on plastids since they are involved in either synthesis, storage, or both, of all major agricultural products which are derived from assimilated carbon, carbohydrates, fatty acids, or amino acids.

Plastid differentiation

A plant contains several different types of plastids which can be classified according to their morphology and physiological role. In each plant, plastid differentiation starts from undifferentiated proplastids of embryonic and meristematic tissues (Hoober, 2006). The proplastids of meristematic tissues differentiate into chloroplasts, the best studied plastid type, upon exposure to light. Chloroplasts are the most abundant plastid type in most photosynthesizing plants and harness the energy of light to fix inorganic carbon. If light is withheld experimentally or by soil cover in the field, a developmental program causes proplastid differentiation into etioplasts defined by their developmental state and particular ultrastructure which subsequently develop into chloroplasts if light exposure eventually occurs (Holm and Deng 1999). Chloroplasts are the only plastid type which is autotrophic for organic carbon, energy and reducing

equivalents whereas all other plastid types depend on the import of precursor metabolites, energy, and reducing power. Chloroplast themselves can differentiate even further into two subtypes in plant species which employ C4 photosynthesis to fix inorganic carbon (Hall et al. 1998; Cribb et al. 2001). Both proplastids and chloroplasts can differentiate into chromoplasts. Chromoplasts are defined based on their color as yellow or reddish plastids which contain enhanced carotenoid levels (Kirk and Tilney-Bassett 1978). The differentiation of proplastids into chromoplasts has been studied in the natural cauliflower mutant *Orange* (Crisp et al. 1975) which is caused by a mutation in a DnaJ cysteine rich domain protein targeted to colorless plastids (Lu et al. 2006). Although the exact mechanism of chromoplast differentiation remains unresolved it likely involves a protein identified as promoting chromoplast development in bell pepper (Hugueney et al. 1995) which is upregulated in the Orange mutant of cauliflower (Lu et al. 2006). Upon differentiation of chloroplasts into chromoplasts extensive vesiculation of the envelope occurs (Hugueney et al. 1995). Leucoplasts are defined as colorless mature plastids and include amyloplasts in roots. These plastids have the capability to green upon light exposure which induces the formation of thylakoids (Ljubicic et al. 1998). Plastids in storage organs are generally white heterotrophic plastids which either store starch (in plants which rely on starch as the reserve material in the developing seedling) or produce fatty acids for triglyceride production (e.g. in *Brassica napus* seeds) or amino acids (for plants which mainly store reserves as proteins). Aside from any specific roles depending on the plastid subtype, e. g. color production and display in chromoplasts or providing rapidly dividing small plastids ready to differentiate in proplastids (Kirk and Tilney-Bassett 1978), all functional plastids can provide their cells with essential metabolites.

Biosynthetic pathways of plastids

All plastids are semiautonomous organelles and cannot be generated de novo but only by division of plastids. They entered the plant lineage in a single endosymbiosis event and have since lost most of their gene content to the nucleus. However most of the anabolic pathways remain localized in the plastids of plant cells. A possible reason for retaining a large number of pathways inside the plastid may be that plastids provide a subcompartment with different reductant and ATP concentrations compared to the remainder of the cell (Heineke et al. 1991).

Chloroplasts are the only autotrophic plastid type as they can transform light energy into chemical energy. The electron transfer chain located within an internal membrane system, the thylakoids, generates energy in the form of ATP and reducing power in the form of NADPH. Both are used to fix inorganic carbon and nitrogen into organic molecules. Energy and reducing power generated and stored in organic molecules is the basis for the anabolic reactions of plants.

Most of the amino acids produced in plants are produced in the plastids. Branched chain amino acids are synthesized in plastids whereas turnover and breakdown is restricted to the mitochondria and the peroxisomes (Binder et al. 2007). Aromatic amino acids are also exclusively synthesized in the plastid via a common intermediate chorismate (Herrmann and Weaver 1999). Chorismate and the aromatic amino acids are precursors for many products of the secondary metabolism such as phylloquinones (Shimada et al. 2005) and– via phenylalanine - salicylic acid, flavonoids and lignins (Weaver and Herrmann 1997). Arginine synthesis is also considered to be a plastidic pathway (Slocum 2005) as is lysine synthesis (Hudson et al. 2005). A number of essential

plant metabolites synthesized in the plastid are derived from 5-ribosyl 1-pyrophosphate. The amino acids tryptophan (Zhao and Last 1995) and histidine (Stepansky and Leustek 2006) are synthesized in the plastid as are purines and pyrimidines with the exception of one reduction step in pyrimidine biosynthesis (Zrenner et al. 2006). Fatty acids are also mainly synthesized in the plastids (Rawsthorne 2002) and serve as precursors for the synthesis of a number of membrane lipids (Benning et al. 2006). The production of a large range of very different metabolites in plastids raises the question how the plastids efficiently import the necessary precursors and export the products. For many agriculturally important products, the elucidation of their biosynthetic pathways has initially been focused on studying the soluble components rather than the transport processes connecting the metabolic network.

Plastid architecture

Plastids are derived from an ancient endosymbiotic event. They can only multiply by division of existing plastids and contain a rudimentary genome as well as transcription and translation machinery. Most of the plastidic proteins are synthesized on cytosolic ribosomes and imported posttranslationally (Leister, 2003). Plastids contain an aqueous phase, the stroma, which is separated from the cytosol by two membranes, the outer envelope membrane which faces the cytosol and the inner envelope membrane which faces the stroma. The stroma contains an additional membrane system, which is well developed in chloroplasts and called thylakoids or a set of vesicles in most other plastid types. The outer envelope is believed to be relatively permeable for solutes smaller than 10kDa (Weber et al. 2005; Weber and Fischer 2007) by way of a number of porins with broad substrate specificity (Pohlmeyer et al. 1997; Pohlmeyer et al. 1998; Bolter et al.

1999; Goetze et al. 2006). Proteins targeted to the plastids posttranslationally pass the outer envelope through an elaborate protein translocation system called Translocon outer envelope of chloroplast or Toc complex and pass the inner envelope through the Translocon inner envelope of chloroplasts or Tic complex (Schnell 2000; Jarvis and Robinson 2004; Soll and Schleiff 2004). Small solutes pass the inner envelope by way of a number of specific metabolite transport proteins (Weber 2004; Weber et al. 2005). The only plastid envelope which has been analyzed comprehensively is the chloroplast envelope of dicot plants such as spinach and Arabidopsis (Ferro et al. 2002; Miras et al. 2002; Ferro et al. 2003; Froehlich et al. 2003). Studies of other differentiated plastid types have been focused on the soluble proteome although occasionally envelope proteins have also been identified (Baginsky et al. 2004; Kleffmann et al. 2004; von Zychlinski et al. 2005; Siddique et al. 2006; Kleffmann et al. 2007).

Project summary and rationale

The long term goal of studying plastid envelopes is to tie the plastid into the different regulatory and metabolic networks of the cell since the ultimate barrier of both metabolites and signaling molecules which originate in the stroma or at the thylakoids are the inner and outer envelope membranes. Evidence based on proteome projects focused on either the thylakoid proteome of chloroplasts, the soluble proteome of chloroplasts or the entire plastid has indicated that, in addition to housekeeping proteins, the plastid carries a protein complement adapted to its specific role (Baginsky et al. 2004; Kleffmann et al. 2004; von Zychlinski et al. 2005; Siddique et al. 2006; Kleffmann et al. 2007). Proteome analysis has mostly been focused on the soluble or thylakoid proteome of different plastid types (Majeran et al. 2005; Majeran et al. 2008) since both are easier

to access and analyze than the envelope proteome. Envelope proteomes have only been analyzed in chloroplasts of C3 plants. In this work, I report about dynamic envelope membranes with respect to both protein content and protein localization. A new reference chloroplast envelope of *Pisum sativum* chloroplasts is introduced and compared qualitatively to a proplastid envelope yielding new insights in proplastid metabolism. The reference chloroplast proteome is also compared semi-quantitatively to the envelope of C4 chloroplasts which reveals that protein amount is adjusted depending on the metabolic fluxes required for each plastid type (for a detailed description of C4 photosynthesis, please see chapter 2). The characterization of accumulation patterns for known metabolite transport proteins of the chloroplast envelope allows the determination of several candidate transport proteins for known metabolite fluxes. One of these candidate transport proteins named Mep1 was characterized in more detail. Finally, *Pisum sativum* chloroplast envelopes and associated membrane as well as a light microsome fraction are used to gain insight in the dynamics of protein localization in the membrane systems surrounding chloroplasts.

Chapter 2 reviews the current status of knowledge about transport processes in C4 photosynthesis. C4 photosynthesis requires two different chloroplast types to achieve a more efficient photosynthesis compared to that of C3 plants and has been targeted in bioengineering efforts to achieve higher yields, for example in rice plants. The review identifies the transport proteins of one of the understudied fields in C4 research which may hamper engineering efforts.

Chapter 3 reports the establishment of a novel model for plastid envelope proteomics, *Pisum sativum*. Prior to this study, chloroplast proteomics was mainly

performed with species that already had a sequenced genome (e.g. Arabidopsis or rice) or transcriptome (e.g. maize) available. Pea as a model for plastid proteomics has several advantages over the models used previously, because large quantities of high quality isolation material can be grown in a short time. There are isolation protocols not only for leaf chloroplasts but also for root leucoplasts available. Moreover, the outer and the inner envelope of pea chloroplasts can be separated efficiently. Although it has been generally assumed that cross species identification is sufficient for organellar proteomics (Baginsky et al. 2004; Siddique et al. 2006), I identify cross species comparisons as a major obstacle to proteome analysis and analyze the parameters limiting it. In addition, I determine that a species-specific transcriptome database suitable for organellar proteome analysis can be generated by pyrosequencing.

In chapter 4, I report the first envelope proteome analysis of proplastids and determine the metabolic role of proplastids based on their envelope protein content. Proplastid envelopes have never been studied since material of meristems is limited in model organism such as Arabidopsis or rice. Electron micrographs of proplastids show that they contain a rudimentary internal membrane system of unknown composition and frequently a small starch granule (Kirk and Tilney-Bassett 1978; Journet and Douce 1985). Proteome analysis of BY-2 plastids which represent a meristem-like system with undifferentiated plastids indicates that these plastids are enriched in enzymes for amino acid biosynthesis (Baginsky et al. 2004). To address proplastid biology, I analyze proplastid envelopes by proteomics and compare them to chloroplast envelopes. The proteins identified are different from those in pea envelopes and point towards the proplastid as an active cellular factory which produces amino acids, fatty acids and

nucleotide precursors for the remainder of the proliferating cell. They contain not only a specific subset of transport proteins but different subsets of thylakoid proteins and components of the protein import complex compared to leaf chloroplast envelopes. The surprising identification of the triosephosphate phosphate translocator at least partially redefines the role of the glucose-6-phoshate phosphate translocator and the oxidative branch of the pentosephosphate pathway in heterotrophic plastids.

Chapter 5 reports the identification of transport protein amount as one of the ways to control metabolite traffic across the inner and outer envelope. C4 photosynthesis is a highly compartmented process which involves chloroplasts in two different cell types. It causes high volumes of metabolite traffic in excess of those observed in C3 plants. When this work was started, the adaptations of the chloroplast envelopes to the increased volume of traffic were unknown. To address this question, the first detailed semi-quantitative picture of a C3 chloroplast envelope was produced and compared to the chloroplast envelope of a C4 plant. The comparison indicated major changes in quantity of membrane transport proteins of both the inner and the outer envelope and provided the first indication that membrane flux is regulated at least in part by the amount of available transport protein. It also yielded a number of candidate transport proteins for metabolite flux across the envelope of C4 chloroplasts.

Chapter 6 describes the analysis of a membrane protein identified as one of the candidate proteins in chapter 5. At the onset of the work nothing was known about the function of the protein and I hypothesized that a detailed analysis of the Arabidopsis homologue with the tools of molecular biology and reverse genetics may have the power to gain insights into the function of the protein in vivo. To this end, I characterized the

expression pattern and confirmed the localization. I isolated an insertion mutant and characterized it as a null mutant which can be complemented. The null mutant displayed a visible phenotype which is predated by a unique biochemical phenotype indicating the Arabidopsis mutant has limited capacity to transport glycerate and glycolate. All attempts to characterize the protein biochemically failed.

Chapter 7 reports the study of the proteomes of not only mixed envelope membranes but also of inner and outer envelope membranes and chloroplast associated membranes in comparison with an ER enriched microsome fraction. The protein composition of outer envelopes has only been studied by electrophoretic methods (Cline et al. 1981) and by a limited proteome analysis without a dedicated sequence database (Schleiff et al. 2003). Recently, a new class of membrane associated with chloroplasts and labeled by a ER GFP marker protein has been identified (Andersson et al. 2007). The comparative proteome analysis was designed to both identify new residents of chloroplast envelopes and localize them to one of the membrane systems. To address this question, I collaborated with Henrik Tjellstroem and Anna Stina Sandelius. The comparative proteome analysis reveals unexpected fluidity of the membrane systems and defines the proteins content of a novel organelle, the plastid associated membranes. GFP fusion proteins support the new hypothesis that the protein localization in the membrane systems of chloroplasts is more dynamic than previously assumed.

References

Andersson MX, Goksor M, Sandelius AS (2007) Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts J Biol Chem 282, 1170-1174

Baginsky S, Siddique A, Gruissem W (2004) Proteome analysis of tobacco bright yellow-2 (BY-2) cell culture plastids as a model for undifferentiated heterotrophic plastids J Proteome Res 3, 1128-1137

Benning C, Xu CC, Awai K (2006) Non-vesicular and vesicular lipid trafficking involving plastids Curr Opin Plant Biol 9, 241-247

Binder S, Knill T, Schuster J (2007) Branched-chain amino acid metabolism in higher plants Physiol Plant 129, 68-78

Bolter B, Soll J, Hill K, Hemmler R, Wagner R (1999) A rectifying ATP-regulated solute channel in the chloroplastic outer envelope from pea EMBO J 18, 5505-5516

Cline K, Andrews J, Mersey B, Newcomb EH, Keegstra K (1981) Separation And Characterization Of Inner And Outer Envelope Membranes Of Pea-Chloroplasts Proc Natl Acad Sci USA 78, 3595-3599

Cribb L, Hall LN, Langdale JA (2001) Four mutant alleles elucidate the role of the G2 protein in the development of C-4 and C-3 photosynthesizing maize tissues Genetics 159, 787-797

Crisp P, Walkey DGA, Bellman E, Roberts E (1975) Mutation Affecting Curd Color In Cauliflower (Brassica-Oleracea L Var Botrytis Dc) Euphytica 24, 173-176

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana Mol Cell Proteomics 2, 325-345

Ferro M, Salvi D, Riviere-Rolland H, Vermat T, Seigneurin-Berny D, Grunwald D, Garin J, Joyard J, Rolland N (2002) Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters Proc Natl Acad Sci U S A 99, 11487-92.

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW, Gage DA, Phinney BS (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis J Proteome Res 2, 413-25

Goetze TA, Philippar K, Ilkavets I, Soll J, Wagner R (2006) OEP37 Is a New Member of the Chloroplast Outer Membrane Ion Channels J Biol Chem 281, 17989-17998

Hall LN, Rossini L, Cribb L, Langdale JA (1998) GOLDEN 2: A novel transcriptional regulator of cellular differentiation in the maize leaf Plant Cell 10, 925-936

Heineke D, Riens B, Grosse H, Hoferichter P, Peter U, Flugge UI, Heldt HW (1991) Redox Transfer Across The Inner Chloroplast Envelope Membrane Plant Physiol 95, 1131-1137

Herrmann KM, Weaver LM (1999) The shikimate pathway Annu Rev Plant Physiol Plant Mol Biol 50, 473-503

Holm M, Deng XW (1999) Structural organization and interactions of COP1, a light-regulated developmental switch Plant Mol Biol 41, 151-158

Hudson AO, Bless C, Macedo P, Chatterjee SP, Singh BK, Gilvarg C, Leustek T (2005) Biosynthesis of lysine in plants: evidence for a variant of the known bacterial pathways Biochim Biophys Acta-Gen Subj 1721, 27-36

Hugueney P, Bouvier F, Badillo A, Dharlingue A, Kuntz M, Camara B (1995) Identification Of A Plastid Protein Involved In Vesicle Fusion And/Or Membrane-Protein Translocation Proc Natl Acad Sci U S A 92, 5630-5634

Jarvis P, Robinson C (2004) Mechanisms of protein import and routing in chloroplasts Curr Biol 14, R1064-R1077

Journet EP, Douce R (1985) Enzymic Capacities Of Purified Cauliflower Bud Plastids For Lipid-Synthesis And Carbohydrate-Metabolism Plant Physiol 79, 458-467

Kirk JTO, Tilney-Bassett RAE (1978) 'The plastids: Their chemistry, structure, growth and inheritance.' (Elsevier: Amsterdam/Oxford)

Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S (2004) The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions Curr Biol 14, 354-362

Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, Baginsky S (2007) Proteome dynamics during plastid differentiation in rice Plant Physiol 143, 912-923

Ljubicic JM, Wrischer M, Ljubesic N (1998) Formation of the photosynthetic apparatus in plastids during greening of potato microtubers Plant Physiology And Biochemistry 36, 747-752

Lu S, Van Eck J, Zhou X, Lopez AB, O'Halloran DM, Cosman KM, Conlin BJ, Paolillo DJ, Garvin DF, Vrebalov J, Kochian LV, Kupper H, Earle ED, Cao J, Li L (2006) The cauliflower or gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation Plant Cell 18, 3594-3605

Majeran W, Cai Y, Sun Q, van Wijk KJ (2005) Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics Plant Cell 17, 3111-3140

Majeran W, Zybailov B, Ytterberg AJ, Dunsmore J, Sun Q, van Wijk KJ (2008) Consequences of C4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells Mol Cell Proteomics 10.1074/mcp.M800016-MCP200, M800016-MCP200

Miras S, Salvi D, Ferro M, Grunwald D, Garin J, Joyard J, Rolland N (2002) Non-canonical transit peptide for import into the chloroplast J Biol Chem 277, 47770-47778

Pohlmeyer K, Soll J, Grimm R, Hill K, Wagner R (1998) A High-Conductance Solute Channel in the Chloroplastic Outer Envelope from Pea Plant Cell 10, 1207-1216

Pohlmeyer K, Soll J, Steinkamp T, Hinnah S, Wagner R (1997) Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane Proc Natl Acad Sci U S A 94, 9504-9509

Rawsthorne S (2002) Carbon flux and fatty acid synthesis in plants Prog Lipid Res 41, 182-196

Schleiff E, Eichacker LA, Eckart K, Becker T, Mirus O, Stahl T, Soll J (2003) Prediction of the plant beta-barrel proteome: A case study of the chloroplast outer envelope Protein Sci 12, 748-759

Schnell DJ (2000) Functions and origins of the chloroplast protein-import machinery. In 'Molecular Trafficking' pp. 47-59)

Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto M, Takamiya K (2005) Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phylloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of Arabidopsis Plant J 41, 627-637

Siddique MA, Grossmann J, Gruissem W, Baginsky S (2006) Proteome analysis of bell pepper (Capsicum annum L.) chromoplasts Plant Cell Physiol, 33-39

Slocum RD (2005) Genes, enzymes and regulation of arginine biosynthesis in plants Plant Physiology And Biochemistry 43, 729-745

Soll J, Schleiff E (2004) Protein import into chloroplasts Nature reviews Molecular cell biology 5, 198-208

Stepansky A, Leustek T (2006) Histidine biosynthesis in plants Amino Acids 30, 127-142

von Zychlinski A, Kleffmann T, Krishnamurthy N, Sjolander K, Baginsky S, Gruissem W (2005) Proteome analysis of the rice etioplast - Metabolic and regulatory networks and novel protein functions Mol Cell Proteomics 4, 1072-1084

Weaver LM, Herrmann KM (1997) Dynamics of the shikimate pathway in plants Trends Plant Sci 2, 346-351

Weber APM (2004) Solute transporters as connecting elements between cytosol and plastid stroma Curr Opin Plant Biol 7, 247-253

Weber APM, Fischer K (2007) Making the connections - The crucial role of metabolite transporters at the interface between chloroplast and cytosol FEBS Lett 581, 2215-2222

Weber APM, Schwacke R, Flugge UI (2005) Solute transporters of the plastid envelope membrane Ann Rev Plant Biol 56, 133-164

Zhao J, Last RL (1995) Immunological Characterization and Chloroplast Localization of the Tryptophan Biosynthetic Enzymes of the Flowering Plant Arabidopsis thaliana J Biol Chem 270, 6081-6087

Zrenner R, Stitt M, Sonnewald U, Boldt R (2006) Pyrimidine and purine biosynthesis and degradation in plants Ann Rev Plant Biol 57, 805-836

Chapter 2

Transport processes – connecting the reactions of C4 photosynthesis

This chapter has been accepted for publication in the series Advances of Photosynthesis and Respiration: Andrea Bräutigam and Andreas P. M. Weber: Transport processes – connecting the reactions of C4 photosynthesis. (2009) Advances in Photosynthesis and Respiration. Series Editor R. Govindjee, Springer

Abbreviations

2-OG	2-oxoglutarate
3-PGA	3-phosphoglycerate
AAC	ATP ADP carrier
DHAP	dihydroxyacetone phosphate
DIC	dicarboxylate carrier
DiT	dicarboxylate translocator
DTC	di- and tricarboxylate carrier
MCF	mitochondrial carrier family
MDH	malate dehydrogenase
NAD-ME	NAD-malic enzyme
NADP-ME	NADP-malic enzyme
OAA	oxaloacetate
PCA	photosynthetic carbon assimilation
PCR	photosynthetic carbon reduction
PEP	phosphoenolpyruvate
PEPC	phosphoenolpyruvate carboxylase
PPT	phosphoenolpyruvate phosphate translocator
PPDK	PEP phosphate dikinase
PEP-CK	PEP carboxykinase
Rubisco	ribulosebisphosphate carboxylase oxygenase
TPT	triosephosphate phosphate translocator

Abstract

The C4 cycle requires immense metabolite fluxes. The spatial separation of initial carbon fixation by phospho*enol*pyruvate carboxylase and entry in the photosynthetic carbon reduction cycle through Rubisco requires metabolites to shuttle not only between cells but also across intracellular membranes. C4 photosynthesis is a highly compartmentalized process. Atmospheric CO2 is fixed into C4 acids (photosynthetic carbon assimilation, PCA) in one domain and C4 acids donate CO2 to Rubisco in another domain (photosynthetic carbon reduction, PCR). PCA occurs in chloroplasts and cytosol; PCR occurs in chloroplasts, and depending on the subtype of C4 photosynthesis may involve the mitochondria and cytosol. Intercellular transport likely occurs symplastically but the intracellular transport processes across the organellar membranes are at least in part mediated by specific transport proteins. These transport processes are of particular interest because metabolites have to be transported at the rate of carbon assimilation; each carbon which is shuttled as a C4 acid necessitates distinct transport processes as does the C3 acid which returns to recycle the initial carbon acceptor. Currently, it is not fully understood how the organellar membranes accommodate the high volume and velocity of the necessary flux.

This chapter will review the different types of C4 cycle reactions and the transport processes required for each sub-type based on the localization of the enzymes involved in the C4 cycle. For each transport process the current knowledge about the transport proteins involved is stated in detail including discussion of candidate transport proteins characterized in C3 systems. Finally, novel strategies for identifying and characterizing

molecular candidates for transport proteins and their importance for engineering a C4 cycle in C3 crop plants are described.

Introduction

One of the highest steady state fluxes of metabolites across organellar membranes known to date originates from the process of C4 photosynthesis. In C4 photosynthesis, the production of the primary carbon acceptor phospho*enol*pyruvate (PEP) and fixation of atmospheric CO2 by phospho*enol*pyruvate carboxylase (PEPC) occurs in the photosynthetic carbon assimilation (PCA) compartment. In the photosynthetic carbon reduction (PCR) compartment CO2 is released from C4 acids and enters the Calvin Cycle (reviewed in Hatch (1987) and Edwards et al. (2001a)). In most C4 species, the PCA and PCR compartments are located in two distinct cell types, with PCA cells being leaf mesophyll tissue and PCR cells surrounding the veins as a bundle sheath. This arrangement is called Kranz anatomy. Metabolite transport between the cells is not completely understood but it likely involves symplastic connections (Craig and Goodchild 1977; Hattersley and Browning 1981; Sowinski et al. 2008). In contrast, intracellular transport requires specific transport proteins in the membranes surrounding the organelles since the reactions of the C4 cycle occur in three subcellular compartments, the cytosol, the chloroplasts, and the mitochondria (Hatch 1987). Both chloroplasts and mitochondria are separated from the cytosol by two membranes. In chloroplasts, the outer envelope membrane which faces the cytosol is relatively permeable to small metabolites because it contains a number of porins with broad substrate specificities (Pohlmeyer et al. 1997; Pohlmeyer et al. 1998; Bolter et al. 1999; Goetze et al. 2006; Murcha et al. 2007). The inner envelope membrane between the intermembrane space and the stroma is the diffusion barrier and contains specific transport proteins (Weber 2004; Weber et al. 2005; Weber and Fischer 2007). In

mitochondria, the outer membrane is also relatively permeable because of a number of porins (Benz 1994) and the inner membrane represents the selectivity filter. Most specific transport proteins residing in the inner envelopes catalyze the movement of two molecules, operating either in counter-exchange or co-transport mode (Weber et al. 2005). When the transport of a solute is coupled to that of another ion or solute it is not only possible to avoid creating electrochemical gradients but the solute can even be transported against its concentration gradient at the expense of the gradient of the second molecule.

For each carbon atom that is assimilated and reduced in C4 species, the C4 cycle performs at least one full turn, and the metabolites involved in C4 photosynthesis are transported across different organellar membranes at the rate of carbon assimilation (Figure 2-1). Overcycling to enrich CO2 within the PCR tissue, which has been estimated at 10% to 40% of the apparent rate of CO2 assimilation (Henderson et al. 1992; Laisk and Edwards 2000; Kubasek et al. 2007), further increases the load on the metabolite transport systems. Compared to the most abundant transport protein in C3 chloroplasts, the triosephosphate phosphate translocator (TPT), each metabolite transport protein involved in the C4 cycle has to carry at least three times more cargo per unit of time. In *Zea mays*, a NADP malic enzyme (NADP-ME) type C4 plant, for each carbon assimilated, at least four transport processes across the PCA chloroplast envelope and at least three transport processes across the PCR chloroplast envelope are required (for a detailed description of transport processes in all C4 photosynthesis subtypes see below). For the three carbon atoms contained in one molecule of triosephosphate, this adds up to at least 21 transport processes. In a C3 plant, however, for three carbon atoms

assimilated, at most one transport process across the chloroplast envelope, (i.e., the export of one molecule of triosephosphate from the chloroplast) is required. It is currently not fully understood how the at least twentyfold higher metabolite flux across the organellar membranes of C4 plants is mastered, especially since a number of transport proteins that carry only minor fluxes in C3 plants are likely playing a more prominent role in C4 photosynthesis because they have to carry a major flux in these plants. In addition to the metabolites which are shuttled to support CO2 assimilation, other pathway intermediates also have to be shuttled between PCR and PCA chloroplasts. Chloroplasts in PCR tissues of C4 plants frequently have reduced water splitting activity to minimize the oxygen concentration around Rubisco (reviewed in Walker and Edwards (1983); Meierhoff and Westhoff 1993). In consequence, reactions such as the reduction of Calvin Cycle intermediates, fatty acid synthesis, and nitrogen or sulfur reduction depend on the reducing power generated in chloroplasts of the PCA tissue. Since pyridine or pyrimidine nucleotides cannot be directly exported or imported into chloroplasts, reduction power can be either shuttled in the form of oxidized and reduced metabolite pairs, such as 3-phosphoglycerate (3-PGA) and triosephosphate, or the intermediates of the pathways themselves will have to be shuttled between PCR and PCA type tissues. To our knowledge, the C4 cycle causes one of the highest steady state fluxes of metabolites occurring across any organellar envelope.

Despite the high load of transported metabolites, C4 photosynthesis likely did not require the invention of novel transport proteins. C4 photosynthesis independently evolved at least 50 times (Sage 2004; Muhaidat et al. 2007), making it highly unlikely that the invention of metabolite transport proteins occurred each time in a convergent

manner. Rather, as is the case for the soluble enzymes involved in C4 photosynthesis (Matsuoka et al. 2001; Sage 2004) pre-existing transport proteins may have been recruited for C4 photosynthesis by increasing their abundance and/or changing their expression patterns. Possibly, minor transport substrates of C3 plant transporters have become major transport substrates of the corresponding C4 plant transporter. Therefore, whenever possible, evidence from C3 plants is considered when transport proteins for C4 species are discussed.

Intercellular fluxes

All C4 plants that rely on two different cell types to enrich CO2 in the vicinity of RubisCO need to transport the metabolites of the C4cycle between these cells types. The two photosynthetic cell types involved are connected by an unusually high number of plasmodesmata (Craig and Goodchild 1977; Evert et al. 1977; Hattersley and Browning 1981). Based on the frequency of plasmodesmata and the diffusion surface area it was calculated that concentration gradients of 10 mM are necessary to drive efficient transport of metabolites by diffusion (Hatch 1987). Indeed, concentration gradients of 5-10 mM for many metabolites involved in the C4 cycle appear to be present in C4 species (Leegood 1985; Stitt and Heldt 1985; Furbank and Hatch 1987). The transport of metabolites between cells is thus dependent on Brownian motion which is limited by cytoplasmic viscosity and temperature (Leegood and Edwards 1996). The concentration gradient for pyruvate, however, which was experimentally determined between whole mesophyll and bundle sheath cells in *Zea mays*, is not steep enough to drive efficient diffusion (Flügge et al. 1985; Stitt and Heldt 1985). It was concluded that the sequestration of metabolites into subcellular compartments such as chloroplasts may provide the driving force for

intercellular transport (Flügge et al. 1985). Studies of plasmodesmatal architecture (Overall et al. 1982; Ding et al. 1992; Roberts and Oparka 2003) have also questioned the assumptions of the simple diffusion model (Sowinski et al. 2008). In particular, the plasmodesmata contain internal structures, most notably the desmotubule, which connects the endoplasmatic reticulum of the adjacent cells and limits the cross section available for diffusion. Plasmodesmata themselves have different size exclusion limits for molecules that can traverse them depending on the cells they connect (Erwee and Goodwin 1985). Recent estimates for the concentration gradients needed to drive metabolite transport are three orders of magnitude higher than previously assumed (Sowinski et al. 2008). Sowinski et al. (2008) propose alternative transport mechanisms such as diffusion through the desmotubule or vesicle transport of metabolites but the membrane continuity between the endoplasmatic reticulum and chloroplasts or mitochondria required for either of the alternatives proposed has never been conclusively demonstrated. Although symplastic transport by diffusion remains the accepted model for intercellular transport (Sowinski et al. 2008), it is currently not known how the steep metabolite gradients needed to drive diffusion are generated.

Transport processes in the NADP malic enzyme type

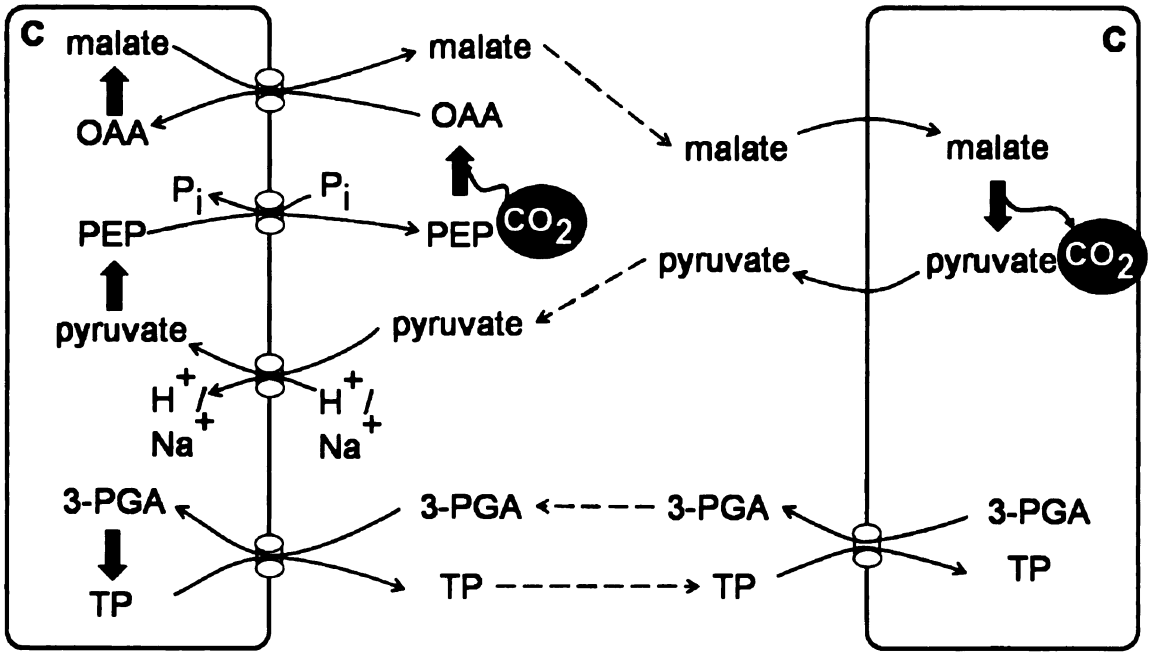


Figure 2-1: Scheme outlining the reactions of C4 photosynthesis and the connecting transport processes for the NADP malic enzyme type of C4 photosynthesis in both the PCA (left) and the PCR (right) compartment; the cell wall is represented by a grey bar; intracellular transport by arrows; intercellular transport by dashed arrows; metabolic reactions by broad arrows; C chloroplast, for additional abbreviations see text;

Operation of the C4 cycle in the NADP-ME type involves the cytosol and chloroplasts of the PCA tissue and the chloroplasts of PCR tissue (Figure 2-1). This type of the C4cycle necessitates at least seven metabolite transport steps for each molecule of CO2 assimilated (Figure 2-1). The CO2 acceptor PEP is first generated within chloroplasts of the PCA tissue and then exported to the cytosol by a phospho*enol*pyruvate/phosphate translocator (PPT). Oxaloacetate (OAA) is produced from PEP and CO2 in the carboxylation reaction in the cytosol of PCA cells and then imported into the chloroplasts of these cells. There it is reduced to malate and then exported to cytosol again. The import and export of these C4 acids are assumed to occur through a single transport protein, the OAA/malate exchanger (Hatch et al. 1984b;

Taniguchi et al. 2004). The reduction of OAA to malate enables the C4 acid to carry not only CO2 but also one reducing equivalent to the PCR tissue. After import into the PCR type chloroplasts, malate is oxidatively decarboxylated to yield pyruvate, NADPH, and CO2, the latter two entering the photosynthetic carbon reduction (or Calvin) cycle. Pyruvate is exported back out of the PCR chloroplasts and returns to the PCA tissue where it is imported back into the chloroplasts to regenerate the CO2 acceptor PEP (reviewed in Hatch (1987)). The transport processes are summarized in Table 2-1. The best studied C4 species are of the NADP–ME subtype and they include species from the genera *Zea*, *Sorghum*, and *Flaveria*.

Table 2-1: Summary of transport processes necessary in NADP-ME plants, for abbreviations see text

domain	substrate	co-substrate	name	evidence	references
PCA	PEP	P_i	PPT	biochemical and molecular, supported by proteomics	Day and Hatch (1981); Fischer et al. (1997); Bräutigam et al. (2008); Majeran et al. (2008)
PCA	OAA	malate	OAT	biochemical	Hatch et al. (1984a)
			DiT	biochemical and molecular, supported by proteomics	Weber et al. (1995); Taniguchi et al. (2002, 2004); Renne et al. (2003)
PCR	malate	?	?	inferred from localization of soluble proteins	-
PCR	pyruvate	?	?	inferred from localization of soluble proteins	-
PCA	pyruvate	H^+	?	biochemical	Flügge et al. (1995); Aoki et al. (1992)
		Na^+	?	biochemical	Aoki et al. (1992)
PCA	3-PGA	DHAP	TPT	biochemical and molecular, supported by proteomics	Day and Hatch (1981); Rumpho und Edwards (1985); Bräutigam et al. (2008)
PCR				inferred from localization of soluble proteins	Majeran et al. (2008)

PEP export from PCA type chloroplasts

The primary CO2 acceptor PEP is regenerated from pyruvate by pyruvate phosphate dikinase (PPDK). This enzyme is localized in PCA type chloroplasts, both in single-cell and Kranz-type C4 photosynthesis, necessitating the export of PEP across the chloroplast envelope (Figure 2-1) (Hatch 1987; Voznesenskaya et al. 2001; Voznesenskaya et al. 2002). It was initially characterized biochemically using isolated PCA chloroplasts of *Zea mays* and *Digitaria sanguinalis* (Huber and Edwards 1977b; Day and Hatch 1981). Both inorganic phosphate (Pi) and 3-PGA but not pyruvate applied externally lead to export of PEP from chloroplasts *in vitro* in *D. sanguinalis* (Huber and Edwards 1977b). The transport protein has high affinity for Pi with an apparent *K*M of 200μM and an equally high affinity for PEP, determined as the apparent *K*i=450μM for inhibition of Pi transport by PEP. Likewise, PEP formation of intact chloroplasts in the light could be stimulated by external application of Pi but not pyruvate, and presumably Pi serves both for export of PEP and as a substrate for ATP formation needed as a substrate for PPDK (Huber and Edwards 1977b). In *Zea mays*, the exchange of PEP and 3-PGA with PEP, 3-PGA, and Pi but not pyruvate or malate and the exchange of 3-PGA for dihydroxyacetonephosphate (DHAP) across intact chloroplast envelopes was demonstrated (Day and Hatch 1981). 3-PGA could exchange with DHAP and the exchange of PEP for DHAP was not tested. From the biochemical data it was concluded that all substrates are exchanged by the same phosphate translocator protein. The first transporter capable of exchanging phosphorylated sugars for phosphate was identified at the molecular level as a triosephosphate phosphate translocator (TPT) in C3 chloroplasts of spinach (Flügge and Heldt 1984). It accepts triosephosphates, 3-PGA, and P_i but not PEP as substrates. Based

on earlier biochemical evidence with intact chloroplasts (Day and Hatch 1981; Rumpho and Edwards 1984; Rumpho and Edwards 1985), tests were performed to determine whether the TPTs of C4 plants, unlike those of the C3 plant spinach, were capable of mediating PEP transport. The TPT from the C4 species *Flaveria trinervia* and *Zea mays* both also accepted PEP in contrast to the spinach phosphate translocator, which has a low affinity for PEP (Fischer et al. 1994). However, the comparison of transport rates of PEP by intact chloroplasts of C4 species revealed substantial differences between the transport capabilities of intact chloroplasts compared to the transport characteristics of C4TPTs (Fischer et al. 1994). A phosphate translocator specific to PEP transport in exchange for Pi, the phosphoenolpyruvate phosphate translocator (PPT), was identified on the molecular level from several tissues and species, including maize endosperm, and was characterized in detail from cauliflower bud envelopes (Fischer et al. 1997). When heterologously expressed, this protein has a high affinity for PEP (K_i=300μM for inhibition of Pi transport) whereas 3-PGA and triosephosphates are only poorly bound and transported (K_i=8,000 and K_i=4,600μM, respectively). RNA gel blot analysis showed that the PPT identified in maize endosperm is expressed at high levels only in non-green tissues (Fischer et al. 1997). Phosphate translocators are evolutionary ancient proteins, and the split into the PPT and TPT subfamilies of transporters occurred already in the common ancestor of the red and the green lineage (Weber et al. 2006). Consequentially, PPTs are present in all plant and algal genomes sequenced to date. The genome of the C3 plant *Arabidopsis thaliana* harbors two PPT genes. The AtPPT1, like the protein from spinach, accepts both PEP and 2-phosphoglycerate whereas AtPPT2 accepts PEP rather than 2-phosphoglycerate (Knappe et al. 2003). In C3 plants, the PPT imports PEP into

chloroplasts as a substrate for the shikimate pathway since chloroplasts lack the activities of enolase and phosphoglyceromutase that are required to produce PEP from 3-phosphoglycerate; therefore PPT provides the substrate for aromatic amino acid biosynthesis (Knappe et al. 2003; Voll et al. 2003). The import of PEP into chloroplasts is a minor flux in C3 plants compared with the export of assimilates, for example, mediated by TPT. PPT activity is also required in bundle sheath chloroplasts of C3 plants to generate a signal that is required for proper mesophyll development (Streatfield et al. 1999; Voll et al. 2003). Recently, a PPT protein abundant in mesophyll chloroplast envelopes and expressed leaf specifically was identified in maize (Bräutigam et al. 2008a) and shown to be mesophyll specific (Majeran et al. 2008) but its activity has not been tested biochemically. Compared to C3 chloroplasts envelopes, this PPT is vastly increased in abundance indicating that the high PEP export rates are maintained at least in part by increasing the amount of transport protein present at the envelope (Bräutigam et al. 2008a). Possibly, based on the results by Huber and Edwards (1977) outlined above, the C4 PPT from *Zea mays* can also exchange 3-PGA for PEP, which is not the case for any of the C3 PPTs characterized to date.

Oxaloacetate and Malate exchange in PCA type chloroplasts

After PEP is exported from the PCA chloroplast, it is carboxylated to OAA by PEP carboxylase in the cytosol. Since carbon is shuttled to PCR cells in the form of malate in NADP-ME type C4 plants, OAA and malate need to be exchanged across the chloroplast envelope. Isolated chloroplasts from both C3 and C4 plants produce malate and evolve oxygen when externally supplied with OAA (Heber 1974; Anderson and House 1979; Day and Hatch 1981). In maize and spinach chloroplasts a high velocity OAA transporter

was characterized biochemically both by OAA-dependent oxygen evolution and radiolabeled precursor uptake studies (Hatch et al. 1984a). In *Zea mays* PCA chloroplasts, this transport protein binds OAA with an apparent *K*M between 53 and 71μM, depending on the method used for measuring, and the corresponding apparent inhibitory binding constants for malate are *K*i=7,300μM and *K*i=7,500μM. In spinach chloroplasts, the *K*M(OAA) is 9μM and *K*i(malate) is 1,400μM. The high affinity for OAA compared to malate is crucial if, like in *Zea mays*, the malate pools are 10-100 times greater than those of OAA (Hatch et al. 1984a). The reaction equilibrium constant of malate dehydrogenase favors the reaction in the direction of malate, rather than OAA (Hatch et al. 1984a). Finally, the velocity of OAA transport in *Zea mays* is sufficiently high to supply the C4 cycle metabolites and orders of magnitude higher than the velocity of transport in C3 chloroplasts of spinach (Hatch et al. 1984a). In C3 plants, the OAA/malate exchanger is hypothesized to be involved in a reducing equivalent shuttle (Scheibe 2004; Scheibe et al. 2005). An OAA/malate exchanger working in concert with malate dehydrogenases in the chloroplasts, the cytosol and the peroxisomes can balance reducing power throughout the cell. The molecular identity of the OAA malate transport protein has not been unequivocally established in either C3 plants or C4 plants. A small family of transport proteins capable of transporting dicarboxylates (dicarboxylate translocators, DiTs) was biochemically identified in spinach chloroplasts (Woo et al. 1987) and later identified at the molecular level (Weber et al. 1995). The protein initially characterized transports malate, fumarate, succinate, and 2-oxoglutarate (2-OG), but not glutamate and was named DiT1 (Weber et al. 1995). Its transport characteristics fit well with the model of two translocators which, working in concert, import 2-OG into the

chloroplast and export glutamate while the counter-substrate malate only cycles (Woo et al. 1987). A second translocator, called DiT2, was also identified at the molecular level and characterized (Taniguchi et al. 2002; Weber and Flügge 2002; Renne et al. 2003). Proteins of the DiT2 family of transporters have a high affinity for glutamate and aspartate in addition to 2-OG (Taniguchi et al. 2002; Renne et al. 2003). In C3 plants, a knock down and a knock out for one member of each family have been analyzed. In tobacco, a knockdown of the sole representative of the DiT1 family of transport proteins causes a photorespiratory phenotype with dramatic metabolic changes in precursors and products of ammonia fixation as well as decreases in photosynthesis rate and sugar pools (Schneidereit et al. 2006). In Arabidopsis, the knockout of one of the two members of the DiT2 family results in a photorespiratory phenotype (Somerville and Ogren 1983; Renne et al. 2003). Changes in OAA and malate pools or redox status have not been reported (Renne et al. 2003; Schneidereit et al. 2006). All DiTs also transport OAA *in vitro*. The affinity for OAA of both, C3 DiT1 and C4 DiT1, is similar (Taniguchi et al. 2002; Taniguchi et al. 2004). For the Arabidopsis protein the apparent *K*M for malate is 700μM and the corresponding apparent *K*i for OAA is 70μM (Taniguchi et al. 2002). For the maize protein the *K*M for malate is 610μM and the corresponding *K*i for OAA is 90 μM (Taniguchi et al. 2004). It was proposed that DiT1 is capable of exchanging OAA and malate *in vivo* (Taniguchi et al. 2004), although the kinetic constants of recombinant reconstituted DiTs are different from those determined with transport experiments with intact isolated chloroplasts (Hatch et al. 1984a). The discrepancies between transport experiments with intact chloroplasts and isolated proteins have yet to be resolved. Proteomic analysis of mesophyll and bundle sheath chloroplast membranes also indicates

DiT1 may indeed be the OAA/malate exchanger. A DiT1 homologue (called OMT in maize) was enriched in mesophyll compared to bundle sheath chloroplast membranes (Supplemental material (Majeran et al. 2008)).The proteome analysis resolves controversial evidence for the expression pattern of this DiT1 family protein which is reported with different expression patterns in *Zea mays* and *Sorghum bicolor* (Renne et al. 2003; Taniguchi et al. 2004; Sawers et al. 2007). Despite mounting evidence, the differences between *in vitro* and *in vivo* data are unresolved and therefore it remains to be determined whether DiT1 exchanges OAA and malate in either C3 or C4 chloroplasts in vivo (Table 2-1).

Malate import into PCR chloroplasts

Malate is produced in PCA cells from OAA, and it moves to PCR cells by diffusion through plasmodesmata. The decarboxylation enzyme, NADP-ME, is localized in chloroplasts and malate is imported into PCR chloroplasts (Figure 2-1). The transport capacities of chloroplasts in PCR tissues are less well understood compared to chloroplasts in PCA tissues. Most species with Kranz-type C4 photosynthesis have re-enforced cell walls around the PCR cells, likely to prevent CO2 from leaking. These cells are therefore less amenable to isolation and characterization in comparison to PCA cells. In *Zea mays*, malate import was studied as a function of CO2 fixation and as a function of pyruvate generation (Boag and Jenkins 1985). Malate import was shown to be affected by application of aspartate (Boag and Jenkins 1985) or aspartate and glutamate (Boag and Jenkins 1986). Application of aspartate lowered the binding constant and increased the maximal velocity of transport. The only candidate transport proteins currently known of being capable of transporting malate into chloroplasts are the DiTs. DiT2 family proteins

from *Zea mays* (Taniguchi et al. 2004) and *Sorghum bicolor* (Renne et al. 2003) are expressed at a higher level in the bundle sheath. A recent proteomic analysis identified a DiT2 family member, ZmDCT2/3, as bundle sheath specific (Majeran et al. 2008). However, DiTs are antiporters that exchange a dicarboxylic acid for malate (Weber et al. 1995; Taniguchi et al. 2002; Renne et al. 2003; Taniguchi et al. 2004). Therefore DiTs cannot catalyze the net import of a C4acid in a uniport mode. The decarboxylation of malate produces the three carbon monocarboxylate pyruvate. Pyruvate has been tested as a counter-substrate for malate transport for a DiT2 protein from *Flaveria bidentis*, a NADP-ME species. It was shown that exchange of pyruvate with malate was negligible *in vitro* (Renne et al. 2003). It is therefore unlikely that DiT2 proteins catalyze the exchange of malate and pyruvate across PCR chloroplast envelopes (Renne et al. 2003). It thus remains currently unknown how the net transport of C4acids across the chloroplast envelopes of PCR cells is achieved.

Pyruvate Export from PCR chloroplasts

After import and decarboxylation of malate in PCR chloroplasts of NADP–ME species, the resulting pyruvate is exported to allow the recycling of the initial carbon acceptor in the PCA compartment. Pyruvate export (Figure 2-1) from PCR type chloroplasts has only been studied in a NAD malic enzyme species, *Panicum miliaceum*, in which the majority of pyruvate is generated in mitochondria. Pyruvate uptake into those PCR chloroplasts was markedly different from pyruvate uptake into PCA chloroplasts both in speed and binding affinity, and was not affected by light (Ohnishi and Kanai 1987b). In NADP-ME species, pyruvate transport across the PCR type chloroplast envelope has not been studied (Flügge et al. 1985; Ohnishi and Kanai 1987a;

Aoki et al. 1992). Possibly, the high concentration of pyruvate that is generated by malate decarboxylation drives the export of pyruvate out of the PCR chloroplasts. It remains to be determined whether the export of pyruvate from the PCR chloroplasts is mediated by the same or by a different transport protein than the one which imports pyruvate into PCA chloroplasts.

Pyruvate import into PCA type chloroplasts

Finally, the pyruvate produced in the PCR compartment needs to be recycled to the CO2 acceptor PEP in the PCA compartment (Figure 2-1). The overall gradient for pyruvate between PCA and PCR tissue opposes the actual direction of transport (Stitt and Heldt 1985), thus leading to the hypothesis that pyruvate is actively transported to sequester it in PCA chloroplasts, hence displacing it from the equilibrium (Flügge et al. 1985). In C4 plants pyruvate import was characterized as a slow process in the dark as a pyruvate anion symport in *Digitaria sanguinalis* (Huber and Edwards 1977a). Later, light-driven active pyruvate transport was characterized in *Zea mays* (Flügge et al. 1985) and *Panicum miliaceum* (Ohnishi and Kanai 1987b). In *Zea mays*, light driven pyruvate transport is dependent on proteins, and it is inhibited by protonophores (Flügge et al. 1985). Transport can be initiated in the dark by applying a pH gradient between the stroma of isolated chloroplasts and the external medium *in vitro* (Aoki et al. 1992). A second, sodium-dependent mode of active pyruvate import into PCA chloroplasts was discovered in *Panicum miliaceum* (Ohnishi and Kanai 1987a). Several other species, such as *Urochloa panicoides* and *Panicum maximum*, but not *Zea mays*, also exhibit sodium-dependent pyruvate transport (Ohnishi et al. 1990). A systematic evaluation of more than forty C4species revealed no correlation between the mode of pyruvate transport (i.e.

sodium- or proton-dependent) and biochemical C4 subtype (Aoki et al. 1992). In all species tested, light driven pyruvate uptake into PCA type chloroplasts could be mimicked either by a sodium or by a proton gradient *in vitro* but not by both for any given species (Aoki et al. 1992).

In C3 plants, pyruvate transport was analyzed biochemically with isolated pea chloroplasts. Pyruvate uptake followed saturation kinetics at low external concentrations and linear kinetics at higher pyruvate concentrations. Based on these results carrier mediated transport was proposed at low substrate concentrations and transport by diffusion at high concentrations (Proudlove and Thurman 1981). Pyruvate transport into the chloroplast of C3 species may be relevant when serving as a substrate for branched chain amino acid biosynthesis or isoprenoid production (Singh and Shaner 1995; Schwender et al. 1996) or for fatty acid production in certain tissues but not in seeds (Andre and Benning 2007; Andre et al. 2007). At the molecular level, the protein or proteins which catalyze pyruvate import remain unknown both in C3 and in C4 species. Possibly, independent evolution of C4 photosynthesis recruited two different types of pyruvate transport proteins, using either a sodium or a proton gradient as the driving force. Passive diffusion of pyruvate across the chloroplast envelope appears unlikely since the metabolite gradient for pyruvate between PCA and PCR tissue requires active pyruvate import into PCA chloroplasts for C4 photosynthesis, and pyruvate transport is dependent on intact proteins in the chloroplast envelope (Flügge et al. 1985). Establishing either sodium or proton gradients to drive pyruvate transport across the chloroplast envelope membrane of PCA cells requires the input of energy. To date, energy requirements for driving metabolite transport have not been included in the overall ATP

balance required for driving the C4 biochemical CO2 pump. Molecular identification and biochemical characterization of the respective transporters will be required to address this question.

The knowledge about transport proteins involved in the C4cycle of NADP-ME plants is summarized in Table 2-1.

Auxiliary transport processes

In addition to the C4cycle metabolites, pyruvate, PEP, OAA, and malate, several other metabolites are also moved between PCA and PCR cells and compartments. The C4cycle serves to enrich the PCR tissue in CO2 and creates an environment that minimizes photorespiration at the site of Rubisco. To further reduce photorespiration, in some species of the NADP malic enzyme subtype, the partial pressure of O2 is minimized. PCR tissue chloroplasts have minimal photosystem II activity (Leegood et al. 1983; Walker and Edwards 1983; Rumpho et al. 1987; Meierhoff and Westhoff 1993) and consequentially minimal production of reducing equivalents through linear electron transport. To supply the reactions in PCR tissue with reducing power, a reducing equivalent shuttle is hypothesized to operate. This shuttle transports 3-PGA from PCR chloroplasts (Figure 2-1) to PCA chloroplasts where 3-PGA is reduced to triosephosphate and moved back to the PCR tissue. It has been proposed that in *Zea mays*, the PCR cycle is compartmented between PCA and PCR chloroplasts with part of the Calvin cycle relegated to the PCA chloroplasts (Majeran et al. 2005). The shuttling of 3-PGA across two chloroplast envelopes into the PCA chloroplasts and the return of triosephosphate to the PCR chloroplasts requires the concerted action of two triosephosphate phosphate translocators. High activity of a transport protein exchanging 3-PGA, triosephosphates,

and P_i has been demonstrated in PCA chloroplasts of *Zea mays* (Day and Hatch 1981; Rumpho and Edwards 1985; Rumpho et al. 1987). In C3 plants the TPT is the most abundant transport protein of the chloroplast envelope and it serves to export triosephosphate from the chloroplast while inorganic phosphate is imported (Flügge and Heldt 1984; Flügge and Heldt 1991). A knock out in TPT causes massive accumulation of transitory starch but no dramatic visible phenotype since the chloroplasts are able to export assimilated carbon during the night in the form of glucose and maltose (Schneider et al. 2002). The transport characteristics determined *in vitro* reveal that the TPT of higher plants is also capable of exchanging 3-PGA for triosephosphates. It has been proposed that TPT functions both as a reducing equivalent shuttle and as the exporter of carbon in C4PCR chloroplasts (Flügge and Heldt 1991). It is unknown whether the inverse direction of 3-PGA transport in comparison to C3 plants (i.e., export of 3-PGA and import of DHAP) requires specific adaptations of the TPT protein, especially since the stromal pH likely causes 3-PGA to have three negative charges.

Both enriching CO2 and minimizing O2 in the vicinity of Rubisco reduces photorespiration in C4 plants, yet it is not absent. Moreover, the separation of the photorespiratory pathway into different tissues has been hypothesized to predate a true C4cycle and indeed be an evolutionary intermediate of true C4 photosynthesis (reviewed in Bauwe, 2009). It has been demonstrated that the final step of the photorespiratory pathway, the phosphorylation of glycerate to 3-PGA by glycerate kinase, is localized exclusively in PCA chloroplasts of different C4 plants belonging to all three subtypes (Usuda and Edwards 1980a). For maize, the localization was independently confirmed by proteomics (Majeran et al. 2005). In *Panicum cappilare*, it was demonstrated that the

PCR tissue photosynthetically produces glycerate which is metabolized in PCA tissue (Usuda and Edwards 1980b). Photorespiration itself is a highly compartmentalized process involving not only the chloroplasts but also peroxisomes and mitochondria (Bauwe, 2008). Even in C3 plants, the transport proteins involved are unknown at the molecular level although biochemical evidence for glycolate and glycerate transport through the same transport protein (Howitz and McCarty 1986; Howitz and McCarty 1991; Young and McCarty 1993). as well as glycine and serine import into mitochondria (Yu et al. 1983)exists. It is currently unknown whether the transport of photorespiratory intermediates in C4 plants occurs through the same transport proteins as in C3 plants especially since in C4 plants the export of glycolate from the chloroplast and the import of glycerate into chloroplasts are located in different tissues.

Enzymes involved in lipid biosynthesis, nitrogen fixation, tetrapyrrol and isoprenoid biosynthesis accumulate preferentially in mesophyll chloroplasts of *Z. mays* whereas enzymes for sulfur import accumulate preferentially in bundle sheath chloroplasts (Majeran et al. 2005). Presumably, the preferential localization of pathways necessitates transport of the pathway products to the other compartment but specific adaptations of transport proteins are not known to date.

Finally, all metabolites entering the chloroplasts do not only need to cross the specificity barrier, the inner envelope, but also the outer envelope. It has recently been demonstrated that specific outer envelope porins are increased in abundance in *Z. mays* chloroplast envelopes compared to *Pisum sativum* chloroplast envelopes (Bräutigam et al. 2008a). The outer envelope porins OEP37 and OEP24 were increased in abundance and OEP21 was decreased. Apparently, the increased amount of metabolite exchange across

the outer envelope is mastered by an increase in specific outer envelope porins demonstrating the importance of the outer envelope for total metabolite traffic to the chloroplast (Bräutigam et al. 2008a).

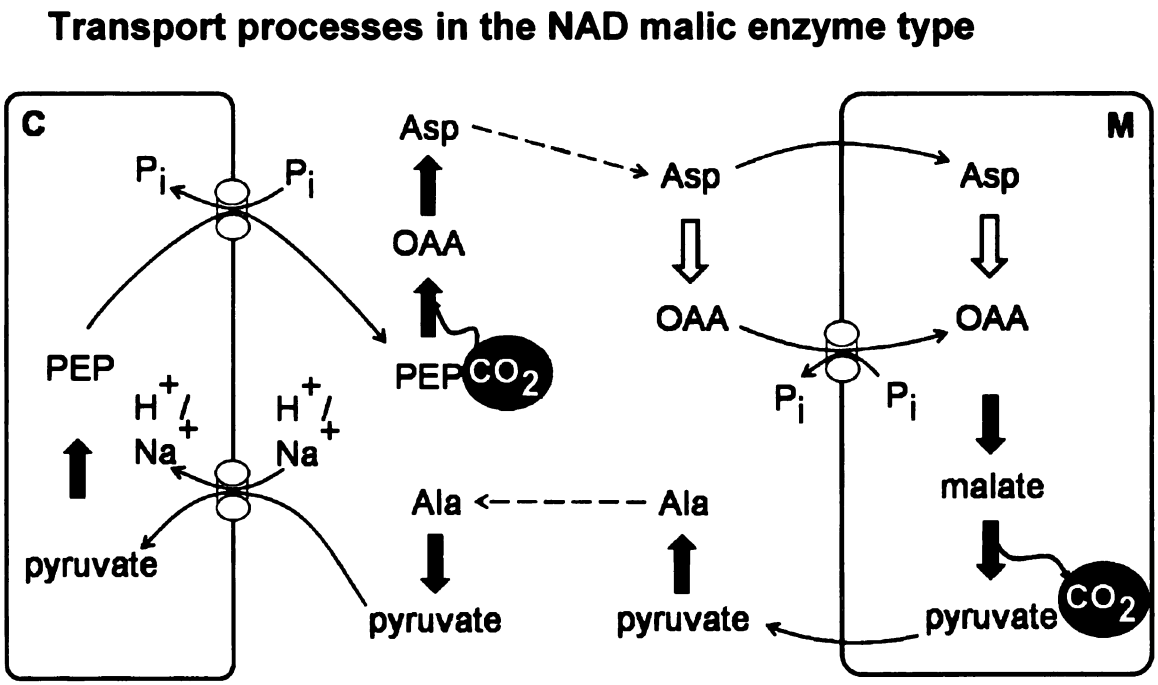


Figure 2-2: Scheme outlining the reactions of C4 photosynthesis and the connecting transport processes for the NAD malic enzyme type of C4 photosynthesis in both the PCA (left) and the PCR (right) compartment; the cell wall is represented by a grey bar; intracellular transport by arrows; intercellular transport by dashed arrows; metabolic reactions by broad arrows; C chloroplast, M mitochondrion; reactions processes which are not resolved are represented by open arrows; for additional abbreviations see text;

The operation of the C4cycle in the NAD malic enzyme (NAD-ME) subtype of C4photosynthesis is compartmentalized between the PCA chloroplasts and cytosol and the PCR mitochondria (Figure 2-2). As is the case in the NADP–ME subtype, the initial carbon acceptor PEP is exported from the chloroplasts and carboxylated in the cytosol. The carboxylation product is OAA, which is subsequently transaminated in the cytosol to yield aspartate. The C4amino acid aspartate then moves to the PCR tissue. The compartmentation of the subsequent transamination to OAA is unresolved (Figure 2-2).

Aspartate may either be imported into the mitochondria and transaminated in the mitochondrial matrix. Alternatively, it may be transaminated to OAA in the cytosol before it is imported. OAA is reduced to malate, probably within the mitochondria (Hatch 1987; Edwards et al. 2001a), although the localization of the corresponding malate dehydrogenase activity has not been demonstrated. Malate is subsequently decarboxylated to yield pyruvate, CO2, and NADH. The resulting CO2 enters the PCR cycle where it is fixed by Rubisco. There are no specific transport proteins for either CO2 or $HCO3^-$ at either the mitochondrial or the chloroplast envelopes known to date. Pyruvate is exported from mitochondria and transaminated to alanine in the cytosol. If the transamination of aspartate to OAA is localized in the mitochondria (Figure 2-2), a 2-oxoglutarate (2-OG) glutamate shuttle is required to connect the transamination of alanine in the cytosol with the one of aspartate in the mitochondria via the amino-donor/ -acceptor pair, glutamate and 2-OG. If the transaminations of aspartate to OAA and pyruvate to alanine both occur in the cytosol (Figure 2-2), they can either be directly coupled or connected by the amino-donor/ -acceptor pair 2-OG and glutamate. Alanine moves back to the PCA cytosol where it is converted to pyruvate in a transamination reaction, and pyruvate is imported back into PCA chloroplasts where the carbon acceptor PEP is regenerated by the action of pyruvate phosphate dikinase (reviewed in Hatch (1987)). In addition to this major route, there is a minor pathway in which malate is formed in and exported from PCA chloroplasts and decarboxylated in PCR mitochondria but this pathway accounts for less than 10% of the total C4 acids moved (Kagawa and Hatch 1975; Hatch et al. 1988). This C4 subtype has also evolved multiple times

independently and includes species from the genera *Amaranthus*, *Cleome*, *Digitaria*, and *Atriplex* (Sage 2004; Muhaidat et al. 2007).

Table 2-2: Summary of transport processes necessary in NAD-ME plants, for abbreviations see text

domain	substrate	co-substrate	name	evidence	references
PCA	PEP	P_i	PPT	biochemical and molecular	Huber and Edwards (1977b); Fischer et al. (1997)
PCR	aspartate	?	?	inferred from presumed localization of soluble proteins	-
	OAA	P_i	DIC	molecular in C3 plants	Palmieri et al. (2008)
PCR	pyruvate	?	?	inferred from localization of soluble proteins	-
PCA	pyruvate	H^+	?	biochemical	Aoki et al. (1992)
		Na^+	?	biochemical	Aoki et al. (1992)

PEP export from PCA chloroplasts

Like in NADP-ME plants the primary carbon acceptor PEP is recycled in PCA chloroplasts by PPDK (Hatch 1987). PEP export is thought to be catalyzed by a member of the same transport protein class, the PPTs, in NAD-ME plants. A detailed description can be found in IIIA.

Dicarboxylate transport in PCR mitochondria

PEP is carboxylated and the resulting OAA transaminated to aspartate in the cytosol of PCA cells. Aspartate diffuses into the PCR domain. Unlike in NADP-ME plants, in NAD-ME plants CO2 is liberated from malate in the mitochondria (Figure 2-2). It remains unclear whether aspartate or OAA is the C4acid imported into mitochondria (Figure 2-2) but like in chloroplasts of NADP-ME plants a net import and not exchange

of C4 acids is necessary. Generally, it is assumed that aspartate is the imported C4 acid in NAD-ME plants (Hatch 1987; Hatch et al. 1988; Edwards et al. 2001b; Hatch 2002) although the metabolite fluxes across the mitochondrial membrane have not been established for any NAD-ME species. In C3 plants, metabolite transport of malate, OAA, aspartate, 2-OG, and glutamate across the mitochondrial membrane has been demonstrated (Desantis et al. 1976; Zoglowek et al. 1988; Hanning et al. 1999) and several proteins catalyzing di- and tricarboxylate exchange have been characterized biochemically and at the molecular level in C3 plants (Laloi 1999; Picault et al. 2002; Picault et al. 2004; Palmieri et al. 2008). The di- and tricarboxylate carrier (DTC) transports a broad spectrum of organic acids. While 2-OG, malate, succinate, OAA, citrate, isocitrate, and sulfate are transported at high levels by DTC from both tobacco and Arabidopsis, both transport proteins do not accept pyruvate or glutamate as a substrate (Picault et al. 2002). DTC depends on a strict counter exchange of two molecules with each other (Picault et al. 2002).The malate transport protein initially identified as the malate transporter from *Panicum miliaceum* (Taniguchi and Sugiyama 1996) is an ortholog of the DTCs characterized from the C3 plants Arabidopsis and tobacco (Picault et al. 2002). Although the spatial and temporal expression pattern of this transporter agrees well with a role in C4photosynthesis (Taniguchi and Sugiyama 1997), the substrate specificity of DTC makes it an unlikely candidate for a major role in C4photosynthesis. The dicarboxylate carrier (DIC) transports a similar spectrum of dicarboxylic acids. DICs from the C3 plant Arabidopsis transport 2-OG, OAA, succinate, malate, phosphate, and sulfate (Palmieri et al. 2008). Like DTCs, DICs do not accept three carbon organic acids such as PEP and pyruvate, or amino acids such as glutamate or

aspartate (Palmieri et al. 2008). In contrast to DTC, the DIC is capable of transporting inorganic anions such as phosphate and DIC can therefore catalyze the net import of a C4 acid into the mitochondrial matrix by exchanging phosphate for a dicarboxylic acid (Palmieri et al. 2008). The transport creates a phosphate imbalance which could be compensated by a phosphate/proton transporter in the mitochondrial inner membrane (McIntosh and Oliver 1994). Neither DTC nor DIC accepts glutamate or aspartate as a substrate. Based on the transport specificities of the proteins known to transport dicarboxylates, we posit that OAA is the main metabolite imported into mitochondria in NAD-ME plants.

Pyruvate Export from PCR mitochondria

Like the import of the C4 acid into mitochondria for decarboxylation, the export of the resulting C3 acid, pyruvate, has not been characterized biochemically in C4 plants of the NAD-ME type (Figure 2-2). In C3 plants, pyruvate is imported into rather than exported from mitochondria since it is one of the substrates for mitochondrial respiration. Isolated mitochondria of the C3 plant pea import pyruvate with saturation kinetics (Proudlove and Moore 1982) and the transport depends on a pH gradient, as protonophores efficiently inhibit transport (Proudlove and Moore 1982). The molecular identity of the mitochondrial pyruvate carrier from plants and other eukaryotes is unknown. Free diffusion of protonated pyruvic acid is an alternative to carrier mediated transport. Small organic acids including pyruvic acid can diffuse through biomembranes (Bakker and Vandam 1974; Proudlove and Thurman 1981; Benning 1986). Possibly, the irreversible decarboxylation of malate creates sufficient amounts of pyruvate within C4 PCR mitochondria to drive export by free diffusion of pyruvate, which is further promoted by

active uptake of pyruvate into C4PCA chloroplasts (Flügge et al. 1985). Alternatively, pyruvate might be transaminated to alanine in the mitochondria already and the neutral amino acid may be exported since neutral amino acids are known to permeate the mitochondrial membrane without specific transport mechanisms (Halling et al. 1973; Wiskich 1977).

Pyruvate import into PCA chloroplasts

After pyruvate is exported from PCR mitochondria and has moved to the PCA compartment, it reenters the PCA chloroplasts to serve as the substrate for PEP regeneration. For a detailed discussion of the transport protein involved, see earlier.

In addition to the metabolites involved in the C4 cycle of NAD-ME plants (summarized in Table 2-2) several additional metabolites likely have to be moved. The liberation of CO2 in the mitochondria and the localization of Rubisco in the chloroplasts possibly results in the need for an inorganic carbon transport protein at either or both organellar membranes. Unlike in NADP-ME plants (Majeran et al. 2008), knowledge about the spatial distribution of reactions like nitrogen and sulfur fixation or isoprenoid biosynthesis in NAD-ME species is limited and consequently, no specific hypothesis about additional transport proteins can be put forward.

Transport processes in the PEP carboxykinase (PEP-CK) type

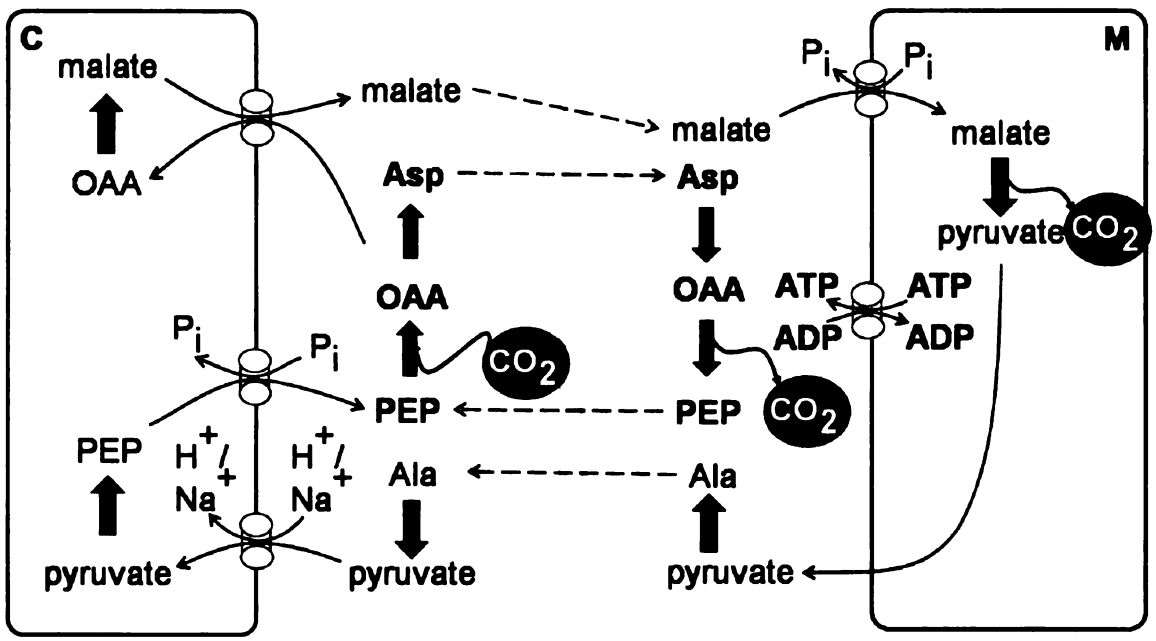


Figure 2-3: Scheme outlining the reactions of C4 photosynthesis and the connecting transport processes for the PEP-CK type of C4 photosynthesis in both the PCA (left) and the PCR (right) compartment; bold print represents the C4 cycle of PEP-CK, normal print represents the C4 cycle of NAD-ME; the cell wall is represented by a grey bar; intracellular transport is represent by arrows; intercellular transport by dashed arrows; metabolic reactions by broad arrows; C chloroplast, M mitochondrion, for additional abbreviations see text;

The core reactions of the C4 cycle involving PEP-CK as the decarboxylation enzyme require the least amount of transport processes (Figure 2-3, black part) (Hatch 1987). In this C4 cycle, the initial carbon acceptor PEP is in part recycled in the cytosol of the PCR tissue. In the PCA cytosol, PEP is carboxylated to OAA, which is subsequently transaminated to aspartate. This C4amino acid moves to the PCR tissue where another transamination takes place and the resulting OAA is decarboxylated by PEP-CK in the cytosol. The CO2 produced enters the PCR cycle of the chloroplasts possibly by free diffusion; whereas PEP moves back to the PCA tissue where it can again serve as the primary carbon acceptor. The cycling metabolites aspartate and PEP transport one aminogroup to PCR cells during each turn of the cycle. The activity of PEP-

CK in the cytosol requires ATP, which is produced in the mitochondria through NADH oxidation and exported from the mitochondria in counter-exchange with ADP (Hatch 1987; Hatch et al. 1988) (Figure 2-3). It has been demonstrated that PEP-CK is not the sole decarboxylation enzyme in PEP-CK type plants. NAD malic enzyme is proposed to contribute about equally to C4 photosynthesis (Figure 2-3, grey part) (Burnell and Hatch 1988a; Burnell and Hatch 1988b; Hatch et al. 1988). Malate is probably the C4acid for this part of the cycle. It is probably produced in PCA chloroplasts and exported in a manner similar to that in NADP-ME plants (Figure 2-2and 2-3). After diffusing to the PCR compartment, malate enters the mitochondria for decarboxylation. The C3 acid produced through NAD malic enzyme in the mitochondria is exported, either as pyruvate or possibly after being transaminated to alanine, and re-imported as pyruvate into PCA chloroplasts to serve as the substrate for recycling the primary carbon acceptor PEP. The cycling metabolites malate and alanine transport one amino group to the PCA tissue in each cycle and therefore balance the amino group moved by aspartate and PEP if both pathways contribute about 50% of the total decarboxylation reactions. The C4 cycle involving PEP-CK does not rely on intracellular transport except for the provision of ATP, but the C4 cycle involving the NAD-ME in mitochondria relies heavily on intracellular transport.

Table 2-3: Summary of transport processes necessary in PEP-CK plants, for abbreviations see text

domain	substrate	co-substrate	name	evidence	references
PCA	PEP	P_i	PPT	inferred from NADP-ME and NAD-ME plants	-
PCA	OAA	malate	OAT	inferred from NADP-ME and NAD-ME plants	-
			DiT	inferred from NADP-ME and NAD-ME plants	-
PCR	malate	P_i	DIC	molecular in C3 plants	Palmieri et al. (2008)
PCR	pyruvate	?	?	inferred from localization of soluble proteins	-
PCR	ATP	ADP	NTT (chloroplasts)	molecular in C3 plants	Neuhaus et al. (1997)
PCR	ATP	ADP	AAC (mitochondria)	molecular in C3 plants	Emmermann et al. (1991); Winning et al. (1991)
PCA	pyruvate	H^+	?	inferred from NADP-ME and NAD-ME plants	-
		Na^+	?	inferred from NADP-ME and NAD-ME plants	-

PEP export from PCA chloroplasts

PEP export is reduced in PEP-CK plants by at least half (Burnell and Hatch 1988b; Burnell and Hatch 1988a; Hatch et al. 1988) compared to other C4 species since part of the PEP is recycled from the cytosol of PCR tissue where it is generated by PEP-CK. PEP recycled in the PCA chloroplasts by PPDK is likely exported through the PPT described earlier.

Oxaloacetate malate exchange in PCA chloroplasts

Up to 50% of the fixed carbon is transported to the PCR tissue as malate (Burnell and Hatch 1988b; Burnell and Hatch 1988a; Hatch et al. 1988) which is probably produced in PCA chloroplast from OAA. The candidate transport protein is described in detail earlier.

Malate import into PCR mitochondria

Like in NAD-ME plants part of the CO2 liberated in PCR cells is produced in the mitochondria by NAD dependent malic enzyme (Burnell and Hatch 1988b; Burnell and Hatch 1988a; Hatch et al. 1988). Since malate is the C4 acid transported to PCR cells no transamination reactions are required and malate can directly be imported in counter-exchange with an inorganic anion such as phosphate or sulfate. For a detailed analysis of the transport process, see earlier.

ATP/ADP translocation to supply PEP-CK with ATP

The import of malate may not only serve to supply NAD-ME with its substrate but also to drive the tricarboxylic acid cycle which produces ATP for the cell. The ATP demand of PCR cells in PEP-CK plants is high since PEP-CK uses one molecule of ATP to liberate CO2 and thus uses ATP at the rate of carbon fixation. The mitochondrial ATP/ADP carrier mediates the exchange of ATP and ADP across the mitochondrial membrane (Emmermann et al. 1991; Winning et al. 1991). Alternatively, ATP may be derived from cyclic electron flow and transported by the plastidic ATP/ADP carrier to the cytosol (Neuhaus et al. 1997; Weber 2004).

The transport processes in PEP-CK plants are summarized in Table 2-3.

Transport processes in single cell C4metabolism

Plants which operate the C4 cycle within one cell do not require intercellular transport through plasmodesmata. In these plants, the initial carbon fixation by PEPC and the subsequent decarboxylation and refixation by Rubisco are spatially separated within one cell, which underscores the importance of intracellular transport processes necessary for enriching CO2 at the site of Rubisco. Examples include *Bienertia cycloptera*, *Suada aralocaspica* (formerly *Borszczowia aralocaspica*) and *Hydrilla verticillata* (Magnin et al. 1997; Voznesenskaya et al. 2001; Voznesenskaya et al. 2002). In *B. cycloptera*, the PCA compartment consists of grana-deficient chloroplasts in the cytoplasm around the periphery of the cell which is devoid of mitochondria. This PCA compartment is separated by the vacuole from the central PCR compartment, which contains mitochondria and typical granal chloroplasts. Both compartments are connected by cytoplasmic channels, which cross the vacuole. *B. cycloptera* is a NAD malic enzyme type plant (Voznesenskaya et al. 2002) and requires all intracellular transport steps described for typical NAD malic enzyme type plants with Kranz anatomy. In *S. aralocaspica*, the PCA and PCR compartments localize to the proximal and distal ends of the same cell, respectively, and are separated by a large vacuole. Again, mitochondria are localized to the PCR compartment only and, like *B. cycloptera*, *S. aralocaspica* is a NAD malic enzyme species (Voznesenskaya et al. 2001). *H. verticillata* is a facultative C4 species in which C4 photosynthesis of the NADP malic enzyme type is induced under CO2 limiting conditions (Magnin et al. 1997). Although single cell C4 plants do not need to transfer metabolites symplastically between different cells, they have to move metabolites efficiently from the PCA domain to the PCR domain and back. It is currently

not known whether metabolite flow is assisted by specialized structures within the cell. The intracellular transport requirements are the same as in C4 plants with Kranz anatomy.

Future prospects

Discovering the molecular identity of C4adapted transport proteins

For virtually all transport proteins the molecular identity in C4 species is unknown although there are a number of good candidates which can be inferred from proteins characterized in C3 systems. For example, ever since it was discovered in *Zea mays* endosperm tissue and characterized from cauliflower, PPT has been assumed to be the exporter of the primary carbon acceptor PEP. A PPT isoform expressed at moderate to high levels in C4 leaves was recently discovered by proteomic analysis of maize mesophyll chloroplast envelope membranes (Bräutigam et al., 2008). Where candidate transport proteins have been identified it is far from clear how the high metabolite flow across the envelope is sustained by the transport protein. Possibly, the C4 isoforms of transport proteins have altered kinetic characteristics compared to C3 homologues or the increased flow is simply accomplished by increased amounts of transport proteins, thereby increasing the *V*max of transport.

Most of the biochemical characterizations of organelle transport proteins have been limited to one C4 species, *Zea mays*. New tools relying on collecting data from non-model species in a high throughput manner may allow studies with the goal of discovering and characterizing the remaining transport proteins. Transcriptomics approaches, if they involve *de novo* sequencing of cDNAs, show great promise in both generating sequence information for downstream applications (Bräutigam et al. 2008b) as well as generating quantitative information on transcript abundance (Weber et al. 2007).

After sequence information has been generated, proteomics of organellar membranes and the soluble proteomes can reveal specific adaptations of C4 organelles to sustain transport capacity (Majeran et al. 2005; Bräutigam et al. 2008a; Majeran et al. 2008). The comparison of data from different C4 species should provide information on differences and similarities in transport properties of specific subtypes of C4 photosynthesis. Since the different C4subtypes employ different enzymes as well as different transport proteins, the identification and characterization of molecular candidate proteins for many of the metabolite fluxes is likely possible. In this sense, maybe the question of which model system best serves to study the C4 syndrome, maize, Flaveria, or Cleome (Brown et al. 2005), could be best answered with Zea *mays*, Flaveria species and Cleome, possibly also including in addition a PEP-CK C4photosynthesis plant and one or several single-cell C4 species. Detailed knowledge about the transport proteins and their specific adaptations will assist in engineering C4 photosynthesis-like metabolism in crop plants, irrespective as to whether the engineering aims to generate single-cell or multiple cell C4photosynthesis. The discovery and characterization of novel metabolite transport proteins may also inform studies of metabolite transport in C3 species. For example, the identification of a pyruvate transporter may help to understand the contributions of pyruvate to fatty acid synthesis during seed filling and to branched chain amino acid biosynthesis, fatty acid biosynthesis, and isoprenoid biosynthesis in leaf tissues.

Prospects for engineering C4photosynthesis into C3 crop species

Metabolite flow across the membrane has been shown to be limiting for metabolic pathways (Häusler et al. 2000; Zhang et al. 2008).For example, antisense repression of TPT in tobacco leads to over-accumulation of transitory starch in chloroplasts during the

day since the export of triosephosphate is compromised (Häusler et al. 2000). A simultaneous increase of glucose-6-phosphate and ATP import into potato tuber amyloplasts by overexpression of both, the glucose-6-phosphate and ATP translocators increases sink strength and yield in potatoes because carbon can be relocated to the amyloplasts with greater efficiency (Zhang et al. 2008). Both examples illustrate how metabolite flow through a pathway can be restricted or increased by decreasing and increasing metabolite flow across membranes. Numerous studies aiming to engineer or alter metabolic pathways have uncovered substrate availability as one of the major limitations of pathway engineering (reviewed in Kunze et al. (2002)). Since metabolite flow across membranes during C4 photosynthesis exceeds all metabolite flows across membranes known in C3 plants, increasing membrane transport capacity will play a key role in engineering C4photosynthesis in C3 crops. The over-expression of PPDK alone in C3 species such as Arabidopsis, potato, and rice has minimal physiological impacts despite over-expression levels of up to 40-fold have been achieved in rice (Fukayama et al. 2001; reviewed in Matsuoka et al. (2001) and in Miyao (2003)). It has been discussed that the limited impact of PPDK overexpression is due to free reversibility of the reaction (Burnell and Hatch 1985) in combination with low activities of inorganic pyrophosphatase and adenylate kinase in C3 plants (Matsuoka et al. 2001).However, the limitation may well be both the lack of pyruvate import as well as PEP export from the chloroplast, which limit substrate availability and cause product accumulation. To substantially increase the metabolite flow through PPDK both pyruvate import and PEP export need to be increased since both pyruvate and PEP transport are minor activities in C3 plants and the chloroplast envelope is not adapted to allow major changes without

genetic modifications. Recently, single cell C4 metabolism has been considered an alternative model system for engineering C3 plants to perform C4 photosynthesis since its creation would avoid major alterations to leaf architecture and cell-to-cell connectivity. Intracellular transport, however, is as crucial to single cell C4 photosynthesis as it is to C4 photosynthesis in species with Kranz anatomy.

Acknowledgements

Work in the author's lab is supported by grants of the National Science Foundation (USA) and the German Research Foundation (DFG, Germany). AB is grateful to the Barnett-Rosenberg Foundation and the Deutsche Studienstiftung for financial support.

The authors wish to thank Dr. R. Sage and two anonymous reviewers for helpful comments on improving the manuscript.

References

Anderson JW, House CM (1979) Polarographic Study Of Oxaloacetate Reduction By Isolated Pea-Chloroplasts Plant Physiol 64, 1058-1063

Andre C, Benning C (2007) Arabidopsis seedlings deficient in a plastidic pyruvate kinase are unable to utilize seed storage compounds for germination and establishment Plant Physiol 145, 1670-1680

Andre C, Froehlich JE, Moll MR, Benning C (2007) A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in Arabidopsis Plant Cell 19, 2006-2022

Aoki N, Ohnishi J, Kanai R (1992) 2 Different Mechanisms For Transport Of Pyruvate Into Mesophyll Chloroplasts Of C-4 Plants-A Comparative-Study Plant Cell Physiol 33, 805-809

Bakker EP, Vandam K (1974) Movement Of Monocarboxylic Acids Across Phospholipid Membranes - Evidence For An Exchange Diffusion Between Pyruvate And Other Monocarboxylate Ions Biochim Biophys Acta 339, 285-289

Benning C (1986) Evidence Supporting A Model Of Voltage-Dependent Uptake Of Auxin Into Cucurbita Vesicles Planta 169, 228-237

Benz R (1994) Permeation Of Hydrophilic Solutes Through Mitochondrial Outer Membranes - Review On Mitochondrial Porins Biochim Biophys Acta 197, 167-196

Boag S, Jenkins CLD (1985) CO2 Assimilation And Malate Decarboxylation By Isolated Bundle Sheath Chloroplasts From Zea-Mays Plant Physiol 79, 165-170

Boag S, Jenkins CLD (1986) The Involvement Of Aspartate And Glutamate In The Decarboxylation Of Malate By Isolated Bundle Sheath Chloroplasts From Zea-Mays Plant Physiol 81, 115-119

Bolter B, Soll J, Hill K, Hemmler R, Wagner R (1999) A rectifying ATP-regulated solute channel in the chloroplastic outer envelope from pea EMBO J 18, 5505-5516

Bräutigam A, Hofmann-Benning S, Weber APM (2008a) Comparative Proteomics of Chloroplast Envelopes from C3 and C4 Plants Reveals Specific Adaptations of the Plastid Envelope to C4 Photosynthesis and Candidate Proteins Required for Maintaining C4 Metabolite Fluxes Plant Physiol 148, 568-579

Bräutigam A, R.P. S, Whitten D, Wilkerson CG, Carr KM, Froehlich JE, Weber APM (2008b) Comparison of the use of a species-specific database generated by

pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes J Biotechnol 136, 44-53

Brown NJ, Parsley K, Hibberd JM (2005) The futured C-4 research - Maize, Flaveria or Cleome? Trends Plant Sci 10, 215-221

Burnell JN, Hatch MD (1985) Light Dark Modulation Of Leaf Pyruvate, P(I) Dikinase Trends Biochem Sci 10, 288-291

Burnell JN, Hatch MD (1988a) Photosynthesis In Phosphoenolpyruvate Carboxykinase-Type-C4 Plants - Pathways Of C-4 Acid Decarboxylation In Bundle Sheath-Cells Of Urochloa-Panicoides Arch Biochem Biophys 260, 187-199

Burnell JN, Hatch MD (1988b) Photosynthesis In Phosphoenolpyruvate Carboxykinase-Type-C4 Plants - Photosynthetic Activities Of Isolated Bundle Sheath-Cells From Urochloa-Panicoides Arch Biochem Biophys 260, 177-186

Craig S, Goodchild DJ (1977) Leaf Ultrastructure Of Triodia-Irritans - C4 Grass Possessing An Unusual Arrangement Of Photosynthetic Tissues Aust J Bot 25, 277-290

Day DA, Hatch MD (1981) Transport Of 3-Phosphoglyceric Acid, Phosphoenolpyruvate, And Inorganic-Phosphate In Maize Mesophyll Chloroplasts, And The Effect Of 3-Phosphoglyceric Acid On Malate And Phosphoenolpyruvate Production Arch Biochem Biophys 211, 743-749

Desantis A, Arrigoni O, Palmieri F (1976) Carrier-Mediated Transport Of Metabolites In Purified Bean Mitochondria Plant Cell Physiol 17, 1221-1233

Ding B, Turgeon R, Parthasarathy MV (1992) Substructure Of Freeze-Substituted Plasmodesmata Protoplasma 169, 28-41

Edwards GE, Franceschi VR, Ku MSB, Voznesenskaya EV, Pyankov VI, Andreo CS (2001a) Compartmentation of photosynthesis in cells and tissues of C4 plants J Exp Bot 52, 577-590

Edwards GE, Furbank RT, Hatch MD, Osmond CB (2001b) What Does It Take to Be C4? Lessons from the Evolution of C4 Photosynthesis Plant Physiol 125, 46-49

Emmermann M, Braun HP, Schmitz UK (1991) The ADP ATP Translocator From Potato Has A Long Amino-Terminal Extension Curr Genet 20, 405-410

Erwee MG, Goodwin PB (1985) Symplast Domains In Extrastellar Tissues Of Egeria-Densa Planch Planta 163, 9-19

Evert RF, Eschrich W, Heyser W (1977) Distribution And Structure Of Plasmodesmata In Mesophyll And Bundle-Sheath Cells Of Zea-Mays-L Planta 136, 77-89

Fischer K, Arbinger B, Kammerer B, Busch C, Brink S, Wallmeier H, Sauer N, Eckerskorn C, Flügge UI (1994) Cloning and in-Vivo Expression of Functional Triose Phosphate/Phosphate Translocators from C-3-Plants and C-4-Plants - Evidence for the Putative Participation of Specific Amino-Acid-Residues in the Recognition of Phosphoenolpyruvate Plant J 5, 215-226

Fischer K, Kammerer B, Gutensohn M, Arbinger B, Weber A, Häusler RE, Flügge UI (1997) A new class of plastidic phosphate translocators: A putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter Plant Cell 9, 453-462

Flügge UI, Heldt HW (1984) The Phosphate-Triose Phosphate-Phosphoglycerate Translocator of the Chloroplast Trends Biochem Sci 9, 530-533

Flügge UI, Heldt HW (1991) Metabolite Translocators of the Chloroplast Envelope Annu Rev Plant Physiol Plant Mol Biol 42, 129-144

Flügge UI, Stitt M, Heldt HW (1985) Light-Driven uptake of Pyruvate into mesophyll Chloroplasts from maize FEBS Lett 183, 335-339

Fukayama H, Tsuchida H, Agarie S, Nomura M, Onodera H, Ono K, Lee BH, Hirose S, Toki S, Ku MSB, Makino A, Matsuoka M, Miyao M (2001) Significant accumulation of C-4-specific pyruvate,orthophosphate dikinase in a C-3 plant, rice Plant Physiol 127, 1136-1146

Furbank RT, Hatch MD (1987) Mechanism Of C-4 Photosynthesis - The Size And Composition Of The Inorganic Carbon Pool In Bundle Sheath-Cells Plant Physiol 85, 958-964

Goetze TA, Philippar K, Ilkavets I, Soll J, Wagner R (2006) OEP37 Is a New Member of the Chloroplast Outer Membrane Ion Channels J Biol Chem 281, 17989-17998

Halling PJ, Brand MD, Chappell JB (1973) Permeability Of Mitochondria To Neutral Amino-Acids FEBS Lett 34, 169-171

Hanning I, Baumgarten K, Schott K, Heldt HW (1999) Oxaloacetate transport into plant mitochondria Plant Physiol 119, 1025-1031

Hatch M (2002) C4 photosynthesis: discovery and resolution Photosyn Res 73, 251-256

Hatch M, Droscher L, Flügge UI, Heldt HW (1984a) A specific Translocator for Oxaloacetate Transport in Chloroplasts FEBS Lett 178, 15-19

Hatch MD (1987) C-4 Photosynthesis - a Unique Blend of Modified Biochemistry, Anatomy and Ultrastructure Biochim Biophys Acta 895, 81-106

Hatch MD, Agostino A, Burnell JN (1988) Photosynthesis In Phosphoenolpyruvate Carboxykinase-Type C-4 Plants - Activity And Role Of Mitochondria In Bundle Sheath-Cells Arch Biochem Biophys 261, 357-367

Hatch MD, Droscher L, Flügge UI, Heldt HW (1984b) A Specific Translocator for Oxaloacetate Transport in Chloroplasts FEBS Lett 178, 15-19

Hattersley PW, Browning AJ (1981) Occurrence Of The Suberized Lamella In Leaves Of Grasses Of Different Photosynthetic Types.1. In Parenchymatous Bundle Sheaths And Pcr (Kranz) Sheaths Protoplasma 109, 371-401

Häusler RE, Schlieben NH, Flügge UI (2000) Control of carbon partitioning and photosynthesis by the triose phosphate/phosphate translocator in transgenic tobacco plants (Nicotiana tabacum). II. Assessment of control coefficients of the triose phosphate/phosphate translocator Planta 210, 383-390

Heber U (1974) Metabolite Exchange Between Chloroplasts And Cytoplasm Annu Rev Plant Physiol Plant Mol Biol 25, 393-421

Henderson SA, Voncaemmerer S, Farquhar GD (1992) Short-Term Measurements Of Carbon Isotope Discrimination In Several C4 Species Aust J Plant Physiol 19, 263-285

Howitz KT, McCarty RE (1986) D-Glycerate Transport By The Pea Chloroplast Glycolate Carrier - Studies On [1-C-14] D-Glycerate Uptake And D-Glycerate Dependent O-2 Evolution Plant Physiol 80, 390-395

Howitz KT, McCarty RE (1991) Solubilization, Partial-Purification, And Reconstitution Of The Glycolate Glycerate Transporter From Chloroplast Inner Envelope Membranes Plant Physiol 96, 1060-1069

Huber SC, Edwards GE (1977a) Transport in C4 Mesophyll Chloroplasts - Characterization of Pyruvate Carrier Biochim Biophys Acta 462, 583-602

Huber SC, Edwards GE (1977b) Transport in C4 Mesophyll Chloroplasts - Evidence for an Exchange of Inorganic-Phosphate and Phosphoenolpyruvate Biochim Biophys Acta 462, 603-612

Kagawa T, Hatch MD (1975) Mitochondria As A Site Of C4 Acid Decarboxylation In C4-Pathway Photosynthesis Arch Biochem Biophys 167, 687-696

Knappe S, Lottgert T, Schneider A, Voll L, Flügge UI, Fischer K (2003) Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis-AtPPT1 may be involved in the provision of signals for correct mesophyll development Plant J 36, 411-420

Kubasek J, Setlik J, Dwyer S, Santrucek J (2007) Light and growth temperature alter carbon isotope discrimination and estimated bundle sheath leakiness in C-4 grasses and dicots Photosyn Res 91, 47-58

Kunze R, Frommer WB, Flügge UI (2002) Metabolic engineering of plants: The role of membrane transport Metab Eng 4, 57-66

Laisk A, Edwards GE (2000) A mathematical model of C-4 photosynthesis: The mechanism of concentrating CO2 in NADP-malic enzyme type species Photosyn Res 66, 199-224

Laloi M (1999) Plant mitochondrial carriers: an overview Cell Mol Life Sci 56, 918-944

Leegood RC (1985) The Intercellular Compartmentation of Metabolites in Leaves of Zea-Mays-L Planta 164, 163-171

Leegood RC, Crowther D, Walker DA, Hind G (1983) Energetics Of Photosynthesis In Zea-Mays.1. Studies Of The Flash-Induced Electrochromic Shift And Fluorescence Induction In Bundle Sheath-Cells Biochim Biophys Acta 722, 116-126

Leegood RC, Edwards GE (1996) 'Photosynthesis and the Environment.' Kluwer Academic Publishers: Dordrecht

Magnin NC, Cooley BA, Reiskind JB, Bowes G (1997) Regulation and localization of key enzymes during the induction of Kranz-less, C-4-type photosynthesis in Hydrilla verticillata Plant Physiol 115, 1681-1689

Majeran W, Cai Y, Sun Q, van Wijk KJ (2005) Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics Plant Cell 17, 3111-3140

Majeran W, Zybailov B, Ytterberg AJ, Dunsmore J, Sun Q, van Wijk KJ (2008) Consequences of C4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells Mol Cell Proteomics 7, 1609-1638

Matsuoka M, Furbank RT, H. F, M. M (2001) Molecular Engeneering of C4 Photosynthesis Annu Rev Plant Physiol Plant Mol Biol 52, 297-314

McIntosh CA, Oliver DJ (1994) The Phosphate Transporter From Pea Mitochondria - Isolation And Characterization In Proteolipid Vesicles Plant Physiol 105, 47-52

Meierhoff K, Westhoff P (1993) Differential Biogenesis Of Photosystem-Ii In Mesophyll And Bundle-Sheath Cells Of Monocotyledonous Nadp-Malic Enzyme-Type C-4 Plants - The Nonstoichiometric Abundance Of The Subunits Of Photosystem-Ii In The Bundle-Sheath Chloroplasts And The Translational Activity Of The Plastome-Encoded Genes Planta 191, 23-33

Miyao M (2003) Molecular evolution and genetic engineering of C4 photosynthetic enzymes J Exp Bot 54, 179-189

Muhaidat R, Sage RF, Dengler NG (2007) Diversity of Kranz anatomy and biochemistry in C-4 eudicots Am J Bot 94, 362-381

Murcha MW, Elhafez D, Lister R, Tonti-Filippini J, Baumgartner M, Philippar K, Carrie C, Mokranjac D, Soll J, Whelan J (2007) Characterization of the Preprotein and Amino Acid Transporter Gene Family in Arabidopsis Plant Physiol 143, 199-212

Neuhaus HE, Thom E, Mohlmann T, Steup M, Kampfenkel K (1997) Characterization of a novel eukaryotic ATP/ADP translocator located in the plastid envelope of Arabidopsis thaliana L Plant J 11, 73-82

Ohnishi J, Flügge UI, Heldt HW, Kanai R (1990) Involvement Of Na+ In Active Uptake Of Pyruvate In Mesophyll Chloroplasts Of Some C4 Plants - Na+/Pyruvate Cotransport Plant Physiol 94, 950-959

Ohnishi JI, Kanai R (1987a) Na-(+)-Induced Uptake Of Pyruvate Into Mesophyll Chloroplasts Of A C-4 Plant, Panicum-Miliaceum FEBS Lett 219, 347-350

Ohnishi JI, Kanai R (1987b) Pyruvate Uptake By Mesophyll And Bundle Sheath Chloroplasts Of A C-4 Plant, Panicum-Miliaceum L Plant Cell Physiol 28, 1-10

Overall RL, Wolfe J, Gunning BES (1982) Inter-Cellular Communication In Azolla Roots.1. Ultrastructure Of Plasmodesmata Protoplasma 111, 134-150

Palmieri L, Picault N, Arrigoni R, Besin E, Palmieri F, Hodges M (2008) Molecular identification of three Arabidopsis thaliana mitochondrial dicarboxylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization Biochem J 410, 621-629

Picault N, Hodges M, Paimieri L, Palmieri F (2004) The growing family of mitochondrial carriers in Arabidopsis Trends Plant Sci 9, 138-146

Picault N, Palmieri L, Pisano I, Hodges M, Palmieri F (2002) Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria - Bacterial

expression, reconstitution, functional characterization, and tissue distribution J Biol Chem 277, 24204-24211

Pohlmeyer K, Soll J, Grimm R, Hill K, Wagner R (1998) A High-Conductance Solute Channel in the Chloroplastic Outer Envelope from Pea Plant Cell 10, 1207-1216

Pohlmeyer K, Soll J, Steinkamp T, Hinnah S, Wagner R (1997) Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane Proc Natl Acad Sci U S A 94, 9504-9509

Proudlove MO, Moore AL (1982) Movement Of Amino-Acids Into Isolated Plant-Mitochondria FEBS Lett 147, 26-30

Proudlove MO, Thurman DA (1981) The Uptake of 2-Oxoglutarate and Pyruvate by Isolated Pea-Chloroplasts New Phytol 88, 255-264

Renne P, Dressen U, Hebbeker U, Hille D, Flügge UI, Westhoff P, Weber APM (2003) The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2 Plant J 35, 316-331

Roberts AG, Oparka KJ (2003) Plasmodesmata and the control of symplastic transport Plant, Cell Environ 26, 103-124

Rumpho ME, Edwards GE (1984) Inhibition Of 3-Phosphoglycerate-Dependent O-2 Evolution By Phosphoenolpyruvate In C-4 Mesophyll Chloroplasts Of Digitaria-Sanguinalis (L) Scop Plant Physiol 76, 711-718

Rumpho ME, Edwards GE (1985) Characterization Of "4,4'-Diisothiocyano-2,2'-Disulfonic Acid Stilbene Inhibition Of 3-Phosphoglycerate-Dependent O-2 Evolution In Isolated-Chloroplasts - Evidence For A Common Binding-Site On The C-4 Phosphate Translocator For 3-Phosphoglycerate, Phosphoenolpyruvate, And Inorganic-Phosphate Plant Physiol 78, 537-544

Rumpho ME, Wessinger ME, Edwards GE (1987) Influence Of Organic-Phosphates On 3-Phosphoglycerate Dependent O-2 Evolution In C-3 And C-4 Mesophyll Chloroplasts Plant Cell Physiol 28, 805-813

Sage RF (2004) The evolution of C-4 photosynthesis New Phytol 161, 341-370

Sawers RJH, Liu P, Anufrikova K, Hwang JTG, Brutnell TP (2007) A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf BMC genomics 8:12

Scheibe R (2004) Malate valves to balance cellular energy supply Physiol Plant 120, 21-26

Scheibe R, Backhausen JE, Emmerlich V, Holtgrefe S (2005) Strategies to maintain redox homeostasis during photosynthesis under changing conditions J Exp Bot 56, 1481-1489

Schneider A, Häusler RE, Kolukisaoglu U, Kunze R, van der Graaff E, Schwacke R, Catoni E, Desimone M, Flügge UI (2002) An Arabidopsis thaliana knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished Plant J 32, 685-699

Schneidereit J, Häusler RE, Fiene G, Kaiser WM, Weber APM (2006) Antisense repression reveals a crucial role of the plastidic 2-oxoglutarate/malate translocator DiT1 at the interface between carbon and nitrogen metabolism Plant J 45, 206-224

Schwender J, Seemann M, Lichtenthaler HK, Rohmer M (1996) Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga Scenedesmus obliquus Biochem J 316, 73-80.

Singh BK, Shaner DL (1995) Biosynthesis of Branched-Chain Amino-Acids - from Test-Tube to Field Plant Cell 7, 935-944

Somerville SC, Ogren WL (1983) An Arabidopsis-Thaliana Mutant Defective in Chloroplast Dicarboxylate Transport Proc Natl Acad Sci USA 80, 1290-1294

Sowinski P, Szczepanik J, Minchin PEH (2008) On the mechanism of C4photosynthesis intermediate exchange between Kranz mesophyll and bundle sheath cells in grasses J Exp Bot 59, 1137-1147

Stitt M, Heldt HW (1985) Generation and Maintenance of Concentration Gradients between the Mesophyll and Bundle Sheath in Maize Leaves Biochim Biophys Acta 808, 400-414

Streatfield SJ, Weber A, Kinsman EA, Häusler RE, Li JM, Post-Beittenmiller D, Kaiser WM, Pyke KA, Flügge UI, Chory J (1999) The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression Plant Cell 11, 1609-1621

Taniguchi M, Sugiyama T (1996) Isolation, characterization and expression of cDNA clones encoding a mitochondrial malate translocator from Panicum miliaceum L Plant Mol Biol 30, 51-64

Taniguchi M, Sugiyama T (1997) The expression of 2-oxoglutarate/malate translocator in the bundle-sheath Mitochondria of Panicum miliaceum, a NAD-malic enzyme-type C-4 plant, is regulated by light and development Plant Physiol 114, 285-293

Taniguchi M, Taniguchi Y, Kawasaki M, Takeda S, Kato T, Sato S, Tahata S, Miyake H, Sugiyama T (2002) Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in Arabidopsis thaliana Plant Cell Physiol 43, 706-717

Taniguchi Y, Nagasaki J, Kawasaki M, Miyake H, Sugiyama T, Taniguchi M (2004) Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize Plant Cell Physiol 45, 187-200

Usuda H, Edwards GE (1980a) Localization Of Glycerate Kinase And Some Enzymes For Sucrose Synthesis In C-3 And C-4 Plants Plant Physiol 65, 1017-1022

Usuda H, Edwards GE (1980b) Photosynthetic Formation Of Glycerate In Isolated Bundle Sheath-Cells And Its Metabolism In Mesophyll-Cells Of The C-4 Plant Panicum-Capillare L Aust J Plant Physiol 7, 655-662

Voll LM, Häusler RE, Hecker R, Weber APM, Weissenböck G, Fiene G, Waffenschmidt S, Flügge UI (2003) The phenotype of the Arabidopsis cue1 mutant is not simply caused by a general restriction of the shikimate pathway Plant J 36, 301ff

Voznesenskaya EV, Franceschi VR, Kiirats O, Artyusheva EG, Freitag H, Edwards GE (2002) Proof of C4photosynthesis without Kranz anatomy in Bienertia cycloptera (Chenopodiaceae) Plant J 31, 649-62

Voznesenskaya EV, Franceschi VR, Kiirats O, Freitag H, Edwards GE (2001) Kranz anatomy is not essential for terrestrial C4plant photosynthesis Nature 414, 543-6

Walker DA, Edwards GE (1983) 'C3, C4: mechanisms, and cellular and environmental regulation, of photosynthesis.' Blackwell Scientific publications: Oxford

Weber A, Menzlaff E, Arbinger B, Gutensohn M, Eckerskorn C, Flügge UI (1995) The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells Biochemistry 34, 2621-7.

Weber APM (2004) Solute transporters as connecting elements between cytosol and plastid stroma Curr Opin Plant Biol 7, 247-253

Weber APM, Fischer K (2007) Making the connections - The crucial role of metabolite transporters at the interface between chloroplast and cytosol FEBS Lett 581, 2215-2222

Weber APM, Flügge UI (2002) Interaction of cytosolic and plastidic nitrogen metabolism in plants J Exp Bot 53, 865-874

Weber APM, Linka M, Bhattacharya D (2006) Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor Eukaryot Cell 5, 609-612

Weber APM, Schwacke R, Flügge UI (2005) Solute transporters of the plastid envelope membrane Ann Rev Plant Biol 56, 133-164

Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the Arabidopsis Transcriptome with Massively Parallel Pyrosequencing Plant Physiol 144, 32-42

Winning BM, Day CD, Sarah CJ, Leaver CJ (1991) Nucleotide-Sequence Of 2 Cdnas Encoding The Adenine-Nucleotide Translocator From Zea-Mays L Plant Mol Biol 17, 305-307

Wiskich JT (1977) Mitochondrial Metabolite Transport Annu Rev Plant Physiol Plant Mol Biol 28, 45-69

Woo KC, Flügge UI, Heldt HW (1987) A 2-Translocator Model for the Transport of 2-Oxoglutarate and Glutamate in Chloroplasts During Ammonia Assimilation in the Light Plant Physiol 84, 624-632

Young XK, McCarty RE (1993) Assay Of Proton-Coupled Glycolate And D-Glycerate Transport Into Chloroplast Inner Envelope Membrane-Vesicles By Stopped-Flow Fluorescence Plant Physiol 101, 793-799

Yu C, Claybrook DL, Huang AHC (1983) Transport Of Glycine, Serine, And Proline Into Spinach Leaf Mitochondria Arch Biochem Biophys 227, 180-187

Zhang L, Häusler RE, Greiten C, Hajirezaei M, Haferkamp I, Neuhaus HE, Flügge U-I, Ludewig F (2008) Overriding the co-limiting import of carbon and energy into tuber amyloplasts increases the starch content and yield of transgenic potato plants Plant Biotechnol J, 6, 453-464

Zoglowek C, Kromer S, Heldt HW (1988) Oxaloacetate And Malate Transport By Plant-Mitochondria Plant Physiol 87, 109-115

Chapter 3

Comparison of the use of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes

This chapter has been previously published in the Journal of Biotechnology: Bräutigam A, Shrestha RP, Whitten D, Wilkerson CG, Carr KM, Froehlich JE and Weber APM (2008) Comparison of the use of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes Journal of Biotechnology 136, 44-53

I would like to acknowledge Kevin Carr for contribution of the pea data to Figure 3-1.

Abstract

Proteomics is a valuable tool for establishing and comparing the protein content of defined tissues, cell types, or subcellular structures. Its use in non-model species is currently limited because the identification of peptides critically depends on sequence databases. In this study, a low-quality preliminary cDNA database for the non-model species *Pisum sativum* created by a small number of massive parallel pyrosequencing (MPSS) runs was compared to comprehensive cDNA databases from *Medicago truncatula* and *Arabidopsis thaliana* created by Sanger sequencing. Each database was used to identify proteins from a pea leaf chloroplast envelope preparation. It is shown that the pea database identified more proteins with higher accuracy, although the sequence quality was low and the sequence contigs were short compared to databases from model species. Although the number of identified proteins in non-species specific databases could potentially be increased by lowering the threshold for successful protein identifications, this strategy markedly increases the number of wrongly identified proteins. The identification rate with non-species specific databases correlated with spectral abundance but not with the predicted membrane helix content, and strong conservation is necessary but not sufficient for protein identification with a non-species specific database. It is concluded that massively-parallel sequencing of cDNAs substantially increases the power of proteomics in non-model species.

Introduction

The first sequenced plant model species, *Arabidopsis thaliana* (thale cress), was chosen not only for its relatively small genome, but also for its small size and rapid life cycle that make it amenable to genetics (Meinke et al., 1998; Somerville and Somerville, 1999; TAGI, 2000). In addition, a large collection of mutants is available, including sequence-indexed insertion mutants that facilitate both forward and reverse genetics approaches (Jander et al., 2002; Parinov and Sundaresan, 2000). However, due to its small size, the presence of a range of secondary metabolites, and the lack of established protocols for isolation of subcellular organelles, Arabidopsis does not always represent the ideal model system for, e.g., organellar proteomics. Proteomics is as a valuable tool for establishing the protein complement of cells and subcellular structures (Baginsky and Gruissem, 2006; Baginsky et al., 2004; Dunkley et al., 2006; Heazlewood et al., 2004; Ito et al., 2007; Kleffmann et al., 2006; Lilley and Dupree, 2006; Peltier et al., 2004; von Zychlinski et al., 2005; Ytterberg et al., 2006), especially since the prediction capability of bioinformatics approaches proved insufficient for large scale annotation of organelle proteomes (Jarvis, 2004; Millar et al., 2006; Reyes-Prieto et al., 2007). In addition, multiple targeting of proteins has been documented frequently (Duchene et al., 2005; Millar et al., 2006; Taira et al., 2004) and recently also non-canonical targeting of proteins to, e.g., chloroplast via the secretory system (Miras et al., 2002; Miras et al., 2007; Radhamony and Theg, 2006; Villarejo et al., 2005). In contrast to Arabidopsis, the garden pea (*Pisum sativum*), is excellently suited for organelle isolation and biochemical studies of enzymes and established protocols for organelle isolation are available in the literature (e.g., Corpas et al., 1999; Miflin and Beevers, 1974; Tobin, 1996).

Unfortunately, little is known about the power of proteomics in non-model species for which no extensive sequence database is available. Current peptide identification technology relies on the generation of ideal mass spectra from theoretical libraries. A sequence database is translated in six frames and the resulting protein sequences are *in silico* digested with trypsin. The resulting peptides are used to calculate an ideal mass spectrum. If an observed spectrum matches a theoretically predicted spectrum with a certain probability the corresponding peptide is called "identified". This method of identification demands a perfect sequence match between the sample peptide and the database peptide, although some programs, such as implementations of the X!-Tandem software (Craig and Beavis, 2004), allow the inclusion of single amino acid mismatches. Allowing more than one mismatch increases the error rate and the time required for the search. With increasing evolutionary distance, perfect matches become less likely even between highly conserved proteins, in particular since conservative changes, such as aspartate to glutamate will already cause a spectral mismatch. In contrast, low quality databases, such as the one discussed in this communication, limit the identification of peptides either by sequencing and assembling errors causing amino acid changes in predicted peptides or by not providing enough peptide coverage for correct identifications due to short contigs. *De novo* sequencing of peptides is considered too slow and limited by computing time for high throughput applications (Baginsky and Gruissem, 2006; Pevtsov et al., 2006). Currently, the identification of proteins from non-model species with limited sequence coverage frequently relies on databases generated from closely related species (Schmidt et al., 2007) or indeed all sequences that are available in public

databases (Taylor et al., 2005) although this method will especially limit the identification of less conserved proteins.

It is has been recently proposed to use massively-parallel pyrosequencing to fully explore the potential of proteomics in non-model species such as pea (Weber et al., 2007). In this study, we systematically assessed the potential and limitations of massively-parallel pyrosequencing to support proteomics applications. To this end, we compared proteomics based on a low-coverage transcriptome sequence database of the garden pea consisting of many short sequence contigs with frequent frame shift errors with a conventionally created and fairly comprehensive cDNA database of a closely related model species (*Medicago truncatula*), and with a high-quality, virtually error-free database generated from a completely sequenced model species (*Arabidopsis thaliana*). We established the limitations of each database and we tested how the degree of conservation, the abundance of mass spectra generated from a particular protein, and the number of transmembrane domains influences the odds for successful protein identification using a non-species specific database. Finally we discuss the consequences of interpreting the proteomics sample based on the different database results.

Material and Methods

Massively-parallel pyrosequencing and generation of sequence databases

Three different databases were generated for proteome analyses. For the generation of the pea transcriptome database, one non-normalized and several normalized libraries were generated and sequenced using massively-parallel pyrosequencing technology (Margulies et al., 2005). The preparation of cDNA libraries was conducted as

described previously (Weber et al., 2007), with the exception that some libraries were normalized to decrease the proportion of highly abundant transcripts. To this end, 1 μg of double-stranded cDNA was normalized using a commercial kit (Trimmer-kit, Evrogen, Moscow, Russia) that is based on Kamchatka crab duplex-specific nuclease (Zhulidov et al., 2004). Following normalization, the cDNA populations were amplified by PCR and 3 μg of the resulting PCR-amplified cDNA was used per sequencing reaction. One cDNA library was generated from leaves and one from hypocotyls. Five preparations generated from the leaf library (four normalized, one non-normalized) and one preparation from the hypocotyl library (normalized) were sequenced using a Roche/454 GS20 instrument. This technology delivered an average read length of 100 nucleotides. Two additional libraries from etiolated and de-etiolated (4h light) leaves were sequenced on a half plate each, using a GS FLX instrument, which allowed for average read lengths of 250 nucleotides. The preliminary pea database used in the reported work was assembled from approximately 2 million pyrosequencing reads with a total of about 230 million nucleotides. The pyrosequencing reads were combined with publicly available pea cDNA sequences from Genbank and the IPK Gatersleben EST database. All sequences, except the full-length and partial mRNA sequences from GenBank, were subjected to QA using the SeqClean EST trimming and validation tool cDNAs and ESTs were clustered and assembled using the TGI Clustering Tools (TGICL) in a multi stage pipeline. The sequence reads were loosely clustered with a modified version of megablast (Zhang et al., 2000) and subsequently the clusters were submitted to the CAP3 sequence assembling program (Huang and Madan, 1999). All programs are available at http://compbio.dfci.harvard.edu/tgi/software. 1,570,251 reads (80% of the total) were

assembled into 135,250 contigs. For the proteome analysis, only contigs longer than 100 base pairs were used. The pea cDNA database is currently undergoing further development and a detailed description of the pea cDNA sequence database and its assembly will be reported in an upcoming manuscript (R.P. Shrestha et al., in preparation). The Medicago database was based on the *M. truncatula* gene index of tentative consensus sequences assembled by TIGR, currently maintained at the Dana-Faber Cancer Institute and Harvard Medical School of Health (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago). For the Arabidopsis database the latest Arabidopsis proteome annotation from the TAIR7 (www.arabidopsis.org) genome release was used.

Processing of proteomics samples

As a proteome sample, chloroplast envelope membranes were isolated from 10-14 days old pea plants as described previously (Douce and Joyard, 1979; Keegstra and Yousif, 1986). Envelope membrane samples (approx. 100 µg of protein) were mixed with SDS–PAGE loading buffer, incubated for 20 minutes on a reaction tube shaker at 15° C, and subsequently separated by 12.5% SDS-PAGE. After staining with Coomassie Brilliant Blue, each gel lane was cut into ten equally-sized slices. Proteins contained in the gel slices were subjected to tryptic cleavage as described by Shevchenko et al. (Shevchenko et al., 1996). Extracted peptides were automatically loaded onto a Waters Symmetry C18 peptide trap (5 µm, 180 µm x 20mm) at a flow rate of 4 µL/min in 2% Acetonitrile/0.1%Formic Acid for 5 minutes by a Waters nanoAcquity Sample Manager. The peptides were eluted onto a Waters BEH C18 nanoAcquity column (1.7 µm, 100 µm x 100mm) and eluted over 90 minutes using a Waters nanoAcquity UPLC into a

ThermoElectron LTQ-FTICR mass spectrometer with a flow rate of 300 nL/min (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 99.9% Acetonitrile/0.1% Formic Acid: gradient of 5% B to 40% B from 0 to 63 minutes, 40% B to 90% B from 63 to 71 minutes and 5% B from 71 to 90 minutes). The top ten ions of each survey scan, which were taken at a resolution of 50,000, were subjected to automated low energy collision induced dissociation. The resulting MS/MS spectra were converted to a peak list using BioWorks Browser v 3.2. This empirical mass spectra library of 27,443 queries was compared to three databases with sequences from *Arabidopsis thaliana*, *Medicago truncatula* and *Pisum sativum*, respectively, using the Mascot searching algorithm, v 2.2 (www.matrixscience.com). Carbamidomethyl cysteine was set as a fixed peptide modification and oxidation of methionine was allowed. Up to two missed tryptic sites were allowed. The peptide tolerance was set to +/-10ppm and the MS/MS tolerance to 0.8kDa.

Analysis of proteomics data

The Mascot output was loaded into the program Scaffold (www.proteomesoftware.com) which calculates protein identification probabilities based on PeptideProphet and ProteinProphet (Keller et al., 2002; Nesvizhskii et al., 2003). The protein identification threshold was set to 99% probability for all databases and at least two unique peptides for each protein ("stringent criteria") or to 99% probability for at least one database with lesser probabilities allowed for the other databases on the same protein and one unique peptide for each protein ("relaxed criteria"). For further analysis the data was exported to MS Excel. For each protein identified, the sequences in both the pea database as well as the Medicago database were mapped to the Arabidopsis proteome

using BlastX (Altschul et al., 1997) and the Arabidopsis gene identifier (AGI) of the closest homologue was noted. Based on its closest Arabidopsis homologue, the protein was annotated and the predicted number of membrane spanning helices was retrieved from TAIR. In cases where the AGIs obtained from the three different identification strategies did not match, the identifications were manually inspected and priority was given to the identification with the highest probability score. For identical probability scores the highest number of unique peptides mapped to a protein was declared the correct identification. When multiple sequences were matched to the same spectra, identifications were either called "mistaken" if none of the matches were identical to the protein identified correctly in the other databases or "multiple" if one of the proteins was identical to the correct match determined as outlined earlier. The complete list of proteins was collapsed into a non-redundant list based on the AGIs with sequences yielding the same hit in the BlastX hit being summed as the same protein.

Results

Properties of the cDNA sequence databases

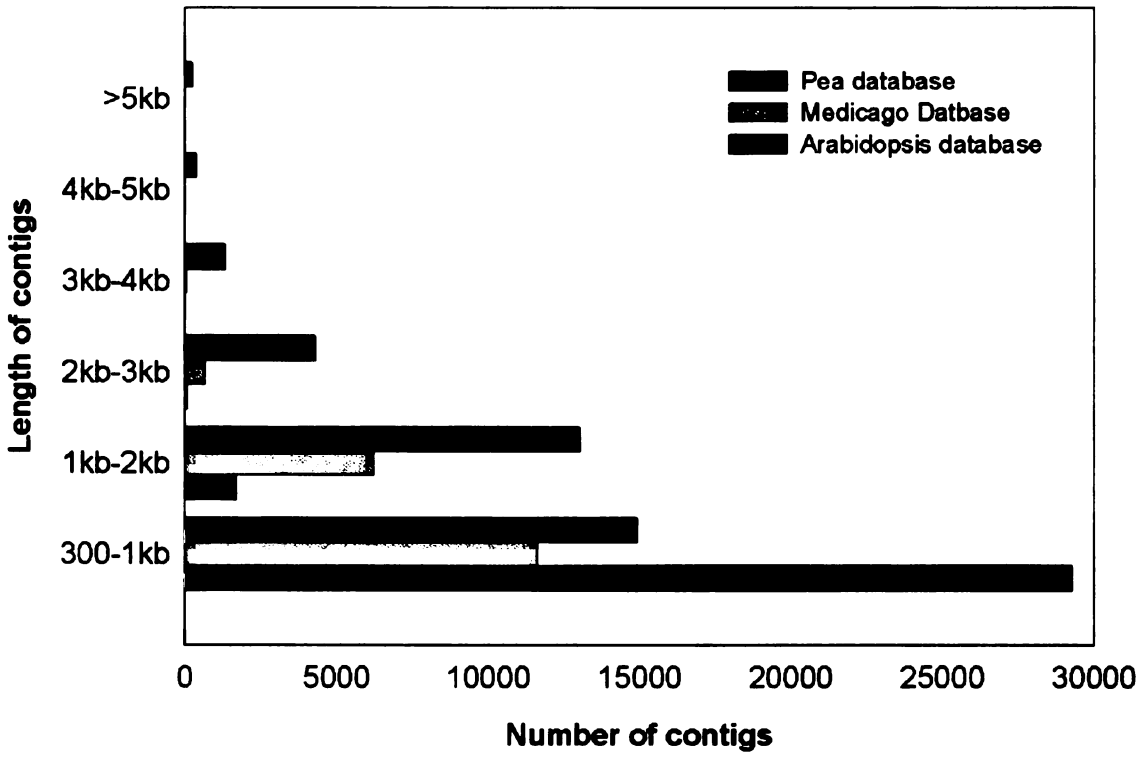


Figure 3-1. The length distribution of contigs (assembled transcripts) in databases used for proteome analyses.

The characteristics of a low-coverage pea cDNA database generated by limited pyrosequencing was compared to two databases generated with conventional sequencing technology. The pea cDNA sequence database contained more than 31,000 relatively short contigs. About 29,000 contigs were between 300 and 1,000 nucleotides in length, about twice as many in this length category as in the Medicago (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago) and an Arabidopsis transcriptome databases (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=arab) (Figure 3-1). In contrast, contigs between one and two kb

were massively underrepresented (1697 contigs) in the pea database in comparison to the Medicago database (6,224 long contigs) and the Arabidopsis database (13,033 long contigs) and contigs longer than 2000 nucleotides were almost absent from the pea database (Figure 3-1). In addition to having shorter contigs, many pea contigs, when translated in all six frames, apparently contained frame shifts that are likely due to assembly or base calling errors (see Supplementary File of pea contigs).

The Medicago database consisted of the tentative consensus sequences (TCs) assembled from EST projects from *M. truncatula*. Medicago is being developed as a model legume species and is hypothesized to serve as a research template for the garden pea. The Medicago TCs were comprised of about 36,000 unique sequences. About one half of these sequences were tentative consensus sequences that represent contigs consisting of several ESTs and the other half were singletons. The length distribution of the Medicagos TCs was more similar to that of the Arabidopsis transcriptome database and the sequence quality judged from open reading frame annotation software was high (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago). While both the Arabidopsis and Medicago databases contain similar numbers of cDNA sequences that are shorter than 1kb, the Medicago database contains only half as many sequences between one and two kb, and only one tenth for transcripts longer than two kb (Figure 3-1). The Arabidopsis proteome derived from the TAIR7 build of the completely sequenced genome served as an example for a completely sequenced genome. The capability of the new pea database to reliably identify proteins despite its obvious shortcomings was compared with the capabilities of the databases of *M. truncatula* and *A. thaliana*.

Performance of the databases in proteomics

Table 3-1: Summary of protein identifications with different databases

Database	stringent criteria			relaxed criteria		
	number of identifications	total number of spectra	avg. number of spectra per ID	number of identifications	total number of spectra	avg. number of spectra per ID
combined	255	8222	32	283	8892	31
Pisum sativum	221	5012	23	255	5139	20
Medicago truncatula	125	1977	16	203	2198	11
Arabidopsis thaliana	82	1233	15	165	1555	9

A pea chloroplast envelope membrane proteome sample was analyzed using the sequence databases described above. To identify the most advantageous database and program parameters for protein identification, we employed a factorial approach. An empirical library of uninterpreted mass spectra was generated from a single proteomics experiment of pea chloroplast envelope membranes. This particular sample was chosen because it represents a relatively minor share of the total cellular proteome and because it contains a large number of highly hydrophobic membrane proteins and thus represents a challenging target for protein identification. In addition, several proteomic studies of chloroplast envelope membranes from a range of plant species have been published, thus providing good references for comparison. All three databases described above were matched to the empirical spectra library using either stringent or relaxed criteria. For both criteria, the largest number of proteins could be identified with the pea database, followed

by Medicago and Arabidopsis. Applying stringent criteria, a total of 8,222 spectra were matched to 255 non-redundant proteins using a combination of all three databases. A list of all identifications can be found in Supplementary Table 3-1 and a non-redundant list of identified proteins in Supplementary Table 3-2. Under stringent conditions the pea database allowed matches of 5,012 spectra against 221 non-redundant proteins (86% of the total). On average each protein was identified by 23 spectra with up to 362 spectra per protein (Table 3-1). The Medicago database yielded 1,977 matched spectra on 125 proteins (49% of the total). With only 32% or 82 proteins the Arabidopsis database allowed the fewest matches. With relaxed criteria the total number of proteins identified using a combination of all three databases increased to 283 and 8,892 spectra were matched. In the pea database, the relaxed criteria allowed for 5,139 matching spectra, but a large number of proteins that were identified with only a few spectra caused the average number of spectra per protein to drop to 20. The Medicago TC database yielded 2,198 spectra mapping to 203 proteins, almost 72% of the total. The *A. thaliana* database yielded the lowest number of identified proteins, with only 165 proteins or 58% of the total (Table 3-1).

There were 36 proteins which could not be identified with the pea database but yielded significant identifications with at least one of the other databases. Eight of these proteins are encoded on an organelle genome (Supplemental Table 3-2). Six proteins were very similar to closely related proteins, which were present in the pea database and were identified. For nine of these proteins corresponding sequences were absent from the pea database. The remaining twelve proteins remained unidentified with the pea database

although corresponding sequences were contained in the database, as indicated by Blast searches against the database.

Abundance of mass spectra is positively correlated with correct protein identification in non-species specific databases

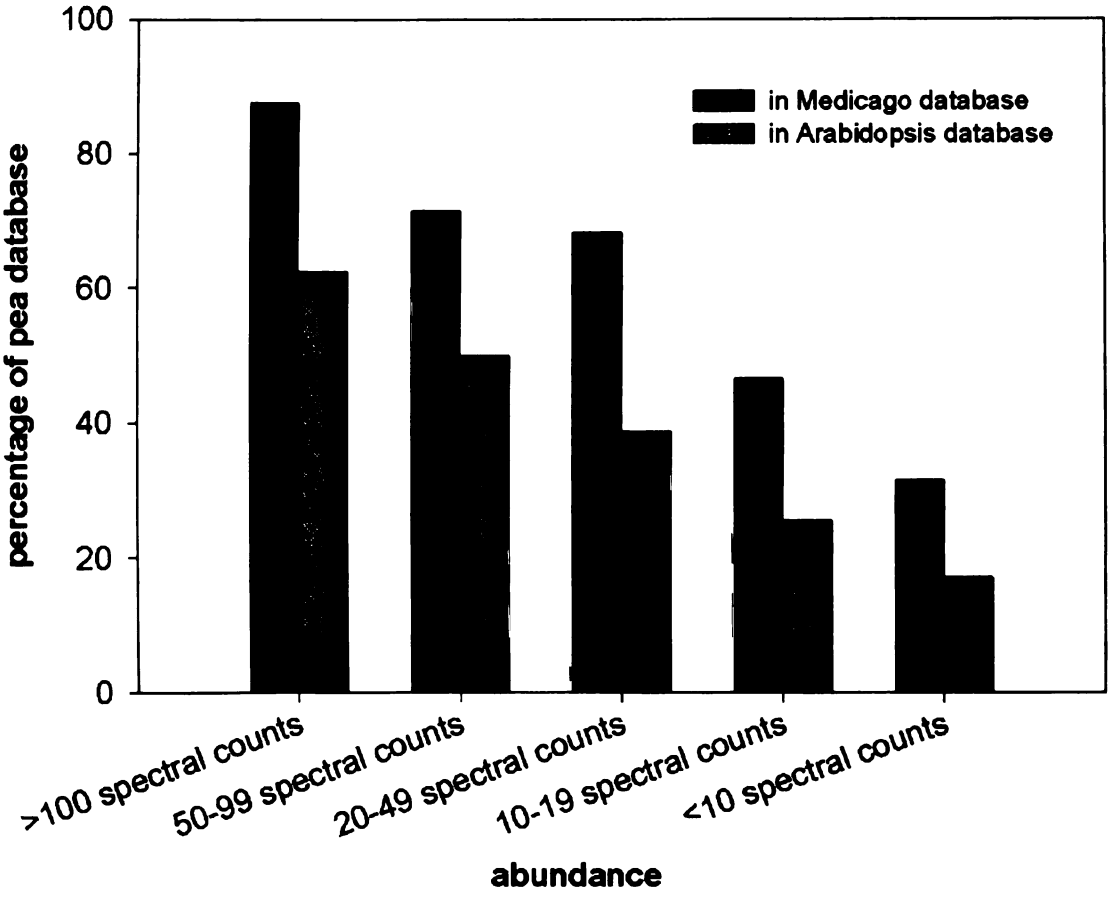


Figure 3-2. The number of spectral counts detected for a protein is correlated with the rate of identification. Proteins in the pea database were grouped by spectral abundance and the percentage of proteins also identified with the non-species specific databases was plotted. The following absolute numbers equal 100%: >100 spectral counts eight proteins, 50-99 spectral counts 14 proteins, 20-49 spectral counts 44 proteins, 10-19 spectral counts 43 proteins and <10 spectral counts 111 proteins.

With both non-species specific databases 51% or 68% of the proteins could not be identified when stringent criteria were applied. We next investigated whether there are

protein characteristics that might serve as predictors of identification probability in the non-species specific databases. Although the correlation between the spectral count and the abundance of a protein in not absolute, the spectral count has been used successfully to estimate protein abundance (Liu et al., 2004; Lu et al., 2007; Zybailov et al., 2005). The proteins in the pea database were arranged into five groups based on the abundance of their matching mass spectra. More than half of the protein identifications using the pea database were based on two to ten spectra, whereas only 3.6% of the proteins were identified with more than one hundred spectra each. For each abundance class, the percentage of proteins that could also be identified with the Medicago and the Arabidopsis databases were plotted (Figure 3-2). The abundance of spectra as determined by the spectral count obtained with the pea database correlated with the identification rate with non-species specific databases (R(squared)>0.95). With the Medicago database almost all of the proteins represented by more than 100 mass spectra could be identified whereas only 38% of the proteins with a spectral count below 10 were identified. With the Arabidopsis database about two thirds of the proteins with a high spectral count were identified and the identification rate dropped to 17% for proteins that had spectral counts of less than 10 (Figure 3-2).

Protein hydrophobicity is a poor predictor for correct protein identification

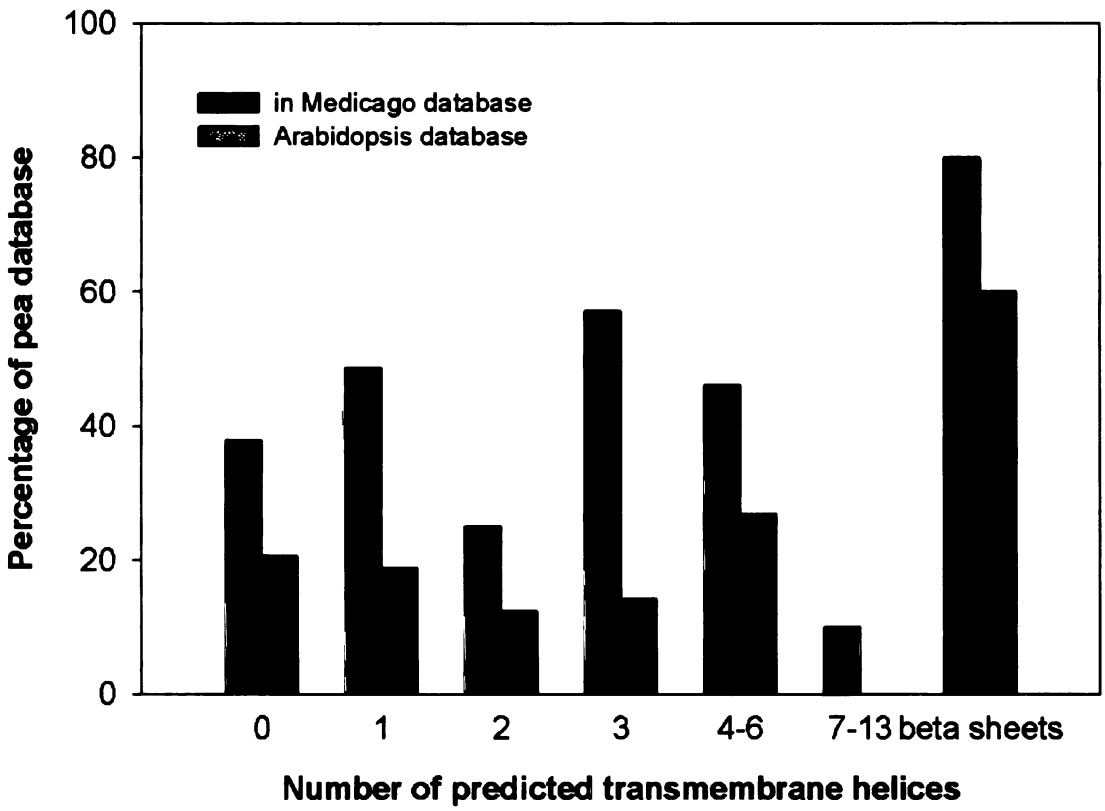


Figure 3-3. The number of predicted membrane spanning helices for a protein is not correlated with the rate of identification. Proteins in the pea database were grouped by predicted membrane spanning helices and the percentage of proteins also identified with the non-species specific databases was plotted. The following absolute numbers equal 100%: no predicted transmembrane helices 111 proteins, one predicted transmembrane helix 37 proteins, two predicted transmembrane helices 24 proteins, three predicted transmembrane helices seven proteins, four to six predicted transmembrane helices 26 proteins, seven to thirteen predicted transmembrane helices ten proteins, predicted beta sheets five proteins

We also tested whether the presence of transmembrane helices adversely affected the probability that a protein was identified using a non-species specific database as has been proposed by (Eichacker et al., 2004). Proteins that were identified using the pea database were grouped according to the number of predicted membrane spanning helices (none, one, two, three, four to six, more than seven) and the presence of beta sheets. More than half of the proteins identified with the pea database were not predicted to contain

transmembrane helices and the other groups contained 2.3-16.8% of the proteins. For each group the percentage of proteins that could also be identified with the non-species specific databases was plotted (Figure 3-3). Unlike protein abundance as deduced from spectral counts, the number of predicted transmembrane helices is of less predictive value for identification rate since no correlation was observed between the number of transmembrane domains and the probability for identification using a non-species specific database. However, among the proteins with more than seven predicted membrane spanning helices only one (Medicago database) or none (Arabidopsis database) was identified out of ten proteins identified with the pea database.

Analysis of abundant non-identified and of wrongly identified proteins

Although proteins with more matching spectra in the pea database had a higher probability to be also identified with the non-species specific databases, many of the fifteen proteins with the highest number of spectra could only be identified with the species-specific database (Table 3-2). The Medicago database did not allow the identification of the small subunit of RubisCO and a chaperonin, and a hydroperoxide lyase and the triosephosphate phosphate translocator (TPT) did not pass the stringent threshold for identification probability ($p<0.01$). With the Arabidopsis database both the hydroperoxide lyase and the small subunit of RubisCO remained unidentified whereas the TPT, two components of the protein import complex, Toc159 and ToC34, a carbonic anhydrase and an unknown protein did not pass the probability threshold (Table 3-2). The sequences of TPT, which was not identified with the Medicago and the Arabidopsis databases, and of malate dehydrogenase (NAD-MDH), which was identified with all three databases, were aligned and the tryptic fragments were determined. PsTPT is highly

similar to both AtTPT (77.5% identity and 14.5% similarity) and MtTPT (92.8% identity and 14.5% similarity). The same is true for PsNAD-MDH and AtNAD-MDH (73% identity and 18.6% similarity) and PsNAD-MDH and MtNAD-MDH (91.7% identity and 3.9% similarity). Despite the high degree of sequence identity AtTPT shares only one and MtTPT only six tryptic peptides with the pea protein. The TPT peptides from the empirical spectra library were mapped onto the aligned sequences. All but one of the TPT peptides that generated mass spectra contained at least one amino acid exchange with respect to Medicago or Arabidopsis. In contrast, 16 peptides of the Medicago NAD-MDH exactly matched the pea peptides. For thirteen of those 16 theoretically predicted peptides mass spectra were detected experimentally, thus permitting reliable identification of the protein using the Medicago database. Only four tryptic peptides of the Arabidopsis NAD-MDH matched the pea protein sequence, but since for two of them mass spectra were experimentally detected the protein could still be reliably identified.

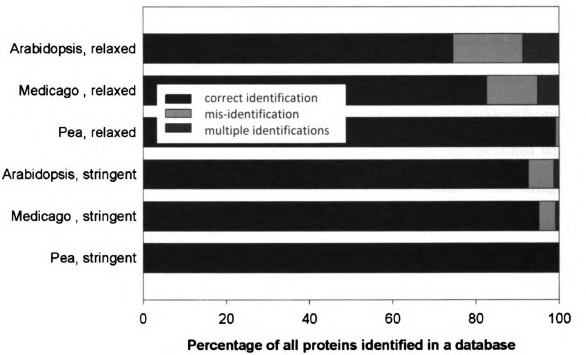


Figure 3-5. The non-species specific databases yield a significant amount of mis-identifications and multiple identifications

When the successful protein identifications were compared between the three databases, several non-matching identifications were revealed (Figure 3-5). Manual inspection of the peptide sequences, the peptide probability scores, and the protein coverage indicated that one of the identifications is likely correct. With stringent criteria none of the identifications with the pea database appeared to be invalid compared to the non-species specific databases. Using relaxed criteria, two proteins were erroneously identified. Peptides for the large subunit of RubisCO were mistakenly annotated as part of a RubisCO-like protein of unknown function encoded on the mitochondrial genome. A second misidentification is of the root glutamate synthase (GLU2) instead of the leaf glutamate synthase (GLU1). Using the Medicago database, 24 proteins were erroneously identified with relaxed criteria and in three cases the peptide with matching spectra is part

of more than one protein and therefore yields multiple identifications of proteins only one of which is correct. Sixteen of the misidentified proteins are closely related to the correct protein and eight identifications corresponded to completely different protein. Unlike with the pea database, even under stringent conditions, fifteen misidentifications persisted. With the Arabidopsis database 29 proteins were mistakenly identified when relaxed criteria were used and eight multiple identifications existed. More stringent criteria still resulted in five misidentifications and in one peptide, which identified multiple proteins, but all persisting misidentifications identify closely related proteins (for details see Supplementary Table 3-2).

Discussion

The results presented in this paper indicate that the prospects for identifying proteins from a species with limited sequence resources by proteomics can be massively increased by generating a species-specific transcriptome database by MPSS, even if the resulting database is of low quality, compared to sequence databases generated by conventional sequencing. When non-species specific databases are used, the odds for protein discovery are limited, and the probability to identify a protein can be predicted by its abundance but not by its content of membrane spanning helices. Strong sequence conservation is necessary but not sufficient to identify a protein with a non-species specific database. Especially when the identification criteria are relaxed to allow imperfect matches and therefore more protein identifications with non-species specific databases, the identifications are more prone to erroneous identifications.

Our initial expectation was that both short contig length and the relatively high rate of sequencing errors that are characteristic to low-coverage MPSS projects would severely limit the prospects for successful protein identification. Unexpectedly, however, the pea database developed in this study was superior to the tested non-species specific databases with regard to the rate of protein discovery and the quality of identifications despite being of low quality and low coverage, as compared to conventional databases. Not only more proteins were identified but also the average number of spectra mapping to each protein was higher (Table 3-1). The advantage of species specificity clearly outweighs the quality issues of the database. Only 10-14% of the total proteins identified remained unidentified with this database. Detailed analysis of these non-identified proteins revealed that organelle-encoded proteins are overrepresented (22%) among those not identified with the pea database, as compared to the overall proportion of identified proteins that are encoded in organelles (4%) (Supplemental Table 3-2). Since the mRNA-isolation protocol used in this study included two consecutive rounds of poly-A+ purification and because reverse translation of mRNA into first strand cDNA was primed by oligo-dT, it is likely that transcripts from organellar genomes are underrepresented in the sample, as reported previously (Weber et al., 2007). For nuclear encoded proteins, sequences could not be identified for nine proteins in the pea transcriptome database and for seven additional proteins the corresponding nucleotide sequences were either short or fragmented into multiple unassembled short contigs, thus demonstrating how an unfinished MPSS-generated sequence database limits proteomic identification technology if the sequence contigs are too short. Incidentally, most of the contigs in the pea database are shorter than one kb with the majority being shorter than 400 nucleotides (Figure 3-1).

This translates into a stretch of approximately 140 amino acids, given no 5' or 3'-UTRs are contained in the sequence. Depending on how the tryptic fragments map onto the short sequence, the identification probability can probably be too low to identify proteins reliably. This problem is currently being addressed by generating additional sequence coverage and new assembling methods for the short and midrange sequence reads obtained by the MPPS technology. The recently released GS FLX instrument produces sequence reads that are 2.5-times longer as the sequence reads of the GS 20. In addition, preliminary tests showed that including a second, less stringent clustering step in the assembly pipeline might serve to produce longer contigs that are more suitable for proteomics applications although this may aggravate the problem of frame shift errors during assembly. In conclusion, the analysis of the proteome sample with several databases served to determine the quality of the sequence database for proteomics. Since it was less likely for proteins to be identified with non-species specific databases the presence of such identifications indicated that the species-specific database could still be improved.

Considering the multitude of short contigs (Figure 3-1) and the presence of sequencing errors that resulted in frame shifts (Supplementary File) it was surprising that, with both databases from model species, fewer proteins were identified than with the novel pea database (Table 3-1). The Arabidopsis database represents a completely sequenced genome that is well annotated and presumably complete. Although the evolutionary split between the genome of *P. sativum* and *A. thaliana* occurred about 100 million years ago (Wikström et al., 2001), we hypothesized that the degree of conservation might be sufficient to identify most proteins, albeit with a lower peptide

count, especially given that the database is complete. Surprisingly, the Arabidopsis database only allowed the identification of 32% of the proteins relative to all databases combined. Notably, in a previous chloroplast envelope proteomics study that used spinach chloroplasts and mostly non-species specific sequences for identification only 50 proteins (25% compared to this study and 15% compared to (Froehlich et al., 2003)) were identified in total, amongst them 21 membrane proteins with more than 4 membrane helices (Ferro et al., 2002). The phosphate translocators that were identified in this previous study were limited to those for which species-specific sequences were available in public databases at that time. In addition to the Arabidopsis database a more closely related model species database was tested for its performance in a proteomics application. *M. truncatula* and *P. sativum* are close relatives in the subfamily Papilionoideae which separated about 25 million years ago (Lavin et al., 2005), which is equivalent to the evolutionary distance between Arabidopsis and the Brassica species (Yang et al., 1999), and both microsynteny and macrosynteny between the genomes have been demonstrated (Gualtieri et al., 2002; Kalo et al., 2004). Medicago is being developed as the model legume species, also serving as a research template for pea. Since the evolutionary distance between Medicago and pea is smaller than between Arabidopsis and pea, we hypothesized that the Medicago database would allow for the identification of more proteins. Indeed, the Medicago database allowed the detection of half of the proteins under stringent conditions, although the average number of spectra detected for each protein was only 50% of those obtained with the pea database. Although the number of sequence contigs longer than 1kb is significantly lower in the Medicago database than in the Arabidopsis database (Figure 3-1), indicating that a large portion of the available

sequences do not represent full length cDNAs, the Medicago database is more successful than the Arabidopsis database in identifying chloroplast envelope proteins (compare to Table 3-1). Based on the quality of the databases and the number of proteins that could be identified, we conclude that evolutionary distance imposes a higher penalty for protein identification than does the quality of the sequence database. This underscores the inability of current peptide identification software for proteomics applications to tolerate amino acid mismatches between the theoretically and experimentally generated spectra especially when stringent identification criteria are applied.

We hypothesized that relaxed identification criteria may allow imperfect matches for peptides and might thus enhance the identification rate for both the Medicago and the Arabidopsis database. Relative to the total number of proteins identified, 58% of proteins could be identified with relaxed criteria compared to 32% with stringent criteria with the Arabidopsis database and 72% compared to 49% with the Medicago database. Unfortunately, the relaxed identification criteria also lead to a high number of mistakenly identified proteins and the identification of multiple family members that shared a single conserved peptide. In conclusion, although allowing protein identifications with a single peptide match and low scoring peptides could increase the coverage, the high number of mistaken and multiple identifications make this strategy inadvisable for increasing the number of protein identifications.

The experiment clearly established the value of a species-specific database, even if it was of low quality (Figure 3-1), for proteomics identification technologies. Since many projects will have to rely on sequence databases of related species, we tested several parameters to determine what limits the identification. Theoretically, for any

successful identification under stringent criteria, a protein must yield at the least two fragmented peptides whose spectra match theoretical spectra from the library. Frequently a higher number of unique peptides are identified for more abundant proteins and the likelihood, that a least two completely conserved peptides are present among them, increases. Accordingly, the abundance of spectra and the identification rate were indeed closely correlated (Figure 3-2). The correlation was not absolute though; we found that extremely abundant proteins, such as the TPT, were only identified with the pea database (Table 3-2) whereas about one third (Medicago database) to one fifth (Arabidopsis database) of low abundance proteins could still be identified with a non-species specific database (Figure 3-2). Conserved structural features of proteins may also limit identification. Proteins with comparatively high numbers of membrane spanning helices and underrepresented recognition sites for trypsin are harder to identify than soluble proteins since there are frequently less spectra available for matching to the respective database (Eichacker et al., 2004). When the proteins identified in this study were grouped according to their predicted membrane helix content and the ratio of proteins identified with the non-species specific databases relative to the pea database was plotted, however, no correlation was observed (Figure 3-3). Possibly, the sample size with two databases and a limited number of membrane proteins was too small to reveal a correlation. Since highly hydrophobic proteins with more than 7 transmembrane helices, such as the TPT (Weber et al., 2005), could not be identified with the non-species specific databases, any analysis of a membrane proteome will likely critically depend on the availability of a species-specific sequence database.

Finally, we studied several of the abundant proteins that were identified in all three databases and the proteins that remained unidentified in the Medicago and the Arabidopsis database. As an example for an abundant membrane protein, TPT, which remained unidentified in the non-species specific databases, was compared with the soluble protein NAD-MDH that was identified with all three databases. Although the TPT orthologues are slightly more conserved than the NAD-MDH orthologues, the mass spectra generated from TPT could not be mapped onto the sequences provided by the Medicago and Arabidopsis databases. For Arabidopsis, only a single tryptic peptide is completely conserved between the corresponding pea and Arabidopsis proteins; hence the protein could not be identified with confidence. Although the Medicago sequence shares six tryptic peptides with the pea sequence, only one of these peptides was matched to the empirical spectra library, which is not sufficient for a significant identification with stringent criteria. Two of the conserved peptides are longer than 23 amino acids and two are very short, which might be the reason why spectra corresponding to these peptides were not experimentally detected. NAD-MDH could be identified with all databases since matching spectra were generated from the protein. A high degree of conservation is necessary but not sufficient for a successful identification of a protein with a non-species specific database since a single amino acid exchange will preclude a peptide match.

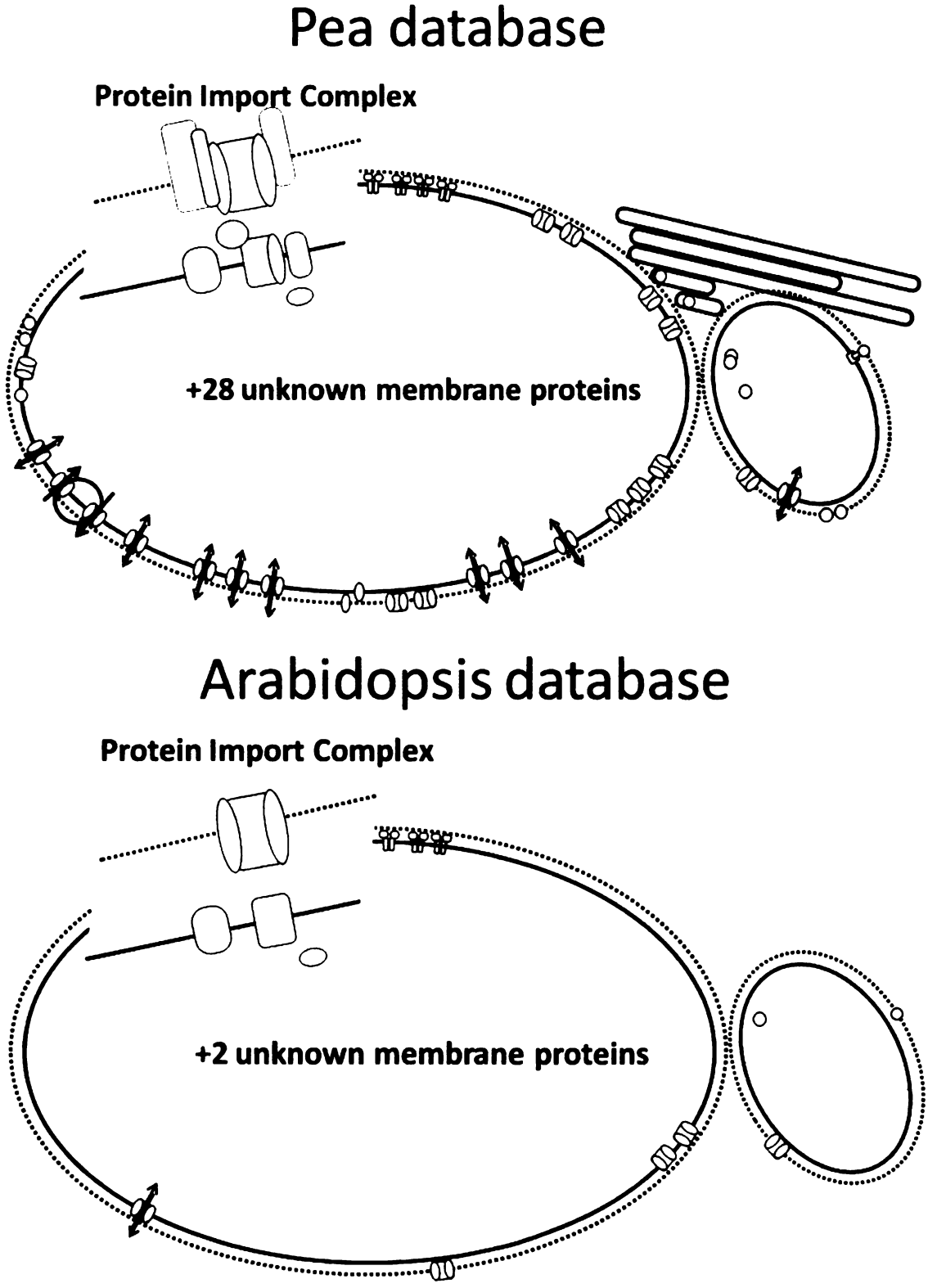


Figure 3-6. Non-species specific databases limit the discovery of new proteins and the interpretation of known proteins in the pea chloroplast envelope proteome sample.

The reduced information caused by using a non-species specific database hampers the interpretation of the data. To visualize this fact, proteins identified with each database were drawn in identical positions onto a schematic representation of a chloroplast. The likely mitochondrial and endomembrane system contaminants were also drawn on the corresponding structures (Figure 3-6). As expected from the reduced number of proteins identified with the Arabidopsis and Medicago databases, the total number of membrane proteins and contaminating proteins is also lower. The Arabidopsis database identified no protein likely to reside in the endomembrane system. Hence one would conclude that there is no contamination from this source, although both the Medicago and the pea database reveal that there is at least one protein present that is believed to reside in the endomembrane system. In the extreme case of the Arabidopsis database used to analyze the pea chloroplast envelope proteome, none of the transport proteins catalyzing the major metabolite fluxes across the envelope (see Weber, 2004; Weber and Fischer, 2007, for recent reviews) could be identified. With the pea database all transporters for phosphorylated sugars that are predicted to reside in the inner envelope of chloroplasts were identified: the triosephosphate, the phosphoenolpyruvate, and the pentosephosphate/phosphate translocators (Flugge et al., 2003; Weber et al., 2004) whereas only the pentose phosphate translocator could be identified with the Medicago database. The two translocator system for importing 2-oxoglutarate and exporting glutamate could also be identified only with the pea database (Renné et al., 2003; Reumann and Weber, 2006; Schneidereit et al., 2006; Weber and Flugge, 2002; Weber et al., 1995). A plastidic ADP/ATP translocator (Möhlmann et al., 1998; Neuhaus et al., 1997; Reiser et al., 2004) could be identified with all three databases as well as several

members of the mitochondrial carrier family that have previously been shown to be targeted to chloroplasts (Bedhomme et al., 2005; Bouvier et al., 2006; Picault et al., 2004). At least one member of the potassium proton exchanger family (Maser et al., 2001) was identified with all databases but a magnesium transporter (Li et al., 2001) and a putative mechanosensitive channel (Haswell and Meyerowitz, 2006) were only seen in the pea database. Two ABC transporters, PAA1 (Shikanai et al., 2003) and HMA1 (Seigneurin-Berny et al., 2006), which import copper and possibly other metal ions into the chloroplast, could be identified with the pea database but not with either of the non-species specific databases. Three other ABC type transporters were sufficiently conserved for identification with all three databases. For the import apparatus, all canonical components that are known to date (Gutensohn et al., 2006) were identified with the pea database except for Tic21 (Teng et al., 2006), which was recently assigned the function of an iron transporter (Duy et al., 2007). With the Medicago database the majority of import complex components could be identified but with the Arabidopsis database only four components were identified. Of the plastid division machinery, PDV2 and MinE were found with the pea database (Glynn et al., 2007; Miyagishima et al., 2006), whereas another component, Arc 6 that has been identified previously (Froehlich et al., 2003), was not identified, but was also only present as a highly fragmented sequence in the pea database. Several likely membrane associated proteins involved in fatty acid and membrane lipid metabolism were identified. For fatty acid synthesis and modification, acetyl co-enzyme A carboxylase, a fatty acid desaturase (McConn et al., 1994) and a long chain fatty acid coenzyme A ligase (Schnurr et al., 2002) were identified as well as proteins involved in lipid metabolism such as 1,2-diacylglycerol 3-

beta-galactosyltransferase for the synthesis of galacto- and UDP-sulfoquinovose:DAG sulfoquinovosyltransferase for synthesis of sulfolipids (SQD2) (Jarvis et al., 2000; Yu et al., 2002). Most proteins were identified with the pea database but SQD2 was only identified with the non-species specific databases. SQD2 was likely not identified because its sequence is fragmented into several short contigs in the pea database (for a complete list of identified proteins and the capabilities of each database please refer to Supplementary Table 3-1). Based on the analysis we conclude that any proteome analysis relying on protein identifications based on a non-species specific database is limited in its conclusions about the presence of proteins and that generating a species specific database even if it is of low quality can massively enhance protein discovery.

Acknowledgements

We thank Shari Tjugum-Holland and Jeff Landgraff of the Michigan State University Research Technology Support Facility for assistance with RNA and DNA analysis and DNA sequencing. This work was supported by a Strategic Partnership Grant (Next Generation Sequencing Center) of the Michigan State University Foundation (to A.P.M.W), NSF-grants IOB-0548610 (to A.P.M.W) and MCB-0519740 (to A.P.M.W), and by an Arabidopsis Functional Genomics Network (WE 2231/4-1) award of the Deutsche Forschungsgemeinschaft (to A.P.M.W).

Supplementary Material

Table S3-1 Redundant list of all proteins identified in the experiment

Table S3-2 Non-redundant list of all proteins identfified in the experiment

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402

Bedhomme, M., Hoffmann, M., McCarthy, E.A., Gambonnet, B., Moran, R.G., Rebeille, F., Ravanel, S. (2005) Folate metabolism in plants - An Arabidopsis homolog of the mammalian mitochondrial folate transporter mediates folate import into chloroplasts. J. Biol. Chem. 280, 34823-34831

Bouvier, F., Linka, N., Isner, J.C., Mutterer, J., Weber, A.P.M., Camara, B. (2006) Arabidopsis SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. Plant Cell 18, 3088-3105

Craig, R., Beavis, R.C. (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466-1467

Douce, R., Joyard, J. (1979) Isolation and Properties of the Envelope of Spinach Chloroplasts. *In* Reid, E, ed, Plant Organelles. Ellis Horwood Publishers, pp 47-59

Duy, D., Wanner, G., Meda, A.R., von Wiren, N., Soll, J., Philippar, K. (2007) PIC1, an Ancient Permease in Arabidopsis Chloroplasts, Mediates Iron Transport. Plant Cell 19, 986-1006

Eichacker, L.A., Granvogl, B., Mirus, O., Muller, B.C., Miess, C., Schleiff, E. (2004) Hiding behind Hydrophobicity: Transmembrane Segments In Mass Spectrometry. J. Biol. Chem. 279, 50915-50922

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., Rolland, N. (2002) Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. Proc. Natl. Acad. Sci. U. S. A. 99, 11487-11492.

Flugge, U.I., Hausler, R.E., Ludewig, F., Fischer, K. (2003) Functional genomics of phosphate antiport systems of plastids. Physiol. Plant. 118, 475-482

Froehlich, J.E., Wilkerson, C.G., Ray, W.K., McAndrew, R.S., Osteryoung, K.W., Gage, D.A., Phinney, B.S. (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J. Proteome Res. 2, 413-425

Glynn, J.M., Miyagishima, S.Y., Yoder, D.W., Osteryoung, K.W., Vitha, S. (2007) Chloroplast Division. Traffic 8, 451-461

Gualtieri, G., Kulikova, O., Limpens, E., Kim, D.J., Cook, D.R., Bisseling, T., Geurts, R. (2002) Microsynteny between pea and Medicago truncatula in the SYM2 region. Plant Mol. Biol. 50, 225-235

Gutensohn, M., Fan, E., Frielingsdorf, S., Hanner, P., Hou, B., Hust, B., Klosgen, R.B. (2006) Toc, Tic, Tat et al.: structure and function of protein transport machineries in chloroplasts. J. Plant Physiol. 163, 333-347

Huang, X.Q., Madan, A. (1999) CAP3: A DNA sequence assembly program. Genome Res. 9, 868-877

Jarvis, P., Dormann, P., Peto, C.A., Lutes, J., Benning, C., Chory, J. (2000) Galactolipid deficiency and abnormal chloroplast development in the Arabidopsis MGD synthase 1 mutant. Proc. Natl. Acad. Sci. U. S. A. 97, 8175-8179

Kalo, P., Seres, A., Taylor, S.A., Jakab, J., Kevei, Z., Kereszt, A., Endre, G., Ellis, T.H.N., Kiss, G.B. (2004) Comparative mapping between Medicago sativa and Pisum sativum. Mol. Genet. Genomics 272, 235-246

Keegstra, K., Yousif, A.E. (1986) Isolation and Characterization of Chloroplast Envelope Membranes. Meth. Enzymol. 118, 316-325

Keller, A., Nesvizhskii, A.I., Kolker, E., Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383-5392

Lavin, M., Herendeen, P.S., Wojciechowski, M.F. (2005) Evolutionary Rates Analysis of Leguminosae Implicates a Rapid Diversification of Lineages during the Tertiary. Syst. Biol. 54, 575 – 594

Liu, H., Sadygov, R.G., Yates, J.R. (2004) A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. Anal. Chem. 76, 4193-4201

Lu, P., Vogel, C., Wang, R., Yao, X., Marcotte, E.M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature Biotechnol. 25, 117-124

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu,

P., Begley, R.F., Rothberg, J.M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380
Maser, P., Thomine, S., Schroeder, J.I., Ward, J.M., Hirschi, K., Sze, H., Talke, I.N., Amtmann, A., Maathuis, F.J.M., Sanders, D., Harper, J.F., Tchieu, J., Gribskov, M., Persans, M.W., Salt, D.E., Kim, S.A., Guerinot, M.L. (2001) Phylogenetic Relationships within Cation Transporter Families of Arabidopsis. Plant Physiol. 126, 1646-1667

McConn, M., Hugly, S., Browse, J., Somerville, C. (1994) A Mutation at the fad8 Locus of Arabidopsis Identifies a Second Chloroplast [omega]-3 Desaturase. Plant Physiol. 106, 1609-1614

Miyagishima, S.Y., Froehlich, J.E., Osteryoung, K.W. (2006) PDV1 and PDV2 mediate recruitment of the dynamin-related protein ARC5 to the plastid division site. Plant Cell 18, 2517-2530

Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646-4658

Picault, N., Hodges, M., Paimieri, L., Palmieri, F. (2004) The growing family of mitochondrial carriers in Arabidopsis. Trends Plant Sci. 9, 138-146

Schmidt, U.G., Endler, A., Schelbert, S., Brunner, A., Schnell, M., Neuhaus, H.E., Marty-Mazars, D., Marty, F., Baginsky, S., Martinoia, E. (2007) Novel tonoplast transporters identified using a proteomic approach with vacuoles isolated from cauliflower buds. Plant Physiol. 145, 216-229

Schnurr, J.A., Shockey, J.M., de Boer, G.-J., Browse, J.A. (2002) Fatty Acid Export from the Chloroplast. Molecular Characterization of a Major Plastidial Acyl-Coenzyme A Synthetase from Arabidopsis. Plant Physiol. 129, 1700-1709

Seigneurin-Berny, D., Gravot, A., Auroy, P., Mazard, C., Kraut, A., Finazzi, G., Grunwald, D., Rappaport, F., Vavasseur, A., Joyard, J., Richaud, P., Rolland, N. (2006) HMA1, a New Cu-ATPase of the Chloro plast Envelope, Is Essential for Growth under Adverse Light Conditions. J. Biol. Chem. 281, 2882-2892

Shevchenko, A., Wilm, M., Vorm, O., Mann, M. (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. Anal. Chem. 68, 850-858

Shikanai, T., Muller-Moule, P., Munekage, Y., Niyogi, K.K., Pilon, M. (2003) PAA1, a P-Type ATPase of Arabidopsis, Functions in Copper Transport in Chloroplasts. Plant Cell 15, 1333-1346

Taylor, N.L., Heazlewood, J.L., Day, D.A., Millar, A.H. (2005) Differential Impact of Environmental Stresses on the Pea Mitochondrial Proteome. Mol. Cell. Proteomics 4, 1122-1133

Teng, Y.S., Su, Y.S., Chen, L.J., Lee, Y.J., Hwang, I., Li, H.M. (2006) Tic21 Is an Essential Translocon Component for Protein Translocation across the Chloroplast Inner Envelope Membrane. Plant Cell 18, 2247-2257

Weber, A.P.M., Schneidereit, J., Voll, L.M. (2004) Using mutants to probe the in vivo function of plastid envelope membrane metabolite transporters. J. Exp. Bot. 55, 1231-1244
Weber, A.P.M., Schwacke, R., Flügge, U.I. (2005) Solute transporters of the plastid envelope membrane. Annu. Rev. Plant Biol. 56, 133-164

Weber, A.P.M., Weber, K.L., Carr, K., Wilkerson, C., Ohlrogge, J.B. (2007) Sampling the Arabidopsis Transcriptome with Massively Parallel Pyrosequencing. Plant Physiol. 144, 32-42
Wikström, N., Savolainen, V., Chase, M.W. (2001) Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond., B, Biol. Sci. 268, 2211-2220

Yang, Y.W., Lai, K.N., Tai, P.Y., Li, W.H. (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J. Mol. Evol. 48, 597-604

Yu, B., Xu, C.C., Benning, C. (2002) Arabidopsis disrupted in SQD2 encoding sulfolipid synthase is impaired in phosphate-limited growth. Proc. Natl. Acad. Sci. U. S. A. 99, 5732-5737
Zhang, Z., Schwartz, S., Wagner, L., Miller, W. (2000) A greedy algorithm for aligning DNA sequences. J. Comp. Biol. 7, 203-214

Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Vagner, L.L., Khaspekov, G.L., Kozhemyako, V.B., Matz, M.V., Meleshkevitch, E., Moroz, L.L., Lukyanov, S.A., Shagin, D.A. (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. 32, e37
Zybailov, B., Coleman, M.K., Florens, L., Washburn, M.P. (2005) Correlation of Relative Abundance Ratios Derived from Peptide Ion Chromatograms and Spectrum Counting for Quantitative Proteomic Analysis Using Stable Isotope Labeling. Anal. Chem. 77, 6218-6224

Chapter 4

Proplastid specific protein patterns revealed by cauliflower curd plastid envelope proteomics

Andrea Bräutigam and Andreas P. M. Weber

This chapter has been submitted to the Journal of Experimental Botany for publication

Abstract

Proplastids are undifferentiated plastids of meristematic tissues. Although undifferentiated, they provide the meristematic cells with plastid synthesized amino acids for protein synthesis, with fatty acids for membrane lipid production and purines and pyrimidines for DNA and RNA synthesis, and. Both the precursor metabolites as well as the products need to cross the plastid envelope. While proplastids are metabolically highly active, they do depend on supply with reductants and energy from the remainder of the cell. They only contain a rudimentary internal membrane system. Little is known about proplastid membrane composition since meristematic tissues are generally limited to the shoot apex and the root tip and therefore do not provide sufficient material for organellar proteome analysis. To overcome this difficulty, cauliflower curd tissue was chosen as a source for isolation of proplastid envelope membranes. The proplastid envelope has a specific membrane protein composition different from those of chloroplasts which is geared to importing precursor metabolites and exporting product metabolites for the rapidly dividing cell. Moreover, the protein import complex has a different composition and the internal membrane system is remarkably different from chloroplast thylakoids. Cauliflower curd tissue is an excellent model system for the study of plastid differentiation since not only white proplastidic cauliflower but also varieties producing chromoplasts as well as green plastids are known.

Introduction

The apical and lateral meristems of plants serve as reservoirs for undifferentiated stem cells that permit indeterminate plant growth. In all meristems only very few cells are actual stem cells while the remainder of the meristem consists of cells which will continue to divide for a finite number of divisions. Meristematic cells are usually surrounded by a thin primary cell wall and they lack a central vacuole and differentiated features. Since meristems do not contain photosynthetically active chloroplasts but undifferentiated proplastids, they are sink tissues which rely on the source tissues for supply with reduced carbon. Although undifferentiated, proplastids support cell growth and division by providing building blocks for the cells, branched chain and aromatic amino acids, fatty acids and lipids as well as nucleotide precursors. Proplastids can differentiate into all plastid subtypes, such as chloroplasts, chromoplasts, and leukoplasts. Since meristematic tissues are usually small, they have not been amenable to analysis by methods requiring large amounts of tissue such as proteomics. In contrast, cauliflower (*Brassica oleracea* ssp. *botrytis*) is an excellent model for proteomics of meristematic tissues. The head of cauliflower, named the 'curd', represents a highly branched abnormally proliferated meristem before floral transition (Kieffer et al., 1998). The branch primordia only grow for a short time until they themselves become apical meristems and produce additional branch primordia (Kieffer et al., 1998). This phenotype has been genetically at least partially tied to floral identity gene mutations (Smith and King, 2000) although most of the phenotypic variability of *B. oleracea* remains unexplained (Labate et al., 2006). Only much later in development do these abnormally proliferating meristems initiate floral development (Kieffer et al., 1998). Before initiation

of floral development, the curd remains an excellent source of meristematic or near meristematic tissues and organelles. The plastids of cauliflower curd tissue (Journet and Douce, 1985) resemble proplastids of other meristematic tissues (summarized in (Kirk and Tilney-Bassett, 1978). These plastids within the meristematic tissue appear to be functional proplastids as they are able to undergo differentiation into chloroplasts as demonstrated by 'greening' of the curds if the cauliflower is not properly covered either by its own leaves or foil (commercial vegetable production guide, Oregon State University, 2004). Cauliflower curd proplastids can also differentiate into chromoplasts (Crisp et al., 1975; Lu et al., 2006).

Like the differentiated plastid types, proplastids are surrounded by two envelope membranes, an outer and an inner envelope. Within these boundaries a protein-rich matrix, the stroma, and a rudimentary internal membrane system are contained (Journet and Douce, 1985). The internal membrane system will differentiate into a prolamellar body in etioplasts or into a thylakoid system in chloroplasts. Although the phloem stream is capable of supplying a number of metabolites to the meristem, the only model for proplastids, plastids of BY-2 cells, are enriched in enzymes for amino acid biosynthesis indicating a metabolically active plastid (Baginsky et al., 2004). Biochemical analysis of cauliflower curd proplastids have revealed that they rely on import of carbon as glucose-6-phosphate and that store carbon as starch (Journet and Douce, 1985; Emes and Neuhaus, 1997; Neuhaus and Emes, 2000). A large supply of carbon is needed not only to prepare the proplastid for differentiation but also for anabolic reactions such as fatty acid or amino acid biosyntheses. Organic nitrogen is also delivered to the meristem by the phloem but it is unknown how proplastids take up organic nitrogen (Emes and

Neuhaus, 1997; Neuhaus and Emes, 2000). Since plastids and proplastids are the site of biosynthesis for branched chain and aromatic amino acids as well as lysine, arginine and histidine, a high rate of nitrogen uptake into the proplastids is necessary to provide the dividing and growing cell with sufficient amounts of amino acids. Since the proplastid is the source for a number of essential metabolites for the dividing and growing meristematic cells, efficient exchange of metabolites across the envelope necessitates the presence of metabolite transport proteins. Whole tobacco BY-2 plastids have been analyzed as a model system for undifferentiated heterotrophic plastids (Baginsky et al., 2004).

In this work the focus is on the envelope proteome of cauliflower curd proplastids to understand the adaptations of the envelope to proplastid-specific metabolism. The transport protein complement indicates that proplastids are active cell factories which imports precursor metabolites and exports products such as amino acids, fatty acids and nucleotides. The proplastids not only contain a distinct set of transport proteins but also a subset of protein import complex components as well as a distinct subset of thylakoid resident proteins very unlike the thylakoid contamination found in chloroplast envelope proteomics. The RNA accumulation pattern of proteins identified in this work is shifted towards shoot apex expressed genes compared to proteins identified in leaf chloroplast envelope proteomics. In summary, the proplastid envelope is specifically adapted to both produce building blocks for the cell as well as differentiate upon external cues.

Material and Methods

Isolation and treatment of membranes

Cauliflower curd plastids were isolated according to (Journet and Douce, 1985). The plastids were lysed and plastid envelopes were prepared as described in (Douce and Joyard, 1979; Keegstra and Yousif, 1986). The envelope membranes were separated into chloroform/methanol (1:1, v/v) soluble and insoluble fractions by diluting a resuspended pellet of envelope membranes with ten volumes of solvent. After incubation on ice for twenty minutes, phases were separated by centrifugation (20,000g; 20 min). The chloroform-soluble fraction was then transferred to a fresh tube whereas the insoluble fraction was recovered as a pellet after removal of the supernatant. Both fractions were dried, washed with hexane to remove lipids, carefully dried again at room temperature and redissolved in SDS-PAGE loading buffer. All fractions were separated by 12.5% SDS-PAGE

Protein identification

Gels were briefly stained with Coomassie Brilliant Blue and then cut into ten equally sized slices. Proteins within each slice were modified and digested with trypsin as described by (Shevchenko et al., 1996). After extraction, the peptides were loaded by a Waters nanoAcquity Sample Manager onto a Waters Symmetry C18 peptide trap (5 μm, 180 μm x 20mm) at a flow rate of 4 μL/min in 2% Acetonitrile/0.1%Formic Acid for 5 minutes. Separation of peptides occurred on a Waters BEH C18 nanoAcquity column (1.7 μm, 100 μm x 100mm) for 90 minutes and the peptides were injected into a ThermoElectron LTQ-FTICR mass spectrometer with a flow rate of 300 nL/min (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 99.9% Acetonitrile/0.1% Formic Acid:

gradient of 5% B to 40% B from 0 to 63 minutes, 40% B to 90% B from 63 to 71 minutes and 5% B from 71 to 90 minutes). The top ten ions of each survey scan (resolution 50,000) were automatically dissociated by low energy collision. The resulting MS/MS spectra were converted to a peak list by the BioWorks Browser version v3.2. Peptides were identified by comparing all mass spectra libraries to a sequence database from *Arabidopsis thaliana*, TAIR 8, (Swarbreck et al., 2008) using the Mascot search algorithm v 2.2 (www.matrixscience.com). Oxidation of methionine was permitted and carbamidomethylcysteine was set as fixed peptide modifications. Two missed tryptic sites were allowed and the MS/MS tolerance was set to 0.8kDa and the peptide tolerance to +/- 10ppm. The results of the Mascot search algorithm were loaded into Scaffold® and analyzed with peptide and protein prophet algorithms (Keller et al., 2002). Parameters were set to 99% confidence for protein identification, 95% confidence for peptide identification and at least two peptides identified for each protein. When multiple matches were reported, the matches were analyzed manually and lower scoring matches were discarded. For proteins discussed in detail, multiple matches to different members of protein families were excluded by protein alignment and for plastid transport proteins as well as import complex components, no identical tryptic peptides were identified in the alignment and the experiment. Results were exported to MS Excel for further analysis.

Protein analysis

Annotation and classification of proteins was assigned based on information in TAIR (Swarbreck et al., 2008), ARAMEMNON (Schwacke et al., 2003), and manual curation of the literature. The predicted location, the number of transmembrane helices and the prediction of betafold structures was retrieved from TAIR. Comparison to data

from other proteome projects was conducted using plprot (Kleffmann et al., 2006). For comparison, data from pea chloroplast envelopes described in (Bräutigam et al., 2008a), which were produced exactly as described for cauliflower proplastid envelopes, was extracted from the public repository PRIDE and reanalyzed according to the same specifications (Supplemental Table 2)

Gene Expression Analysis

For each protein, the RNA expression data was extracted from the AtGenExpress Developmental series (Schmid et al., 2005). The shoot apex expression level was calculated as the arithmetic mean from experiments ATGE_6, ATGE_8, ATGE_29 and ATGE_46 to ATGE_52. The leaf expression level was calculated as the average expression from experiments ATGE_10 to ATGE_21. The ratio of expression was calculated as the log(2) of the quotient expression(apex) divided by expression(shoot) and compared to data from pea chloroplast envelope proteins (Supplemental Table 2).

Results and Discussion

Overall membrane proteome analysis

Table 4-1: Comparison of membrane protein identifications in proplastid and chloroplast envelope; data for chloroplast envelopes extracted from (Bräutigam et al., 2008b) and reanalyzed

	total	membrane proteins	known transport proteins	putative transport proteins
proplastid envelope	174	56	16	7
chloroplast envelope	321	110	16	21

Proteomic analysis of cauliflower curd plastid envelope membranes identified 174 distinct proteins (Table 4-S1). This number is significantly lower in comparison to what has been reported for maize or pea chloroplast envelope membranes (Bräutigam et al., 2008a) (Table 4-1). Using a non-species specific database significantly lowers the identification probability especially for proteins expressed at a low level (Bräutigam et al., 2008b) but the drawback of lower identification probability is superseded by superior tissue availability. The proplastid envelope sample allowed the identification of sixteen known and seven putative membrane proteins, 9.2% and 4% of the total proteins identified, whereas the chloroplast envelope proteome sample allowed only the detection of 5% and 6.5% respectively. The overall proportion one third of proteins with predicted transmembrane helices is similar between both proteome samples. It was concluded that the proplastid envelope proteome sample is suitable for transport protein analysis.

Table 4-2: Transport proteins of the proplastid envelope and their function; question marks denote unknown transport proteins or metabolites

	transporter	imports...	exports...	transport inward	transport outward	pathways
transport proteins mainly acting as importers	GPT	carbon skeletons		G-6-P	phosphate	starch synthesis
	GPT	carbon skeletons		G-6-P	triosephosphate	pentosephosphate pathway reduction equivalent synthesis
	TPT	reduction power		triosephosphate	3-phosphoglycerate	reduction equivalent synthesis
	NTT	energy		ATP	ADP	all anabolic reactions
	NTT	energy		ATP	ADP	all anabolic reactions
	PHT3.1	-		?	phosphate	balances phosphate pool
	PPT	carbon skeletons		phosphoenol-pyruvate	phosphate	shikimate pathway branched chain amino acid synthesis (from pyruvate) fatty acid synthesis (from pyruvate)
	DiT2	organic nitrogen		glutamate	malate	amino acid synthesis
	DiT1	-		2-oxoglutarate	malate	amino acid synthesis
	SAMC	methyldonor		S-adenosyl methionine	S-adenosyl homocystein	methyl requiring reactions
	?	organic sulfur?/ sulfate?		?	?	sulfur requiring compounds
	TGD complex	membrane lipid precursors		-	phosphatidic acid	membrane lipid synthesis
transport proteins mainly acting as exporters	*OEP16-homologue*		amino acids	-	amino acids	protein biosynthesis
	Brittle1		nucleotide precursors	-	adenylates	DNA and RNA synthesis
	?		pyrimidin pathway intermediates	?	pyrimidin derivates	DNA and RNA synthesis
	LACS9		membrane lipid precursors	-	fatty acids	membrane lipid synthesis

Transport proteins

Compared to the membrane transport proteins identified in chloroplast envelope proteome samples (Ferro et al. 2003; Froehlich et al. 2003) different membrane transport proteins are identified in the proplastid envelope proteome sample (Tables 4-S1 and 4-S2). Generally, one can divide the transport proteins in the proplastid envelope proteome sample in those that import precursor metabolites for synthesis as well as those that provide reduction power and energy and those which export products to the cytosol (Table 4-2).

Transport proteins for carbon skeleton import

The phosphate translocators act as importers in proplastids (Table 4-2). Among the phosphate translocators, two isoforms of the glucose-6-phosphate phosphate translocator (GPT), the phosphoenolpyruvate phosphate translocator (PPT) and the triosephosphate phosphate translocator (TPT) were identified in the proteome sample (Table 4-S1). Multiple matches of the same peptide to different family family members were excluded by aligning the proteins and checking for identical tryptic peptides. The phosphate translocator proteins are distinct from each other. The GPTs are not present in sufficient quantities in leaves of either Arabidopsis (Ferro et al., 2003; Froehlich et al., 2003), *Pisum sativum* (Bräutigam et al., 2008a; Bräutigam et al., 2008b) or *Zea mays* (Bräutigam et al., 2008a) to be detected by chloroplast envelope proteomics but at least one isoform can be detected in tobacco BY-2 cell plastids (Baginsky et al., 2004). In heterotrophic plastids, this transporter can supply the plastid with G6P for starch biosynthesis as well as reduction power (Emes and Neuhaus, 1997; Kammerer et al., 1998; Neuhaus and Emes, 2000) and carbon skeletons through the oxidative pentose

phosphate pathway. The GPT exchanges glucose-6-phosphate for either phosphate or triosephosphate in a strict counter-exchange mode (Kammerer et al., 1998). Since two GPT isoforms are identified (Table 4-S1), it is possible that, in vivo, GPT isoforms may have higher affinity for one of the counter substrates. Reduction power is generated in the oxidative branch of the pentosephosphate pathway (OPPP) by oxidizing G6P to gluconolactone which subsequently is converted into triosephosphate by a series of reactions. The intermediates of this pathway serve as carbon skeletons for anabolic reactions or the final product can serve as the counter-exchange substrate for the import of G6P.

Transport proteins for amino acid production

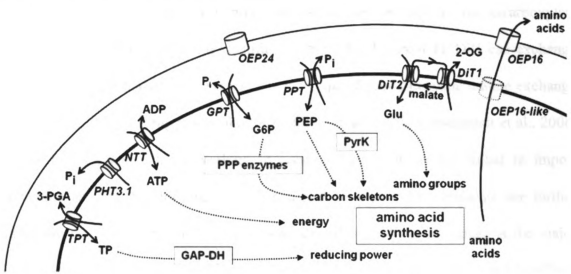


Figure 4-1: The proplastid envelope enables efficient exchange of metabolites needed for amino acid biosynthesis; the transport proteins are printed italic, TPT triosephosphate phosphate translocator, PHT phosphate transporter, NTT ATP ADP exchanger, GPT glucose-6-phosphate phosphate translocator, PPT phosphoenolpyruvate phosphate translocator, DiT dicarboxylate transporter, OEP outer envelope porin; enzymatic reactions are printed in boxes, PyrK pyruvate kinase PPP pentosephosphate pathway, GAP-DH glyceraldehyde dehydrogenase complex;

The phosphoenolpyruvate phosphate translocator (PPT) is also present in the proplastid envelope proteome sample (Table 4-2, Figure 4-1). The proplastid is the sole cellular source of aromatic amino acids for the rapidly growing and dividing cells (Neuhaus and Emes, 2000; Baginsky et al., 2004). Phosphoenolpyruvate transported by PPT is a direct precursor of aromatic amino acids (Knappe et al., 2003; Voll et al., 2003) together with erythrose-4-phosphate produced in the pentosephosphate pathway. PEP is also the source for pyruvate needed for branched chain amino acid biosynthesis as the plastids contain substantial pyruvate kinase activity (Andre and Benning, 2007; Andre et al., 2007). In addition to branched chain and aromatic amino acids the plastids are also the sole source for arginine and lysine as well as histidine and threonine. The amino acids not only require carbon, but also amino groups for their production. The dicarboxylate carriers DiT1 and DiT2 are both identified (Table 4-S1, Figure 4-1). Dit1 can exchange 2-oxoglutarate for malate whereas DiT2 is more specific for glutamate malate exchange (Weber et al., 1995; Taniguchi et al., 2002; Renne et al., 2003; Schneidereit et al., 2006; Weber and Fischer, 2007). Working in concert these proteins are suited to import glutamate as a source for organic nitrogen and to export 2-oxoglutarate for further metabolic reactions, for example in mitochondria (Figure 4-1). Glutamate is the major amino acid transported in the phloem (Weibull and Melin, 1990; Lohaus and Moellers, 2000). The identification of both dicarboxylate carriers solves the question how nitrogen is imported into heterotrophic proplastids (Figure 4-1) (Neuhaus and Emes, 2000).

Transport proteins for energy, reducing power and anabolic precursors

TPT has also been identified with a high spectral count (Table 4-2, Table 4-S1). This protein has been considered the hallmark transport protein of chloroplasts (Flugge and Heldt, 1984; Schneider et al., 2002; Flugge et al., 2003; Weber et al., 2005) and has been assumed to be absent from heterotrophic plastids (Flugge et al., 2003). In chloroplasts, TPT exports the main product of the reductive pentosephosphate pathway, triosephosphate, into the cytosol in exchange for inorganic phosphate. In vitro characteristics indicate that TPT can also exchange 3-PGA for triosephosphate (Loddenkotter et al., 1993). Since cauliflower proplastids lack a functional photosynthetic thylakoid membrane system (Table 4-S1) (Journet and Douce, 1985), they are unlikely to produce large amounts of triosephosphate for export through photosynthetic carbon assimilation. Possibly, the TPT acts as a reduction equivalent shuttle, as has been proposed for TPT in chloroplasts of C4plants (Bräutigam et al., 2008a) and for C3 leaves (Heineke et al., 1991). In this case, a GAP-DH in the chloroplast produces NADPH by oxidizing triosephosphate to 3-PGA which can be exported to the cytosol again (Figure 4-1). The resulting reducing power can be used for anabolic reactions such as amino acid synthesis (Figure 4-1).

Many if not all of the anabolic reactions of starch, amino acid and nucleotide synthesis require energy in form of ATP. Since cauliflower curd proplastids lack chlorophyll (Journet and Douce, 1985) they are unlikely to produce substantial amounts of ATP by photophosphorylation. Two ATP ADP exchangers (NTTs) are detected with high absolute spectral counts among the transport proteins (Table 4-S1, Figure 4-1). These proteins provide the proplastids with energy by importing ATP in counter-

exchange for ADP. The consumption of ATP produces inorganic phosphate which is not transported by NTTs. This phosphate imbalance may be countered by the putative phosphate transport protein PHT3.1 of proplastids (Figure 4-1). This putative phosphate transport protein is also present in BY2 cell plastids as well as chloroplast envelopes (Baginsky et al., 2004; Kleffmann et al., 2004; Bräutigam et al., 2008a).

The proplastid envelope also harbors a S-adenosylmethionine (SAM) carrier which provides the plastid with SAM for a multitude of methylation reactions (Bouvier et al., 2006).The main biochemical effect of loss of SAM in the plastids is decreased leaf chlorophyll content leading to stunted growth (Bouvier et al., 2006) but it likely also effects meristematic tissues directly since it is present on proplastid envelopes.

Transport proteins for export of products

Although the plastids are the sole source of a number of amino acids, the amino acid exporter at the inner envelope has not been characterized at the molecular level. However, the transport pore through the outer envelope is known, the outer envelope porin OEP16 (Pohlmeyer et al., 1997). OEP 16 can transport a broad spectrum of amino acids but excludes sugars and sugar phosphates (Pohlmeyer et al., 1997) and it is one of the proteins with the highest spectral count in the proteome sample (Table 4-S1). Two OEP16 homologues, AGI1 and AGI2 (Murcha et al., 2007), are also identified with a high spectral count (Table 4-S1). At least one of these proteins is localized in the inner envelope (Brautigam et al. submitted) and is considered a candidate protein for amino acid transport at the inner envelope (Figure 4-1).

Apart from amino acids for protein biosynthesis, the cytosol also depends on purins as well as pyrimidins for DNA synthesis since, for both pathways, the majority of

the enzymes is localized in the plastids (Zrenner et al., 2006). Purin derivatives, more specifically, adenylates can be exported through an adenylate uniporter, Brittle1, in dicots (Kirchberger et al., 2008). Brittle1 of Arabidopsis is mainly expressed in the roots with the most intense expression being in the root tip in the region of the meristem. GUS activity was not readily visible in the shoot meristem (Kirchberger et al., 2008) although publicly available expression data (Schmid et al., 2005) visualized in the eFP browser (Winter et al., 2007) indicates strong expression in the shoot apex confirming the proplastid envelope proteome results (Table 4-S1). For pyrimidine biosynthesis, dihydro-orotate has to be exported to be reduced to orotic acid at the mitochondrial membrane; the transport protein is no known (Zrenner et al., 2006). The exporter for the pyrimidine pathway intermediates is also unknown. Possibly, one of the membrane proteins of unknown function (Table 4-1) is either of the unknown transport proteins involved in pyrimidine metabolism. None of the proteins is similar to the characterized purine/pyrimidine transport proteins residing outside of the plastid (Gillissen et al., 2000). In addition to nucleotide precursors for DNA and RNA synthesis and amino acids for protein biosynthesis, fatty acids and membrane lipids are needed to sustain rapid cell proliferation. Fatty acids and membrane lipids are produced in a complex metabolic network localized in multiple compartments (Benning et al., 2006). Within the proteome sample, one of three components of the TGD transporter which imports phosphatidic acid from the ER (Awai et al., 2006) was identified as well as a long acyl chain CoA synthase implicated in fatty acid transport (Schnurr et al., 2002) (Table 4-2).

The protein import complex

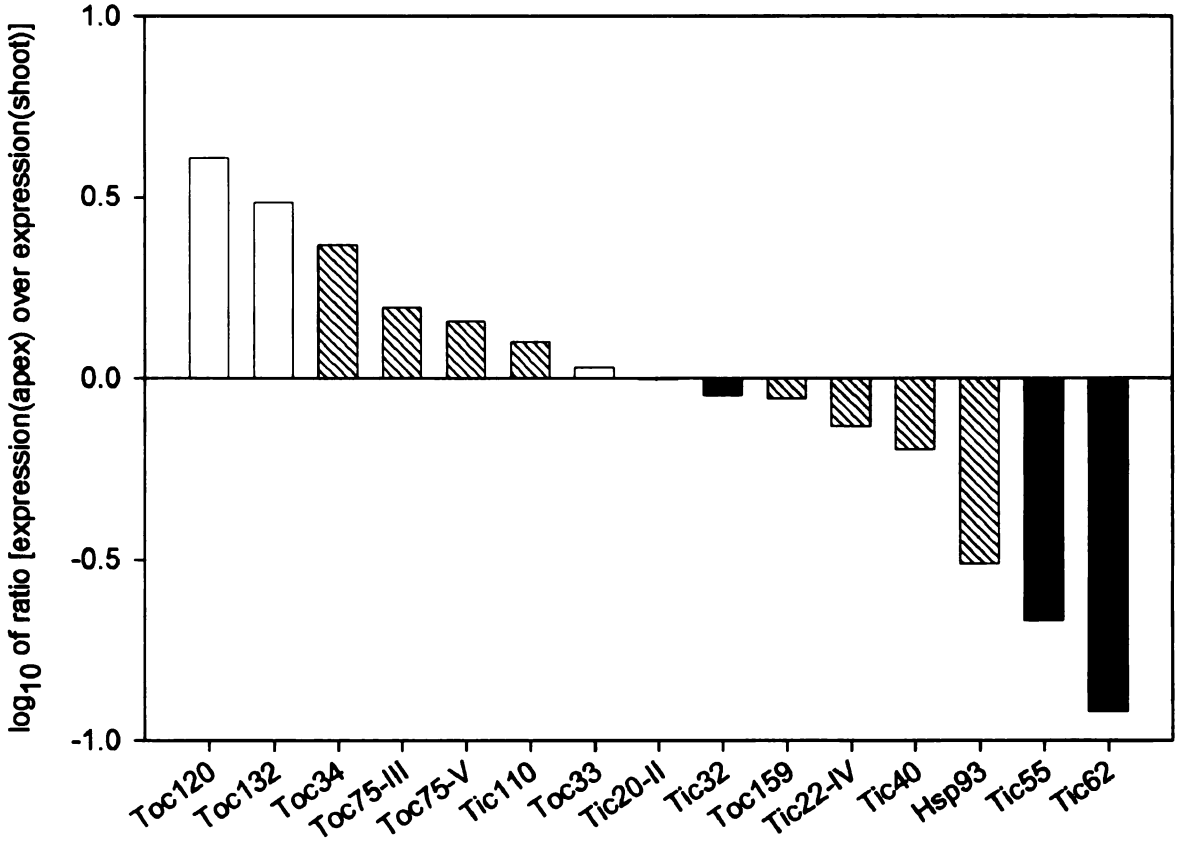


Figure 4-2: Expression of the protein import complex components varies between leaves and the shoot apex; white bars, identified in proplastids only, black bars, identified in chloroplasts only, hatched bars, identified in both samples; for each protein, the gene expression of the corresponding Arabidopsis gene was extracted from the AtGenExpress database; the expression values were calculated as log10 of the ratio between expression in the shoot apex and expression in the leaves, i.e. positive values indicate a higher expression in the shoot apex;

The protein import complex components identified in proplastid envelopes include all three import receptors, Toc120, Toc132 and Toc159, two smaller GTPases ToC33 and ToC34 and the main channel through the outer envelope Toc75-III as well as Toc75-V (Figure 4-2). Of the inner envelope complex, Tic110, Tic20, Tic40 and Tic22 as

well as the import chaperone Hsp93 (also called ClpC) were found (Figure 4-2). In contrast, in chloroplast envelope proteomics, Toc120 and Toc132 have not been identified so far (Ferro et al., 2003; Froehlich et al., 2003; Kleffmann et al., 2004; Kleffmann et al., 2007; Bräutigam et al., 2008a; Bräutigam et al., 2008b). For the inner envelope complex, additional components, Tic32, Tic55 and Tic62, can be identified in chloroplast proteomics (Table 4-S2) e.g.(Bräutigam et al., 2008b). The expression pattern of the protein import complex components was analyzed in detail and compared to the identification of components in chloroplast and proplastid proteomics. Expression values for the genes were extracted from publicly available microarray data in Arabidopsis and the ratio of shoot apex to leaf expression was calculated and normalized by logarithmic transformation. The proteins with the highest shoot apex/leaf expression ratio, the alternative import receptors Toc120 and Toc132, are both identified only in the envelope proteome sample of cauliflower (Figure 4-2). In contrast, the proteins with the lowest ratio, i.e. the proteins expressed more highly in the leaf, are the import complex components Tic55 and Tic62, which both could only be identified in pea leaf envelopes (Figure 4-2, Table 4-S2). Toc33 identified in proplastids only and Tic32, which was only identified in chloroplast envelopes, is of approximately equal expression between both tissues. Toc120 and Toc132 are alternative import receptors to Toc159 (Jackson-Constan and Keegstra, 2001; Ivanova et al., 2004; Kubis et al., 2004). If both are knocked out, plants are severely compromised in their ability to establish growth, contain almost no chlorophyll and have damaged chloroplasts (Kubis et al., 2004). It has been proposed that Toc132 and Toc120 are essential for the of import housekeeping proteins into plastids whereas Toc159 is specialized for the import of proteins involved in photosynthesis

(Ivanova et al., 2004; Kubis et al., 2004; Kessler and Schnell, 2006), although the distinction of imported products is not absolute (Kubis et al., 2004). Since to date neither Toc132 nor Toc120 have been identified by proteomics in either chloroplast or chloroplast envelope studies, it may be concluded that the protein levels of both proteins are rather low in leaves, precluding identification by proteomics, whereas their higher expression in meristematic tissue (Figure 4-2) allows identification. Indeed based on their spectral abundance, Toc159, Toc132 and Toc120 appear to be equally expressed (Table 4-S1). Toc75, Tic110, Tic20, Tic22 and Tic40 are identified in proplastids as well as chloroplast envelopes in line with their housekeeping function (Figure 4-2). The members of the redox sensing complex however, Tic32, Tic62 and Tic55 cannot be identified in proplastid envelope proteomics. This may either be due to a lower expression level for example of Tic55 and Tic62 (Figure 4-2) or to a lack of membrane association of any of the components. Possibly, the undifferentiated proplastid without a completely developed thylakoid system is not yet poised to sense and relay the redox status mostly generated by photosynthesis.

Overall proteome analysis

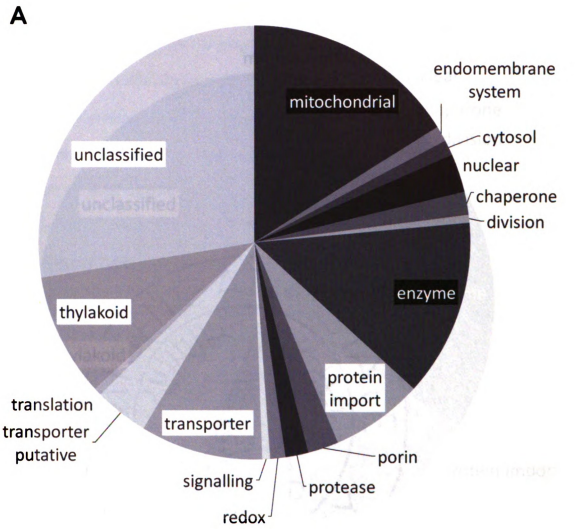
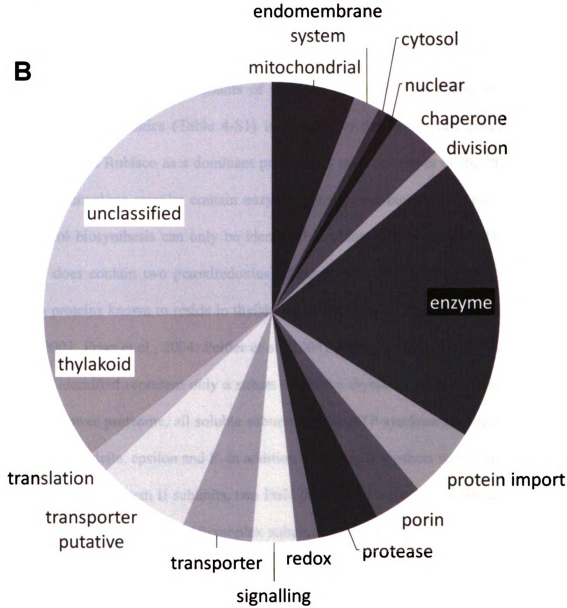


Figure 4-3: Overall proteome contents in plastid envelopes; all proteins were annotated and classified (Tables4- S1 and 4-S2), (A) proplastid envelope proteome, (B) chloroplast envelope proteome

Figure 4-3 (cont'd)



After establishing patterns of the transport proteins for metabolites and protein import complex in the proplastid envelope, we compared the remaining proteins identified in proplastid envelopes (Table 4-S1) with those identified in chloroplast envelopes (Table 4-S2).

The annotation and classification of proteins identified revealed approximately similar proportions of enzymes in both proplastid and chloroplast envelopes (Figure 4-3A and 4-3B). However, the types of enzymes are different. The proplastid envelope sample does not contain sufficient amounts of enzymes of the PPP including Rubisco to be identified by proteomics (Table 4-S1) in contrast to the chloroplast envelope sample which contains Rubisco as a dominant protein and seven enzymes of the PPP (Table 4-S2). Both envelope samples contain enzymes for pigment biosynthesis but enzymes for tocopherol biosynthesis can only be identified in chloroplast envelopes. The proplastid sample does contain two peroxiredoxins (Dietz et al., 2002; Horling et al., 2003) and sixteen proteins known to reside in thylakoids in chloroplasts (Peltier et al., 2000; Peltier et al., 2002; Friso et al., 2004; Peltier et al., 2004) (Figure 4-3A). The thylakoid resident proteins identified represent only a subset of known thylakoid proteins (Table 4-S3). In the cauliflower proteome, all soluble subunits of the ATP synthase are identified, alpha, beta, gamma, delta, epsilon and F. In addition only very few others were found, PsbP and PsbS, two photosystem II subunits, two FtsH proteases, FtsH1 and FtsH8, a CAAX type protease and a cytochrome b6f complex subunit, cyt6 (Table 4-S3). In contrast, in other envelope projects the spectrum of thylakoid proteins identified with the envelope is much broader and also includes many subunits of the photosystems and light harvesting complex proteins (Table 4-S2) (see for example (Bräutigam et al., 2008a)). Proteins of the thylakoid membrane system always appear in envelope proteome analysis (Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008a; Bräutigam et al., 2008b). The only proteins abundant enough to be identified by proteomics in cauliflower proplastid envelopes are components of the ATPase, two PS II components and no PS I

components. One of the two PSII components that can be detected are involved in redox protection. PsbS plays a key role in non-photochemical quenching (Li et al., 2000) whereas PsbP is involved in PS II assembly and stability (Ishihara et al., 2007). Both proteins accumulate in etioplasts compared to chloroplasts (Kanervo et al., 2008). The absence of enzymes of the PPP and the limitation of thylakoid resident proteins confirm the absence of carbon fixation in proplastids already indicated by the absence of a developed thylakoid system and the absence of chlorophyll (Journet and Douce, 1985). On the other hand, the proplastids appear poised to differentiate; they are equipped to both dissipate the proton gradient and handle the redox stress generated during photosynthetic electron transfer.

The annotation and classification of all identified proteins (Figure 4-3A) also revealed that about 20% of the proteins can be assigned to a location other than the plastid, such as the mitochondria, the endomembrane system, the cytosol or the nucleus. Hence, the proportion of putative contaminants in cauliflower curd envelope membranes is approximately two-fold higher as compared to pea envelope proteome samples (Figure 4-3B). It has been noted earlier that cauliflower curd plastids carry a substantial mitochondrial contamination, as determined by marker enzyme assays (Journet and Douce, 1985), which is confirmed by our proteomics study. Putative contaminants identified in cauliflower bud plastids differ from those that are routinely found in leaf chloroplasts preparations, such as glycine decarboxylase complex subunits (Bräutigam et al., 2008a), which is absent from the proplastid preparation. In addition to the proteins typically encountered in plastid isolations, such as histones and abundant proteins of the mitochondrial respiratory chain and ATPases, the cauliflower proteome sample contained

unusual contaminants, e. g. a prohibitin and a Mcm4-like protein (Table 4-S1) not identified in other plastid envelope proteome. Both of these are highly expressed in the shoot apex (Table 4-S1), which may explain their presence in this sample. The total expression analysis of the proteins identified in proplastid and chloroplast envelopes also showed a shift towards shoot apex expressed genes (Figure 4-S1).

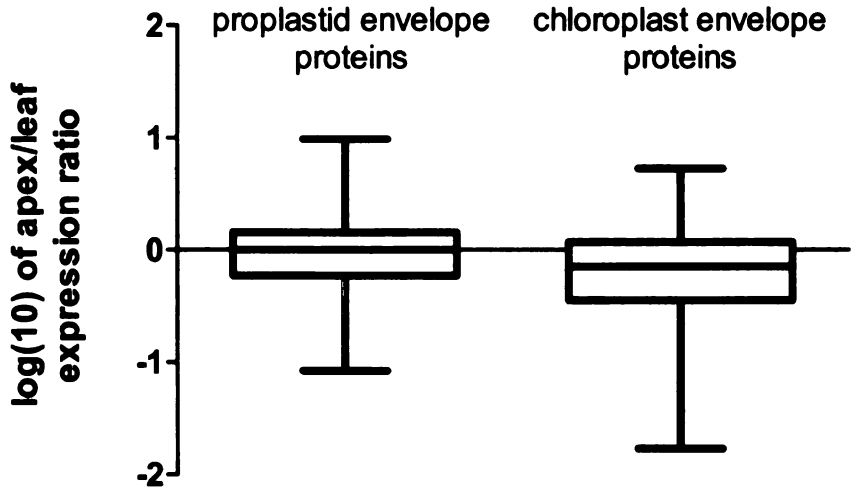


Figure 4-S1: RNA accumulation patterns between shoot apex and leaves for orthologous genes in Arabidopsis visualized by a whisker plot; for each orthologue of both proplastid and chloroplast identified proteins, the $\log_{10}$ of the expression ratio between shoot apex and leaf was calculated, i.e. higher values indicate a shift of expression towards the shoot apex

Conclusion

Proteomics of the proplastid membrane system identified proplastid-specific patterns: The transport proteins of the envelopes are specifically adapted to supply the proliferating cells of the meristems with the necessary building blocks and the proteins import complex is geared to import proteins involved in metabolism rather than photosynthesis. Overall proteome analysis indicates proplastids are not only equipped to support cellular growth but also adapted to differentiate upon light exposure. Cauliflower

curd will be well suited to study plastid differentiation. Untreated cauliflower curd contains proplastids, differentiation into chloroplasts can be induced by exposure to light and differentiation into chromoplasts can be studied in the orange variety of cauliflower allowing the study of three different plastid types in the same system.

Supplementary material

Supplementary table 4-1: Proteins identified in proplastid envelopes from cauliflower curd tissue

Supplementary table 4-2: Proteins identified in chloroplast envelopes from pea leaves (Bräutigam et al., 2008a) were extracted from PRIDE and reanalyzed

Supplementary table 4-3: Thylakoid proteins identified in envelopes of proplastids and chloroplasts

Acknowledgements

We would like to acknowledge expert help from the Michigan State University Research Technology facility.

References

Andre, C and Benning, C. 2007. Arabidopsis seedlings deficient in a plastidic pyruvate kinase are unable to utilize seed storage compounds for germination and establishment. Plant Physiology 145, 1670-1680.

Andre, C, Froehlich, JE, Moll, MR and Benning, C. 2007. A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in Arabidopsis. Plant Cell 19, 2006-2022.

Awai, K, Xu, CC, Tamot, B and Benning, C. 2006. A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking. Proceedings of the National Academy of Sciences of the United States of America 103, 10817-10822.

Baginsky, S, Siddique, A and Gruissem, W. 2004. Proteome analysis of tobacco bright yellow-2 (BY-2) cell culture plastids as a model for undifferentiated heterotrophic plastids. Journal of Proteome Research 3, 1128-1137.

Benning, C, Xu, CC and Awai, K. 2006. Non-vesicular and vesicular lipid trafficking involving plastids. Current Opinion in Plant Biology 9, 241-247.

Bouvier, F, Linka, N, Isner, JC, Mutterer, J, Weber, APM and Camara, B. 2006. Arabidopsis SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. Plant Cell 18, 3088-3105.

Bräutigam, A, Hofmann-Benning, S and Weber, APM. 2008a. Comparative Proteomics of Chloroplast Envelopes from C3 and C4 Plants Reveals Specific Adaptations of the Plastid Envelope to C4 Photosynthesis and Candidate Proteins Required for Maintaining C4 Metabolite Fluxes. Plant Physiology 148, 568-579.

Bräutigam, A, Shrestha, RP, Whitten, D, Wilkerson, CG, Carr, KM, Froehlich, JE and Weber, APM. 2008b. Comparison of the use of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes. Journal of Biotechnology 136, 44-53.

Crisp, P, Walkey, DGA, Bellman, E and Roberts, E. 1975. Mutation Affecting Curd Color In Cauliflower (Brassica-Oleracea L Var Botrytis Dc). Euphytica 24, 173-176.

Dietz, KJ, Horling, F, Konig, J and Baier, M. 2002. The function of the chloroplast 2-cysteine peroxiredoxin in peroxide detoxification and its regulation. Journal of Experimental Botany 53, 1321-1329.

Douce, R and Joyard, J. 1979. Isolation and Properties of the Envelope of Spinach Chloroplasts: Ellis Horwood.

Emes, MJ and Neuhaus, HE. 1997. Metabolism and transport in non-photosynthetic plastids. Journal of Experimental Botany 48, 1995-2005.

Ferro, M, Salvi, D, Brugiere, S, Miras, S, Kowalski, S, Louwagie, M, Garin, J, Joyard, J and Rolland, N. 2003. Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. Molecular & Cellular Proteomics 2, 325-345.

Flugge, UI, Hausler, RE, Ludewig, F and Fischer, K. 2003. Functional genomics of phosphate antiport systems of plastids. Physiologia Plantarum 118, 475-482.

Flugge, UI and Heldt, HW. 1984. The Phosphate-Triose Phosphate-Phosphoglycerate Translocator of the Chloroplast. Trends in Biochemical Sciences 9, 530-533.

Friso, G, Giacomelli, L, Ytterberg, AJ, Peltier, JB, Rudella, A, Sun, Q and van Wijk, KJ. 2004. In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana chloroplasts: New proteins, new functions, and a plastid proteome database. Plant Cell 16, 478-499.

Froehlich, JE, Wilkerson, CG, Ray, WK, McAndrew, RS, Osteryoung, KW, Gage, DA and Phinney, BS. 2003. Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. Journal of Proteome Research 2, 413-25.

Gillissen, B, Burkle, L, Andre, B, Kuhn, C, Rentsch, D, Brandl, B and Frommer, WB. 2000. A new family of high-affinity transporters for adenine, cytosine, and purine derivatives in arabidopsis. Plant Cell 12, 291-300.

Heineke, D, Riens, B, Grosse, H, Hoferichter, P, Peter, U, Flugge, UI and Heldt, HW. 1991. Redox Transfer Across The Inner Chloroplast Envelope Membrane. Plant Physiology 95, 1131-1137.

Horling, F, Lamkemeyer, P, Konig, J, Finkemeier, I, Kandlbinder, A, Baier, M and Dietz, KJ. 2003. Divergent light-, ascorbate-, and oxidative stress-dependent regulation of expression of the peroxiredoxin gene family in Arabidopsis. Plant Physiology 131, 317-325.

Ishihara, S, Takabayashi, A, Ido, K, Endo, T, Ifuku, K and Sato, F. 2007. Distinct functions for the two PsbP-like proteins PPL1 and PPL2 in the chloroplast thylakoid lumen of arabidopsis. Plant Physiology 145, 668-679.

Ivanova, Y, Smith, MD, Chen, KH and Schnell, DJ. 2004. Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. Molecular Biology of the Cell 15, 3379-3392.

Jackson-Constan, D and Keegstra, K. 2001. Arabidopsis genes encoding components of the chloroplastic protein import apparatus. Plant Physiology 125, 1567-1576.

Journet, EP and Douce, R. 1985. Enzymic Capacities Of Purified Cauliflower Bud Plastids For Lipid-Synthesis And Carbohydrate-Metabolism. Plant Physiology 79, 458-467.

Kammerer, B, Fischer, K, Hilpert, B, Schubert, S, Gutensohn, M, Weber, A and Flugge, UI. 1998. Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. Plant Cell 10, 105-17.

Kanervo, E, Singh, M, Suorsa, M, Paakkarinen, V, Aro, E, Battchikova, N and Aro, EM. 2008. Expression of protein complexes and individual proteins upon transition of etioplasts to chloroplasts in pea (Pisum sativum). Plant And Cell Physiology 49, 396-410.

Keegstra, K and Yousif, AE. 1986. Isolation And Characterization Of Chloroplast Envelope Membranes. Methods in Enzymology 118, 316-325.

Keller, A, Nesvizhskii, AI, Kolker, E and Aebersold, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical Chemistry 74, 5383-5392.

Kessler, F and Schnell, DJ. 2006. The function and diversity of plastid protein import pathways: A multilane GTPase highway into plastids. Traffic 7, 248-257.

Kieffer, M, Fuller, MP and Jellings, AJ. 1998. Explaining curd and spear geometry in broccoli, cauliflower and 'romanesco': quantitative variation in activity of primary meristems. Planta 206, 34-43.

Kirchberger, S, Tjaden, JH and Neuhaus, E. 2008. Characterization of the Arabidopsis Brittle1 transport protein and impact of reduced activity on plant metabolism. The Plant Journal 56, 51-63.

Kirk, JTO and Tilney-Bassett, RAE. 1978. The plastids: Their chemistry, structure, growth and inheritance. Amsterdam/Oxford: Elsevier.

Kleffmann, T, Hirsch-Hoffmann, M, Gruissem, W and Baginsky, S. 2006. plprot: A comprehensive proteome database for different plastid types. Plant And Cell Physiology 47, 432-436.

Kleffmann, T, Russenberger, D, von Zychlinski, A, Christopher, W, Sjolander, K, Gruissem, W and Baginsky, S. 2004. The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Current Biology 14, 354-362.

Kleffmann, T, von Zychlinski, A, Russenberger, D, Hirsch-Hoffmann, M, Gehrig, P, Gruissem, W and Baginsky, S. 2007. Proteome dynamics during plastid differentiation in rice. Plant Physiology 143, 912-923.

Knappe, S, Lottgert, T, Schneider, A, Voll, L, Flugge, UI and Fischer, K. 2003. Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis-AtPPT1 may be involved in the provision of signals for correct mesophyll development. Plant Journal 36, 411-420.

Kubis, S, Patel, R, Combe, J, Bedard, J, Kovacheva, S, Lilley, K, Biehl, A, Leister, D, Rios, G, Koncz, C and Jarvis, P. 2004. Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. Plant Cell 16, 2059-2077.

Labate, JA, Robertson, LD, Baldo, AM and Bjorkman, T. 2006. Inflorescence identity gene alleles are poor predictors of inflorescence type in broccoli and cauliflower. Journal Of The American Society For Horticultural Science 131, 667-673.

Li, XP, Bjorkman, O, Shih, C, Grossman, AR, Rosenquist, M, Jansson, S and Niyogi, KK. 2000. A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature 403, 391-395.

Loddenkotter, B, Kammerer, B, Fischer, K and Flugge, UI. 1993. Expression of the Functional Mature Chloroplast Triose Phosphate Translocator in Yeast Internal Membranes and Purification of the Histidine-Tagged Protein by a Single Metal-Affinity Chromatography Step. Proceedings of the National Academy of Sciences of the United States of America 90, 2155-2159.
Lohaus, G and Moellers, C. 2000. Phloem transport of amino acids in two Brassica napus L. genotypes and one B-carinata genotype in relation to their seed protein content. Planta 211, 833-840.

Lu, S, Van Eck, J, Zhou, X, Lopez, AB, O'Halloran, DM, Cosman, KM, Conlin, BJ, Paolillo, DJ, Garvin, DF, Vrebalov, J, Kochian, LV, Kupper, H, Earle, ED, Cao, J and Li, L. 2006. The cauliflower or gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. Plant Cell 18, 3594-3605.

Murcha, MW, Elhafez, D, Lister, R, Tonti-Filippini, J, Baumgartner, M, Philippar, K, Carrie, C, Mokranjac, D, Soll, J and Whelan, J. 2007. Characterization of the Preprotein and Amino Acid Transporter Gene Family in Arabidopsis. Plant Physiology 143, 199-212.

Neuhaus, HE and Emes, MJ. 2000. Nonphotosynthetic metabolism in plastids. Annual review of plant physiology and plant molecular biology 51, 111-140.

Peltier, JB, Emanuelsson, O, Kalume, DE, Ytterberg, J, Friso, G, Rudella, A, Liberles, DA, Soderberg, L, Roepstorff, P, von Heijne, G and van Wijk, KJ. 2002. Central

functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. Plant Cell 14, 211-36.

Peltier, JB, Friso, G, Kalume, DE, Roepstorff, P, Nilsson, F, Adamska, I and van Wijk, KJ. 2000. Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12, 319-41.

Peltier, JB, Ytterberg, AJ, Sun, Q and van Wijk, KJ. 2004. New functions of the thylakoid membrane proteome of Arabidopsis thaliana revealed by a simple, fast, and versatile fractionation strategy. Journal of Biological Chemistry 279, 49367-49383.

Pohlmeyer, K, Soll, J, Steinkamp, T, Hinnah, S and Wagner, R. 1997. Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane. Proceedings of the National Academy of Sciences of the United States of America 94, 9504-9509.

Renne, P, Dressen, U, Hebbeker, U, Hille, D, Flugge, UI, Westhoff, P and Weber, APM. 2003. The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. Plant Journal 35, 316-331.

Schmid, M, Davison, TS, Henz, SR, Pape, UJ, Demar, M, Vingron, M, Scholkopf, B, Weigel, D and Lohmann, JU. 2005. A gene expression map of Arabidopsis thaliana development. Nature Genetics 37, 501-506.

Schneider, A, Hausler, RE, Kolukisaoglu, U, Kunze, R, van der Graaff, E, Schwacke, R, Catoni, E, Desimone, M and Flugge, UI. 2002. An Arabidopsis thaliana knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. Plant Journal 32, 685-699.

Schneidereit, J, Hausler, RE, Fiene, G, Kaiser, WM and Weber, APM. 2006. Antisense repression reveals a crucial role of the plastidic 2-oxoglutarate/malate translocator DiT1 at the interface between carbon and nitrogen metabolism. Plant Journal 45, 206-224.

Schnurr, JA, Shockey, JM, de Boer, G-J and Browse, JA. 2002. Fatty Acid Export from the Chloroplast. Molecular Characterization of a Major Plastidial Acyl-Coenzyme A Synthetase from Arabidopsis. Plant Physiology 129, 1700-1709.

Schwacke, R, Schneider, A, van der Graaff, E, Fischer, K, Catoni, E, Desimone, M, Frommer, WB, Flugge, UI and Kunze, R. 2003. ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. Plant Physiology 131, 16-26.

Shevchenko, A, Wilm, M, Vorm, O and Mann, M. 1996. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. Analytical Chemistry 68, 850-858.

Smith, LB and King, GJ. 2000. The distribution of BoCAL-a alleles in Brassica oleracea is consistent with a genetic model for curd development and domestication of the cauliflower. Molecular Breeding 6, 603-613.

Swarbreck, D, Wilks, C, Lamesch, P, Berardini, TZ, Garcia-Hernandez, M, Foerster, H, Li, D, Meyer, T, Muller, R, Ploetz, L, Radenbaugh, A, Singh, S, Swing, V, Tissier, C, Zhang, P and Huala, E. 2008. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Research 36, D1009-D1014.

Taniguchi, M, Taniguchi, Y, Kawasaki, M, Takeda, S, Kato, T, Sato, S, Tahata, S, Miyake, H and Sugiyama, T. 2002. Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in Arabidopsis thaliana. Plant And Cell Physiology 43, 706-717.

Voll, LM, Häusler, RE, Hecker, R, Weber, APM, Weissenböck, G, Fiene, G, Waffenschmidt, S and Flügge, UI. 2003. The phenotype of the Arabidopsis cue1 mutant is not simply caused by a general restriction of the shikimate pathway. Plant Journal 36, 301ff.

Weber, A, Menzlaff, E, Arbinger, B, Gutensohn, M, Eckerskorn, C and Flugge, UI. 1995. The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. Biochemistry 34, 2621-7.

Weber, APM and Fischer, K. 2007. Making the connections - The crucial role of metabolite transporters at the interface between chloroplast and cytosol. FEBS Letters 581, 2215-2222.
Weber, APM, Schwacke, R and Flugge, UI. 2005. Solute transporters of the plastid envelope membrane. Annual Review of Plant Biology 56, 133-164.

Weibull, J and Melin, G. 1990. Free Amino-Acid Content Of Phloem Sap From Brassica Plants In Relation To Performance Of Lipaphis-Erysimi (Hemiptera, Aphididae). Annals Of Applied Biology 116, 417-423.

Winter, D, Vinegar, B, Nahal, H, Ammar, R, Wilson, GV and Provart, NJ. 2007. An Electronic Fluorescent Pictograph Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE 2, e718.

Zrenner, R, Stitt, M, Sonnewald, U and Boldt, R. 2006. Pyrimidine and purine biosynthesis and degradation in plants. Annual Review of Plant Biology 57, 805-836.

Chapter 5

Comparative proteomics of chloroplasts envelopes from C3 and C4 plants reveals specific adaptations of the plastid envelope to C4 photosynthesis and candidate proteins required for maintaining C4 metabolite fluxes

This chapter has been previously published in Plant Physiology: Bräutigam A, Hofmann-Benning S and Weber APM (2008) Comparative Proteomics of Chloroplast Envelopes from C3 and C4 Plants Reveals Specific Adaptations of the Plastid Envelope to C4 Photosynthesis and Candidate Proteins Required for Maintaining C4 Metabolite Fluxes Plant Physiol 148, 568-579

I would like to acknowledge Dr. Susanne Hoffmann-Benning for assistance in proteomics.

Abstract

C4 plants have up to ten-fold higher apparent CO2 assimilation rates than the most productive C3 plants. This requires higher fluxes of metabolic intermediates across the chloroplast envelope membranes of C4 plants in comparison to that of C3 plants. In particular, the fluxes of metabolites involved in the biochemical inorganic carbon pump of C4 plants, such as malate, pyruvate, oxaloacetate, and phosphoenolpyruvate must be considerably higher in C4 plants because they exceed the apparent rate of photosynthetic CO2 assimilation whereas they represent relatively minor fluxes in C3 plants. While the enzymatic steps involved in the C4 biochemical inorganic carbon pump have been studied in much detail, little is known about the metabolite transporters in the envelope membranes of C4 chloroplasts. In this study, we have used comparative proteomics of chloroplast envelope membranes from the C3 plant *Pisum sativum* and mesophyll cell chloroplast envelopes from the C4 plant *Zea mays* to analyze the adaptation of the mesophyll cell chloroplast envelope proteome to the requirements of C4 photosynthesis. We show that C3 and C4-type chloroplasts have qualitatively similar but quantitatively very different chloroplast envelope membrane proteomes. In particular, translocators involved in the transport of triose phosphate and phospho*enol*pyruvate as well as two outer envelope porins are much more abundant in C4 plants. Several putative transport proteins have been identified that are highly abundant in C4 plants, but relatively minor in C3 envelopes. These represent prime candidates for the transport of C4 photosynthetic intermediates, such as pyruvate, oxaloacetate, and malate.

Introduction

C4 photosynthesis allows fast biomass accumulation with high nitrogen and water use efficiency (Leegood and Edwards, 1996; Sage, 2004) and is a desired trait to increase productivity of crop plants (Matsuoka et al., 1998). To facilitate C4 photosynthesis, in maize, a C4 plant of the NADP-malic enzyme type, the primary fixation and the reduction of carbon are spatially separated between two different cell types. Primary carbon fixation occurs in the mesophyll cells (Hatch, 1987). The mesophyll surrounds the bundle sheath cells (BSC) where CO2 is enriched around Rubisco and the reduction of carbon takes place. The chloroplasts of mesophyll and bundle sheath tissues are adapted to their respective roles (Slack et al., 1969; Edwards et al., 2001; Majeran et al., 2005). In addition to carbon fixation and reduction several other pathways, such as nitrogen reduction and assimilation, are partitioned between mesophyll and bundle sheath chloroplasts (Renné, 2003; Majeran et al., 2005) and the adaptation of the soluble chloroplast proteome to C4 photosynthesis has been studied in considerable detail (Majeran et al., 2005). In maize, initial carbon assimilation in the mesophyll cell cytoplasm is accomplished by PEP carboxylase (PEPC), yielding oxaloacetate (OAA). OAA is then imported into the chloroplasts where it is reduced to malate, and subsequently exported to the cytosol again. After diffusion into BSC, malate is decarboxylated in the chloroplasts, yielding CO2, NADPH, and pyruvate. While CO2 and NADPH enter the Calvin cycle in BSC, pyruvate is returned to the mesophyll, where it is imported into the chloroplasts and converted to PEP by phosphoenolpyruvate phosphate dikinase (PPDK), thus regenerating the primary CO2 acceptor, which is exported to the cytosol to enter a new round of CO2 assimilation (Figure 5-1). In maize,

carbon fixation is optimized beyond simply concentrating CO2 in the vicinity of Rubisco. The bundle sheath chloroplasts have limited photosystem II activity (Meierhoff and Westhoff, 1993) and produce less O2, which further reduces the oxygenation reaction of Rubisco. However, absence of PSII activity prevents operation of linear electron transport, limiting the production of reduction equivalents in the bundle sheath. Since CO2 assimilation in the Calvin cycle requires NADPH, this necessitates shuttling of reduction equivalents between mesophyll and bundle sheath by a 3-phosphoglycerate (3-PGA)/triosephosphate (TrioseP) shuttle (Figure 5-1). Despite the detailed knowledge about the soluble proteins involved in and necessary for C4 photosynthesis, the adaptation of integral and associated membrane proteins remains largely unknown. In this work, we focus on analyzing the quantitative and qualitative differences between chloroplast envelope membranes of C3 and C4 plants.

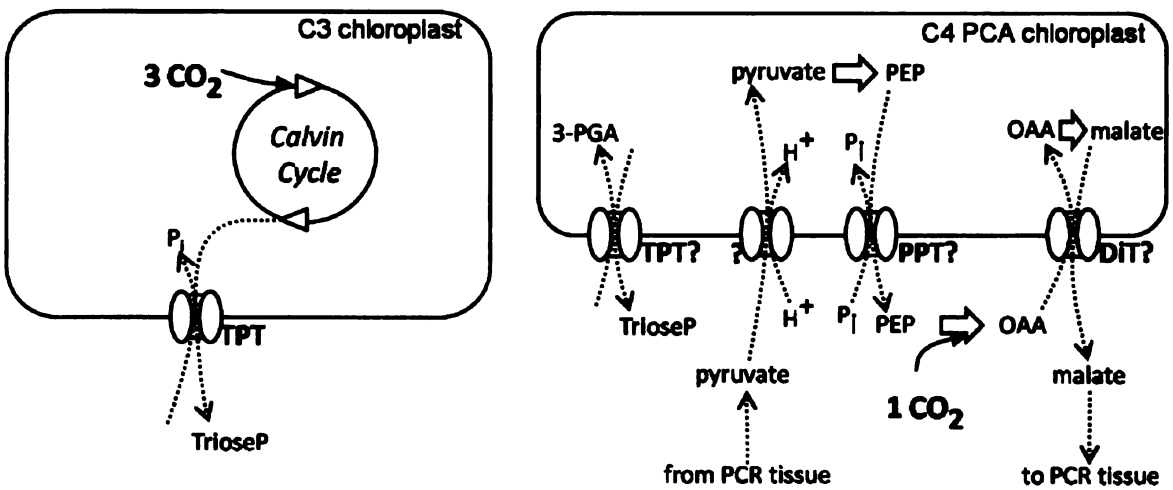


Figure 5-1: Schematic representation of central carbon metabolism and associated transport processes in C3 chloroplasts and C4 PCA type chloroplasts: in C3 chloroplasts, for three carbons fixed, at most one transport process is required; in C4 PCA type chloroplasts, for three carbons fixed, at least twelve transport processes are required, for abbreviations see text

The plastids of green plants are separated from the cytosol by two membranes. Metabolite transport across the outer envelope is controlled by substrate-specific pore-forming proteins (Pohlmeyer et al., 1997; Pohlmeyer et al., 1998; Bolter et al., 1999; Goetze et al., 2006). Solute transport across the inner envelope membrane is catalyzed by a large range of specific metabolite transporters (Weber, 2004; Weber et al., 2005; Weber and Fischer, 2007), some of which are capable of transporting metabolites against a concentration gradient. The spatial separation of initial carbon fixation and subsequent reduction in C4 plants requires a very high metabolite flow across the chloroplast envelope of both mesophyll and bundle sheath chloroplasts that exceeds the apparent rate of carbon assimilation (Laisk and Edwards, 2000). Pea fixes about 17 micromole carbon per square meter leaf area and second (Grodzinski et al., 1998) and requires at most one transport process for three carbons fixed. Maize fixes about 27 micromoles carbon per square meter and second (Grodzinski et al., 1998) and requires at least four transport processes (Figure 5-1) for each carbon fixed. Consequently, the total metabolite transport rate across the chloroplast envelopes in C4 plants exceeds the one in C3 plants by a factor of at least eighteen. High velocity transport of all four metabolites involved in core C4 photosynthesis across the mesophyll chloroplast envelope has been demonstrated using isolated chloroplasts (Huber and Edwards, 1977; Hatch et al., 1984; Flügge et al., 1985; Aoki et al., 1992). While most transport proteins involved in core C4 photosynthesis in mesophyll chloroplasts have not yet been unequivocally identified at the molecular level, good candidates exist for PEP export, triose phosphate shuttling, and for oxaloacetate/malate transport. The molecular nature of the pyruvate transporter, however, is unknown. Likewise, it is unknown whether pyruvate transport across the mesophyll

and the bundle sheath chloroplast envelope is mediated by the same transport protein or different transport proteins. Adaptations of additional membrane proteins as a consequence of the spatial separation of photosynthesis, similar to what has been demonstrated for soluble proteins, are unknown. Since increasing the capacity for metabolite transport across the chloroplast envelope membrane is likely a key adaptation to C4 photosynthesis (Edwards et al., 2001), engineering efforts for introducing C4 photosynthesis in a C3 crop plant will likely critically depend on engineering not only the C4 pathway but also metabolite flux.

In this work, the protein complements of envelopes membranes of C3 chloroplasts and C4 mesophyll chloroplasts are analyzed qualitatively and semi-quantitatively. We hypothesized that analyzing chloroplasts with different modes of photosynthesis, such as the C3 and C4-types of carbon dioxide assimilation, will reveal the adaptations of the chloroplast envelope proteome to increased metabolite flow. Unfortunately, routine methods are not available to compare membrane proteins of different species quantitatively or even semi-quantitatively. Membrane proteins are not amenable to two dimensional gel electrophoresis since extremely hydrophobic proteins, such as metabolite transporters, do not focus in the first dimension (Choe et al., 2005) and quantitative comparison relying on identical peptides such as affinity tagging are also not applicable since there is considerable evolutionary distance between the C3 model pea and the C4 model maize. A direct quantification method, the total spectral count of proteins (the number of mass spectra which map to one protein) has been used to compare and even quantify proteins on a large scale (Liu et al., 2004; Lu et al., 2007) (Majeran et al., 2008). This method has recently also been applied to yeast membrane proteins, with results

comparable to SILAC (Zybailov et al., 2005). We applied this strategy to compare the relative abundance of proteins in the chloroplast envelopes of C3 and C4 plants. We demonstrate that the massive metabolite fluxes across the chloroplast envelope required for maintaining the high photosynthetic rates of C4 plants are associated with significant increases of the relative abundance of several metabolite transporters, thus pinpointing apparent bottlenecks in metabolite flux across the chloroplast envelope membrane.

Material and Methods

Preparation of chloroplast envelope protein samples

Chloroplast envelope membranes were isolated from pea (*Pisum sativum* var. Little Marvel) plants as described previously (Douce and Joyard, 1979; Keegstra and Yousif, 1986) and from maize (*Zea mays*) plants grown on field sites. Briefly, fully expanded maize leaves were harvested, stored on ice, and cut into small pieces using razor blades. The leaves were homogenized in a Waring Blender and the resulting slurry was filtered through several layers of miracloth to remove the bundle sheath strands. Chloroplasts and chloroplast envelopes were isolated as described in (Douce and Joyard, 1979; Keegstra and Yousif, 1986). Envelope membranes were diluted in ten volumes of ice cold 1:1 (v/v) chloroform/methanol and stored on ice for twenty minutes. Insoluble proteins were sedimented by centrifugation at 20,000 g for 20 minutes ("pellet fraction") and both the protein pellet and the soluble fraction were dried and washed with hexane to remove residual membrane lipids. Envelope membrane samples and fractionated samples were mixed with SDS–PAGE loading buffer, incubated for 20 minutes on a reaction tube shaker at 15° C, and subsequently separated by 12.5% SDS-PAGE.

Protein identification

After staining with Coomassie Brilliant Blue, each gel lane was cut into ten equally-sized slices. Proteins contained in the gel slices were subjected to tryptic cleavage as described by Shevchenko (Shevchenko et al., 1996). Peptides were extracted and loaded onto a Waters Symmetry C18 peptide trap (5 μm, 180 μm x 20mm) at a flow rate of 4 μL/min in 2% Acetonitrile/0.1%Formic Acid for 5 minutes, using a Waters nanoAcquity Sample Manager. Using a Waters nanoAcquity UPLC system, the peptides were separated on a Waters BEH C18 nanoAcquity column (1.7 μm, 100 μm x 100mm) over 90 minutes and fed into a ThermoElectron LTQ-FTICR mass spectrometer with a flow rate of 300 nL/min (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 99.9% Acetonitrile/0.1% Formic Acid: gradient of 5% B to 40% B from 0 to 63 minutes, 40% B to 90% B from 63 to 71 minutes and 5% B from 71 to 90 minutes). Survey scans were taken at a resolution of 50,000 and the top ten ions were dissociated by automated low energy collision. The BioWorks Browser version v3.2 was used for converting the resulting MS/MS spectra to a peak list.

All mass spectra libraries were compared to sequence databases from *Pisum sativum* (Bräutigam et al., 2008) and *Zea mays* (ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Zea_mays), respectively, using the Mascot search algorithm, v 2.2 (www.matrixscience.com). Carbamidomethyl cysteine was set as a fixed peptide modification and oxidation of methionine was permitted. Up to two missed tryptic sites were allowed. The peptide tolerance was set to +/-10ppm and the MS/MS tolerance to 0.8kDa.

Protein annotation and bioinformatics

Mascot Results were analyzed with using an implementation of the peptide and protein prophet algorithms (Keller et al., 2002) (Scaffold®) with parameters set to 99% confidence for protein identification, requiring at least two unique peptides for each protein and 95% confidence for all peptides counted. Where Scaffold reported multiple proteins identified for the same peptides, each match was manually inspected and low scoring matches were discarded. The results were then exported to MS Excel for further analysis. Cross-identifications in previous chloroplast proteomics projects was determined using BlastX against the plprot database (Kleffmann et al., 2006) and the corresponding Arabidopsis proteins were identified by BlastX (Altschul et al., 1997) in TAIR. Functional annotation and classification presented in this manuscript was deduced from information in TAIR (Swarbreck et al., 2008), ARAMEMNON (Schwacke et al., 2003), membranetransport.org (Ren et al., 2007), and manual curation of the pertinent literature. The predicted location and the number of transmembrane helices were retrieved from ARAMEMNON. Tentative consensus sequences that identified the same Arabidopsis protein were aligned. If the sequences were identical, one of the identifications was discarded. If the sequences overlapped only partially, the annotations were unified and the number of peptides summed up to generate a list of non-redundant identifications (Tables 5-S1 and 5-S2). Before semi-quantitative analysis the spectral counts in each fraction were corrected for loading. The original data can be downloaded from PRIDE (http://www.ebi.ac.uk/pride/) (Martens et al., 2005).

Semi-quantitative analysis of protein abundance

The semi-quantitative analysis of protein abundance was based on the spectral count, i.e., the number of mass spectra mapping to a given protein in a single experiment. In the first experiment for each envelope preparation all proteins in the sample were separated by SDS-PAGE and identified by LC-ESI-MS/MS without prior fractionation ("whole envelopes"). In a second experiment, the proteins were first fractionated into a chloroform/methanol soluble and an insoluble fraction. Proteins from both fractions were then separated by SDS-PAGE and subsequently identified by LC-ESI-MS/MS. The spectral counts for each protein in both fractions were summed up to yield the "sum" fraction. For all four experiments the spectral count for each protein was normalized to the total number of spectra within the experiment ("percentage of the total spectral count") (Table 5-S1 and 5-S2). The robustness of the semi-quantitative analysis was tested by introducing a number of disturbances into the experiment: omitting all proteins with a spectral count lower than 10 spectra identified, and by including and excluding putative extraplastidial contaminations. The results were robust.

Accession numbers: All proteomics data reported in this manuscript have been submitted to the PRIDE data repository (http://www.ebi.ac.uk/pride/) (Martens et al., 2005) and can be downloaded from PRIDE using the following PRIDE experiment accession numbers:

Maize C4 PCA type chloroplast envelopes - pellet fraction' (PRIDE Experiment Accession # 3370)
Maize C4 PCA type chloroplast envelopes - soluble fraction' (PRIDE Experiment Accession # 3371)
Maize C4 PCA type chloroplast envelopes - whole' (PRIDE Experiment Accession # 3372)
Pea C3 type chloroplast envelopes - pellet fraction' (PRIDE Experiment Accession # 3376)
Pea C3 type chloroplast envelopes - soluble fraction' (PRIDE Experiment Accession # 3377)
Pea C3 type chloroplast envelopes - whole' (PRIDE Experiment Accession # 3378)

Results

Envelope proteome coverage and purity

The pea chloroplast envelope proteome was chosen to represent the envelope proteome of a C3 chloroplast. The proteins of the protein import complex found in our study were compared to those identified in earlier efforts (Froehlich et al., 2001; Ferro et al., 2003; Froehlich et al., 2003). We were able to identify all import complex components found by Froehlich et al. (2003) with the exception of Toc33 and we identified three additional import complex components that were not previously found. In comparison to Ferro et al. (2003) we also identified two additional components while Toc33 was again missing. With regard to a major metabolite transport protein family, the phosphate translocators (Knappe et al., 2003), we identified all members predicted to be present in the inner envelope of C3 chloroplasts (Eicks et al., 2002; Flügge et al., 2003). Based on this data and further analysis (not shown), we conclude that the proteome of pea chloroplast envelope has similar qualitative composition as the proteomes from other C3 plants analyzed previously.

The maize chloroplast isolation protocol applied in this study was optimized for isolation of maize mesophyll chloroplasts and thus C4 mesophyll chloroplast envelope membranes. Based on the virtual absence of Rubisco and complete absence of malic enzyme, two markers for bundle sheath chloroplasts, and the relative abundance of mesophyll marker enzymes, such as PPDK and PEPC, the maize chloroplast envelope samples indeed represents a highly mesophyll enriched preparation (Table 5-S1).

For each of the envelope proteome samples, the relative spectral abundance likely resulting from extraplastidial sources such as mitochondria, endomembrane system,

cytosol, and nucleus were determined. The level of contamination based on this measure was low; for the samples from maize it was below 2.2% and for the samples from pea below 5.2%. In maize, no mitochondrial contamination was detected and extraplastidial proteins were mostly residents of the cytosol and the endomembrane system. In pea, the main contaminant was mitochondrial proteins. The complete list of extraplastidial proteins can be extracted from Tables 5-S1 and 5-S2. Relative abundance comparisons were performed with and without removing the contaminations from the samples and the results were robust and therefore independent of the level and source of the contamination. We concluded that the samples are suitable for comparing a C3 type with a C4 mesophyll chloroplast envelope.

The envelope proteomes of C3 and C4 chloroplasts are qualitatively similar

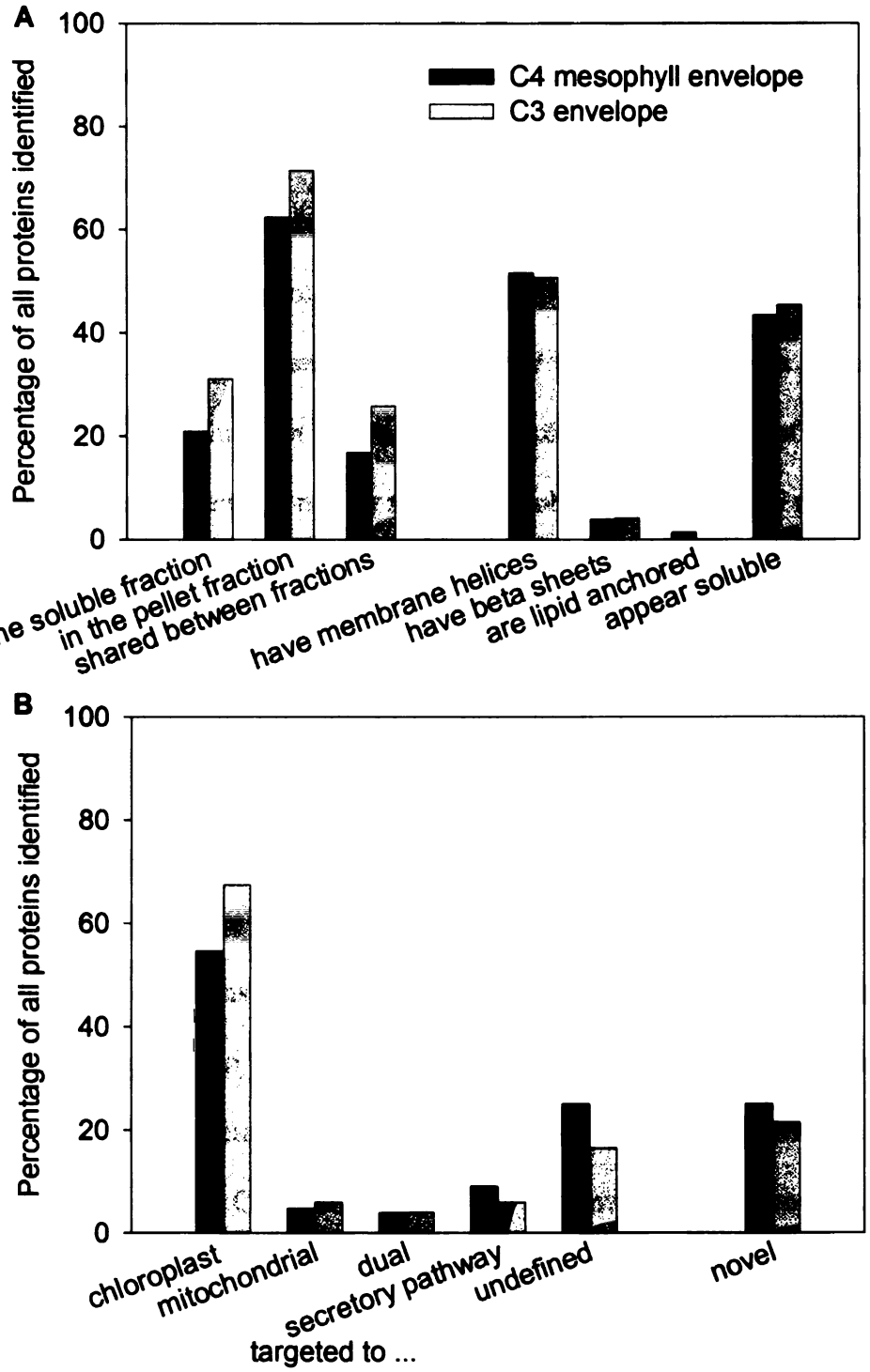


Figure 5-2: The envelope proteomes are similar when analyzed qualitatively. The percentage of proteins within one proteome is plotted. They are similar with regard to their physico-chemical properties (solubility in organic solvents and presence of membrane attachment structures) **(A)** and with regard to their predicted targeting and percentage of novel proteins **(B)**.

In the C4 mesophyll chloroplast envelope proteome, 231 non-redundant proteins were identified and, in the pea chloroplast envelope proteome, 322 non-redundant proteins were identified. Taken together, 420 unique proteins were identified of which 368 (87.6%) are traditional chloroplast residents. In both samples a similar percentage of proteins are soluble or insoluble in chloroform/methanol, respectively, with about one third of all proteins being soluble in organic solvents to at least some degree (Figure 5-2A). Likewise, a similar share of proteins could be detected in both fractions (Figure 5-2A). The fraction soluble in organic solvents contains a number of proteins with high membrane helix content but neither the hydrophobicity index nor the number of predicted membrane helices were strongly correlated with the solubility in organic solvent (data not shown).

Little more than one half of the proteins in both samples contain recognizable structures for membrane attachment (Figure 5-2A). Most of these proteins have predicted alpha helices which can span a membrane, some have demonstrated or predicted beta sheets, and very few are predicted to be anchored to the lipid bilayer by prenylation. The other half of the proteins has no obvious domains for membrane attachment or insertion.

The proteins in both envelope preparations are also very similar when their bioinformatically generated targeting predictions are compared. Most of the proteins identified in both the C4 mesophyll and the C3 envelope proteome sample possess a canonical target peptide for the protein import complex of chloroplasts (Figure 5-2B) (Emanuelsson et al., 1999; Emanuelsson et al., 2000; Schwacke et al., 2003). Less than ten percent are predicted being targeted to the mitochondria (Figure 5-2B). For a surprisingly large group, no targeting signal can be identified within the N-terminus of

the protein sequence and a number of proteins have strong bioinformatics support for targeting to the secretory pathway (Figure 5-2B). These proteins include well-known residents of the chloroplast envelope such as Toc64 and Toc159 as well as two outer envelope porins, OEP21 and OEP24.

Both envelope proteome samples yielded a comparable proportion of proteins not previously identified by proteomics, with 58 novel proteins from the C4 maize mesophyll envelope proteome samples and 69 novel proteins from the C3 type pea envelope proteome sample.

Proteins with similar relative abundance between samples

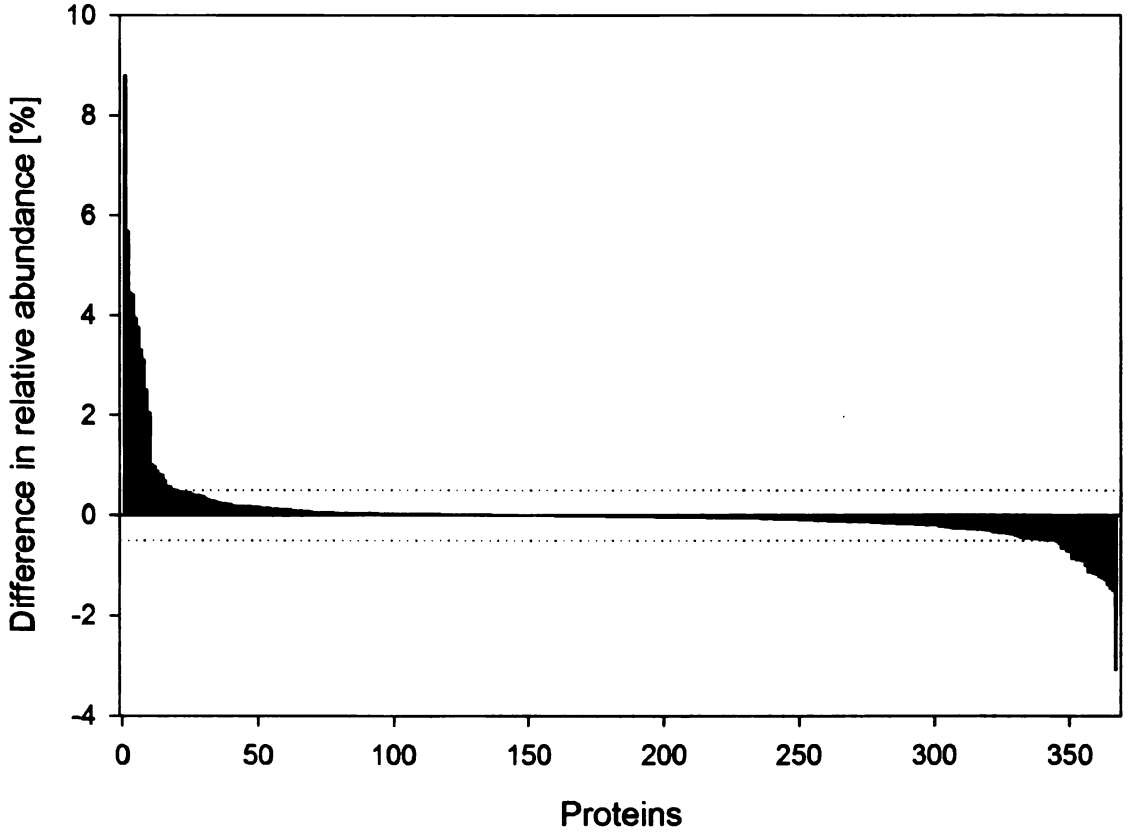


Figure 5-3: Quantitatively the envelope proteomes differ in selected proteins. The relative abundance in C3-type envelopes was subtracted from the relative abundance in C4 PCA-type envelopes and the difference was plotted for each protein; dotted lines represent differences of +/- 0.5%

To visualize the compositional differences between C3 and C4 envelope membrane proteomes, the differences between the percentages of the total spectral count for each protein (spectral count percentage in pea was subtracted from the spectral count percentage in maize) were plotted. Proteins that were identified in only a subset of the experiments were set to zero in the remaining experiments. Plotting the difference between the relative spectral abundance in C4 mesophyll and C3 chloroplast envelopes revealed that the majority of proteins do not differ more than 0.5% in their relative spectral abundance (Figure 5-3). This large group can be broken down into smaller groups of proteins. Each group differs for the reason it belongs to the proteins which are similar and selected examples are shown in Figure 5-4A. In a first group, both relative spectral counts are high or intermediate, as is the case for proteins of the protein import complex components (Figure 5-4A). As examples, part of the inner envelope pore, Tic110, and two outer envelope components, Toc34 and Toc64, are shown (Figure 5-4A). A long acyl chain CoA synthase and a protein of unknown function also belong to this group of proteins with high relative spectral abundance in all samples. A fourth import complex component, Toc159, inexplicably was reduced in one of two replicate experiments for C4 PCA type envelopes (Figure 5-4A). Proteins that showed large variance between replicate experiments were considered unreliable and were therefore not further considered. The putative glucose transporter pGlcT, a putative ATP-dependent transporter, and an enzyme of chlorophyll biosynthesis are of intermediate relative abundance in both samples. There are also proteins, which have a low absolute spectral count in one or both samples and therefore do not generate large differences, such as the transcription factor CIL or two proteins of unknown function (Figure 5-4A

and Tables 5-S1 and 5-S2). A complete list of proteins with similar relative spectral abundance can be found in Supplementary Table 5-S3.

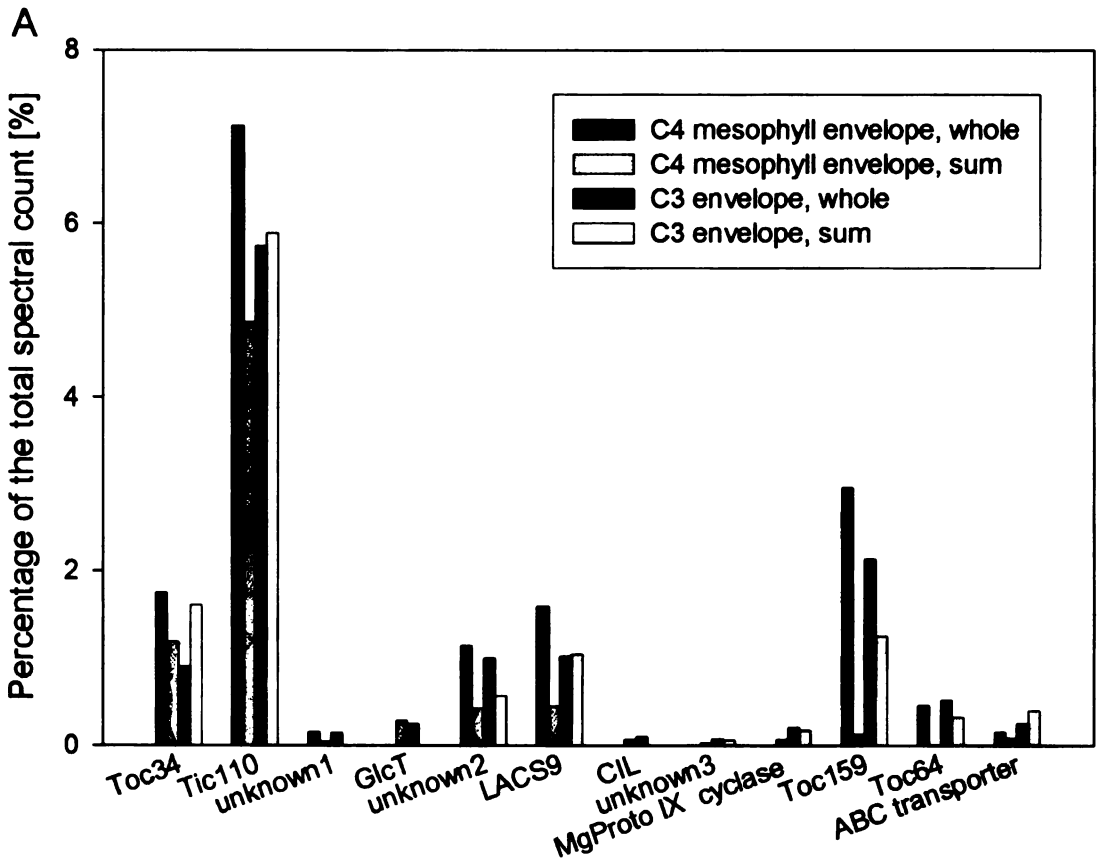
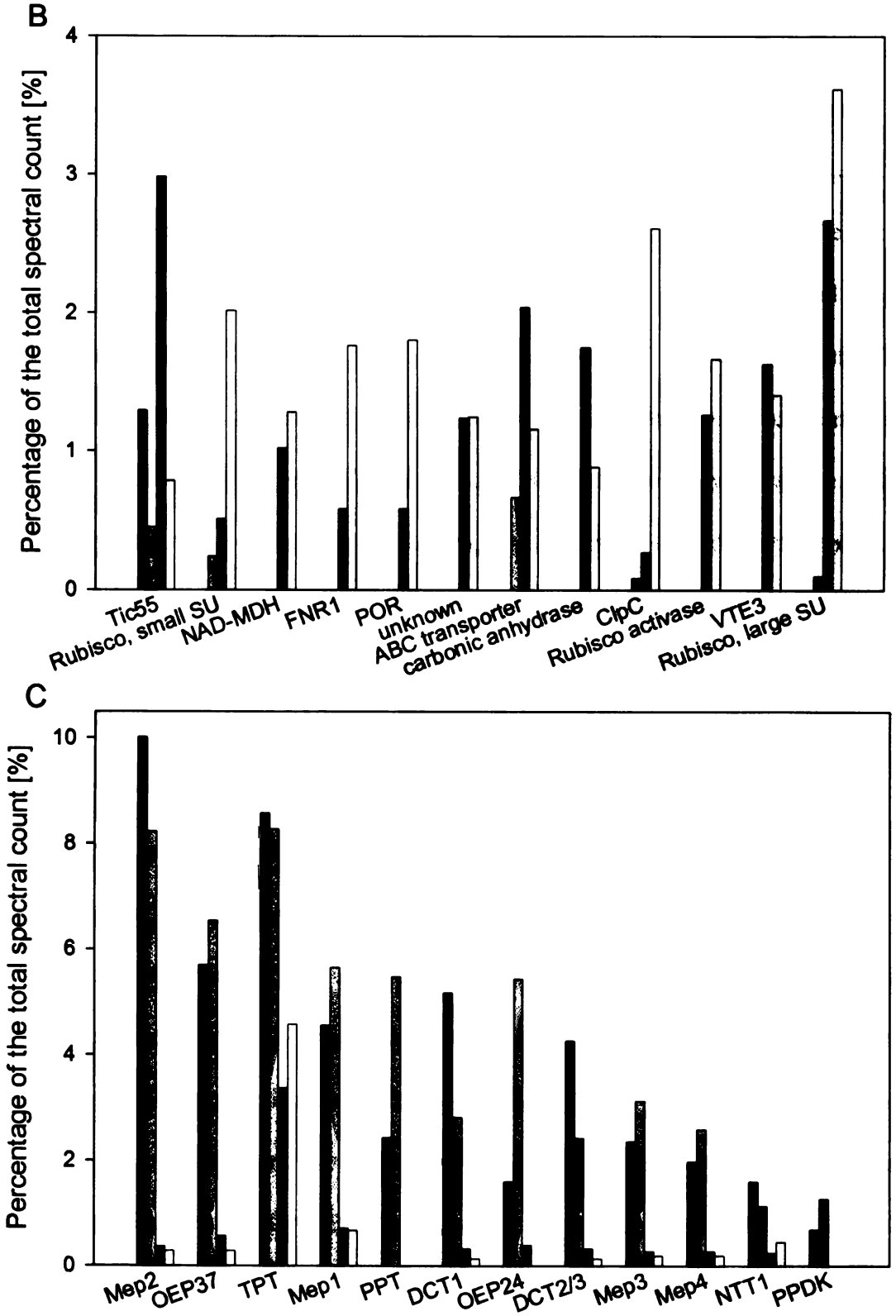


Figure 5-4: For the extremes and selected unchanged proteins out of Figure 5-3 detailed results were plotted; selected proteins that do not change significantly in relative abundance **(A);** the twelve proteins which are lowest in relative abundance in C_4-PCA type envelopes compared to C3-type envelopes **(B);** the twelve proteins which are highest in relative abundance in C_4-PCA type envelopes compared to C3-type envelopes**(C);**

Figure 5-4 (cont'd)



Proteins with different relative spectral abundance in C4 mesophyll envelopes compared to C3 type envelopes

Among the proteins with markedly decreased relative spectral abundance, only four out of twelve contained membrane-spanning helices, whereas of the proteins with increased abundance all but one were integral membrane proteins. Most of the proteins that were underrepresented in C4 mesophyll envelopes could not be detected at all in either of the replicate C4 experiments. The twelve proteins with the highest relative decreases are plotted in Figure 5-4B. There are four proteins involved in carbon fixation for PCR cycle, Rubisco large and small subunits, the Rubisco activase, and a carbonic anhydrase, with Rubisco large subunit showing the highest relative decrease. In addition there are three protein associated with the protein import complex, Tic55, the ferredoxin:NADP reductase, and the import chaperone Hsp93/ClpC (Soll and Schleiff, 2004). There are also three enzymes, VTE3, a methyltransferase involved in vitamin E and plastoquinone biosynthesis (Cheng et al., 2003), protochlorophyllide reductase (Beale, 1999), and a NAD-dependent malate dehydrogenase. Finally, there are two proteins of unknown function one of which is a putative ATP-dependent transporter.

The proteins that occupy a larger percentage of the spectral count in maize have high amplitudes of up to 9%, whereas proteins that occupy a larger percentage in the pea envelope have a lower amplitude of up to 3% (Figure 5-3). Most of the following proteins which show major relative increases in maize belong to the classes of known and putative transport proteins, except for PPDK, the enzyme required for regenerating the CO2 acceptor PEP (Figure 5-4C). The known transport proteins are two phosphate translocators, phospho*enol*pyruvate phosphate translocator (PPT) (Fischer et al., 1997)

and triosephosphate phosphate translocator (TPT) (Flügge and Heldt, 1984), two dicarboxylate translocators (DiTs), DCT1 and DCT2/3, and the ATP/ADP translocator (NTT1) (Neuhaus et al., 1997). There are also two outer envelope proteins (OEPs), OEP24 and OEP37 (Pohlmeyer et al., 1998; Goetze et al., 2006),. Finally, we identified four proteins of unknown function in this group (Mesophyll envelope protein 1-4). Mep1 is predicted to have twelve membrane spanning helices and Mep2 is predicted to have a single membrane spanning helix. Mep3 and Mep4 are paralogues, which map to the same Arabidopsis ortholog and both are predicted to have four membrane spanning helices. Of the twelve proteins with the highest difference in spectral count compared to C3 envelopes, ten have differential protein accumulation patterns between mesophyll and bundle sheath (Majeran et al., 2008). Eight accumulate to a higher level in mesophyll membranes and two accumulate to a higher level in bundle sheath membranes (Table 5-S4).

Discussion

Pea was chosen to represent C3 plants because it has served as a model for C3 chloroplasts for a long time and high purity chloroplast envelopes can be isolated with relative ease. Maize was chosen to represent C4 plants since most of the biochemical work on transport proteins has been published for maize chloroplasts compared to other C4 models (Huber and Edwards, 1977a, 1977b; Hatch et al., 1984; Ohnishi et al., 1990; Aoki et al., 1992).We established that the C3 envelope proteome from pea is comparable to earlier envelope proteomes prepared from the C3 plant Arabidopsis and confirmed the presence and absence, respectively, of marker proteins of C4 photosynthesis for the C4 mesophyll envelope sample such as PPDK, PEPC, and Rubisco. The level of

contamination was at most 5.2%. In the samples isolated from pea, the biggest contributors were mitochondrial outer envelope proteins such as porins. It is well known that, to foster metabolite exchange for photorespiration, chloroplasts of C3 plants are closely associated with mitochondria and peroxisomes (Schumann et al., 2007), thus explaining mitochondria being the major source of contaminating proteins. In the maize sample mitochondrial contaminants were virtually absent. The contaminant with the highest relative spectral count was PEPC, a cytosolic enzyme that is required for initial CO2 fixation. Considering that the isolation protocols for pea and maize are almost identical, the marked difference in extraplastidial contaminants may result from the altered requirements in organelle association. Mesophyll chloroplasts do not photorespire since Rubisco is virtually absent and these chloroplasts therefore do not require a close association of chloroplasts, mitochondria and peroxisomes. Both envelope preparations also contain a number of proteins identified in previous thylakoid proteome studies (Tables 5-S1 and 5-S2) (Peltier et al., 2000; Peltier et al., 2002; Peltier et al., 2004).Currently it remains unknown whether these proteins are trapped en route to the internal membrane system or whether they are a contamination introduced during the envelope isolation (Ferro et al., 2002; Froehlich et al., 2003).Relative abundance comparisons were performed with and without removing the contaminations from the samples and the results were robust and therefore independent of the level and source of the contamination.

We analyzed whether detailed qualitative comparisons were possible. Solid judgments about the significance of presence or absence of proteins require proteomics to be saturated to avoid false negative calls. To determine whether the proteome

identifications in either sample were saturated or whether a substantial number of proteins remained unidentified, the well understood pathways of glycolipid biosynthesis were analyzed. They provide a number of housekeeping proteins that are expected to be identified in envelope proteomics studies if saturation was reached, such as two enzymes necessary for sulfolipid biosynthesis and two known enzymes and a three-partite transport protein involved in galactolipid biosynthesis (Benning et al., 2006). As in earlier efforts, the envelope proteomes in this study only identify a subset of proteins in each pathway indicating that saturating coverage of the chloroplast envelope proteome remains to be achieved. It remains to be determined whether enzymes of glycolipid biosynthesis are difficult to detect with mass spectrometry or whether they are of too low absolute abundance. When additional replicates of the pea envelope proteome were tested, we also observed that enzymes with a low absolute spectral count disappeared and the enzyme catalyzing the next step appeared from replicate samples (data not shown). This may indicate that when proteins with a low number of spectra are analyzed, small variations during peptide separation, ionization, and detection may have determined whether or not they are present in any given sample. As a consequence, the envelope proteomes and not single proteins were the basis for the qualitative comparison.

We identified 231 and 322 non-redundant proteins in the C4 and C3 chloroplast envelopes, respectively. The higher number of proteins identified in the pea sample likely results from two reasons: (i) the total envelope sample from C4 mesophyll envelopes yielded a lower total spectral count with fewer proteins identified (Table 5-S1) although the relative abundances for each protein remained similar (Figure 5-4). Many proteins with a low absolute spectral count in the other experiments might have escaped detection.

(ii) The C4 mesophyll envelope sample contains some proteins with a high relative spectral count compared to the C3 envelope sample, with up to 9% difference in relative abundance (Figure 5-3). The peptides belonging to these proteins may have suppressed peptides of lesser abundance during ionization or detection in the mass spectrometer. The pre-fractionation by organic solvent extraction permitted the detection of additional proteins that could not be detected in a whole envelope preparation as many of the proteins yielding high relative spectral counts fractionated into the organic solvent soluble fraction, thus removing the main source for ion suppression. Yet total coverage did not reach the level obtained with C3 envelopes.

The analysis of the physico-chemical properties revealed that the C4 mesophyll and the C3 envelope proteome are remarkably similar. The fractionation pattern into soluble and insoluble in organic solvent was reproducible, as was the proportion of integral membrane proteins (Figure 5-2A). Little more than one half of the proteins in both samples contain recognizable structures for membrane attachment. In both envelope proteomes this group of proteins included a number of proteins for which a close association with the membrane has previously been demonstrated, such as the membrane lipid synthesizing and modifying enzymes (Jarvis et al., 2000; Froehlich et al., 2001; Sanda et al., 2001; Yu et al., 2002). It cannot be excluded that the remaining seemingly soluble proteins also are closely associated with the chloroplast envelope, similar to what has been demonstrated for glycolytic enzymes at the mitochondrial membranes (Graham et al., 2007).

The proteins in both envelope preparations are also very similar when their bioinformatically generated targeting predictions are compared. About half of the

proteins identified in both envelope proteome sample possess a chloroplast target peptide for the protein import complex (Figure 5-2B) (Emanuelsson et al., 1999; Emanuelsson et al., 2000; Schwacke et al., 2003), as was expected based on earlier results (Ferro et al., 2002; Froehlich et al., 2003). Some of the proteins that are predicted possessing a mitochondrial target peptide might be erroneously annotated as mitochondrial proteins by the prediction algorithm; a well documented case in point is the most abundant metabolite transport protein on the C3 envelope, the TPT, which is predicted to be targeted to the mitochondria. Alternatively, it may be due to contamination of the envelope preparation with true mitochondrial proteins. This, however, is unlikely at least for the C4 mesophyll envelope sample since no *bona fide* mitochondrial proteins could be identified in this preparation. For a surprisingly large group, no targeting signal can be identified within the N-terminus of the protein sequence and a number of proteins have strong bioinformatics support for targeting to the secretory pathway (Figure 5-2B), including well established plastid residents. Especially for proteins of the import complex it has been established that not all of them require the canonical import machinery. Many of the proteins without classical chloroplast targeting peptides have been identified in multiple independent plastid proteomics studies (Tables 5-S1 and 5-S2). The proteins identified in this and other studies might represent candidates for novel protein import pathways, as have recently been reported for a carbonic anhydrase (Villarejo et al., 2005) and outer envelope proteins (Bae et al., 2008).

Both envelope proteome samples yielded a comparable proportion of proteins not previously identified in plastid proteome projects. Some of the novel identifications may be due to the instrumentation used in our study, since the ultra high pressure HPLC

coupled to the FT-ICR is capable of protein identification with very high resolution. Some proteins may have been identified since the sample was fractionated prior to proteome analysis and some proteins, especially from the maize envelope sample, may have been identified since the chloroplast envelope is adapted to C4 photosynthesis and C4 chloroplast envelopes have not yet been analyzed by proteomics.

A semi-quantitative view of the envelope proteomes

For several reasons a semi-quantitative approach was needed to understand the differences between a C4 mesophyll and a C3 type chloroplast envelope. As pointed out earlier, qualitative analysis is hampered by unsaturated proteome identification and hence some uncertainty is associated with the identification of proteins with low absolute spectral counts. The proteome sample from maize was compared to previous proteome samples and more than 70% of the proteins identified in maize have been previously found in the plastid proteomes from other species (Figure 5-2B), indicating a large portion of the plastid envelope proteome is shared between different plastid species. Based on these results and on the adaptations of soluble proteins to C4 photosynthesis, we hypothesized that the differences between the C3 and the C4 chloroplast envelope are quantitative rather than qualitative. Unfortunately, no quantitative tools for comparing proteomes of different species are available. To overcome this limitation, we introduced percentage of the total spectral count as a measure for quantitative composition of the envelope proteome. This percentage is normalized to the total number of spectra identified within one single experiment, similar to the normalization procedures used for interpretation of RNA hybridization experiments. This method enables comparisons between evolutionary distant species. It is based on the assumption that orthologous

proteins from different species have similar physico-chemical properties and thus behave similarly throughout separation and identification when contained in similar samples, such as chloroplast envelopes. Although the percentage of total spectral counts is not an absolute measure of protein abundance, it is capable of capturing the relative contribution of a protein to the total, which enables comparison of non-related samples. The compositional differences between C3 and C4 envelope membrane proteomes were visualized by plotting the differences between the percentages of the total spectral count for each protein (spectral count percentage in pea was subtracted from the spectral count percentage in maize). We chose to compare the difference in relative abundance over the fold-change between the samples. Fold changes are likely a good measure if the proteins to be compared have high absolute spectral counts, which would allow a wide range of comparable values. In contrast, the envelope samples mainly consist of proteins of up to ten absolute spectral counts each (Tables 5-S1 and 5-S2) similar to results reported earlier (Bräutigam et al., 2008). A comparison based on fold-changes would lead to many proteins of low absolute spectral count to be erroneously identified as differentially expressed between C3 and C4 mesophyll envelopes and would fail to identify the protein with the second highest difference in relative spectral count, the TPT (data not shown). We thus restricted analysis to the proteins with the highest relative change in expression, which yielded comparable results in both experiments in each sample.

Marker Enzymes

A number of the proteins that were reduced or absent in the C4 sample are associated with functions that are expected to be absent from C4 mesophyll chloroplasts, such as the Rubisco large and small subunits, Rubisco activase, and carbonic anhydrase

for photosynthetic carbon reduction (Figure 5-4B). Since C4 mesophyll tissue has strongly reduced or absent Rubisco activity, the enzyme itself and its activase are also reduced. In mesophyll cells, a carbonic anhydrase, which quickly equilibrates CO2 and hydrogen carbonate, is needed in the cytosol for PEPC rather than in the chloroplast. The only soluble protein, which is massively increased in the C4 mesophyll envelope samples, is PPDK (Figure 5-4C). The detection of this soluble enzyme, which occupies a large percentage of the spectral count within the C4 mesophyll envelope proteome sample, may result from its high abundance resulting from its involvement in C4 photosynthesis and/or a close association with the membranes. Likely, it is absent from the pea sample because in contrast to C4 plants it represents a minor plastidic and cytosolic protein (Parsley and Hibberd, 2006) in C3 plants.

Proteins of the Protein Import Complex

At least two of the proteins which form the protein import complex seem to be housekeeping proteins, Tic110 and Toc75, which have a high relative spectral abundance in both samples. They form the pore in the inner and the outer envelope (Soll and Schleiff, 2004). The import receptor Toc159 was excluded from analysis since its relative abundance varied considerably between the biological replicates conducted on the C3 envelope. Two protein import complex components, Tic55 and ClpC/Hsp93, the import chaperone or protease subunit, were found among the proteins with a lower relative spectra abundance in C4 mesophyll envelopes. The remaining proteins, which are believed to be involved in redox regulation of protein import, Tic32 and Tic62 (Kuchler et al., 2002), were identified in the C3 envelope sample and could not be identified from the C4 mesophyll envelope. Taken together with the results of import complex

components, the two proteins involved in reduction equivalent synthesis and balancing, the FNR (ferredoxin:NADP+ reductase) and a NAD-dependent malate dehydrogenase may indicate a different mode of redox-dependent import in C3 as suggested by (Kuchler et al., 2002) compared to C4 mesophyll envelopes. This difference may be explained by the spatial separation of reduction equivalent production between mesophyll and bundle sheath chloroplasts, which may result in a change of redox status regulation.

Transport proteins

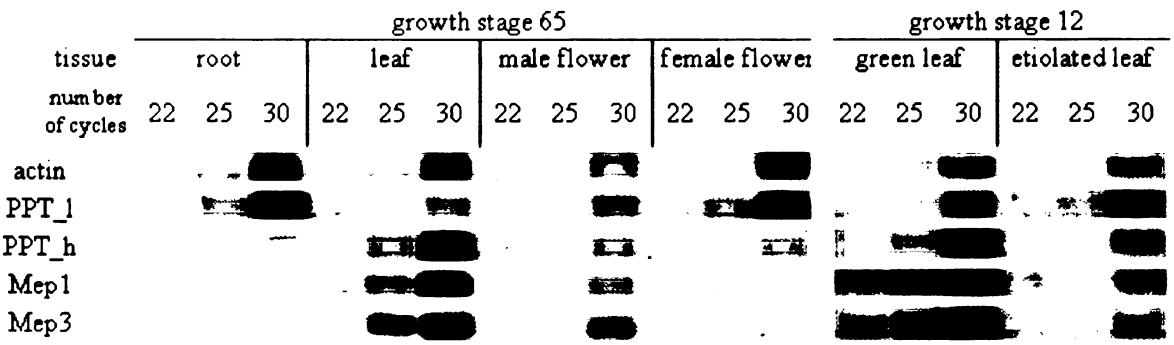


Figure 5-S1: For selected proteins, mRNA accumulation patterns in different tissues were analyzed

At the beginning of the light period tissue from roots, leaves, male and female flowers (dehusked) was harvested from maize stage 65. Total RNA was prepared and 0.5µg total RNA was reverse transcribed with Superscript III (Invitrogen) according to the manufacturer's instructions. The cDNA for each gene was amplified with GoTaq (Promega) according to the manufacturer's instructions. Primers were chosen to span 300bp (actin fwd:aagtacccgattgagcatgg rev:acctgaccatcaggcatctc, PPT_low fwd:CTTCCAGTCAAGGAATGTGCT rev:ATTGCCCACTGAATGTGTGA, PPT_high fwd:GGCCTCTGGTACCTGTTCAA rev:CAGAAAAGAAAGGCTCCATAGC, Mep1 fwd:TCGGCATGTTCTGTGTCTTC rev:GATGGTGCAAGAAGCTCTGTC and Mep3 fwd:TGCAGAGGATTGGAGCTTTT rev:AAGCACATTGCACTCAGCAG) with the resulting PCR product spanning an intron. Amplifications were carried out with 22, 25 and 30 cycles respectively and an extension time of 30 seconds. PPT_l (TC327380) is highly expressed in roots as demonstrated by Fischer et al. (1997) and female flowers and expressed more intensively in etiolated tissue compared to green leaves. In contrast PPT_h (TC349162) is mainly expressed in leaves as are the candidate genes Mep1 (TC319539) and Mep3 (TC343640) and their expression is very low in etiolated tissue.

Phosphate translocators The transport protein TPT is one of the proteins with the highest relative spectral abundance in both envelope samples but it is two-fold more abundant in C4 mesophyll than in C3 type envelopes (Figure 5-4C). TPT is the most abundant envelope transport protein in C3 chloroplast envelopes because it carries the major flux of carbon out of the C3 chloroplast during the day. In C4 mesophyll chloroplasts, the carbon fixation by Rubisco occurs in the bundle sheath plastids and carbon export can therefore not be the reason for the high relative abundance of TPT. However, since the bundle sheath chloroplasts are deficient in reduction equivalents due to limited photosystem II activity (Meierhoff and Westhoff, 1993), the reduction of 3-phosphoglycerate to triosephosphate occurs in mesophyll chloroplasts (Majeran et al., 2005). Since one exchange of the reduced for the oxidized form is necessary for each carbon fixed by Rubisco, the flux through the C4 TPT is at least threefold as compared to the C3 TPT, where only one exchange is required for three fixed carbon units for export (Figure 5-1). Compared to the C3 TPT, the C4 TPT exchanges 3-PGA rather than phosphate for triosephosphate (Figure 5-1) and may thus be specifically adapted to its new role. Interestingly, this protein is reported to be more abundant in C4 mesophyll chloroplast membranes compared to bundle sheath chloroplast membranes (Majeran et al., 2008) although TPT has to also export carbon from bundle sheath chloroplasts. In C4 PCA-type chloroplasts, phospho*enol*pyruvate (PEP) has to be exported from the chloroplast with a rate slightly exceeding the rate of carbon fixation (Laisk and Edwards, 2000). In C3 chloroplasts, PEP transport is a minor flux and the PEP phosphate translocator (PPT) was initially identified in maize endosperm and characterized from

cauliflower buds (Fischer et al., 1997). This PPT identified in maize is highly expressed in roots and the female flower and not higher expressed in green leaves compared to etiolated tissue (Figure 5-S1). Orthologues of this PPT are expressed in leaf tissue of the C3 plant *Arabidopsis thaliana* (Knappe et al., 2003; Voll et al., 2003), albeit at low levels. The maize mesophyll chloroplast envelope samples contain an isoform of PPT that is among the three most abundant proteins in this sample (Tables 5-S1 and Figure 5-4C), while in pea PPT belongs to the low-abundance group of proteins. In contrast to the PPT identified earlier this PPT is expressed highly in leaves, barely detectable in roots and higher expressed in green than in etiolated leaves (Figure 5-S1).The massive flux of PEP required for CO2 fixation is mediated by a specific PPT in the C4 leaf and it can only be maintained by increasing the amount of PPT in the envelopes, as compared to the envelope of the C3 species pea. This PPT protein is reported to be mesophyll specific (Majeran et al., 2008) in accordance with a role in C4 photosynthesis.

The demands for two of the four high volume fluxes necessary for C4 photosynthesis (triosephosphate vs. 3-phosphoglycerate and PEP vs. Pi) are thus accommodated by increased amounts of the respective transport proteins and hence increased V_{max}. The pentose phosphate translocator XPT could only be identified in the C3 envelope sample and a GPT was not detected in either experiment.

Dicarboxylate translocators The envelope proteome of C4 mesophyll chloroplasts contains a higher percentage of proteins from the dicarboxylate translocator (DiT) family (Weber et al., 1995). Proteins of both the glutamate/malate exchanger type, DiT2 (Taniguchi et al., 2002; Renne et al., 2003), called DCT1 and DCT2/3 in maize (Taniguchi et al., 2004), and of the 2-oxoglutarate/malate exchanger family, DiT1

(Weber et al., 1995), called OMT in maize (Taniguchi et al., 2004), are enriched. These transport proteins connect cytosolic and plastidic nitrogen metabolism through a two-translocator mechanism in C3 plants (Woo et al., 1987; Weber and Flügge, 2002; Renne et al., 2003) with DiT1 and DiT2 also playing a major role in photorespiration (Taniguchi et al., 2002; Renne et al., 2003; Schneidereit et al., 2006). Their function in C4 chloroplasts, such as those of maize mesophyll cells, is less well understood. There is controversial evidence with respect to their mRNA accumulation patterns in mesophyll cells (Renne et al., 2003; Taniguchi et al., 2004; Sawers et al., 2007). The protein accumulation pattern indicates higher expression of OMT and DCT1 in mesophyll chloroplast membranes and higher expression of DCT2/3 in bundle sheath chloroplast membranes (Majeran et al., 2008) (Table 5-S4) The DiT family members have been proposed to play a role in central nitrogen metabolism (Renne et al., 2003) and, for OMT of the DiT1 family, to be the oxaloacetate/malate shuttle that is needed for core C4 photosynthesis (Figure 5-1) (Taniguchi et al., 2004). Currently, the *in vivo* function of DiTs in C4 plants remains unclear, although their higher abundance in C4 compared to C3 may suggest C4 photosynthesis causes higher fluxes of their cargo metabolites. Especially the role of the additional DiT2 family member DCT2/3 present in the bundle sheath of maize has not been elucidated as the connection of nitrogen metabolism only requires two translocators and the function as an oxaloacetate/malate shuttle has only been proposed for OMT of the DiT1 family. In both samples, all members of the respective DiT families were identified.

Outer envelope porins The higher metabolite flux across the inner envelope of C4 chloroplast necessitates a comparably high flow through the outer

envelope. Two outer envelope porins, OEP24 and OEP37, both occupy a larger percentage of the spectral count in C4 mesophyll envelopes. *In vitro*, OEP24 transports triosephosphates and dicarboxylates and is thus perfectly suited to accommodate the metabolite fluxes needed for core C4 photosynthesis (Pohlmeyer et al., 1998). OEP37 has been shown to transport inorganic cations *in vitro* but *in vivo* substrates have not yet been established since the corresponding knockout mutant in the C3 plant Arabidopsis does not display an apparent phenotype (Goetze et al., 2006). In contrast, the regulated outer envelope porin, OEP21 (Bolter et al., 1999), is reduced in relative abundance (Table 5-S3), although not among the top twelve reduced proteins. It may be reduced since the regulation, which allows fine-tuning of the metabolite flow across the C3 envelope by the supply of ATP, 3-phosphoglycerate, and triosephosphate, may hinder metabolite exchange under C4 conditions. OEP16, an outer envelope protein which also forms a channel through the membrane, likely transports amino acids (Pohlmeyer et al., 1997). In contrast to the other outer envelope porins identified, this protein does not differ in relative spectral abundance between C3 and C4 PCA-type chloroplasts. The adaptations of the outer envelope proteins apparently reflect the changes in metabolite flux, indicating that the flux across the outer envelope might be limited and regulated by its proteins.

Other transport proteins Although metabolite transport proteins appeared to be generally increased in C4 mesophyll chloroplast envelope over C3-type envelopes, two proteins with unknown function, such as a putative ABC type transport protein and a protein of unknown function predicted being anchored to the membrane were absent from the C4 sample. In addition to the phosphate and dicarboxylate translocators, the C4

mesophyll envelopes contain more ATP/ADP translocator protein as compared to C3 envelopes from pea. Mesophyll chloroplasts have a high demand for ATP since the regeneration of the primary CO2 acceptor, PEP, from pyruvate requires two ATP for each reaction. Since mesophyll chloroplasts are the source of reduction equivalents for both mesophyll and bundle sheath chloroplasts, cyclic electron transport may be limited in favor of linear electron transport, thus reducing the availability of ATP in the chloroplast stroma. This limitation could be overcome by importing ATP from other sources into PCA-type chloroplasts. Within the group of proteins that are more abundant in maize mesophyll compared to pea C3 chloroplast envelopes are also four proteins of unknown function. Of these proteins one has one, three have four, and one has twelve predicted transmembrane helices. This study, complemented by data from (Majeran et al., 2008) allows us to posit hypothesis about proteins catalyzing additional C4 metabolite fluxes. Mep1, a protein with twelve predicted transmembrane helices, is enriched in C4 mesophyll compared to C3 envelopes and its protein accumulates evenly between mesophyll and bundle sheath. Moreover, its mRNA accumulates mainly in green leaves (Figure 5-S1). This pattern of expression fits to the pyruvate transport protein which carries a higher load in C4 plants compared to C3 plants and is needed in both mesophyll and bundle sheath chloroplasts. Mep3 and Mep4, a pair of closely related proteins with four predicted transmembrane helices of which one accumulated predominantly in the bundle sheath and the other in mesophyll tissue, are also candidates for the pyruvate transporter since they are both elevated in C4 compared to C3. In contrast, Mep2 accumulates mainly in the mesophyll and hence is a candidate for an oxaloacetate/malate shuttle if that function is not fulfilled by a member of the DiT family. The protein

sequence and predicted structure of all candidate proteins is unrelated to any characterized proteins.

Apart from being strong candidates for catalyzing metabolite fluxes across the maize mesophyll chloroplast envelope, which are increased to transfer core C4 photosynthesis metabolites, proteins of unknown function may carry fluxes which are increased as a byproduct of the C4 syndrome. For example, sulfur metabolism seems to be differentially localized in C4 chloroplasts between mesophyll and bundle sheath (Majeran et al., 2005) and therefore may require abundant transfer proteins. A comparison of bundle sheath with C3 chloroplast envelope membranes may be necessary to identify a candidate for the malate importer of bundle sheath chloroplasts.

Conclusions

The comparison of the C4 mesophyll and C3 chloroplast envelopes proteomes has revealed differences beyond the expected changes in metabolite transport proteins needed to support core C4 photosynthesis including major changes in the outer envelope. The molecular nature of the phosphate translocators involved in C4 photosynthesis was established and a number of candidate proteins for the additional fluxes were identified. Similar to what is observed during the transition from C3 to CAM metabolism in *Mesembryanthemum crystallinum* (Häusler et al., 2000), the abundance of chloroplast envelope membrane transporters is adjusted to meet the high metabolic flux rates demanded by C4 photosynthesis. To date, metabolite transport proteins have not been included in efforts to reengineer C4 photosynthesis. This analysis points to a greater role of the chloroplast outer and inner envelope membranes at least in mesophyll tissue for establishing the C4 carbon concentrating mechanism than previously assumed.

Limitations in metabolite exchange across the chloroplast envelope may well have hampered efforts to establish C4 photosynthesis in C3 crop plants.

Supplemental Material

Supplemental Data Table 5-S1. Proteins identified in C4 PCA-type maize chloroplast envelope membranes. The table lists proteins identified, number of spectra mapping to each maize accession number, annotation, classification, number of membrane spanning domains, targeting prediction, and previous identifications in other proteomics studies.

Supplemental Data Table 5-S2. Proteins identified in C3-type pea chloroplast envelope membranes. The table lists proteins identified, number of spectra mapping to each maize accession number, annotation, classification, number of membrane spanning domains, targeting prediction, and previous identifications in other proteomics studies.

Supplemental Data Table 5-S3. Percentage of total spectral counts for each protein identified in C4 PCA-type and C3-type chloroplasts of maize and pea, respectively.

Acknowledgements

We are grateful to the MSU Proteomics Core Facility for help with proteomic identification of chloroplast envelope membrane proteins. We thank the Bioinformatics Core of the MSU Research Technology Support Facility for assistance with database generation, sequence annotation, and data mining.

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402

Aoki N, Ohnishi J, Kanai R (1992) 2 Different Mechanisms For Transport Of Pyruvate Into Mesophyll Chloroplasts Of C-4 Plants-A Comparative-Study. Plant Cell Physiol 33: 805-809

Bae W, Lee YJ, Kim DH, Lee J, Kim S, Sohn EJ, Hwang I (2008) AKr2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. Nature Cell Biol 10: 220-U101

Beale SI (1999) Enzymes of chlorophyll biosynthesis. Photosyn Res 60: 43

Benning C, Xu CC, Awai K (2006) Non-vesicular and vesicular lipid trafficking involving plastids. Curr Opin Plant Biol 9: 241-247

Bolter B, Soll J, Hill K, Hemmler R, Wagner R (1999) A rectifying ATP-regulated solute channel in the chloroplastic outer envelope from pea. EMBO J 18: 5505-5516

Bräutigam A, Shresta RP, Whitten D, Wilkerson CG, Carr KM, Froehlich JE, Weber APM (2008) Comparison of the use of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes. J Biotechnol doi:10.1016/j.jbiotec.2008.02.007

Cheng ZG, Sattler S, Maeda H, Sakuragi Y, Bryant DA, DellaPenna D (2003) Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. Plant Cell 15: 2343-2356

Choe LH, Aggarwal K, Franck Z, Lee KH (2005) A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells. Electrophoresis 26: 2437-2449

Douce R, Joyard J (1979) Isolation and Properties of the Envelope of Spinach Chloroplasts. *In* E Reid, ed, Plant Organelles. Ellis Horwood Publishers, pp 47-59

Edwards GE, Furbank RT, Hatch MD, Osmond CB (2001) What Does It Take to Be C4? Lessons from the Evolution of C4 Photosynthesis. Plant Physiol 125: 46-49

Eicks M, Maurino V, Knappe S, Flügge UI, Fischer K (2002) The plastidic pentose phosphate translocator represents a link between the cytosolic and the plastidic pentose phosphate pathways in plants. Plant Physiol 128: 512-522

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1004-1016

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Prot Sci 8: 978-984

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N (2003) Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol Cell Prot 2: 325-345

Ferro M, Salvi D, Riviere-Rolland H, Vermat T, Seigneurin-Berny D, Grunwald D, Garin J, Joyard J, Rolland N (2002) Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. Proc Natl Acad Sci U S A 99: 11487-11492.

Fischer K, Kammerer B, Gutensohn M, Arbinger B, Weber A, Häusler RE, Flügge UI (1997) A new class of plastidic phosphate translocators: A putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. Plant Cell 9: 453-462

Flügge UI, Häusler RE, Ludewig F, Fischer K (2003) Functional genomics of phosphate antiport systems of plastids. Physiol Plant 118: 475-482

Flügge UI, Heldt HW (1984) The Phosphate-Triose Phosphate-Phosphoglycerate Translocator of the Chloroplast. Trends Biochem Sci 9: 530-533

Flügge UI, Stitt M, Heldt HW (1985) Light-Driven uptake of Pyruvate into mesophyll Chloroplasts from maize. FEBS Lett 183: 335-339

Froehlich JE, Benning C, Dörmann P (2001) The digalactosyldiacylglycerol (DGDG) synthase DGD1 is inserted into the outer envelope membrane of chloroplasts in a manner independent of the general import pathway and does not depend on direct interaction with monogalactosyldiacylglycerol synthase for DGDG biosynthesis. J Biol Chem 276: 31806-31812.

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW, Gage DA, Phinney BS (2003) Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J Prot Res 2: 413-425

Goetze TA, Philippar K, Ilkavets I, Soll J, Wagner R (2006) OEP37 Is a New Member of the Chloroplast Outer Membrane Ion Channels. J Biol Chem 281: 17989-17998

Graham JWA, Williams TCR, Morgan M, Fernie AR, Ratcliffe RG, Sweetlove LJ (2007) Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. Plant Cell 19: 3723-3738

Grodzinski B, Jiao JR, Leonardos ED (1998) Estimating photosynthesis and concurrent export rates in C-3 and C-4 species at ambient and elevated CO2. Plant Physiol117: 207-215

Hatch MD (1987) C-4 Photosynthesis - a Unique Blend of Modified Biochemistry, Anatomy and Ultrastructure. Biochim Biophys Acta 895: 81-106

Hatch MD, Droscher L, Flügge UI, Heldt HW (1984) A Specific Translocator for Oxaloacetate Transport in Chloroplasts. FEBS Lett 178: 15-19

Häusler RE, Baur B, Scharte J, Teichmann T, Eicks M, Fischer KL, Flügge UI, Schubert S, Weber A, Fischer K (2000) Plastidic metabolite transporters and their physiological functions in the inducible crassulacean acid metabolism plant *Mesembryanthemum crystallinum*. Plant J 24: 285-296

Huber SC, Edwards GE (1977) Transport in C4 Mesophyll Chloroplasts - Characterization of Pyruvate Carrier. Biochim Biophys Acta 462: 583-602

Huber SC, Edwards GE (1977) Transport in C4 Mesophyll Chloroplasts - Evidence for an Exchange of Inorganic-Phosphate and Phosphoenolpyruvate. Biochim Biophys Acta 462: 603-612

Jarvis P, Dörmann P, Peto CA, Lutes J, Benning C, Chory J (2000) Galactolipid deficiency and abnormal chloroplast development in the Arabidopsis MGD synthase 1 mutant. Proc Natl Acad Sci U S A 97: 8175-8179

Keegstra K, Yousif AE (1986) Isolation and Characterization of Chloroplast Envelope Membranes. Meth Enzymol 118: 316-325

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74: 5383-5392

Kleffmann T, Hirsch-Hoffmann M, Gruissem W, Baginsky S (2006) plprot: A comprehensive proteome database for different plastid types. Plant Cell Physiol 47: 432-436

Knappe S, Flügge UI, Fischer K (2003) Analysis of the Plastidic phosphate translocator Gene Family in Arabidopsis and Identification of New phosphate translocator-

Homologous Transporters, Classified by Their Putative Substrate-Binding Site. Plant Physiol 131: 1178-1190

Knappe S, Löttgert T, Schneider A, Voll L, Flügge UI, Fischer K (2003) Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis-AtPPT1 may be involved in the provision of signals for correct mesophyll development. Plant J 36: 411-420

Kuchler M, Decker S, Hormann F, Soll J, Heins L (2002) Protein import into chloroplasts involves redox-regulated proteins. EMBO 21: 6136-6145

Laisk A, Edwards GE (2000) A mathematical model of C-4 photosynthesis: The mechanism of concentrating CO2 in NADP-malic enzyme type species. Photosyn Res 66: 199-224

Leegood RC, Edwards GE (1996) Photosynthesis and the Environment, Vol 5. Kluwer Academic Publishers, Dordrecht

Liu H, Sadygov RG, Yates JR (2004) A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. Anal Chem 76: 4193-4201

Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature Biotechnol 25: 117-124

Majeran W, Cai Y, Sun Q, van Wijk KJ (2005) Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. Plant Cell 17: 3111-3140

Majeran W, Zybailov B, Ytterberg AJ, Dunsmore J, Sun Q, van Wijk KJ (2008) Consequences of C4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. Mol Cell Prot 10.1074/mcp.M800016-MCP200: M800016-MCP800200

Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, Gevaert K, Vandekerckhove J, Apweiler R (2005) PRIDE: The proteomics identifications database. Proteomics 5: 3537-3545

Matsuoka M, Nomura M, Agarie S, Miyao-Tokutomi M, Ku MSB (1998) Evolution of C4 photosynthetic genes and overexpression of maize C4 genes in rice. J Plant Res 111: 333-337

Meierhoff K, Westhoff P (1993) Differential Biogenesis Of Photosystem-II In Mesophyll And Bundle-Sheath Cells Of Monocotyledonous NADP-Malic Enzyme-Type C-4 Plants - The Nonstoichiometric Abundance Of The Subunits Of Photosystem-II In The Bundle-

Sheath Chloroplasts And The Translational Activity Of The Plastome-Encoded Genes. Planta 191: 23-33

Neuhaus HE, Thom E, Möhlmann T, Steup M, Kampfenkel K (1997) Characterization of a novel eukaryotic ATP/ADP translocator located in the plastid envelope of Arabidopsis thaliana L. Plant J 11: 73-82

Ohnishi J, Flugge UI, Heldt HW, Kanai R (1990) Involvement of Na+ in Active Uptake of Pyruvate in Mesophyll Chloroplasts of Some C4 Plants - Na+/Pyruvate Cotransport. Plant Physiol 94: 950-959

Parsley K, Hibberd JM (2006) The Arabidopsis PPDK gene is transcribed from two promoters to produce differentially expressed transcripts responsible for cytosolic and plastidic proteins. Plant Mol Biol 62: 339-349

Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, von Heijne G, van Wijk KJ (2002) Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. Plant Cell 14: 211-236.

Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I, van Wijk KJ (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12: 319-341.

Peltier JB, Ytterberg AJ, Sun Q, van Wijk KJ (2004) New functions of the thylakoid membrane proteome of Arabidopsis thaliana revealed by a simple, fast, and versatile fractionation strategy. J Biol Chem 279: 49367-49383Pohlmeyer K, Soll J, Grimm R, Hill K, Wagner R (1998) A High-Conductance Solute Channel in the Chloroplastic Outer Envelope from Pea. Plant Cell 10: 1207-1216

Pohlmeyer K, Soll J, Steinkamp T, Hinnah S, Wagner R (1997) Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane. Proc Natl Acad Sci U S A 94: 9504-9509

Ren QH, Chen KX, Paulsen IT (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res 35: D274-D279

Renné P, Dressen U, Hebbeker U, Hille D, Flügge UI, Westhoff P, Weber APM (2003) The Arabidopsis mutant *dct* is deficient in the plastidic glutamate/malate translocator DiT2. Plant J 35: 316-331

Sage RF (2004) The evolution of C-4 photosynthesis. New Phytologist 161: 341-370

Sanda S, Leustek T, Theisen MJ, Garavito RM, Benning C (2001) Recombinant Arabidopsis SQD1 converts UDP-glucose and sulfite to the sulfolipid head group precursor UDP-sulfoquinovose in vitro. J Biol Chem 276: 3941-3946.

Sawers RJH, Liu P, Anufrikova K, Hwang JTG, Brutnell TP (2007) A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf. BMC Genomics 8:12

Schneidereit J, Häusler RE, Fiene G, Kaiser WM, Weber APM (2006) Antisense repression reveals a crucial role of the plastidic 2-oxoglutarate/malate translocator DiT1 at the interface between carbon and nitrogen metabolism. Plant J 45: 206-224

Schumann U, Prestele J, O'Geen H, Brueggeman R, Wanner G, Gietl C (2007) Requirement of the C3HC4zinc RING finger of the Arabidopsis PEX10 for photorespiration and leaf peroxisome contact with chloroplasts. Proc Natl Acad Sci U S A 104: 1069-1074

Schwacke R, Schneider A, van der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flügge UI, Kunze R (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. Plant Physiol 131: 16-26

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. Anal Chem 68: 850-858

Slack CR, Hatch MD, Goodchild J (1969) Distribution of Enzymes in Mesophyll and Parenchyma-Sheath Chloroplasts of Maize Leaves in Relation to C4-Dicarboxylic Acid Pathway of Photosynthesis. Biochem J 114: 489

Soll J, Schleiff E (2004) Protein import into chloroplasts. Nature Rev Mol Cell Biol 5: 198

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36: D1009-D1014

Taniguchi M, Taniguchi Y, Kawasaki M, Takeda S, Kato T, Sato S, Tahata S, Miyake H, Sugiyama T (2002) Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in *Arabidopsis thaliana*. Plant Cell Physiol 43: 706-717

Taniguchi Y, Nagasaki J, Kawasaki M, Miyake H, Sugiyama T, Taniguchi M (2004) Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize. Plant Cell Physiol 45: 187-200

Voll LM, Häusler RE, Hecker R, Weber APM, Weissenböck G, Fiene G, Waffenschmidt S, Flügge UI (2003) The phenotype of the Arabidopsis *cue1* mutant is not simply caused by a general restriction of the shikimate pathway. Plant J 36: 301ff

Weber A, Menzlaff E, Arbinger B, Gutensohn M, Eckerskorn C, Flügge UI (1995) The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. Biochemistry 34: 2621-2627.

Weber APM (2004) Solute transporters as connecting elements between cytosol and plastid stroma. Curr Opin Plant Biol 7: 247-253

Weber APM, Fischer K (2007) Making the connections - The crucial role of metabolite transporters at the interface between chloroplast and cytosol. FEBS Lett 581: 2215-2222
Weber A, Flügge UI (2002) Interaction of cytosolic and plastidic nitrogen metabolism in plants. J Exp Bot 53: 865-874

Weber APM, Schwacke R, Flügge UI (2005) Solute transporters of the plastid envelope membrane. Ann Rev Plant Biol 56: 133-164

Woo KC, Flügge UI, Heldt HW (1987) A 2-Translocator Model for the Transport of 2-Oxoglutarate and Glutamate in Chloroplasts During Ammonia Assimilation in the Light. Plant Physiol 84: 624-632

Yu B, Xu CC, Benning C (2002) Arabidopsis disrupted in SQD2 encoding sulfolipid synthase is impaired in phosphate-limited growth. Proc Natl Acad Sci U S A 99: 5732-5737

Zybailov B, Coleman MK, Florens L, Washburn MP (2005) Correlation of Relative Abundance Ratios Derived from Peptide Ion Chromatograms and Spectrum Counting for Quantitative Proteomic Analysis Using Stable Isotope Labeling. Anal Chem 77: 6218-6224

Chapter 6

Identification and characterization of a metabolite transport protein involved in photorespiration: the chloroplastidic glycerate glycolate carrier Mep1

Andrea Bräutigam and Andreas Weber

Introduction

The photorespiratory pathway is the refixation of phosphoglycolate produced by the oxygenation reaction of Rubisco (Ogren, 1984). The active center of Rubisco enables not only the carboxylation of ribulose-1,5-bisphosphate to form two molecules of 3-phosphoglycerate (3-PGA) but also an oxygenation reaction with O2 which produces one molecule of 3-PGA and one of phosphoglycolate (Schneider et al., 1992; Kellogg and Juliano 1997). The side reaction is caused by the stereochemistry of the active center (Wingler et al., 2000). Therefore a recycling process for carbon lost to phosphoglycolate has evolved which is called the photorespiratory pathway. This pathway is a highly compartmented process and uses enzymes located in chloroplasts, peroxisomes and mitochondria (Ogren, 1984). Phosphoglycolate is dephosphorylated in the chloroplasts (Somerville and Ogren, 1979b) and glycolate leaves the chloroplasts and enters the peroxisomes (Ogren, 1984). A candidate transport protein was characterized biochemically (Young and McCarty 1993). The glycolate glycerate carrier transports glycolate and glycerate depending on a pH gradient across the membrane (Young and McCarty, 1993). The transport is not a strict counter-exchange but the presence of the second substrate on the opposite side of the membrane stimulates transport (Howitz and McCarty, 1986; Howitz and McCarty, 1991). After glycolate is imported into peroxisomes it is oxidized to glyoxylate and transaminated to glycine with either serine or glutamate as the donor of the amino group (Somerville and Ogren, 1980; Ogren, 1984). Glycine is moved to the mitochondria where it is decarboxylated by a multi-subunit complex to finally form serine (Somerville and Ogren, 1981; Somerville and Ogren, 1982). At this point photorespiration is coupled to one-carbon metabolism since

the decarboxylation produces methyl-THF (Collakova et al., 2008). Serine is exported from the mitochondria and reenters the peroxisomes where it is deaminated to hydroxypyruvate (Somerville and Ogren, 1980). Photorespiration is coupled to ammonia fixation in the chloroplasts (Weber, 2004; Linka and Weber, 2005). The exchange of 2-oxoglutarate and glutamate across the chloroplast envelope is necessary for efficient photorespiration probably to balance the amount of amino donors and –acceptors as well as free ammonia (Weber, 2004; Linka and Weber, 2005). Hydroxypyruvate is reduced to glycerate and has to be re-imported into chloroplasts where it can be phosphorylated to 3-PGA and reenter photosynthetic carbon reduction (Ogren, 1984). The reentry may be catalyzed by the same protein which mediates glycolate export and is also driven by a pH gradient (Howitz and McCarty, 1986; Howitz and McCarty, 1991; Young and McCarty, 1993). The compartmentation of the enzymes in different compartments as well as the connection to ammonia fixation necessitates a large number of transport steps. Although the last enzyme involved in photorespiration has been identified recently (Boldt et al. 2005) only the transport proteins for 2-oxoglutarate and glutamate which are only peripherally associated with photorespiration have been identified at the molecular level (Somerville and Ogren 1983; Weber et al. 1995; Renne et al. 2003). No transport protein for core photorespiratory metabolites has been identified at the molecular level.

Co-expression of proteins involved in the same pathway is a well known concept in bacterial operons which recently has been applied to eukaryotes (e.g. (Reumann and Weber, 2006)). The operon structure of many genes in procaryotes which leads to coordinated gene expression has enabled researches to identify additional genes is a pathway once the first member of an operon was characterized (Overbeek et al., 2005).

Although eukaryotic genes are generally not spatially arranged into clusters according to metabolic pathways, the expression of genes involved in one pathway is frequently regulated coordinately (Horan et al., 2008).

In this work a transport protein for photorespiratory metabolites at the chloroplast envelope was identified. We describe the identification of a candidate transport protein for photorespiratory metabolites based on publicly available microarray data. The intracellular localization of the candidate transporter was determined and both mutant analysis as well as localization allowed us to pinpoint the molecular function. It was attempted to demonstrate biochemical function.

Materials and Methods

Co-expression analysis

The photorespiratory genes were extracted from TAIR based on the predicted or annotated function and confirmed through literature review whenever possible. For each gene, publicly available microarray data was queried through the CSB.DB (Steinhauser et al. 2004). The lists of genes were manually screened for membrane proteins by comparing Arabidopsis genome identification codes with ARAMEMNON (Schwacke et al. 2003). The membrane proteins were compiled into lists and retested when additional databases became available. Finally, Pearson and Spearman correlation coefficients were extracted from the plant expression database (Horan et al. 2008) for photorespiratory genes and Mep1. In parallel, microarray datasets were assembled and clustered hierarchically according to (Reumann and Weber 2006) to determine how photorespiratory genes cluster compared to the complete genome.

Molecular Biology

The -1.5kb upstream gene region of Mep1 (At1g32080) was amplified and cloned directionally into pENTR-SD-TOPO. The insert was mobilized into pMDC163 with LR clonase (Invitrogen) and the resulting construct was transferred to Arabidopsis by *Agrobacterium tumefaciens* mediated transformation with the floral dip method (Bent, 2000). Transformed plants were selected; the T3 generation was rescreened on hygromycin and subjected to staining at suitable developmental stages. Staining was carried out as described in (Knappe et al. 2003). Plant were fixed with 3.7% formaldehyde in 80% ethanol, destained, and maintained in 80% ethanol.

A C-terminal GFP fusion was created by amplifying the coding region of Mep1 and cloning the product into pDONR221. The insert was mobilized into pMDC83 and transiently expressed in tobacco leaves according to (Waadt and Kudla 2007).

The insertion line was isolated and verified according to the recommendations from the SALK center. Total RNA was prepared according to (Knappe et al. 2003), cDNA synthesis and PCR amplification of targets with the SSIII Polymerase (NEB) was carried out according to the manufacturer's instructions with 35 amplification cycles. The insertion mutant line was maintained at elevated CO2 concentrations for improved growth and seed set.

Metabolite analysis

Plants were grown with a 12h light, 12 hour dark cycle at 3000ppm CO2 with 100uE light intensity for four to six weeks and watered as needed with half strength Hoagland's solution. For metabolite analysis, the complete rosette was harvested, shock frozen in liquid nitrogen and ground to a fine powder before being aliquoted under liquid

nitrogen. For metabolite analysis, approximately 100mg of tissue (fresh weight) was extracted with 700ul methanol and 700ul chloroform. The chloroform fraction was discarded. For LC-MS/MS analysis, the methanolic extract was analyzed according to (Lu et al., 2008) except that standards were also prepared in methanol. For GC-MS analysis, the methanolic extract was dried and derivatized according to (Fiehn et al., 2001). Derivatized metabolites were separated for 30 minutes on HP5 columns. Metabolites were identified by co-elution with standards and by fractionation pattern and quantified with external standard curves of complex standards.

Results

Identification of candidate transport proteins for photorespiratory metabolite flux

Co-expression of genes has been used successfully to identify genes involved in the same pathway (Reumann and Weber, 2006). The enzymes known to be involved in the photorespiratory pathway were used to query publicly available co-expression databases. These genes belong to a large cluster of 100 to 150 genes depending on the clustering method which includes genes involved in photorespiration, photosynthesis and chloroplast function (Horan et al., 2008). The co-regulation coefficients for both the Spearman (S) and the Pearson (P) functions between the photorespiratory enzymes were annotated and a large number of genes involved in photorespiration are indeed co-expressed (Table 6-1).

Table 6-1: Co-expression Spearman coefficients for genes involved in photorespiration and for the new candidate gene Mep1; only the isoform with the best correlation coefficient is shown; correlations with p<0.1 are printed boldface

Enzyme	AGI	Abbre-viation	PGLP1	GOX1	AGT1	GGT1	GLDP1	GLDH1	GLDH3	GLDT1	SHM1	HPR1	GlyK	Mep1
2-PG phosphatase	At5g36790	PGLP1	1.00	-	-	-	-	-	-	-	-	-	-	-
Glycolate oxidase	At3g14420	GOX1	**0.93**	1.00	-	-	-	-	-	-	-	-	-	-
Ser:glyoxylate aminotransferase	At2g13360	AGT1	**0.91**	**0.95**	1.00	-	-	-	-	-	-	-	-	-
Glu:glyoxylate aminotransferase	At1g23310	GGT1	**0.92**	**0.96**	**0.96**	1.00	-	-	-	-	-	-	-	-
GDC P-protein	At4g33010	GLDP1	**0.93**	**0.95**	**0.93**	**0.94**	1.00	-	-	-	-	-	-	-
GDC H-protein	At2g35370	GLDH1	**0.93**	**0.92**	<0.9	<0.9	<0.9	1.00	-	-	-	-	-	-
	At1g32470	GLDH3	**0.91**	<0.9	<0.9	<0.9	<0.9	**0.94**	1.00	-	-	-	-	-
Gly decarboxylase T-protein	At1g11860	GLDT1	**0.92**	**0.92**	<0.9	<0.9	<0.9	**0.92**	**0.95**	1.00	-	-	-	-
Ser hydroxy methyltransferase	At4g37930	SHM1	**0.93**	**0.96**	**0.94**	**0.96**	**0.95**	**0.91**	<0.9	**0.92**	1.00	-	-	-
Hydroxypyruvate reductases	At1g68010	HPR1	**0.93**	**0.97**	**0.96**	**0.97**	**0.94**	<0.9	<0.9	**0.91**	**0.96**	1.00	-	-
Glycerate kinase	At1g80380	GlyK	<0.9	<0.9	<0.9	<0.9	<0.9	<0.9	<0.9	<0.9	<0.9	<0.9	1.00	
AtMep1	At1g32080	Mep1	**0.93**	**0.90**	**0.90**	**0.90**	<0.9	**0.93**	**0.93**	**0.93**	**0.91**	**0.93**	<0.9	1.00

Phosphoglycolate phosphatase is co-expressed with all genes for the pathway except the L-subunit of GDC and glycerate kinase (Table 6-1). For the enzymes where several isoforms of the enzymes exist, only one is strongly co-expressed with other photorespiratory genes. Of the five isoforms encoding enzymes for glycolate oxidation only AtGOX1 is strongly correlated (AtGOX2 is indistinguishable due to the probe on the ATH1 chip) whereas the other four candidate genes show no co-expression with other photorespiratory genes (Table 6-1 and data not shown). Likewise only GGT1 but not GGT2, only one isoform each of the subunits of GDC and only the SHM isoform is co-expressed. After co-expression between the enzymes involved in photorespiration was established, proteins of unknown function which have a co-regulation coefficient of at least 0.9 in either the Pearson or the Spearman correlation were tested for the presence of predicted membrane spanning helices. One protein of unknown function (Mep1) was both co-expressed with genes involved in photorespiration and had twelve predicted membrane spanning helices.

Mep1 is expressed in green tissues only

After the co-expression analysis revealed AtMep1 to be a candidate for a transport protein involved in photorespiration, the expression pattern was tested experimentally. A promoter::GUS fusion expressed stably in Columbia-0 background was stained only in tissues and organs containing chloroplasts in all developmental stages tested. A total of eight lines were analyzed with seven lines showing identical patterns and no GUS activity visible in the eighth line. In the seedling stage, both cotyledons were stained strongly, with the hypocotyl stained weaker and the primary root remained white.

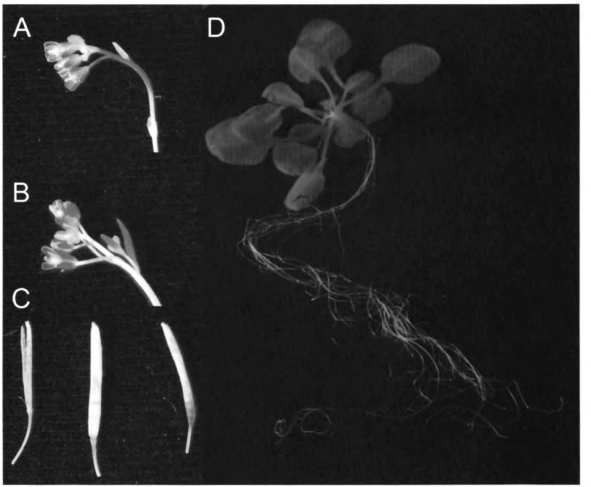


Figure 6-1: GUS expression pattern of a P(Mep1)::GUS fusion – this image is presented in color

In adult plants of the vegetative stage, all rosette leaves stained although the major veins of leaves and the center of the rosette around the meristem show less GUS activity (Figure 6-1D). As in the seedling stage, the roots remain white although small patches of GUS activity were occasionally observed in single plants. Within the flower, GUS activity is confined to green tissues (Figure 6-1A and B). The sepals but not the petals or the stamen show blue staining. In siliques the GUS staining is confined to the tip and the base of the siliques in various stages of ripening but the green silique walls showed no

GUS activity (Figure 6-1C). This expression pattern agrees with data extracted from the AtGenExpress database of publicly available microarrays where the highest expression is detected in rosette leaves and the stem with moderate expression in flowers and very low expression in roots, pollen and siliques (Figure 6-1A and B).

The Mep1 insertion mutant shows a visible phenotype in ambient air

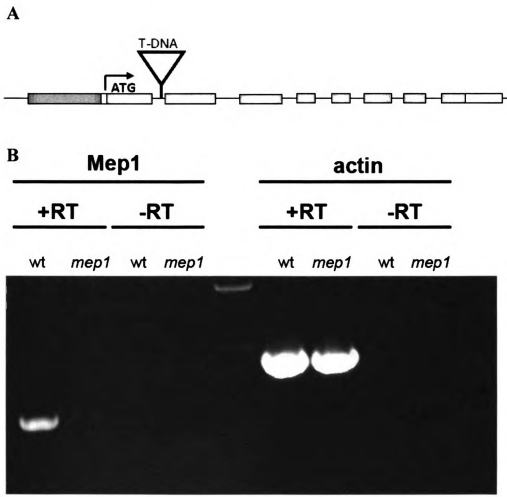
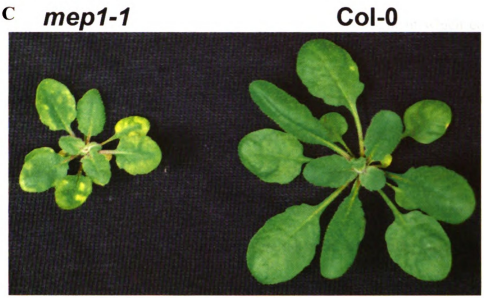


Figure 6-2: Isolation of a Mep1 insertion mutant; Structure of the Mep1 gene (At1g32080) and insertion site of the SALK line (A); Mep1 specific transcripts are undetectable in the insertion mutant of Mep1 *mep1-1* (B); the insertion mutant *mep1-1* shows a distinct phenotype (C); this image is presented in color

Figure 6-2 (cont'd)

C



To elucidate the in planta function of Mep1 an insertion line in the gene was isolated from the SALK collection. Although the SALK insertion is localized in an intron of the gene (Figure 6-2A), no mRNA was detected in saturated RT-PCR (Figure 6-2B). The plants show a strong visible phenotype if grown in ambient air with at least 14 hour light periods (Figure 6-2C). Leaves start to yellow in the regions between the veins and finally bleach out but stay green along the veins. Young leaves are not affected. It was observed that the phenotype is conditional both based on the CO2 content of the surrounding air and the light intensity and light period. The visible phenotype is completely suppressed by supplying the plants with at least 3000ppm CO2 independent of the light conditions. In ambient air, increasing the light intensity or increasing the day length aggravates the phenotype which is always visible at day lengths greater than 14 hours with a light intensity greater than 80 uE after at least five days of exposure. Once the phenotype occurs, it cannot be reversed by decreasing the light intensity or increasing the carbon concentration. Newly emerging leaves however do not display the phenotype

if kept under non-inducing conditions. The visible phenotype of *mep1-1* can be complemented by introducing a 3.5kb genomic DNA fragment which contains the full length gene and its native promoter.

A biochemical phenotype is observed even under non-inducing conditions

To understand the underlying biochemical condition of the visible phenotype and the basis of suppression in air enriched in CO2, steady state metabolite concentrations were determined. Three time points were chosen. Since the phenotype is aggravated by increasing light duration, all samples were taken in the evening based on the hypothesis that metabolites accumulate with increasing day length. Initial measurements were taken of plants grown in CO2 enriched air. Plants were shifted to ambient air at 80 uE and after two days (no phenotype visible) and after five days (at the onset of phenotype) whole rosettes were harvested. The steady state concentrations of amino acids were determined from liquid methanolic extracts, organic acids and simple sugars were measured from dried and derivatized metabolite extracts.

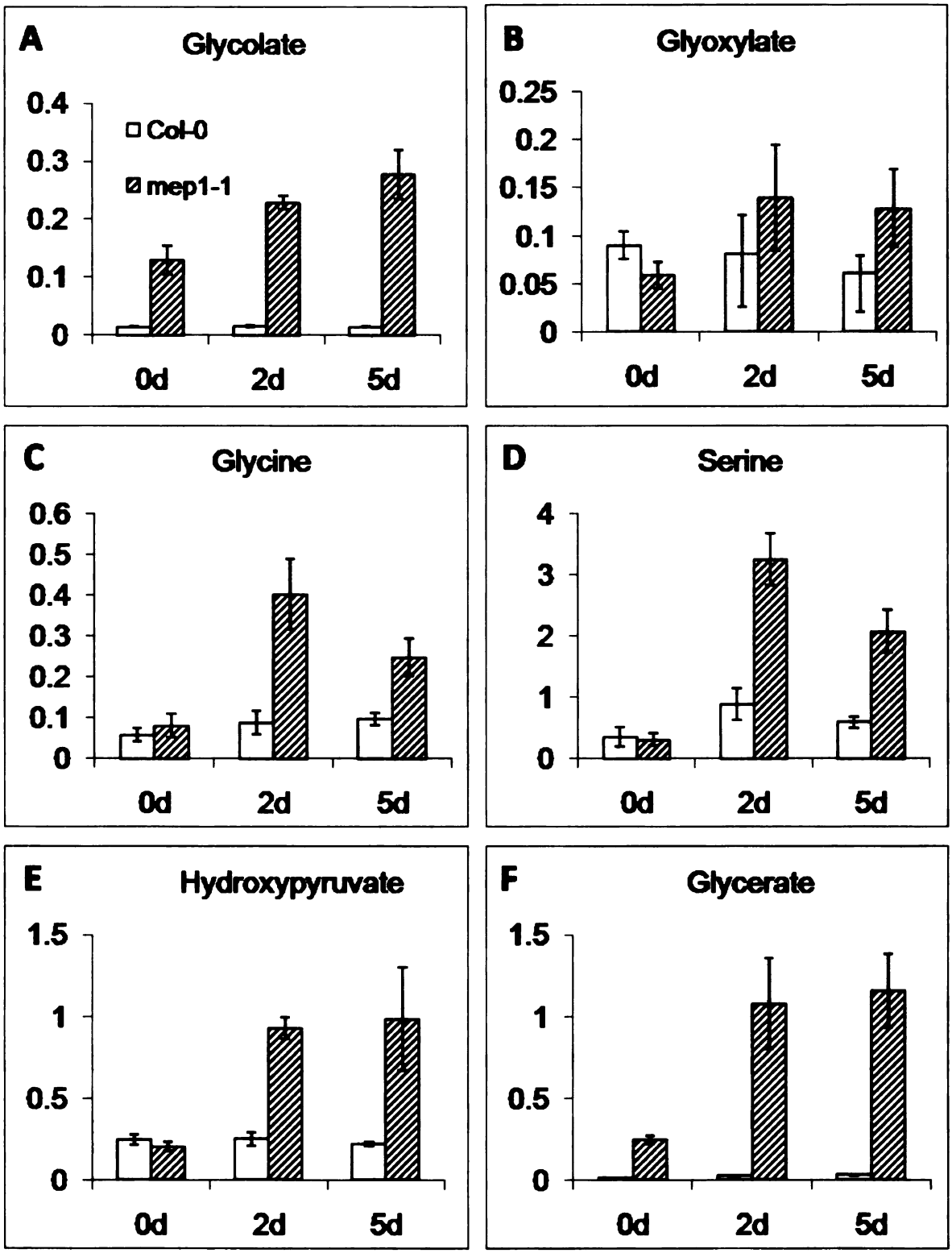


Figure 6-3: Metabolite accumulation in the Mep1 mutant compared to wild type on a shift from 3000ppm CO2 to ambient air; metabolites were measured for plants kept at high CO2 and after two and five day shift to ambient air; all values are measured in umol/ g fresh weight except hydroxypyruvate which is shown as arbitrary units; glycolate (A), glyoxylate (B), glycine (C), serine (D), hydroxypyruvate (E) and glycerate (F);

All amino acids were measured as well as the following organic acids: glycolate, hydroxypyruvate, glycerate, organic acids of the TCA cycle except for oxaloacetate. The five metabolites showing the biggest change are those involved in photorespiration (Figure 6-3). Glycerate and glycolate accumulate to high levels already under high CO2 conditions (Figure 6-3A and F). Glycolate accumulates eight fold and increases to sixteen fold when shifted to ambient air. Glycerate accumulates twenty fold initially and increases to forty fold when shifted. Three other metabolites, glycine, serine and hydroxypyruvate, are rather similar to wild type as long as plants are kept in high CO2 air but start to accumulate to higher levels than wild type once plants are shifted (Figure 6-3C, D and E). Glyoxylate is not significantly different (Figure 6-3B). The highest accumulation of metabolites is measured two days after the shift and metabolite concentrations start to decline at the five day time point. To gain a more dynamic picture of the metabolite accumulation pattern and to exclude that the accumulation of metabolites measured was the result of a pleiotropic effect, a detailed time course of plants grown under high CO2 conditions and of plants shifted to ambient air was measured. Both glycolate and glycerate accumulate in a light dependent pattern already under high CO2 conditions (Figure 6-4 A and E). Glycolate levels at night are comparable to wild type levels and rise in the course of the day to the highest point in the evening (Figure 6-4 A). Glycerate levels remain at a high level even during the night but rise during the day (Figure 6-4 E). As demonstrated in before, glycine, serine and hydroxypyruvate are similar to wild type levels when plants are incubated in high CO2. Serine is the only one of the photorespiratory metabolite which displays changes in wild

type upon shift to ambient air; the levels approximately double. Glycine, serine and hydroxypyruvate levels increase upon shifting the plants but hydroxypyruvate and serine do not display daylight dependent accumulation but continue to accumulate during the initial measurement during the night before they fall, and start accumulating again on the second day of ambient conditions. All plants show no visible phenotype during the experiment and only start to develop yellow leaf areas on day five of the shift to ambient air.

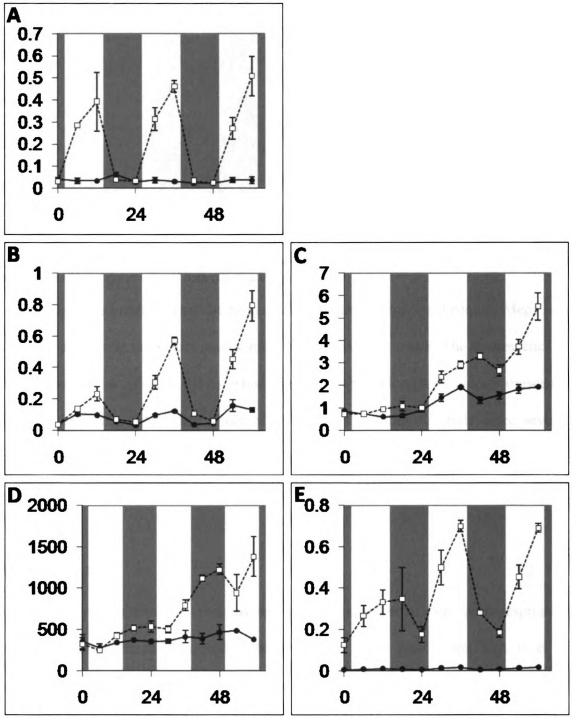


Figure 6-4: Time course of metabolite accumulation in the Mep1 insertion line compared to wild type upon shift from 3000ppm (until 24 hours) to ambient air (30 hours till end); all values are measured in umol/ g fresh weight except hydroxypyruvate which is shown as arbitrary units; glycolate (A), glycine (B), serine (C), hydroxypyruvate (D) and glycerate (E)

Mep1 is localized to the chloroplast envelope

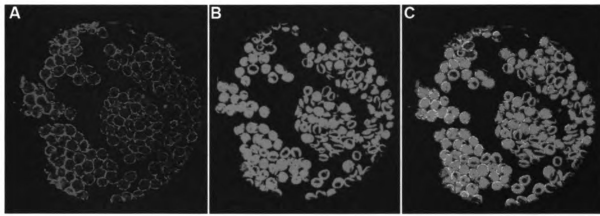


Figure 6-5: Mep1 is localized in the chloroplast envelope; fluorescence of GFP (A); chlorophyll auto-fluorescence (B) and the overlay of both (C); this image is in color

Mep1 was localized as a C-terminal GFP fusions protein. The fusion of Mep1 was expressed transiently in tobacco leaves under a 35S-CaMV promoter. The fluorescence of GFP forms circles (Figure 6-5A) which surround the chlorophyll auto-fluorescence (Figure 6-5B). The fluorescent circle of GFP is not a complete circle since several openings remain. Prolonged overexpression leads to accumulation of fluorescent protein in the cytosol.

Discussion

None of the transport proteins involved in transport of core photorespiratory metabolites has been characterized at the molecular level. Due to the high level of compartmentation, several transport processes are needed at the organellar membranes; at the chloroplast envelope glycolate has to be exported while glycerate is imported, the peroxisomal membrane is crossed by not only glycolate and glycerate but also glycine and serine, and the mitochondria need to import two molecules of glycine while only one

serine is exported (Ogren, 1984). We report the identification of a protein with twelve predicted transmembrane helices which is co-expressed over a wide range of conditions with other photorespiratory genes (Table 6-1). The correlation coefficient is similar to the ones of other photorespiratory genes except glycerate kinase which is not co-expressed at all. All other photorespiratory genes correlate well with each other and with Mep1. However, the photorespiratory genes are part of a larger cluster comprising many genes involved in photosynthesis (Horan et al., 2008) (Weber APM, unpublished). Co-expression analysis was successfully used to identify a candidate transport protein for photorespiratory metabolites. The experimentally determined expression pattern of Mep1 agrees well with expression data gathered in different Affymetrix experiments (Figure 6-1) (Schmid et al., 2005; Winter et al., 2007). The gene is exclusively expressed in green tissues with the strongest expression in leaves (Figure 6-1) (Schmid et al., 2005; Winter et al., 2007).

An insertion mutant line was isolated as a tool to analyze the metabolic consequences if the transport protein is absent from the cell. The visible phenotype is unlike that of known photorespiratory mutants which are frequently lethal when grown under ambient air (Somerville and Ogren, 1979b; Somerville and Ogren, 1979a; Somerville and Ogren, 1981). The visible phenotype of *mep1-1* was compared to the one of *shm1-1* which is impaired in the mitochondrial production of serine during photorespiration (Voll et al., 2006). *Shm1-1* yellows under ambient air but dies before developing bleached leaf regions. But when grown under high light conditions, *shm1-1* mimics the phenotype of *mep1-1*, thus demonstrating that the visible phenotype of photorespiratory mutants is variable with varying light conditions. Comparison of the

visible phenotype with the expression pattern of the gene showed that the bleached intercoastal regions do not correspond to the expression domains of the protein (Figure 6-1) indicating that the bleaching is a secondary effect of the mutation. After establishing that *mep1-1* is a likely photorespiratory mutant, a detailed phenotypic analysis was initiated. To avoid pleiotropic effects caused by the yellowing and bleaching of the leaf regions and the concomitant slower growth, plants were kept under non-inducing conditions and the steady state metabolite content was analyzed after two and five days of shift to ambient air. Although a photorespiratory phenotype was suspected, the metabolite analysis was unbiased covering all amino acids, all TCA cycle organic acids except oxaloacetate and the photorespiratory intermediates glycolate, hydroxypyruvate and glycerate, and the sugars glucose, fructose, sucrose and maltose. The suppression of pleiotropic phenotypes was successful. Of more than forty metabolites, major changes were recorded only for those metabolites which are intermediates of one pathway, the photorespiratory pathway (Figures 7-3 and 7-4). The detailed time course allowed resolving the accumulation kinetics for all metabolites and again, the biggest changes were recorded for photorespiratory intermediates (Figure 6-4). Notably, the biggest change is recorded in the two metabolites which have to cross the chloroplast envelope, the glycolate which needs to be exported from the chloroplast and the glycerate which needs to be re-imported from the cytosol to be converted to 3-phosphoglycerate. The analysis of transiently expressed protein::GFP fusions indicated that Mep1 is localized to the plastid envelope (Figure 6-5). This finding is supported by localization programs such as TargetP which assign a canonical target peptide for the chloroplast at the protein N-terminus and also by various chloroplast envelope proteome projects in which Mep1 is

identified as a chloroplast envelope protein (Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008). The other intermediates involved in photorespiration only accumulate after the flux through the photorespiratory pathway is increased by lowering the CO2 concentration which has been shown also for other photorespiratory mutants (Boldt et al., 2005). The steady state levels of metabolites measured do not reflect the flux through the pathway, they only indicated the pool size of a metabolite at a given time point. In absolute measure, the largest metabolite accumulation by far is in the glycerate pool during the day (Figures 7-3 and 7-4). The main pathway to metabolize photorespiratory glycerate in the leaf is the chloroplast localized glycerate kinase. The insertion mutant in glycerate kinase also shows glycerate accumulation in ambient air (Boldt et al., 2005). The time course of metabolite accumulation indicates that the decreasing pool size of glycerate is accompanied by increasing pool sizes of hydroxypyruvate and serine during the night in ambient air (Figure 6-4) which may indicate that glycerate is metabolized by extraplastidial enzymes such as hydroxypyruvate reductase and that the hydroxypyruvate pool is equilibrated with the serine pool by transamination reactions. Thus the increasing pool sizes of photorespiratory metabolites other than glycolate and glycerate may be caused by a block in glycerate metabolism.

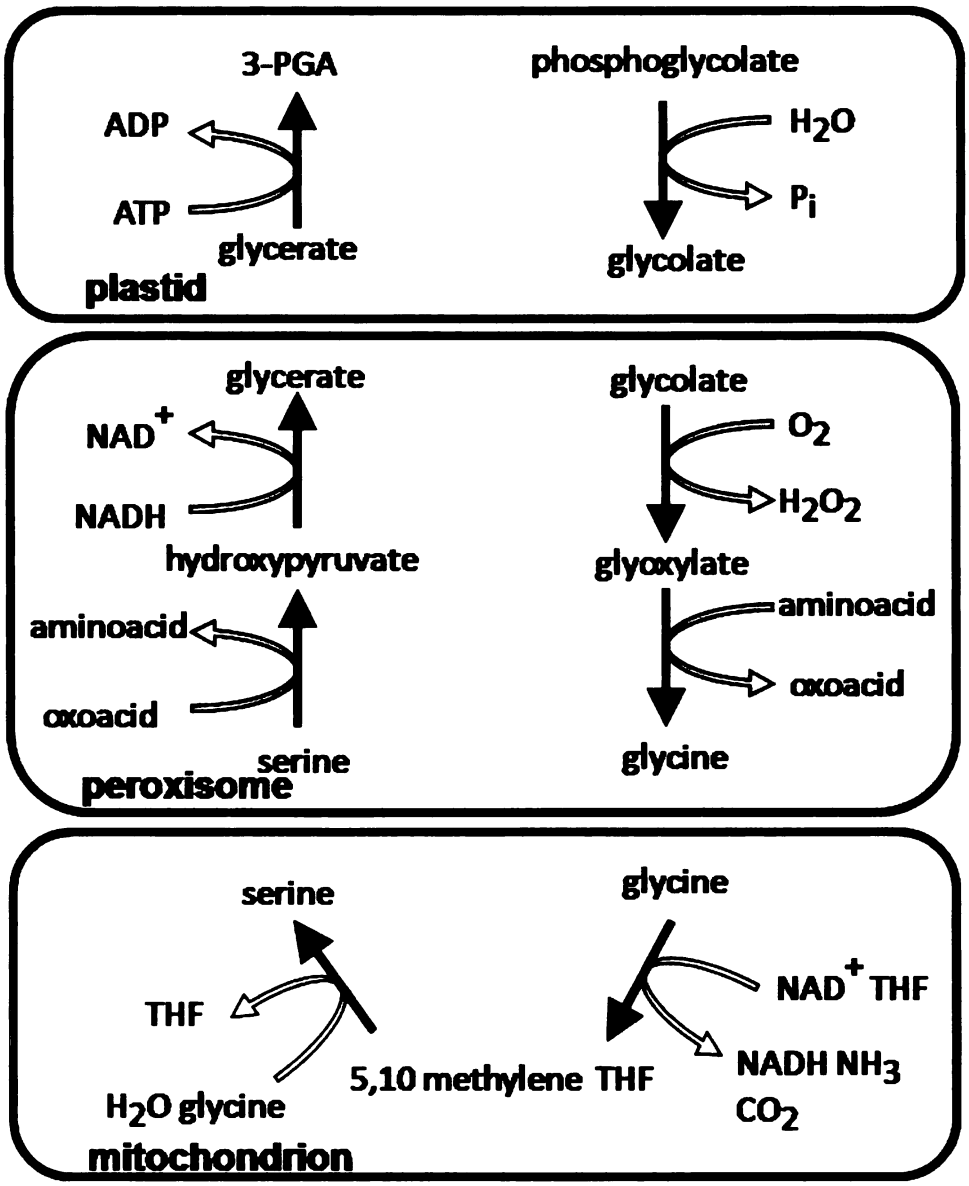


Figure 6-6: Schematic representation of photorespiration.

At first glance, it is counterintuitive why both the initial intermediate leaving the chloroplast, glycolate, and the final intermediate entering the chloroplast, glycerate, accumulate in the *mep1-1* mutant. However with a pathway as compartmented as photorespiration not only the linear pathway but also its compartmentation has to be considered. Mep1 is a chloroplast envelope localized protein which is only expressed in

green tissues. McCarty and coworkers described biochemically a protein which is capable of transporting glycolate and glycerate and which does not depend on stochiometric exchange but rather on a pH gradient across the envelope membrane (Howitz and McCarty, 1986; Howitz and McCarty, 1991; Young and McCarty, 1993). If this protein were mutated, both the export of glycolate as well as the import of glycerate would be disturbed. With many transport protein mutant, mild phenotypes are reported. Unlike enzymes, most transport proteins have comparatively broad substrate specificities but display high affinity for their in vivo substrates. However, if a mutation causes high metabolite accumulation, a transport protein with low affinity may now be able to transport a substrate it did not under wild type conditions. In addition, small acids are known to cross the envelope independently of transport proteins. Hence, glycolate accumulating in the chloroplast due to the absence of Mep1 may accumulate only to the level at which free diffusion or another transport protein moves it out of the chloroplast. Once glycolate is available in the cytosol, photorespiration can occur as in wild type until glycerate is produced in peroxisomes. It can no longer be imported to the chloroplast where the final enzyme involved in photorespiration in localized (Boldt et al., 2005). Unlike glycolate, glycerate accumulates to higher levels in the cell which only decline to a defined limit during the night before accumulating again once flux through photorespiration resumes. The overnight decline in glycerate is accompanied in a rise in its precursors, hydroxypyruvate and serine, which may reflect on the pathway which glycerate takes during the night. The combined evidence of expression pattern, protein localization and phenotype analysis points to Mep1 as the glycolate glycerate carrier of the chloroplast envelope. To conclusively proof its function, a biochemical

characterization is necessary. Four different systems to characterize the protein biochemically were tested. The availability of mutant and wild type plants enables a transport analysis on whole chloroplasts. Previous work in wheat and pea indicates that chloroplasts evolve oxygen if fed with glycerate since the 3-phosphoglycerate formed by glycerate kinase can be reduced to triosephosphate thus consuming reducing power and resupplying linear electron transfer with NADP (Heber et al., 1974). Unfortunately, the assay requires fine tuning of ATP, phosphate, and glycerate concentration supplied to the chloroplasts during the import assay (Heber et al., 1974). Isolation of metabolite import competent chloroplasts was successful as indicated by oxygen evolution of chloroplasts fed with 3-PGA but no conditions could be established where wild type chloroplasts evolved oxygen when fed with glycerate. It was also attempted to express the protein heterologously in different systems for reconstitution assays in artificial liposomes. Many transport proteins were characterized after expression in *Saccharomyces cerevisiae* (Loddenkotter et al., 1993; Fischer et al., 1994; Weber et al., 1995; Kammerer et al., 1998; Eicks et al., 2002). For Mep1 neither expression with a C-terminal nor an N-terminal histidine tag proved successful despite optimizing the translation initiation. Changing the 5' end of the construct as suggested in (Klepek et al., 2005) to improve RNA stability and/or translation initiation did not improve expression. Expression in a bacterial system was also attempted since membrane transport proteins have also been reconstituted successfully from bacterial protein extracts (Fiermonte et al., 2001; Picault et al., 2002; Palmieri et al., 2008). Both C-terminal and N-terminal fusion proteins which did or did not carry a T7 tag for improving expression did not lead to detectable amounts of protein. Expression conditions were varied for length of induction and growing

temperature without any success. Autoinduction has been reported as a possible strategy to overcome protein toxicity in heterologous expression (Studier, 2005). Autoinduction was tested instead of induction by IPTG without any success. Novel bacterial strains optimized for membrane protein production (Miroux and Walker, 1996) did not lead to protein expression either. Finally, an *in vitro* protein expression system was established based on protocols from (Schwarz et al. 2007). In this system, transcription is initiated from a T7 promoter by T7 polymerase and translation is carried out with an *Escherichia coli* extract. Amino acids and nucleotides are supplied to the transcription/translation assay by dialysis (Schwarz et al. 2007). This system has been used successfully to express membrane proteins which had proved recalcitrant to expression before (summarized in Schwarz et al., 2007). Three different membrane proteins were successfully expressed but expression of Mep1 with different tags for detection lead to no detectable protein expression and instead frequently caused the complete assay mixture to precipitate from solution. It has not been possible to analyze Mep1 biochemically since all attempts to produce protein for reconstitution remain unsuccessful. Possibly, the high hydrophobicity of Mep1, a protein with twelve transmembrane helices and virtually no hydrophilic loops causes the protein to aggregate in non-native expression systems. Plant cells may express chaperones to keep highly hydrophobic proteins in solution prior to insertion into the membranes.

References

Bent AF (2000) Arabidopsis in planta transformation. Uses, mechanisms, and prospects for transformation of other species Plant Physiol 124, 1540-7.

Boldt R, Edner C, Kolukisaoglu U, Hagemann M, Weckwerth W, Wienkoop S, Morgenthal K, Bauwe H (2005) D-GLYCERATE 3-KINASE, the last unknown enzyme in the photorespiratory cycle in Arabidopsis, belongs to a novel kinase family Plant Cell 17, 2413-2420

Bräutigam A, Hofmann-Benning S, Weber APM (2008) Comparative Proteomics of Chloroplast Envelopes from C3 and C4 Plants Reveals Specific Adaptations of the Plastid Envelope to C4 Photosynthesis and Candidate Proteins Required for Maintaining C4 Metabolite Fluxes Plant Physiol 148, 568-579

Collakova E, Goyer A, Naponelli V, Krassovskaya I, Gregory JF, Hanson AD, Shachar-Hill Y (2008) Arabidopsis 10-formyl tetrahydrofolate deformylases are essential for photorespiration Plant Cell 20, 1818-1832

Eicks M, Maurino V, Knappe S, Flugge UI, Fischer K (2002) The plastidic pentose phosphate translocator represents a link between the cytosolic and the plastidic pentose phosphate pathways in plants Plant Physiol 128, 512-522

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana Mol Cell Proteomics 2, 325-345

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2001) Metabolite profiling for plant functional genomics Nature biotechnol 19, 173-173

Fiermonte G, Dolce V, Palmieri L, Ventura M, Runswick MJ, Palmieri F, Walker JE (2001) Identification of the human mitochondrial oxodicarboxylate carrier - Bacterial expression, reconstitution functional characterization, tissue distribution, and chromosomal location J Biol Chem 276, 8225-8230

Fischer K, Arbinger B, Kammerer B, Busch C, Brink S, Wallmeier H, Sauer N, Eckerskorn C, Flugge UI (1994) Cloning and in-Vivo Expression of Functional Triose Phosphate/Phosphate Translocators from C-3-Plants and C-4-Plants - Evidence for the Putative Participation of Specific Amino-Acid-Residues in the Recognition of Phosphoenolpyruvate Plant J 5, 215-226

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW, Gage DA, Phinney BS (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis J Proteome Res 2, 413-25

Heber U, Kirk MR, Gimmler H, Schafer G (1974) Uptake And Reduction Of Glycerate By Isolated-Chloroplasts Planta 120, 31-46

Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF, Zhu JK, Cushman JC, Gollery M, Girke T (2008) Annotating genes of known and unknown function by large-scale coexpression analysis Plant Physiol 147, 41-57

Howitz KT, McCarty RE (1986) D-Glycerate Transport By The Pea Chloroplast Glycolate Carrier - Studies On [1-C-14] D-Glycerate Uptake And D-Glycerate Dependent O-2 Evolution Plant Physiol 80, 390-395

Howitz KT, McCarty RE (1991) Solubilization, Partial-Purification, And Reconstitution Of The Glycolate Glycerate Transporter From Chloroplast Inner Envelope Membranes Plant Physiol 96, 1060-1069

Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A, Flugge UI (1998) Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter Plant Cell 10, 105-17.

Kellogg EA, Juliano ND (1997) The structure and function of RuBisCO and their implications for systematic studies Am J Bot 84, 413-428

Klepek YS, Geiger D, Stadler R, Klebl F, Landouar-Arsivaud L, Lemoine R, Hedrich R, Sauer N (2005) Arabidopsis POLYOL TRANSPORTER5, a new member of the monosaccharide transporter-like superfamily, mediates H+-symport of numerous substrates, including myo-inositol, glycerol, and ribosele Plant Cell 17, 204-218

Knappe S, Lottgert T, Schneider A, Voll L, Flugge UI, Fischer K (2003) Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis-AtPPT1 may be involved in the provision of signals for correct mesophyll development Plant J 36, 411-420

Linka M, Weber APM (2005) Shuffling ammonia between mitochondria and plastids during photorespiration Trends Plant Sci 10, 461-465

Loddenkotter B, Kammerer B, Fischer K, Flugge UI (1993) Expression of the Functional Mature Chloroplast Triose Phosphate Translocator in Yeast Internal Membranes and Purification of the Histidine-Tagged Protein by a Single Metal-Affinity Chromatography Step Proc Natl Acad Sci U S A 90, 2155-2159

Lu Y, Savage LJ, Ajjawi I, Imre KM, Yoder DW, Benning C, DellaPenna D, Ohlrogge JB, Osteryoung KW, Weber AP, Wilkerson CG, Last RL (2008) New connections across pathways and cellular processes: Industrialized mutant screening reveals novel associations between diverse phenotypes in Arabidopsis Plant Physiol 146, 1482-1500

Miroux B, Walker JE (1996) Over-production of proteins in Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels J Mol Biol 260, 289-298

Ogren WL (1984) Photorespiration - Pathways, Regulation, And Modification Annu Rev Plant Physiol Plant Mol Biol 35, 415-442

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes Nucl Acids Res 33, 5691-5702

Palmieri L, Picault N, Arrigoni R, Besin E, Palmieri F, Hodges M (2008) Molecular identification of three Arabidopsis thaliana mitochondrial dicarboxylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization Biochem J 410, 621-629

Picault N, Palmieri L, Pisano I, Hodges M, Palmieri F (2002) Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria - Bacterial expression, reconstitution, functional characterization, and tissue distribution J Biol Chem 277, 24204-24211

Renne P, Dressen U, Hebbeker U, Hille D, Flugge UI, Westhoff P, Weber APM (2003) The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2 Plant J 35, 316-331

Reumann S, Weber APM (2006) Plant peroxisomes respire in the light: Some gaps of the photorespiratory C-2 cycle have become filled - Others remain Biochimica Et Biophysica Acta-Molecular Cell Research 1763, 1496-1510

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development Nature Genet 37, 501-506

Schneider G, Lindqvist Y, Branden CI (1992) Rubisco - Structure And Mechanism Annu Rev Biophys Biomolec Struct 21, 119-143

Schwacke R, Schneider A, van der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flugge UI, Kunze R (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins Plant Physiol 131, 16-26

Schwarz D, Klammt C, Koglin A, Lohr F, Schneider B, Dotsch V, Bernhard F (2007) Preparative scale cell-free expression systems: New tools for the large scale preparation of integral membrane proteins for functional and structural studies Methods 41, 355-369

Somerville CR, Ogren WL (1979a) Gas-Exchange Analysis of a Photosynthesis-Photorespiration Mutant of Arabidopsis-Thaliana Plant Physiol 63, 152-152

Somerville CR, Ogren WL (1979b) Phosphoglycolate Phosphatase-Deficient Mutant of Arabidopsis Nature 280, 833-836

Somerville CR, Ogren WL (1980) Photo-Respiration Mutants of Arabidopsis-Thaliana Deficient in Serine-Glyoxylate Aminotransferase Activity Proc Natl Acad Sci USA 77, 2684-2687

Somerville CR, Ogren WL (1981) Photorespiration-Deficient Mutants of Arabidopsis-Thaliana Lacking Mitochondrial Serine Transhydroxymethylase Activity Plant Physiol 67, 666-671

Somerville CR, Ogren WL (1982) Mutants of the Cruciferous Plant Arabidopsis-Thaliana Lacking Glycine Decarboxylase Activity Biochem J 202, 373-380

Somerville SC, Ogren WL (1983) An Arabidopsis-Thaliana Mutant Defective in Chloroplast Dicarboxylate Transport Proc Natl Acad Sci USA 80, 1290-1294

Steinhauser D, Usadel B, Luedemann A, Thimm O, Kopka J (2004) CSB.DB: a comprehensive systems-biology database Bioinformatics 20, 3647-3651

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures Protein Expr Purif 41, 207-234

Voll LM, Jamai A, Renne P, Voll H, McClung CR, Weber APM (2006) The Photorespiratory Arabidopsis shm1 Mutant Is Deficient in SHM1
Plant Physiol 140, 59-66

Waadt R, Kudla J (2007) Cold Spring Harbor Protocols

Weber A, Menzlaff E, Arbinger B, Gutensohn M, Eckerskorn C, Flugge UI (1995) The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells Biochemistry 34, 2621-7.

Weber APM (2004) Solute transporters as connecting elements between cytosol and plastid stroma Curr Opin Plant Biol 7, 247-253

Wingler A, Lea PJ, Quick WP, Leegood RC (2000) Photorespiration: metabolic pathways and their role in stress protection Philos Trans R Soc Lond, B, Biol Sci 355, 1517-1529

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An Electronic Fluorescent Pictograph Browser for Exploring and Analyzing Large-Scale Biological Data Sets PLoS ONE 2, e718

Young XK, McCarty RE (1993) Assay Of Proton-Coupled Glycolate And D-Glycerate Transport Into Chloroplast Inner Envelope Membrane-Vesicles By Stopped-Flow Fluorescence Plant Physiol 101, 793-799

Chapter 7

Dynamics of the membrane systems surrounding the chloroplast stroma revealed by proteomics and GFP fusion proteins

Andrea Bräutigam[1,2], Frederique K. H. Breuers[1], Henrik Tjellström[5], John Froehlich[4], John Ohlrogge[3], Anna Stina Sandelius[5] and Andreas P. M. Weber[1,3]

I would like to acknowledge Henrik Tjellström and Anna Stina Sandelius for preparation of the plastid associated membrane and ER enriched microsome proteome samples. I would also like to acknowledge Frederique K. H. Breuers for cloning and infiltrating the constructs used for protein localization by transient expression tobacco plants.

Abstract

The soluble stroma of plastids is surrounded by a two membrane systems, the outer and the inner envelope. Traditionally, the plastid has been considered completely separated from the remainder of the cell but recent evidence has pointed to tight connections between the chloroplasts and the remainder of the cell. We have developed *Pisum sativum* as a model system for organellar proteomics which now allows us to isolate and compare not only the inner and outer envelopes, but also the newly discovered plastid associated membranes (PLAMs) and a light microsome fraction representing the ER. Comparative proteomics showed expected distribution of marker proteins and also allowed identification of putative markers for PLAMs and revealed remarkable dynamics of outer envelope proteins. The comparative proteome data constitutes a new resource for differential localization data for proteins of the chloroplast envelope. Transient expression of GFP fusions of the candidate PLAM proteins supported PLAM localization for two membrane bound and one soluble protein. Analysis of GFP fusions of outer envelope proteins over time showed induction of stromule and vesicle formation with increasing protein load on the system only for outer envelope proteins but not for inner envelope proteins thereby identifying a tool for systematic analysis of stromule formation and function. No proteomic evidence for a role of PLAMs in lipid transfer between the ER and the chloroplasts was found. However, the data led to a new hypothesis that PLAMs may represent budding autophagosomes at the tip of stromules and at the chloroplast envelope.

Introduction

Chloroplasts are the defining organelles of the Archaeplastida (Adl et al., 2005). Frequently, the chloroplast has been viewed as a distinct entity, separated from the remainder of the cell by the membrane system surrounding the chloroplast, the inner and the outer envelope membranes. The inner envelope is the true barrier to solute import and metabolite and ion exchange which are mediated by a number of specific transport proteins (Weber, 2004; Weber et al., 2005; Weber and Fischer, 2007). The outer envelope is relatively permeable to small solutes by way of a number of porins with a broad metabolite transport spectrum, but impermeable to molecules larger than 10kDa such as most proteins (Weber, 2004; Weber et al., 2005; Weber and Fischer, 2007). Most proteins are imported through a multi-protein complex spanning both the outer and the inner envelope (Toc and Tic complex). Cargo proteins are addressed to a receptor at the outer envelope by a chloroplast target peptide (cTP) and are transferred to the stroma through two protein channels, Toc75 in the outer and Tic110 in the inner envelope (Jarvis and Robinson, 2004; Soll and Schleiff, 2004; Kessler and Schnell, 2006). Membrane proteins of the inner envelope are also imported through this system and inserted either during or after transit (Li and Schnell, 2006; Tripp et al., 2007). In addition to plastid envelope proteomics (Bräutigam et al., 2008a; Bräutigam et al., 2008b) and detailed targeting analysis (Villarejo et al., 2005; Nanjo et al., 2006) have revealed a number of chloroplast resident proteins which do not possess a canonical cTP and enter the plastid by other ways. For example, two soluble enzymes, a carbonic anhydrase and a nucleotide pyrophosphatase reach the chloroplast stroma through the endomembrane system (Villarejo et al., 2005; Nanjo et al., 2006), but the targeting from the Golgi compartment

to the chloroplast remains unresolved. Two outer envelope proteins, OEP64 (formerly called Toc64) and OEP7, are both recognized by a protein receptor, AKR2A, and inserted into the envelope independent of the Toc and Tic complex proteins (Bae et al., 2008). The binding motif, hydrophobic amino acids arranged in an alpha helix followed by a stretch of charged amino acids (Lee et al., 2001; Bae et al., 2008), is similar to the canonical target peptide for the secretory pathway which also consists of a stretch of hydrophobic amino acids (Fourrier et al., 2008). Only two cargo proteins for this pathway are known to date, hence no prediction programs are available for the binding motif.

Although the envelope membranes of plastids have often been viewed as static barriers, they are indeed a remarkably dynamic system. The envelopes can form long structures protruding from the body of the chloroplast (Gray et al., 2001). These 'stromules' are lined by both the inner and the outer envelope and contain stroma but no thylakoids (Gray et al., 2001; Hanson and Sattarzadeh, 2008). They can extend and retract and even shed their tip, resulting in a stroma filled vesicle (Hanson and Sattarzadeh, 2008). Their function is currently not known (Hanson and Sattarzadeh, 2008). The envelope can form vesicles both to the inside as well as to the outside. Plant viruses can induce the formation of invaginations and eventually vesicles at the envelope which harbor the virus replication machinery (Hatta et al., 1973; Prod'homme et al., 2003). Vesicles towards the cytosol are formed when chloroplast proteins are destined for degradation in the vacuole (Ishida et al., 2008) but the mechanism underlying the formation of these vesicles is not understood. Finally, advanced microscopy techniques have revealed that a fraction of the ER is in close contact with the chloroplasts and that even during chloroplast isolation a fraction of the ER remains firmly attached to the

chloroplasts (Andersson et al., 2007). This fraction has been named plastid associated membranes and it has been hypothesized that this may represent a contact site between the endomembrane system and the plastids (Benning et al., 2006; Andersson et al., 2007). The production of fatty acids and lipids is dependent on a close and efficient interaction of two different compartments, the ER and the chloroplast (Benning et al., 2006). Essentially all acyl chains are produced in plastids and in most plants 85% or more of these acyl chains must cross the envelope before assembly into glycerolipids. Furthermore, glycerolipids can be exchanged between both compartments in both directions (Schnurr et al., 2002; Xu et al., 2005; Awai et al., 2006; Benning et al., 2006; Lu et al., 2007) and indeed in terms of lipid traffic, the plastid envelope may be the most active membrane found in biology.

A protocol for separation of outer and inner envelope membranes has been developed for leaf chloroplasts of *Pisum sativum* (Cline et al., 1981). The isolation of outer envelope membranes of suitable quality for proteomics requires large amounts of leaf material making the adaptation of the protocol for the model plant Arabidopsis impractical. In this work, we made use of a novel transcriptome database for *Pisum sativum* specifically developed to enable organellar proteomics (Bräutigam et al., 2008b).The different membrane systems surrounding the chloroplast, the inner and outer envelope, plastid associated membranes (PLAMs) and an ER enriched microsome fraction, were defined by comparative proteomics and the results were integrated with bioinformatic targeting information. Both the outer envelope membranes as well as the plastid associated membranes are considerably less static than previously assumed. Intense protein dynamics through the membrane systems are confirmed by GFP fusion

proteins. Finally, the distribution of proteins involved in fatty acid and membrane lipid biosynthesis as well as novel candidate transport proteins are discussed.

Material and Methods

Preparation and analysis of proteome samples

Four different proteome samples from 10-14 days old pea plants were isolated. Chloroplasts from pea leaves were isolated as described previously (Douce and Joyard, 1979; Keegstra and Yousif, 1986). Inner and outer envelopes were separated as described by (Cline et al. 1981), plastid associated membranes and ER enriched microsomes were isolated as described by (Andersson et al., 2007).Membrane samples were mixed with SDS–PAGE loading buffer and proteins were dissolved by shaking at 15° C for 20 minutes. Proteins were separated by SDS-PAGE and each gel lane was cut into ten equally-sized slices. The proteins of each slice were digested with trypsin and modified according to (Shevchenko et al., 1996) and loaded automatically onto a Waters Symmetry C18 peptide trap (5 μm, 180 μm x 20mm) at a flow rate of 4 μL/min in 2% Acetonitrile/0.1%Formic Acid for 5 minutes by a Waters nanoAcquity Sample Manager. The peptides were separated on a Waters BEH C18 nanoAcquity column (1.7 μm, 100 μm x 100mm) with a 90 minute gradient using a Waters nanoAcquity UPLC and eluted into a ThermoElectron LTQ-FTICR mass spectrometer with a flow rate of 300 nL/min (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 99.9% Acetonitrile/0.1% Formic Acid: gradient of 5% B to 40% B from 0 to 63 minutes, 40% B to 90% B from 63 to 71 minutes and 5% B from 71 to 90 minutes). Survey scans at a resolution of 50,000 were used to identify the top ten ions of each survey scan which were then subjected to

automated low energy collision induced dissociation. A BioWorks Browser v3.2 converted the MS/MS spectra to peaks lists which were compared to a *Pisum sativum* cDNA database using the Mascot searching algorithm, v 2.2 (www.matrixscience.com). Carbamidomethyl cysteine was set as a fixed peptide modification and oxidation of methionine was allowed. Up to two missed tryptic sites were allowed. The peptide tolerance was set to +/-10ppm and the MS/MS tolerance to 0.8kDa. Mascot results were imported into Scaffold® for analysis.

Data analysis

Protein identifications were called if at least two peptides were identified of which each had to be identified itself with 95% probability and the total protein identification probability was 99%. Identifications of the mixed envelope sample were imported from PRIDE (Bräutigam et al., 2008a) and added into the analysis. For all matches, the closest Arabidopsis Blast Hit was retrieved and both annotation and categorization is based on the Arabidopsis protein. Targeting information, structural analysis and prediction of transmembrane helices were retrieved with the bulk data retrieval tool at TAIR (www.arabidopsis.org). A protein was assigned to a membrane system based on its absolute spectral count observed in the proteome samples. Having twice as many counts in one or two categories over the others merited designation with three different confidence classes [class 1: ratio >twofold, absolute count >10; class 2: ratio >twofold, absolute count >5; 3 trend] with class 4 referring to identification in three or more membrane systems.

GFP fusion protein construction and analysis

For selected proteins, the localization was tested using C-terminal GFP fusion constructs. The cDNA of interest was amplified from RNA isolated from different developmental stages of Arabidopsis (young seedling, adult leaves or senescing leaves), using oligonucleotide primers compatible with the GATEWAY BP reaction system. The forward primer contained an optimized KOZAK consensus to ensure efficient translation. After BP reaction, the vectors were isolated and the insert was mobilized into pMDC83 ((Curtis and Grossniklaus, 2003). GFP was observed after transient expression in tobacco leaves (Waadt and Kudla, 2006) and isolation of protoplasts after one, two, three and four days of leaf infiltration with a Pascal confocal microscope. GFP fluorescence was isolated with a bandpass filter BP505-550nm; chloroplast fluorescence was filtered with a longpass filter LP650 and imaged after excitation with an argon laser at 458 and 488nm with 5% intensity. Pictures represent a single optical slice.

Results

The overall proteome analysis reveals a distinct protein localization pattern.

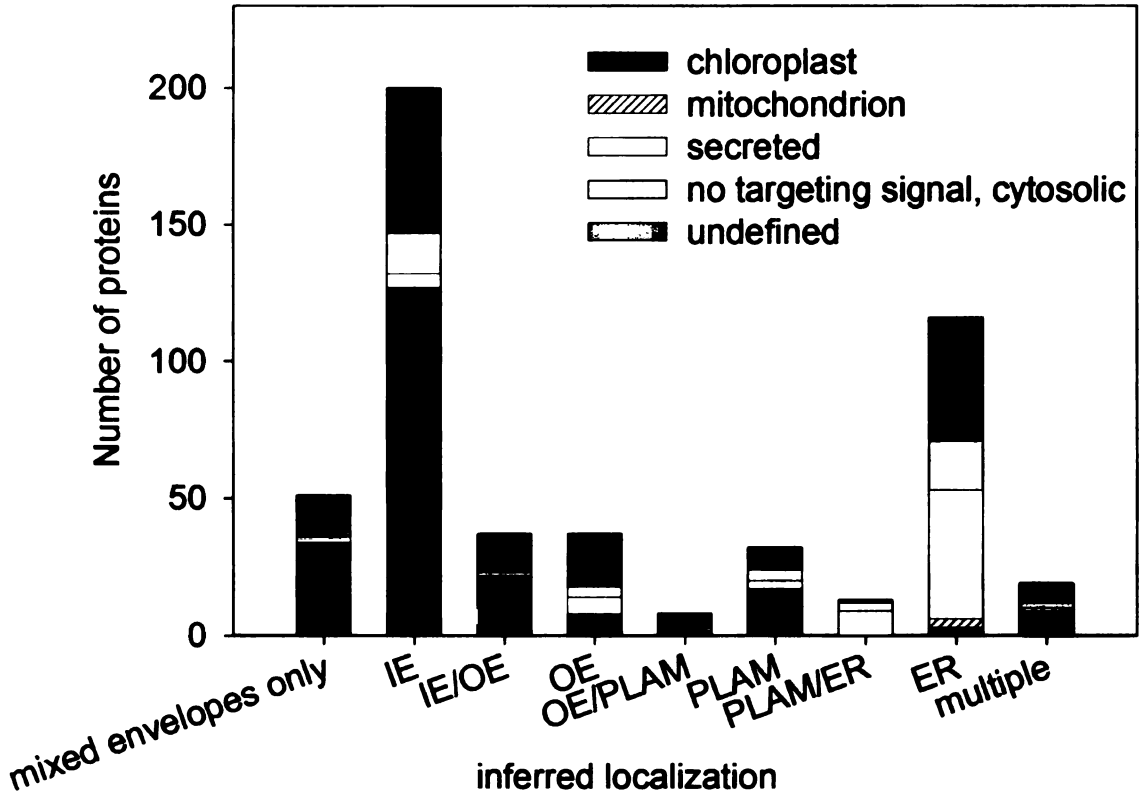


Figure 7-1: Protein distribution across the fractions plotted as number of proteins determined for each membrane system including targeting predictions from TargetP; IE inner envelope, OE outer envelope, PLAM plastid associated membrane; multiple localization cannot be determined because the protein distributes evenly;

Proteomics identified a total of 522 proteins with at least two peptides each (peptide probability, $p<0.05$; protein probability $p<0.01$) (Table 7-S1). All but six could be mapped onto an ortholog in the Arabidopsis proteome (Table 7-S1). The localization of each protein was inferred by comparing its absolute spectral abundance in each sample (for details see material and methods). Not all proteins could be assigned to one of the samples. A number of proteins were equally abundant in two adjacent membrane systems

(i.e., some of the IE/OE localized proteins could not unequivocally be sorted into either of the categories but were identified in both samples). Twenty proteins were identified with similar abundance in all samples. The number of protein identifications was not evenly distributed. Both the inner envelope and the ER enriched microsome fraction harbored more proteins than the outer envelope and the PLAM samples (Figure 7-1) with 203 proteins in the inner envelope and 117 proteins in the ER-enriched microsomes. For each of the proteins the localization was predicted from its sequence by TargetP and plotted onto the overall protein distribution (Figure 7-1). Many of the proteins only identified in mixed envelopes (Figure 7-1) have canonical target peptides for the Tic and Toc complex, as do most of the proteins categorized to the inner envelope fraction, and are identified as chloroplast proteins by prediction programs (Emanuelsson et al., 1999; Emanuelsson et al., 2000). In these categories, very few are predicted to be targeted to the endomembrane system or the cytosol. 26% and 28%, respectively, cannot be sorted by bioinformatics at all and are classified as undefined. These proteins present conflicting evidence for targeting at their N-terminus and can therefore not be predicted by the program (Figure 7-1). The 39 proteins assigned to the outer envelope fraction present a very different prediction pattern. Proteins appear to have very different N-termini most of which cannot be predicted by the program, resulting in an 'undefined' prediction. The 32 proteins of the PLAM fraction present a picture rather like the inner envelope proteins. The second category with well defined bioinformatics predictions is the fraction of proteins residing in the ER based on their selective enrichment in the ER-enriched microsome fraction. This category contains most of the proteins also predicted to be targeted to the secretory pathway though over half of the proteins cannot be assigned to a

location by the prediction program. In addition to a few proteins predicted to enter the chloroplasts, some of the proteins are predicted to localize to the mitochondria. Many of these are proteins that have already been localized to the mitochondria by various methods indicating that the ER-enriched microsomal fraction but none of the others also contains *bona fide* mitochondrial proteins (Table 7-S1).

Known residents of the inner and outer envelope

To analyze whether the preparation of the proteome samples had succeeded in enriching the distinct fractions, the protein import complex (Tic and Toc complex) was analyzed. It has been studied in great detail in the past three decades and the localization of the resident proteins has been well established. The outer envelope components are denoted by Toc for translocon outer envelope of chloroplasts and a number representing their molecular weight at first analysis and the inner envelope components are denoted as Tic for translocon inner envelope of chloroplasts and a number. All proteins residing in the outer envelope based on prior knowledge have their highest spectral abundance in the outer envelope proteome sample (Figure 7-2).

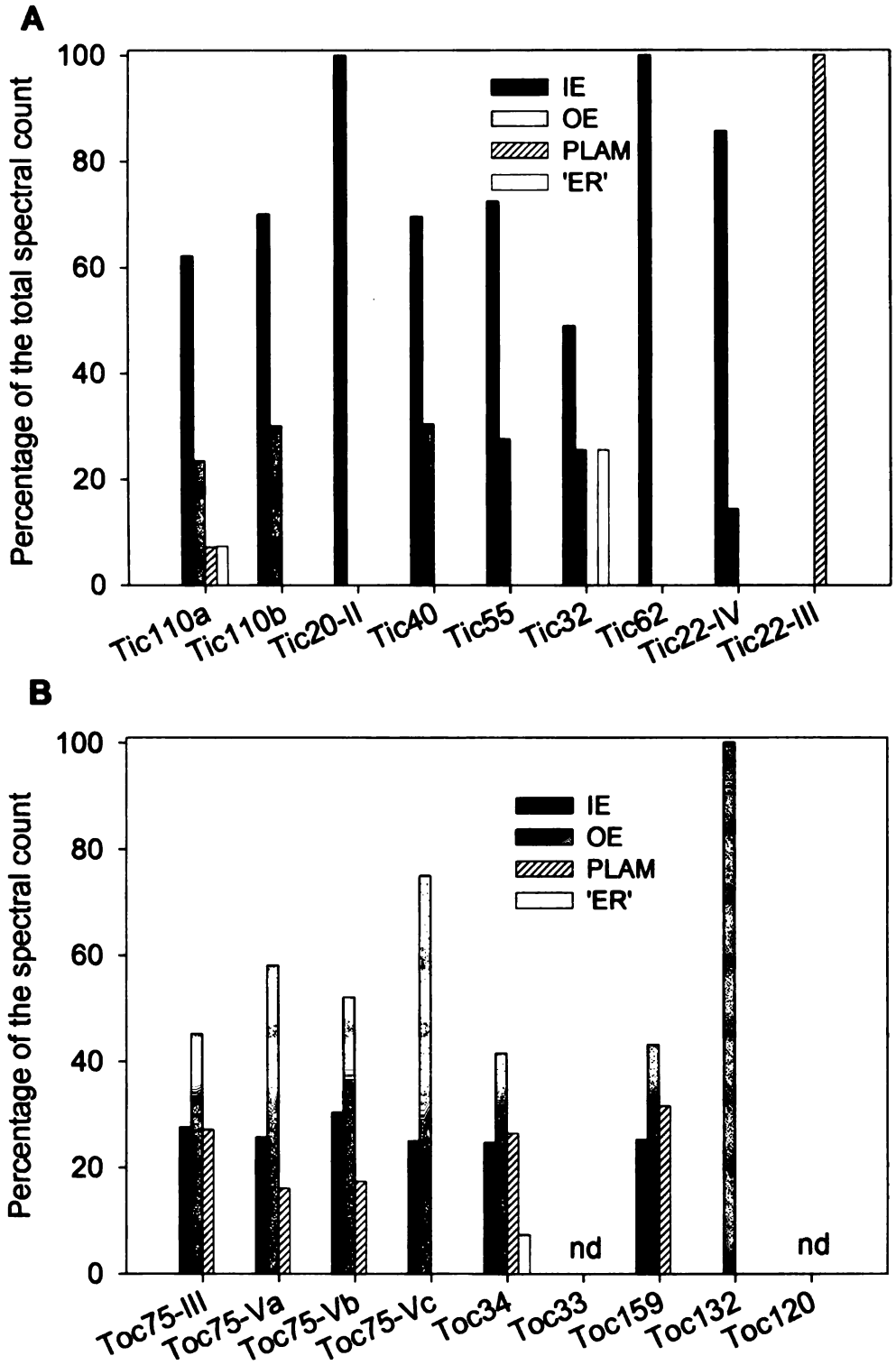


Figure 7-2: Distribution of the spectral count between the proteome samples for proteins of the protein import complex; (A) Proteins of the inner envelope complex; (B) proteins of the outer envelope complex; IE inner envelope sample, OE outer envelope sample, PLAM plastid associated membranes sample, 'ER' light microsome fraction, nd not detected

They include the import receptor Toc159, the two Toc75 isoforms one of which is the protein import channel through the outer envelope. Both Toc64 and ToC34 are also most abundant in the outer envelope sample. The alternative import receptor Toc132 not only localized to the outer envelope but has been identified in a chloroplast proteome sample for the first time. Virtually all of the Tocs can also be found in the PLAM proteome sample with high spectral abundance but not in the ER enriched microsome fraction. In contrast, the inner envelope components distribute mostly to the inner envelope proteome sample. Tic110, the presumed channel through the inner envelope, as well as Tic 40 and Tic55 are mainly in the inner envelope (Figure 7-2). One Tic22 isoform presumably localized to the intermembrane space, but associated with the transport complex at the inner envelope is also identified in the inner envelope fraction but the other Tic22 isoform is exclusively localized in the PLAM fraction (Figure 7-2). In addition to the proteins of the import complex, known and putative transport proteins and porins were checked. Members of the mitochondrial carrier family, the putative and known metabolite transport proteins, and most ATP dependent transport proteins localized to the inner envelope (Table 7-S1). Only WBC7, an ATP dependent transport protein, localized to the outer envelope. All the porins localize to the outer envelope (OEP37, OEP24) or to the PLAMs (OEP24) but a protein related to OEP16 localized to the inner envelope.

Proteins of the outer envelope and the plastid associated membranes

Thirty-seven proteins were assigned to the outer envelope, 8 were assigned to both, the outer envelope and the PLAMs, 32 localized to the PLAM, and 13 were assigned to both the PLAM and the ER (Figure 7-1). The PLAM fraction contains the

largest number of proteins that are predicted to be soluble whereas the fraction of proteins shared between the PLAMs and the ER contains the largest proportion of membrane protein (Table 7-S1). In the outer envelope, the proteins with the highest spectral count are those involved in transport. All known proteins of the Toc complex except Toc120 and all known outer envelope porins for metabolite transport were identified. The only putative transport protein localized to the outer envelope is the ABC transporter WBC7. In addition there are a number of proteins with putative functions in fatty acid and lipid metabolism as well as proteins containing predicted protein-protein interaction domains. Proteins with a metabolic role other than lipid metabolism include an ascorbate peroxidase, a glysosyl hydrolase whose target is unknown, hexokinase and a number of putative enzymes (Table 7-S2). One protein of unknown function consists of only beta strands (Table 7-S1), but overall no trend towards preferential accumulation of beta-fold proteins was identified in either of the outer envelope associated fractions (Table 7-S1). After it became apparent that many fractions also contain known stroma proteins, the known localizations of proteins (endomembrane system, envelopes, stroma, localization not classified and protein of unknown function, see Table 7-1) were determined. Whereas the inner and the outer envelope contain less than ten and less than fourteen percent of stromal proteins, the fraction of proteins seemingly divided between these two compartments contains mostly stromal proteins and proteins of unknown function (Table 7-1).

Table 7-1: Known localizations of proteins assigned to the different fractions; localization was not classified if it is not known or known to be outside the three criteria (e.g. mitochondrial for proteins assigned to the ER fraction)

	percentage of proteins in each fraction						
classical localization	IE	IE/OE	OE	OE/PLAM	PLAM	PLAM/ER	ER
endomembrane system	0.5	0.0	2.8	12.5	6.5	69.2	39.6
envelopes	16.8	8.1	25.0	37.5	9.7	0.0	1.1
stroma	9.9	35.1	13.9	37.5	41.9	0.0	0.0
localization not classified	38.1	32.4	27.8	0.0	16.1	15.4	20.9
protein of unknown function	34.7	24.3	30.6	12.5	25.8	15.4	38.5

Seven proteins are shared between the OE and the PLAM samples, among them four enzymes (Table 7-S3). There is a glucosidase known to localize to the OE (Fourrier et al., 2008), a putative transaldolase, a monodehydroascorbate reductase known to localize to peroxisomes (Lisenbee et al., 2005; Eastmond, 2007), and a putative subunit of the pyruvate dehydrogenase complex. In addition to the enzymes there are also two isoforms of the outer envelope porin OEP16 (Pohlmeyer et al., 1997) and a protein of unknown function carrying a prefoldin domain.

The proteins mainly localized in the PLAMs consist mostly of known stromal proteins and proteins of unknown function (Figure 7-3). There are plastidic ribosomal proteins (eight) and proteins of unknown function (eight). In addition to proteins involved in protein biosynthesis there are also three proteins involved in protein folding, two Hsp60-like proteins and a chaperonin. Three putative enzymes, two different pyruvate dehydrogenase complex subunits and a dephospho-CoA kinase also localize exclusively to the PLAMs. The mitochondrial import protein Tom-20 is part of this group as is one of the Tic22 isoforms identified. Finally, there is a group of proteins with diverse functions,

the outer envelope porin 24, a putative steroid binding protein, Chup1, a protein involved in chloroplast positioning, a subunit of plasmamembrane ATPase, cytochrome b5, two ferritins, and a eukaryotic ribosomal subunit. Many proteins of the PLAM fraction have canonical cTPs (Figure 7-1).

The third group of proteins is shared between the ER-enriched microsomal fraction and the PLAMs. Notably, both stromal and envelope proteins are absent from this fraction (Table 7-2). There are four ATPase subunits, three proteins involved in protein folding, and two proteins of unknown function. A VDAC-like porin, a strictosidine synthase-like protein and another protein involved in red light signaling are also present in this group. Compared to the proteins assigned to the outer envelope (14 predicted to have transmembrane helices, 23 predicted to be soluble), PLAMS are enriched in soluble proteins (27 to five with predicted transmembrane helices). PLAMs contain a mixture of proteins traditionally considered plastid like proteins and of proteins traditionally considered ER proteins as well as a number of proteins of unknown function (Figure 7-1, Table 7-1).

Within the ER fraction, known residents of the endomembrane system, in particular proteins involved in protein folding, are identified. Some of them are exclusively identified in the ER- enriched microsome fraction whereas most are shared with the PLAMs. Only one of the proteins, a calnexin is also identified in the chloroplast envelopes.

GFP localization data

The dynamic localization of a number of proteins localized to the outer envelope or the PLAMs was investigated by fusion to GFP and transient expression in tobacco

leaves. The transient expression driven by the cauliflower mosaic virus 35S promoter allowed monitoring of the cellular distribution of the GFP signal with increasing protein load on the system with time and hence a dynamic distribution of fusion proteins.

The outer envelope protein Toc64 known to be an outer envelope protein served as the control for the outer envelope; likewise Mep1 as the inner envelope control. MDAR (peroxisome), TGD4 (ER) and hexokinase 1 (mitochondrion) were included since they have been reported to localize to compartments other than the chloroplast by previous publications (Damari-Weissler et al., 2007; Eastmond, 2007; Xu et al., 2008). The Tic22 isoform, the EF hand protein, and the kinesin-like protein were included to determine whether they may serve as markers for the PLAMs in future. Two heatshock proteins were evaluated since they are expected to be soluble but were identified in the membrane samples and six proteins of unknown function were tested for their localization.

Table 7-2: GFP localization results for selected outer envelope and PLAM proteins; AGI Arabidopsis genome identifier, cTP chloroplast target peptide, helix+charged targeting motiv of AKR2A mediated protein import present,

AGI	Targeting Signals	Description	Initial localization	Localization after prolonged expression	Localization Class based on pattern	Inferred localization based on proteome analysis
At1g32080	cTP	inner envelope protein	Ring	Ring and Cytosol	Class I	IE
At3g17970	helix+ charged	OEP 64	Ring	Ring and Stromules or Vesicles	Class II	OE
At5g17670	cTP	putative lipase	Vesicles	nd	Class III	OE
At1g64850	no signals	protein with EF domains	Vesicles+Ring	Ring and Stromules	Class III	PLAM
At5g64816	helix+ charged	protein of unknown function	Vesicles+Ring	Ring and Stromules or Vesicles	Class III	OE
At1g77590	helix+ charged	fatty acid transporter	ring+jackets	Ring and Stromules	Class IV	OE
At2g01320	helix+ charged	unknown transporter	ring+jackets	Ring and Stromules	Class IV	OE
At2g32240	none detected	protein with ATPase domain	Cytosol	Cytosol	Class V	multiple
At4g29130	helix+ charged	hexokinase 1	Mitochondria	Mitochondria	Class V	OE
At1g20330	helix+ charged	sterol methyltransferase	spots	spots	Class V	PLAM/ER
At3g27820	helix	monodehydroascorbate reductase	spots	spots, tight chloroplast association	Class V	OE/PLAM
At3g13470	cTP	Hsp60-like protein	nd	budding vesicles	nd	PLAM
At3g06960	none detected	TGD4	nd	circle and spots	nd	IE
At2g44640	weak cTP	TGD4-like	nd	circle	nd	OE
At3g23710	cTP	Tic22-III	nd	spots	nd	PLAM
At1g20560	helix+ charged	ATP dependent synthetase	nd	circle	nd	PLAM

The fluorescence pattern of fifteen fusion proteins showed four distinctive chloroplast fluorescence patterns when expressed transiently in tobacco leaf cells for 1-4 days. Class I proteins initially localized in a circle surrounding the chlorophyll auto-fluorescence (Figure 7-3A) and prolonged overexpression of the protein results in unspecific accumulation in the cytosol (Figure 7-3B). This pattern is displayed by the inner envelope control protein, Mep1 (Table 7-2). Proteins of class II initially display a ring shaped fluorescence surrounding the auto-fluorescence of chlorophyll (Figure 7-3C). OEP64 is the sole member of this class (Table 7-2). Its fluorescence pattern eventually displays extensive stromule formation (Figure 7-3D). Frequently, instead of stromules, an abundance of vesicles of 1-1.5um diameter is visible (Figure 7-3D). Proteins of the third class label a vesicle membrane with distinct openings with a weaker label in a circle surrounding the chlorophyll auto-fluorescence

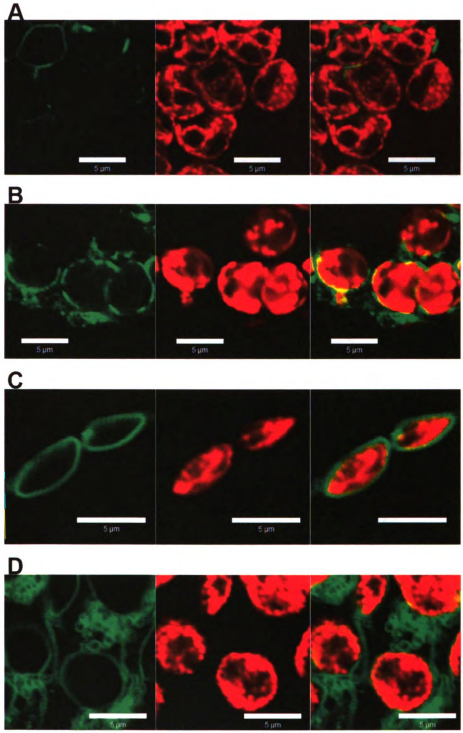


Figure 7-3: GFP localization of fusion proteins after initial (A, C, E and G) and prolonged (B, D, F, H) transient expression in tobacco leaves; bar size is 5um for panels A to G, for panel H 2um; for each picture: left panel chlorophyll auto-fluorescence, middle panel GFP fluorescence and right panel overlay; localization patterns of the inner envelope marker Mep1 (A, B), the outer envelope marker OEP64 (C, D), a protein of unknown function with EF hand domains (E, F) and an ABC transporter of unknown function, WBC7 (G, H); this image is in color

Figure 7-3 (cont'd)

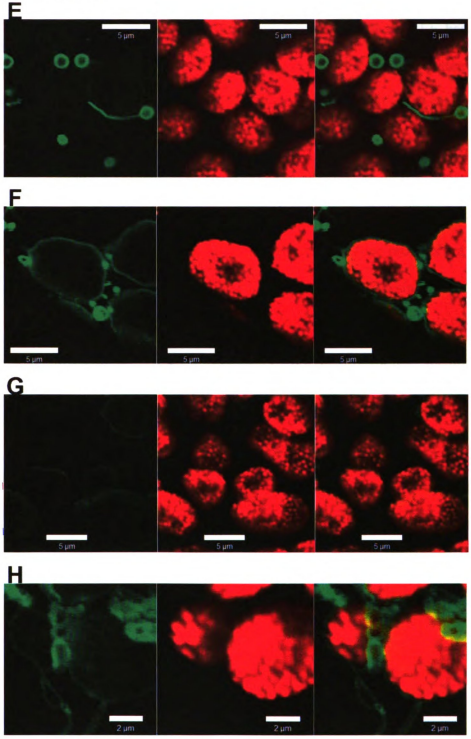


Figure 7-3E). Prolonged expression will result in stronger labeling of the circle and eventually formation of stromules (Figure 7-3F). A protein with EF hands, a putative lipase and a small protein without known domains display this pattern (Table 7-3). Proteins in class IV also label initially a circle around the chlorophyll auto-fluorescence but this circle frequently has loops (Figure 7-3G). Prolonged expression will result in stromule formation (Figure 7-3H) and occasionally vesiculation (as visualized for the Class II protein OEP64, Figure 7-3D, lower panel). The formation of stromules and extensive vesicles are observed for all putative outer envelope proteins but not for Mep1. Class V encompasses all extraplastidial localizations. The kinesin-like protein localized to the cytosol, the hexokinase localized to small dots of a fixed size whereas the monodehydroascorbate reductase and the sterolmethyltransferase localize to spots of varying size.

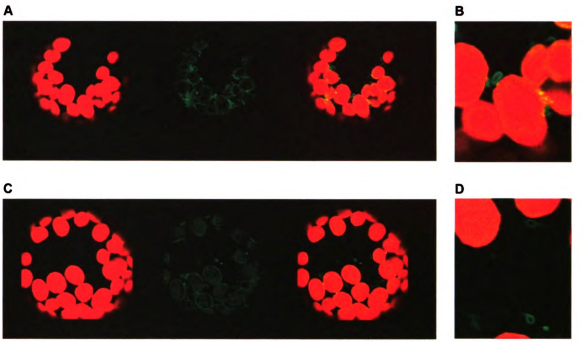


Figure 7-4: GFP expression pattern of Hsp60-like protein after three days of expression in tobacco leaves; (A) GFP fluorescence is strongest in extension of the chloroplast envelopes and vesicles, (B) detail of A, vesicle forming at the envelope, (C) GFP expression is visible in a vesicle forming at the tip of a stromule, (D) detail of (C); this image is in color

Transiently expressed GFP fusion proteins of Hsp60-like co-localizes with chlorophyll auto-fluorescence but is particularly visible in vesicles attached to the chloroplast envelope (Figure 7-4A and B) and forming at the end of stromules (Figure 7-4C and D).

Discussion

Targeting Predictions for Envelope proteins

To understand the network of proteins residing in the membrane system surrounding the plastids, the inner envelope, the outer envelope, and the PLAMs, we analyzed four separate proteome samples, one preparation enriched in inner envelopes, one enriched in outer envelopes, an acid wash of intact chloroplasts to obtain a PLAM

sample, and an ER-enriched microsome fraction. The inner envelope fraction was considerably richer in proteins than the outer envelope, as had been suggested previously based on gel electrophoretic analysis (Cline et al., 1981; Wernerwashburne et al., 1983). This is supported by mixed envelope samples (Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008a), which mainly contain proteins of the inner envelope with very few outer envelope proteins present (Table 7-S1). Bulk data retrieval from TAIR which uses mainly the TargetP program (Emanuelsson et al., 2000) for assignment identifies about two thirds of the proteins assigned to the inner envelope as chloroplast residents based on their N-terminal cTP (Figure 7-1) confirming that most of the proteins of the inner envelope reach it by a conventional targeting mechanism through the protein import complex. Of the remainder of the proteins, most have conflicting targeting information at their N-terminus resulting in 'undefined' targeting. Much of this conflicting targeting information can be resolved by manual inspection of the targeting results, which frequently identify the mitochondrion as the alternative target (data not shown). This kind of targeting is frequent for inner envelope proteins with the most prominent case being the most abundant inner envelope protein, the triosephosphate phosphate translocator. There are also two percent of the proteins which appear to have a targeting signal for the secretory pathway as has been demonstrated in other plastid envelope proteome projects of mixed envelopes. In total chloroplast proteomics, the number of proteins with canonical target peptide was higher (86%) and fewer proteins with target peptides to the secretory pathway (0.6%) were identified (Zybailov et al., 2008). This may reflect more diverse protein targeting pathways of inner envelope proteins compared to stromal proteins. The pattern of targeting predictions for the proteins assigned to the outer

envelope is very different from the pattern of proteins of proteins assigned to the inner envelope. It contains a large proportion of proteins with predicted target peptides for the secretory pathway and many proteins with 'undefined' targeting as well as proteins with canonical cTPs. Manual inspection of the targeting for the proteins with 'undefined' localization by bulk data retrieval with the help of ARAMEMNON reveals that many of these proteins carry weak signals for the secretory pathway (Table 7-S1). Unlike the inner envelope proteins, most of the known outer envelope proteins such as members of the Toc complex do not contain a cTP (Soll and Schleiff, 2004) except for Toc75-III which is targeted by a cTP and imported through the classical protein import complex (Inoue and Potter, 2004). The different N-terminal motifs among the proteins of the outer envelope may indicate that there is a multitude of targeting pathways. Apart from the targeting mechanism for Toc75, the targeting mechanism for two additional outer envelope proteins is known. They are inserted with the help of the protein AKR2A and the recognition motif at the N-terminus consists of a stretch of hydrophobic amino acids followed by positively charged amino acids (Bae et al., 2008). This pattern may be misrecognized as a weak signal for the secretory pathway. Many proteins of the outer envelope fraction hence represent putative additional cargo for the AKR2A mediated pathway. For example, LACS9 and WBC7 both contain a predicted transmembrane helix at the very N-terminus which is followed by charged residues (Table 7-2). Two soluble proteins which reach the chloroplast stroma through the endomembrane system are known to date showing a connection between the Golgi and the chloroplast (Villarejo et al., 2005; Nanjo et al., 2006). Possibly there are also membrane proteins targeted via the endomembrane system. The targeting predictions hint that proteins can take a multitude

of pathways to enter the outer envelope membrane. Therefore, proteome analysis followed by GFP fusion protein analysis as conducted here represents the only method to reliably determine the outer envelope proteome. The ER enriched microsome fraction consists mostly of proteins with a canonical target peptide for the secretory pathway or of undefined targeting. Only very few proteins in this fraction carry an N-terminal extension recognized by prediction programs to target the chloroplast or the mitochondrion. Although the ER enriched microsome fraction represents the proteome sample with the most mitochondrial proteins, the total amount of contamination is very low, emphasizing the purity of the preparations.

Protein Import Complex

With the different proteome samples at hand, not only large scale overall assignment are possible but also detailed localization analysis of proteins with known and putative function as well as analysis of isoforms within gene families. The strength of this method becomes obvious when the protein import complex is considered (Figure 7-2). Eight different proteins of the Toc complex were identified; the main import channel Toc75-III, three isoforms of Toc75-V, the import receptors Toc159 and Toc132 and ToC34 as well as Toc64. All proteins have their highest absolute spectral count in the outer envelope sample but they are also isolated in the inner envelope and PLAM but not the ER-enriched microsome fraction. Cross contamination of inner and outer envelope samples with each other has been observed frequently (Cline et al., 1981; Schleiff et al., 2003) but the strong presence of import complex components in the PLAM sample remains unexplained. For the Tic complex, the localization is almost exclusively to the inner envelope and, to a lesser degree, to the outer envelope sample with two exceptions.

Two isoforms of Tic110, Tic40, one isoform of Tic22, Tic20, Tic62 and Tic55 are clearly inner envelope proteins although Tic22 is described as only peripherally associated with the inner envelope.Tic22-III is exclusively localized in the PLAMs. Transient expression of Tic22-III GFP fusion proteins yielded a fluorescence pattern confirming localization to structures distinct of the chloroplast. The function of Tic22 in the envelope is currently unknown making it impossible to speculate about the function of its isoform in the PLAMs. TiC32 displays a similar pattern of inner/outer envelope sample localization but amazingly also is present in substantial amounts in the ER enriched microsome fraction. Although the import complex largely distributes between the proteome samples as expected, the strong presence of Toc complex components as well as Tic22-III in the PLAMs and the unusual pattern of TiC32 posit new questions.

Purity of the proteome samples

Tagging a number of proteins with GFP allowed us to not only observe dynamics of proteins within the envelope but also identify proteins which localize to other organelles if expressed as GFP fusion proteins. The kinesin-like protein localized to the cytosol (Table 7-2) and thus may have been peripherally associated with the outer envelope, the PLAMs and the ER-enriched microsomes, and thus been identified in all three proteome samples. For other proteins, the identification in the outer envelope or the PLAMs is less easy to explain. Several proteins localized to defined spots (Table 7-2) within the cytosol which may represent localization to mitochondria, peroxisomes or the Golgi. The identification of these proteins in proteomics is probably due to contamination of the preparation with these organelles which however may also represent a tight

association in vivo (Bräutigam et al., 2008a). Once extraplastidal localization was confirmed the proteins were removed from the analysis.

Trafficking of proteins to and from the plastid

The isolation and initial characterization of vesicles attached to the chloroplasts that can be labeled with an ER marker protein (Andersson et al., 2007) prompted us to analyze the proteome of this new vesicle class. In comparison to the envelope fractions and the ER fraction, the PLAM fraction contains proteins with a different targeting pattern, mostly proteins with a canonical TP indicating that the PLAM fraction is not a subfraction of either the ER-enriched microsomes or the outer envelope. The composition also argues against an artificial fraction produced during isolation of the fraction (Table 7-S1). In contrast, PLAMs appear to be a compartment assembled with proteins from both the ER as well as the chloroplast with a large contribution of outer envelope proteins. Not only the targeting information of the proteins from the outer envelope and the PLAM fractions is different, the PLAM fraction also contains a large portion of soluble stromal proteins. Both the pattern of targeting predictions as well as the protein content of the PLAMs marks them as a hybrid compartment of the chloroplast and the ER. Lipid analysis also indicates that PLAMs contain a mixture of lipids typical for membranes of the endomembrane system and of the chloroplast membranes (Andersson et al., 2007). Several of the proteins identified in the outer envelope and the PLAMs were tested for their localization in vivo with transient expression of the respective GFP fusion proteins (Table 7-2).

The analyses of targeting for GFP fusion proteins revealed unexpected patterns with protein localizations that changed depending on the time of observation (Figure 7-

3E and F). Two of the GFP fusion proteins localize to vesicles with bright fluorescence and comparatively dim fluorescence is observed in the envelope initially (Table 7-2). The proteins are both of unknown function, one protein has EF hand domains indicating the potential to bind calcium, the other one is a protein of unknown function. Prolonged expression leads to localization to the envelope in a circle surrounding the chlorophyll auto-fluorescence in addition to the vesicles (Figure 7-3E and F). The temporal pattern of expression which progresses from vesicles to the envelope suggests the proteins mainly localize to vesicles in vivo. The localization to vesicles is supported by the proteome analysis which mainly localized the protein to the PLAM fraction (Table 7-2). To our knowledge, this is the first demonstration of a membrane protein localizing to vesicles associated with the chloroplast and the envelope both by proteomics and by GFP fusion proteins. Both proteins progress from labeling vesicles and weakly labeling the envelope to a strong labeling of the chloroplast envelope in addition to the vesicles. This labeling pattern may indicate that the vesicles in vivo localization is in the membranes of the vesicles but the proteins pass through the envelope on their way to the vesicles since increased protein load leads to partial retention of the GFP signal in the envelope and in stromules forming at the envelope. Neither of the proteins producing this pattern has a canonical cTP (Table 7-2). A third protein mainly localizing to vesicles with dimmer GFP fluorescence in the envelope and the stroma is a Hsp60 (Figure 7-4). The protein is localized to the PLAMs by proteomics and it has a canonical cTP (Table 7-2). The protein can indeed be detected in the stroma (Figure 7-4). However the strongest fluorescent is visible in nascent vesicles forming at the tip of stromules (Figure 7-4B) and at the envelope (Figure 7-4D). Two proteins which were mainly detected in the PLAM

proteome sample were confirmed to localize to vesicles associated with chloroplasts by GFP fusions. The third protein was detected mainly in outer envelope samples also localizes to PLAMs when expressed transiently as a C-terminal GFP fusion. All three proteins can serve as markers for future studies of this new vesicle class inside plant cells. A number of putative outer envelope resident proteins were also investigated. These proteins, which initially localize to vesicles or to the envelope (Table 7-3) with the exception of the inner envelope marker protein Mep1, finally can be observed in either stromules or vesicles (e.g. Figure 7-3D). The formation of vesicles out of stromules is a well known phenomenon and has been described as 'tip-shedding' (Gray et al., 2001) although the function of either stromules or vesicles remains unknown (Gray et al., 2001; Hanson and Sattarzadeh, 2008). Formation of stromules can only be observed by prolonged overexpression and therefore increased protein load but never during initial stages of the experiment (Figure 7-3). In contrast to the known and putative outer envelope protein fusions which reside in defined structures such as stromules or vesicles, the inner envelope marker protein fusion accumulates as diffuse material in the cytosol which may represent proteins not yet imported into the envelope (Figure 7-3B). Formation of stromules upon overexpression of another inner envelope protein, the triosephosphate phosphate translocator, has not been reported (Gray et al., 2001). The difference in localization to the inner or outer envelope, respectively, seems to define the pattern upon prolonged expression but additional experimentation with more inner envelope resident proteins will be used to support the hypothesis. The overexpression of outer envelope proteins thus revealed a novel protein localization pattern of unknown significance. The vesicles labeled by outer envelope GFP fusion proteins are of 1-1.5um

in size (Figure 7-3), the same size observed for autophagosomes in plant cells (Bassham, 2007; Ishida et al., 2008). Although the production of vesicles has not been observed directly, they start to appear at the same time when stromules appear in the cells. Remarkably, the PLAMs are richer in outer envelope proteins than in inner envelope proteins (Table 7-S1), a pattern which is particularly evident when components of the protein import complex are considered (Figure 7-2). The correlation of these two phenomena, comparative proteomics reveals PLAMs are rich in outer envelope membranes and overexpression of outer envelope proteins leads to formation of vesicles, prompted us to speculate that PLAMs may be budding autophagosomes. It is currently not known how the different membranes systems are arranged within the PLAMs, whether these are truly mixed membranes or whether they represent envelope vesicles engulfed by ER-type membrane. It has recently been demonstrated that, if the proton ATPase in unable to acidify compartments, GFP tagged Rubisco and other chloroplast proteins can be detected in autophagy bodies in the vacuole apart from the plastid (Ishida et al., 2008). Moreover, an autophagy protein (ATG8e) colocalizes with stroma filled protuberances from the chloroplast leading to a model of quadruple membrane bound autophagosomes being delivered from the chloroplast to the vacuole for degradation of stromal proteins (Ishida et al., 2008). The PLAM proteome sample does contain several proton ATPase subunits needed for acidification (Table 7-S1). Several proteins assigned to the PLAM fraction are known stromal residents, most notably ribosomal proteins, ferritins and the large subunit of Rubisco which is encoded on the plastid genome as well as a transaldolase. It has been suggested previously that PLAMs represent the contact sites for lipid transfer between the ER and the chloroplasts (Benning et al., 2006;

Andersson et al., 2007) but neither the localization of GFP fusion proteins nor the proteome analysis confirmed the localization of lipid metabolism enzymes or transport proteins specifically in the PLAMs.

We posit that PLAMs (Andersson et al., 2007) represent budding autophagosomes as proposed by (Ishida et al., 2008) based on the following evidence: (a) PLAMs contain a large number of stromal soluble proteins as expected for autophagosomes (b) the protein composition of PLAMs is a hybrid of the endomembrane system and the chloroplast as proposed in the model for autophagosome formation (c) increased protein load of outer envelope but not inner envelope proteins induces the formation of stromules and vesicles (d) proteomics of PLAMs indicate a high proportion of outer envelope proteins in the sample.

This hypothesis is currently being tested by colocalization experiments of stromules and vesicles formed after prolonged expression of outer envelope proteins with markers for autophagosomes such as monodansylcadaverine (Contento et al., 2005), lysotracker (Bassham, 2007) and ATG8, an autophagosomal marker protein (Contento et al., 2005). In addition, co-localization studies of different combinations of proteins which label proteins observed in early or late vesicles are designed to understand whether these vesicles are of the same type. Since cells with excess amounts of stromules and vesicles can now be produced efficiently through the prolonged overexpression of defined proteins, electron microscopic reconstructions are possible. With electron microscopy, the membrane arrangement within the vesicles may possibly be resolved.

Acknowledgements

We would like to thank the Michigan State University Research Technology Support Facility for expert proteome analysis.

Supplemental Material

Supplemental Table 7-1: All proteins which were identified in the proteome experiments

Supplemental Table 7-2: Proteins of the outer envelope fractions

Supplemental Table 7-3: Proteins of the plastid associated membrane fractions

References

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor M (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists J Eukaryot Microbiol 52, 399-451

Andersson MX, Goksor M, Sandelius AS (2007) Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts J Biol Chem 282, 1170-1174

Awai K, Xu CC, Tamot B, Benning C (2006) A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking Proc Natl Acad Sci U S A 103, 10817-10822

Bae W, Lee YJ, Kim DH, Lee J, Kim S, Sohn EJ, Hwang I (2008) AKr2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis Nat Cell Biol 10, 220-U101

Bassham DC (2007) Plant autophagy-more than a starvation response Curr Opin Plant Biol 10, 587-593

Benning C, Xu CC, Awai K (2006) Non-vesicular and vesicular lipid trafficking involving plastids Curr Opin Plant Biol 9, 241-247

Bräutigam A, Hofmann-Benning S, Weber APM (2008a) Comparative Proteomics of Chloroplast Envelopes from C3 and C4 Plants Reveals Specific Adaptations of the Plastid Envelope to C4 Photosynthesis and Candidate Proteins Required for Maintaining C4 Metabolite Fluxes Plant Physiol 148, 568-579

Bräutigam A, Shrestha RP, Whitten D, Wilkerson CG, Carr KM, Froehlich JE, Weber APM (2008b) Comparison of the use of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes Journal of Biotechnology 136, 44-53

Cline K, Andrews J, Mersey B, Newcomb EH, Keegstra K (1981) Separation And Characterization Of Inner And Outer Envelope Membranes Of Pea-Chloroplasts Proc Natl Acad Sci USA 78, 3595-3599

Contento AL, Xiong Y, Bassham DC (2005) Visualization of autophagy in Arabidopsis using the fluorescent dye monodansylcadaverine and a GFP-AtATG8e fusion protein Plant J 42, 598-608

Curtis MD, Grossniklaus U (2003) A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta Plant Physiol 133, 462-469

Damari-Weissler H, Ginzburg A, Gidoni D, Mett A, Krassovskaya I, Weber APM, Belausov E, Granot D (2007) Spinach SoHXK1 is a mitochondria-associated hexokinase Planta 226, 1053-1058

Douce R, Joyard J (1979) 'Isolation and Properties of the Envelope of Spinach Chloroplasts.' (Ellis Horwood)

Eastmond PJ (2007) MONODEHYROASCORBATE REDUCTASE4 is required for seed storage oil hydrolysis and postgerminative growth in Arabidopsis Plant Cell 19, 1376-1387

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence J Mol Biol 300, 1004-16

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites Protein Sci 8, 978-84

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana Mol Cell Proteomics 2, 325-345

Fourrier N, Bedard J, Lopez-Juez E, Barbrook A, Bowyer J, Jarvis P, Warren G, Thorlby G (2008) A role for SENSITIVE TO FREEZING2 in protecting chloroplasts against freeze-induced damage in Arabidopsis Plant J 55, 734-745

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW, Gage DA, Phinney BS (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis J Proteome Res 2, 413-25

Gray JC, Sullivan JA, Hibberd JM, Hansen MR (2001) Stromules: Mobile protrusions and interconnections between plastids Plant Biology 3, 223-233

Hanson MR, Sattarzadeh A (2008) Dynamic morphology of plastids and stromules in angiosperm plants Plant Cell Environ 31, 646-657

Hatta T, Bullivan.S, Matthews RE (1973) Fine-Structure Of Vesicles Induced In Chloroplasts Of Chinese Cabbage Leaves By Infection With Turnip Yellow Mosaic-Virus Journal Of General Virology 20, 37-50

Inoue K, Potter D (2004) The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms Plant J 39, 354-365

Ishida H, Yoshimoto K, Izumi M, Reisen D, Yano Y, Makino A, Ohsumi Y, Hanson MR, Mae T (2008) Mobilization of Rubisco and Stroma-Localized Fluorescent Proteins of Chloroplasts to the Vacuole by an ATG Gene-Dependent Autophagic Process Plant Physiol 148, 142-155

Jarvis P, Robinson C (2004) Mechanisms of protein import and routing in chloroplasts Curr Biol 14, R1064-R1077

Keegstra K, Yousif AE (1986) Isolation And Characterization Of Chloroplast Envelope Membranes Meth Enzymol 118, 316-325

Kessler F, Schnell DJ (2006) The function and diversity of plastid protein import pathways: A multilane GTPase highway into plastids Traffic 7, 248-257

Lee YJ, Kim DH, Kim YW, Hwang I (2001) Identification of a signal that distinguishes between the chloroplast outer envelope membrane and the endomembrane system in vivo Plant Cell 13, 2175-2190

Li M, Schnell DJ (2006) Reconstitution of protein targeting to the inner envelope membrane of chloroplasts J Cell Biol 175, 249-259

Lisenbee CS, Lingard MJ, Trelease RN (2005) Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase Plant J 43, 900-914

Lu BB, Xu CC, Awai K, Jones AD, Benning C (2007) A small ATPase protein of Arabidopsis, TGD3, involved in chloroplast lipid import J Biol Chem 282, 35945-35953

Nanjo Y, Oka H, Ikarashi N, Kaneko K, Kitajima A, Mitsui T, Munoz FJ, Rodriguez-Lopez M, Baroja-Fernandez E, Pozueta-Romero J (2006) Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-Golgi to the chloroplast through the secretory pathway Plant Cell 18, 2582-2592

Pohlmeyer K, Soll J, Steinkamp T, Hinnah S, Wagner R (1997) Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane Proc Natl Acad Sci U S A 94, 9504-9509

Prod'homme D, Jakubiec A, Tournier V, Drugeon G, Jupin I (2003) Targeting of the Turnip Yellow Mosaic Virus 66K replication protein to the chloroplast envelope is mediated by the 140K protein Journal Of Virology 77, 9124-9135

Schleiff E, Eichacker LA, Eckart K, Becker T, Mirus O, Stahl T, Soll J (2003) Prediction of the plant beta-barrel proteome: A case study of the chloroplast outer envelope Protein Sci 12, 748-759

Schnurr JA, Shockey JM, de Boer G-J, Browse JA (2002) Fatty Acid Export from the Chloroplast. Molecular Characterization of a Major Plastidial Acyl-Coenzyme A Synthetase from Arabidopsis Plant Physiol 129, 1700-1709

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels Anal Chem 68, 850-858

Soll J, Schleiff E (2004) Protein import into chloroplasts Nature reviews Molecular cell biology 5, 198-208

Tripp J, Inoue K, Keegstra K, Froehlich JE (2007) A novel serine/proline-rich domain in combination with a transmembrane domain is required for the insertion of AtTic40 into the inner envelope membrane of chloroplasts Plant J 52, 824-838

Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L, Samuelsson G (2005) Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast Nat Cell Biol 7, 1224

Weber APM (2004) Solute transporters as connecting elements between cytosol and plastid stroma Curr Opin Plant Biol 7, 247-253

Weber APM, Fischer K (2007) Making the connections - The crucial role of metabolite transporters at the interface between chloroplast and cytosol FEBS Lett 581, 2215-2222

Weber APM, Schwacke R, Flugge UI (2005) Solute transporters of the plastid envelope membrane Ann Rev Plant Biol 56, 133-164

Wernerwashburne M, Cline K, Keegstra K (1983) Analysis Of Pea Chloroplast Inner And Outer Envelope Membrane-Proteins By Two-Dimensional Gel-Electrophoresis And Their Comparison With Stromal Proteins Plant Physiol 73, 569-575

Xu CC, Fan J, Froehlich JE, Awai K, Benning C (2005) Mutation of the TGD1 chloroplast envelope protein affects phosphatidate metabolism in Arabidopsis Plant Cell 17, 3094-3110

Xu CC, Fan JL, Cornish AJ, Benning C (2008) Lipid trafficking between the endoplasmic reticulum and the plastid in Arabidopsis requires the extraplastidic TGD4 protein Plant Cell 20, 2190-2204

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting Signals, N-Terminal Modifications and Abundance of the Chloroplast Proteome PLoS ONE 3, e1994

Chapter 8

Conclusions and Future Perspectives

The long term goal of studying plastids is to fully understand the plastid localized biological processes as well their connection to the remainder of the cell. Plastids are metabolic factories and provide the cell with many precursor molecules for maintenance, growth and differentiation. They are the site of starch synthesis in higher plants, they produce most of the fatty acids in plant cells, and they are the source of branched chain and aromatic amino acids for protein biosynthesis. Human nutrition critically depends on plants for carbohydrates, proteins and fatty acids as well as vitamins. Either plants serve directly as food or are fed to livestock which is in turn consumed by humankind. The increasing world population and the decreasing area of arable land pose unique challenges to feeding the world and feeding the world's hunger for non-fossil energy. Understanding the metabolite factory of the plant cell, the plastid, may enable us to engineer plants with increased yield for both food and biofuel production or increased nutritional value. Although the metabolic pathways residing in plastids are fairly well understood, there are major limitations in understanding the connection of plastids to the remainder of the cell. We decided to gain insight into plastid biology and differentiation by focusing proteome studies on the envelope which separates the soluble stroma from the remainder of the cell and serves as the interface between the stroma and the cytosol. Initial comprehensive proteome studies of plastids have been focused mainly on green photosynthetic chloroplasts. The internal membrane system which contains the electron transfer chain, the thylakoids, the stroma and also the envelopes of chloroplasts have been studied in detail by proteomics (Ferro et al., 2000; Peltier et al., 2000; Ferro et al., 2002; Peltier et al., 2002; Ferro et al., 2003; Froehlich et al., 2003; Baginsky and

Gruissem, 2004; Friso et al., 2004; Kleffmann et al., 2004; Sun et al., 2004; Baginsky et al., 2005; Zybailov et al., 2008). Studies of other plastid subtypes have been focused on whole plastids (Baginsky et al., 2004; von Zychlinski et al., 2005; Siddique et al., 2006; Kleffmann et al., 2007) which limits the discovery of proteins integral to or associated with the envelope membranes. The goal of my work was to explore new model systems to analyze differentiated plastid types in comparison with classical chloroplasts and to provide detailed insights into envelope biology by using comparative instead of qualitative proteome analysis.

In chapter 3, I report on using an organism without a sequenced genome for proteome analysis. In cases like this, the proteome analysis has been based on sequence information from related species (Baginsky et al., 2004). This new analysis identifies the limitations of this approach for proteomics which are especially severe for proteins of low abundance. This paper not only reports the limitations but also demonstrates a possible strategy to overcome them, the generation of a suitable database by pyrosequencing. Alternative strategies to increase the detection power on non-species specific databases are shown to lead to increased numbers of false positives and thus are not suitable. The generation of new model organism for proteomics beyond the plants with sequenced genomes is important because most of the models were chosen either since they are crop plants or because they are amenable to genetics but never because they are suitable for biochemistry. For the new model *Pisum sativum*, protocols for isolation of both leaf chloroplasts and root amyloplasts (Borchert et al., 1989; Borchert et al., 1993) are in place. For pea chloroplasts, protocols to subfractionate the chloroplast envelope into the inner and outer envelope have been described (Cline et al., 1981).

Mitochondria can also be efficiently purified from pea roots, stems and leaves (Bourguignon et al., 1992; Zancani et al., 1995). Pea plants also allow the purification of endomembrane system components in sufficient quantity for biochemical analysis (Hellgren et al., 1993). The new database enables comparative proteomics of organelles in different tissues of the same plant. The pea chloroplast envelope proteome dataset containing at least semi-quantitative information also serves as a template to which other plastid types can be compared.

In chapter 4, the first analysis of the envelope proteome of a non-green plastid is reported. The primary role of proplastids was considered to act as a reservoir for plastid development rather than being metabolically active themselves (Wise, 2006). The analysis of proplastid-like plastids in tobacco cell cultures had already revealed highly active plastids which contain many proteins for amino acid synthesis (Baginsky et al., 2004) but meristematic proplastids were never analyzed in vivo since meristematic tissue has been limited prior to the introduction of cauliflower as the tissue source. The analysis of the envelope proteome revealed a distinct set of envelope proteins indicating meristematic proplastids are highly active cell factories for amino acids and nucleotides. The presence of a reduction equivalent shuttle at the envelope leads to a reconsideration of the role of glucose-6-phosphate import and the two translocator model could be adapted to proplastids. In addition to candidates for uncharacterized fluxes, the proteome analysis also revealed the adaptation of the protein import complex in proplastids and a unique subset of thylakoid proteins. The successful analysis of meristematic plastids in a Brassicaceae species opens the possibility to study chromoplasts, proplastids and chloroplasts in the same model plant.

In chapter 5, I report on changes in protein abundance in different chloroplast types rather than on differences in the protein pattern as reported for the comparison of chloroplasts and proplastids. The analysis reveals that increasing the transport protein amount in the inner envelope is one of the major adaptations for increasing the metabolite flow across a membrane to accommodate a biochemical pathway like C4 photosynthesis. Surprisingly, the respective outer envelope porins were also increased. This pattern has implication both for engineering efforts of pathways and for C4 photosynthesis research. Although the plastid envelope has evolved to handle the metabolic flow required for its respective role and thus exerts no control over the flow through the pathway, major changes in metabolite flow requires evolutionary adjustments at the plastid envelope. Bioengineering of bulk flow in biochemical pathways thus may require not only the presence of soluble enzymes but also changes in transport protein content if different compartments are involved. Since at least the known transport proteins involved in C4 photosynthesis accommodate the increased metabolite flow by increased abundance, new transport protein candidates for defined flows may be identified based on this feature. The analysis demonstrates the power of comparative 'omics' approaches which are suitable to identify new components involved in pathways with high metabolite flow without introducing bias into the experiments by predetermined expectations.

Chapter 6 reports the analysis of a transport protein candidate.

In chapter 7d, the newly generated pea sequence database was used to study chloroplast envelopes in more detail. Since the envelope contains only a small fraction of the cellular protein, analysis by proteomics was limited with small model species like Arabidopsis. Using pea allowed the separation and analysis of inner and outer envelope

membranes as well as plastid associated membranes and a light microsomal fraction. The outer envelope has received little attention compared to the inner envelope membrane of plastids although it represents the ultimate barrier to the cytosol and therefore also the interface to the remainder of the cell. At the onset of this work, little was known about the dynamics of envelope membranes other than the fact that stromules, stroma filled protuberances, can extend from the envelope in some cell types. The recent discovery of a close association with the endomembrane system (plastid associated membranes or PLAMs) prompted a comparative proteome analysis with the goal to define the new compartment PLAM and to discover new proteins of the envelope as well as localize known proteins to either the outer or the inner envelope. The comparative proteome data already revealed considerable protein dynamics between the membrane systems and marked the PLAMs as a hybrid compartment of the ER and the chloroplast with a large contribution of outer envelope proteins. The dynamic localization patterns were analyzed in vivo by protein GFP fusions transiently expressed in tobacco cells. The GFP fusion proteins confirmed the presence of chloroplast associated vesicles and three proteins, two membrane bound and one soluble, are primarily associated with these vesicles. The analysis of known and putative outer envelope proteins revealed new fluorescence patterns based on the protein load on the system. After prolonged expression, many of the outer envelope protein labeled the envelope, stromules and vesicles, a behavior which has never been observed with either stoma or inner envelope targeted proteins. Considering the GFP data in context of the proteome data lead to a new testable hypothesis about PLAMs. They are budding autophagosomes and their formation can be induced by overexpression of both PLAM and outer envelope proteins. The identification of both an

inducible system and of resident proteins allows testing several hypothesis about PLAMs. Their colocalization with autophagy markers such as monodansylcadaverine and AtATG8e is being tested. We will also start analyzing the budding of vesicles with co-localization experiments to test whether all proteins that label vesicular structures indeed colocalize and to determine a time-resolved pattern of protein localization during the formation of a vesicle. This new approach will yield insights into protein transport out of the plastid and highlight the dynamic nature of the envelope membranes.

In summary, this work has highlighted dynamic features of the envelope and introduced novel models for the study of plastids by proteomics. The plastid envelope turned out to be dynamic in regard to the protein composition; the proplastid envelope has a qualitatively different pattern compared to the chloroplast envelope. When comparing two chloroplast types, the differences were based on protein abundance rather than protein presence and therefore only revealed by semi-quantitative comparison. In all cases, the envelopes reflected the metabolic roles of the respective plastids. The envelopes of chloroplasts themselves are also dynamic. Traditionally considered static with regard to membrane architecture of the envelope membranes, both proteomics and GFP fusion protein analysis reveal vesicles with distinct composition and varying localization patterns of proteins based on their protein load in the system.

References

Baginsky S, Gruissem W (2004) Chloroplast proteomics: potentials and challenges J Exp Bot 55, 1213-1220

Baginsky S, Kleffmann T, von Zychlinski A, Gruissem W (2005) Analysis of shotgun proteomics and RNA profiling data from Arabidopsis thaliana chloroplasts J Proteome Res 4, 637-640

Baginsky S, Siddique A, Gruissem W (2004) Proteome analysis of tobacco bright yellow-2 (BY-2) cell culture plastids as a model for undifferentiated heterotrophic plastids J Proteome Res 3, 1128-1137

Borchert S, Grosse H, Heldt HW (1989) Specific Transport Of Inorganic-Phosphate, Glucose-6-Phosphate, Dihydroxyacetone Phosphate And 3-Phosphoglycerate Into Amyloplasts From Pea Roots FEBS Lett 253, 183-186

Borchert S, Harborth J, Schunemann D, Hoferichter P, Heldt HW (1993) Studies Of The Enzymatic Capacities And Transport-Properties Of Pea Root Plastids Plant Physiol 101, 303-312

Bourguignon J, Macherel D, Neuburger M, Douce R (1992) Isolation, Characterization, And Sequence-Analysis Of A Cdna Clone Encoding L-Protein, The Dihydrolipoamide Dehydrogenase Component Of The Glycine Cleavage System From Pea-Leaf Mitochondria Eur J Biochem 204, 865-873

Cline K, Andrews J, Mersey B, Newcomb EH, Keegstra K (1981) Separation And Characterization Of Inner And Outer Envelope Membranes Of Pea-Chloroplasts Proc Natl Acad Sci USA 78, 3595-3599

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana Mol Cell Proteomics 2, 325-345

Ferro M, Salvi D, Riviere-Rolland H, Vermat T, Seigneurin-Berny D, Grunwald D, Garin J, Joyard J, Rolland N (2002) Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters Proc Natl Acad Sci U S A 99, 11487-92.

Ferro M, Seigneurin-Berny D, Rolland N, Chapel A, Salvi D, Garin J, Joyard J (2000) Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins Electrophoresis 21, 3517-26.

Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, Sun Q, van Wijk KJ (2004) In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana chloroplasts: New proteins, new functions, and a plastid proteome database Plant Cell 16, 478-499

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW, Gage DA, Phinney BS (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis J Proteome Res 2, 413-25

Hellgren L, Morre DJ, Sellden G, Sandelius AS (1993) Isolation Of A Putative Vesicular Intermediate In The Cell-Free Transfer Of Membrane From Transitional Endoplasmic-Reticulum To The Golgi-Apparatus Of Etiolated Seedlings Of Garden Pea J Exp Bot 44, 197-205

Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S (2004) The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions Curr Biol 14, 354-362

Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, Baginsky S (2007) Proteome dynamics during plastid differentiation in rice Plant Physiol 143, 912-923

Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, von Heijne G, van Wijk KJ (2002) Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction Plant Cell 14, 211-36.

Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I, van Wijk KJ (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins Plant Cell 12, 319-41.

Siddique MA, Grossmann J, Gruissem W, Baginsky S (2006) Proteome analysis of bell pepper (Capsicum annum L.) chromoplasts
Plant Cell Physiol, 33-39

Sun Q, Emanuelsson O, van Wijk KJ (2004) Analysis of curated and predicted plastid subproteomes of Arabidopsis. Subcellular compartmentalization leads to distinctive proteome properties Plant Physiol 135, 723-734

von Zychlinski A, Kleffmann T, Krishnamurthy N, Sjolander K, Baginsky S, Gruissem W (2005) Proteome analysis of the rice etioplast - Metabolic and regulatory networks and novel protein functions Mol Cell Proteomics 4, 1072-1084

Wise R (2006) 'The diversity of plastid form and function.' (Springer: Amsterdam)

Zancani M, Macri F, Peruffo ADB, Vianello A (1995) Isolation Of The Catalytic Subunit Of A Membrane-Bound H+-Pyrophosphatase From Pea Stem Mitochondria Eur J Biochem 228, 138-143

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting Signals, N-Terminal Modifications and Abundance of the Chloroplast Proteome PLoS ONE 3, e1994

Appendix

High-throughput colorimetric method for the parallel assay of glyoxylic acid and ammonium in a single extract

This chapter has been previously published in Analytical Biochemistry: Bräutigam A*, Gagneul D* and Weber APM (2007) High-throughput colorimetric method for the parallel assay of glyoxylic acid and ammonium in a single extract Anal Biochem 362, 151-3

* These authors contributed equally to the work.

I would like to acknowledge Dr. David Gagneul for collaboration on developing the extract and contributing figure 2B

Introduction

Glyoxylate and ammonia are signature metabolites of a major plant metabolic pathway, the C_2 oxidative carbon cycle, also known as photorespiration. This complex salvage pathway recycles most of the carbon that is lost from the Calvin cycle in the form of phosphoglycolate as a consequence of the oxygenation reaction of RubisCO and it constitutes a significant metabolic flux in photosynthesizing leaves of C3 plants [1]. Mutations in enzymes of this pathway are usually lethal, underlining the importance of this salvage pathway [2-4]. Although all enzymes and some of the metabolite transporters involved in this highly-compartmentalized pathway have been identified [5-9], information about the majority of the transport proteins and the processes regulating the pathway is still missing [1, 4], possibly because mutations in the corresponding genes cause subtle metabolic phenotypes. Thus, a high throughput method to quantify metabolites known to accumulate in photorespiratory mutants, glyoxylate and ammonium, is desirable. Glyoxylate is also a pathway intermediate of the glyoxylate cycle that is involved in making carbon moieties from fatty acids available to carbohydrate biosynthesis by gluconeogenesis. This pathway is important during seed germination and seedling establishment of oil seed plants and possibly during senescence but does not play a significant role in photosynthesizing leaves. While primary nitrogen assimilation by nitrate and nitrite reductase and the phenylpropanoid pathway are also sources of ammonium in planta [10], photorespiration contributes at least an order of magnitude more ammonium than these pathways [11, 12], making it a suitable marker for the photorespiratory pathway.

Here we present improved versions of two colorimetric methods for the quantification of ammonia and glyoxylate that makes these procedures suitable for high throughput and sensitive quantification of these signature metabolites in plant tissues.

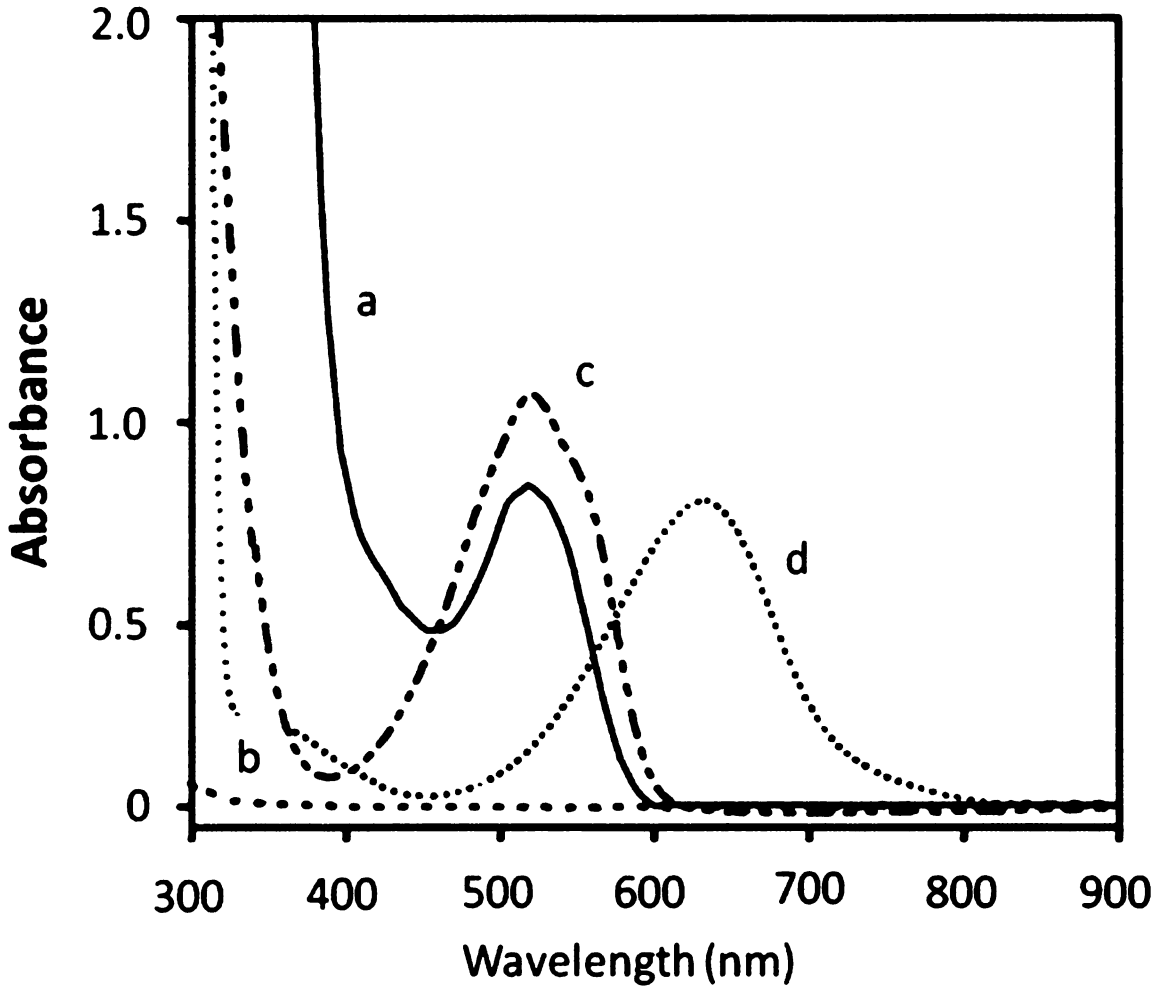


Figure 1: Absorption spectra of a crude acidic extract of *Arabidopsis* leaves (a), of a partially purified acidic extract of *Arabidopsis leaves* (b), of the formazan of glyoxylic acidic (c), and of the indophenol derivative of ammonia (d) as described in the text.

Gas chromatography is not suitable for the determination of glyoxylate, presumably because it is unstable and non-enzymatically decomposes when coming into contact with oxygen. Glyoxylate has been measured using a colorimetric method [5, 13, 14]. In this method, a 1,5-diphenylformazancarboxylic acid derivative of glyoxylic acid is

produced through mild oxidation of a glyoxylic acid hydrazone in the presence of excess phenylhydrazine. The formazan of glyoxylic acid shows maximal absorption at a wavelength of 520 nm (Figure 1). The quantification of ammonium (based on the Berthelot reaction [15, 16]) from the same plant extract as glyoxylate has been described previously in tobacco [5]. Ammonia reacts with hypochlorite and forms a monochloramine, which reacts with phenol to an indophenol derivative that has deep blue colour and displays maximal absorbance at 620 nm (Figure 1). Although both methods have been used previously to determine glyoxylate and ammonia contents in plant tissues [5, 17], it is hampered by a number of interfering compounds that quench the signal. In addition, many plant tissues accumulate anthocyanines, especially under stress conditions, which absorb light at similar wavelengths and therefore introduce noise to the assays. To address these problems and to render the assay suitable for high-throughput analysis, we have developed a new protocol optimized for use in a 96-well format that allows the colorimetric quantification of glyoxylate and ammonia from a single plant extract.

Method

Plant tissue is harvested and snap-frozen in liquid nitrogen, and ground to a fine powder under liquid nitrogen. The modified extraction procedure is as follows: 100 mg of powder is extracted in 1 ml of 100 mM HCl, and 500 µl of chloroform is subsequently added. The samples are rotated for 15 minutes at 4°C and the phases are separated by centrifugation (12,000x*g*, 10 minutes, 8°C). The aqueous phase is transferred to a fresh tube containing 50 mg of acid washed activated charcoal, thoroughly mixed, and

centrifuged again (20,000x*g*, 5 minutes, 8°C). For glyoxylate quantification, 450 µl of the supernatant is mixed with 50 µl of a freshly prepared 1% (v/v) solution of phenylhydrazine in 100 mM HCl and the sample is incubated at 95°C for two minutes and cooled on ice for 6 minutes. The sample is acidified by adding 250 µl concentrated HCl and divided into three aliquots of 225 µl each in cavities of a 96 well plate: Two technical replicates are oxidized with 25 µl of a solution of 1.6% (w/v) K3Fe(CN)6 in water and one aliquot serves as control with 25 µl water added. Colour formation is measured exactly 6 minutes after addition of K3Fe(CN)6 solution since the reaction reaches a plateau after 4 minutes and the coloured reaction product slowly decomposes with time afterwards.

For ammonia quantification the supernatant obtained after charcoal treatment is diluted 1:1 (v/v) in 100 mM HCl..20 µl of this solution is mixed with 100 µl of a 1% (w/v) phenol 0.005% (w/v) sodium-nitroprusside solution in water ("reagent I"), and 100 µl of a 1% (v/v) sodium-hypochlorite 0.5% (w/v) sodium hydroxide solution in water ("reagent II") are added. The samples are incubated at 37°C for 30 minutes and light absorption at 620 nm is measured.

Results and Discussion

To develop the glyoxylate assay we recapitulated the original procedure of Dekker and Maitra [13] with varying concentrations of sodium glyoxylate in water. 10 µl sodium glyoxylate solution was added to 500 µl 0.1% phenylhydrazine in 100 mM HCl as previously described, heated, cooled on ice and acidified with concentrated HCl and oxidized to observe colour formation. We then tested the parameters of the analysis to

determine the optimal incubation temperature (95°C), incubation time (2 minutes), cooling time (6 minutes) and time between addition of oxidizing agent and measurement (6 minutes) that still allowed the processing of 32 samples per experiment. Changing these parameters leads to changed absorption values but still produces linear standard curves. The standard curve was linear from 0.25-200 µM sodium glyoxylate.

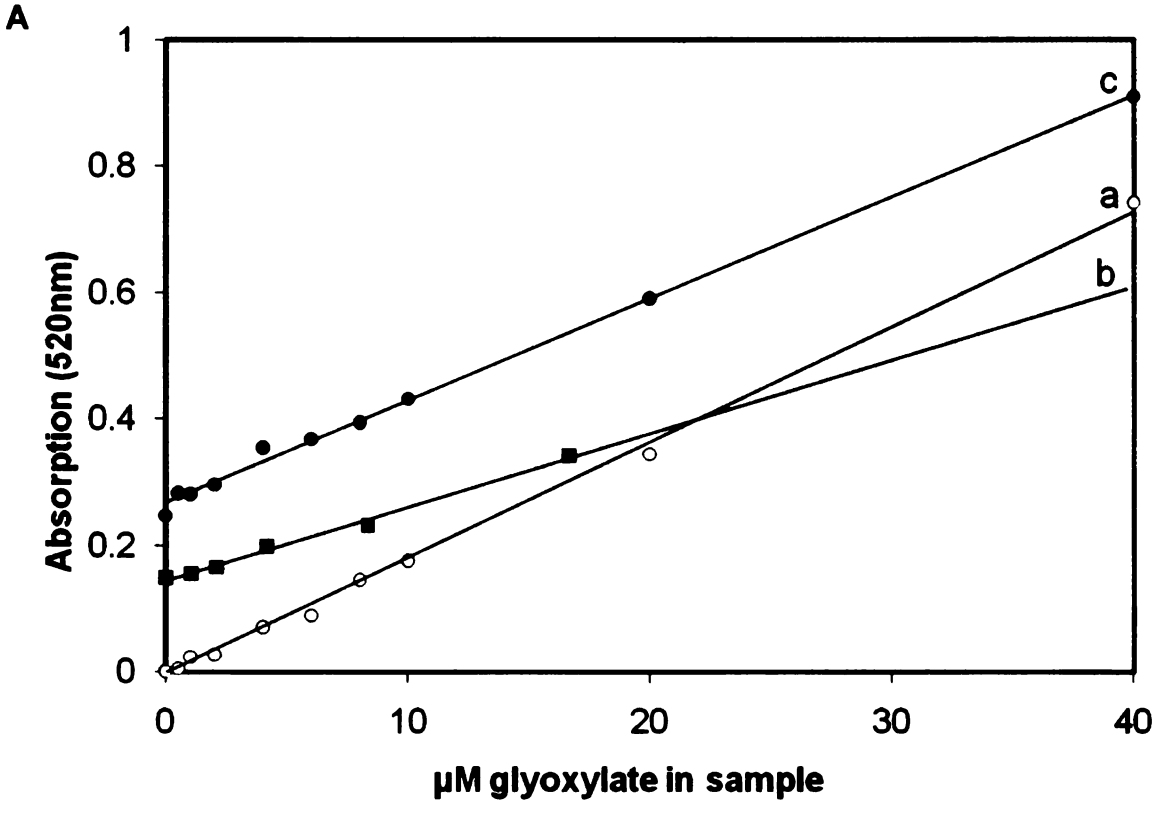
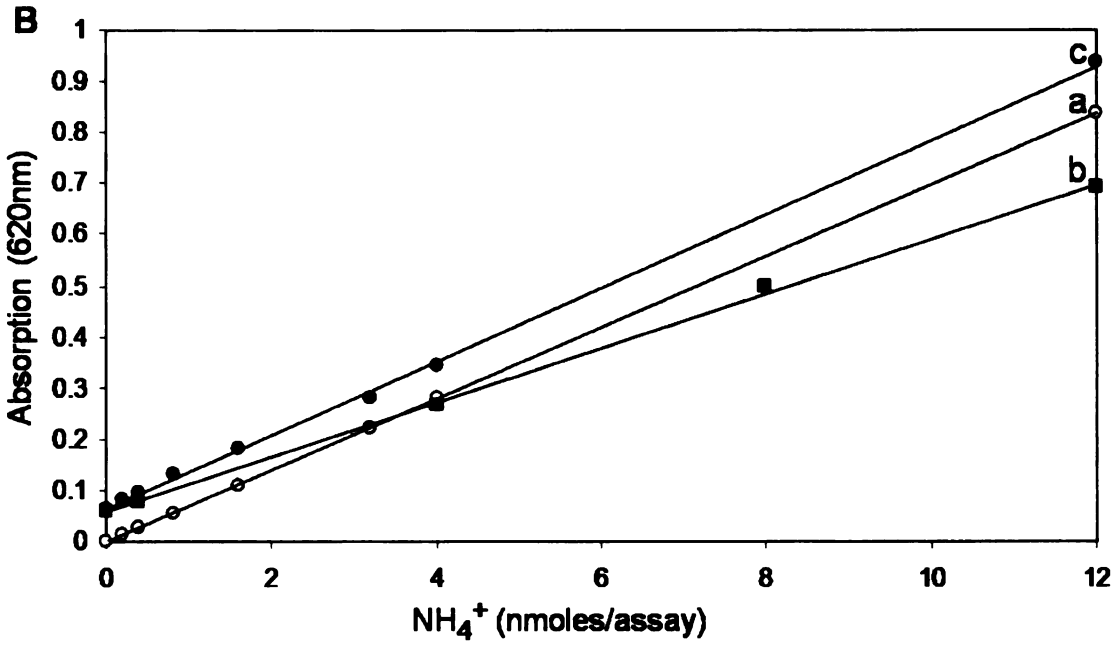


Figure 2: Comparison of standard curves for glyoxylate (A) and ammonia (B) determination. The external standard curve was prepared in 100 mM HCl (a). The internal standard curves were prepared in crude acidic extract of *Arabidopsis* leaves and were either measured (b) or partially purified and measured (c). The glyoxylate and ammonia contents were determined as described in the text. For all curves, *R*(squared) > 0.99.

Figure 2 (cont'd)



We then tested the suitability of the Dekker and Maitra method [13] for plant extracts. Plant tissue was extracted in 0.1% phenylhydrazine in 100 mM HCl and spiked with varying concentrations of sodium glyoxylate. The debris in spiked extracts was precipitated by centrifugation and the supernatant was boiled, cooled, and acidified, as described above. Upon acidification the remaining soluble proteins denatured and turned the sample turbid necessitating an additional centrifugation to clear the extract. One aliquot served as blank and two aliquots were oxidized as described above. When this internal standard curve was compared to the external standard curve prepared without the addition of plant extract, we observed a marked difference in slope between both curves. This indicates the presence of one or more compounds in the plant extract that quench the signal (Figure 2A). Plant tissues also frequently contain varying amounts of anthocyanines, which might obscure the specific signal because they absorb light at a

similar wavelength as the formazan (Figure 1). To remove lipophilic compounds such as chlorophylls and proteins from the extract we added chloroform to powdered plant tissue extracted with 100 mM HCl (derived from [18]). Although after centrifugation proteins and chlorophylls were quantitatively removed from the aqueous phase, the slope of the resulting standard curve remained identical to the initial experiment. We therefore tested several commercially available resins for removal of the interfering compounds. Amberlite XAD7 quantitatively removed anthocyanines from the extract as described in [19] without absorbing glyoxylic acid, but the interfering compound or compounds remained, as judged from the slopes of internal and external standard calibration curves. Finally, we tested whether acid washed activated charcoal would absorb the interfering compound without impacting the recovery of glyoxylic acid. In 100 mM HCl, charcoal did not absorb glyoxylic acid but efficiently absorbs either phenylhydrazine or the reaction product of glyoxylic acid and phenylhydrazine, glyoxylic acid hydrazone. Therefore the extraction of glyoxylic acid and the conjugation to phenylhydrazine had to be temporally separated by crude extraction in 100 mM HCl, extract purification, and subsequent reaction of the glyoxylic acid with phenylhydrazine.

When plant tissue is extracted with 100 mM HCl and chloroform and purified with activated charcoal, both anthocyanines and the unknown interfering compounds are efficiently removed without absorbing glyoxylic acid since the slope of the internal standard curve fits perfectly to the slope of the external standard curve (Figure 2A). The shift between both curves represents the glyoxylic acid content in the plant tissue.

To develop the ammonium assay for plant tissue we first tested the Berthelot method with standard solutions of varying concentrations of ammonium in 100 mM HCl.

The method of Weatherburn [15] was adapted to the 96 well format. 100µl reagent I was mixed with 20 µl sample and 100 µl reagent II. The samples were incubated at 37°C for 30 minutes and measured at 620 nm. The resulting standard curve was linear in the range 0.01-3 mM ammonium. Next we extracted finely ground plant tissue in 100 mM HCl and spiked with varying concentrations of ammonium. Although the internal standard curve was linear in the range 0.01-1 mM ammonium, its slope was markedly different from the external standard curve, indicating the presence of one or more interfering compounds (Figure 2B). To remove the unknown interfering compounds, plant tissue was extracted with 100 mM HCl, spiked with varying concentrations of ammonia, and centrifuged at 14,000x*g* to remove cell debris. The supernatant was incubated with acid washed activated charcoal, as decribed above. To 100 µl reagent I 20 µl purified diluted extract was added followed by 100 µl reagent II. The slope of the resulting internal standard curve fitted perfectly to the slope of the internal standard curve. Finally we ensured that the additional purification step with chloroform needed for glyoxylate determination did not interfere with the determination of ammonium (Figure 2B). It has been reported previously that the Berthelot method is not suitable for determining ammonia concentrations in plant extracts [20]. Interfering compounds result in a difference in slope of the internal standard curve compared to an external standard as shown in Figure 2. To assess whether different amino acid concentrations in the extract interfere with the quantification of ammonium, we doubled and quadrupled the extract concentration in the essay. A twice-concentrated extract will not result in a changed slope whereas quadrupling the extract concentration introduces sufficient interfering compounds to change the slope. The protocol presented in [20] requires elaborate instrumentation

whereas the novel extraction and purification protocol presented here allows the reliable and efficient quantification of ammonium with a spectrophotometer. However, if this new protocol is to be adapted for different tissues or plant species, we recommend optimizing the extract concentration to minimize interference as described above.

In summary, we have developed a novel purification protocol for acidic plant extracts in which interfering compounds and co-absorbing compounds were quantitatively removed by chloroform and charcoal extraction. This new method allows the simultaneous determination of two signature metabolites of photorespiration, glyoxylate and ammonium, from the same plant extract. The method has also been adapted to the high-throughput 96-well format and can potentially be robotized, thus making it suitable for mutant screens.

Acknowledgements

This work was supported by NSF-awards IOB-0548610, MCB-0618335, and MCB-0519740.

References

1. A.P.M. Weber, Synthesis, Export and Partitioning of the End Products of Photosynthesis, in: R. R. Wise, J. K. Hoober (Eds.) The Structure and Function of Plastids, Springer, Dordrecht, 2006, pp. 273-292.

2. C.R. Somerville, An early Arabidopsis demonstration. Resolving a few issues concerning photorespiration, Plant Physiol. 125 (2001) 20-24.

3. A. Wingler, P.J. Lea, W.P. Quick, R.C. Leegood, Photorespiration: metabolic pathways and their role in stress protection, Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci. 355 (2000) 1517-1529.

4. S. Reumann, A.P.M. Weber, Plant peroxisomes respire in the light: Some gaps of the photorespiratory C_2 cycle have become filled - others remain, Biochim. Biophys. Acta 1763 (2006) http://dx.doi.org/10.1016/j.bbamcr.2006.09.008.

5. J. Schneidereit, R.E. Häusler, G. Fiene, W.M. Kaiser, A.P.M. Weber, Antisense repression reveals a crucial role of the plastidic 2-oxoglutarate/malate translocator DiT1 at the interface between carbon and nitrogen metabolism, Plant J. 45 (2006) 206-224.

6. R. Boldt, C. Edner, U. Kolukisaoglu, M. Hagemann, W. Weckwerth, S. Wienkoop, K. Morgenthal, H. Bauwe, D-glycerate 3-kinase, the last unknown enzyme in the photorespiratory cycle in Arabidopsis, belongs to a novel kinase family, Plant Cell 17 (2005) 2413-2420.

7. L.M. Voll, A. Jamai, P. Renne, H. Voll, C.R. McClung, A.P.M. Weber, The photorespiratory Arabidopsis *shm1* mutant is deficient in *SHM1*, Plant Physiol. 140 (2006) 59-66.

8. P. Renné, U. Dreßen, U. Hebbeker, D. Hille, U.I. Flügge, P. Westhoff, A.P.M. Weber, The *Arabidopsis* mutant *dct* is deficient in the plastidic glutamate/malate translocator DiT2, Plant J. 35 (2003) 316-331.

9. M. Linka, A.P.M. Weber, Shuffling ammonia between mitochondria and plastids during photorespiration, Trends Plant Sci. 10 (2005) 461-465.

10. C. Fritz, N. Palacios-Rojas, R. Feil, M. Stitt, Regulation of secondary metabolism by the carbon-nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism, Plant J. 46 (2006) 533-548.

11. A.J. Keys, I.F. Bird, M.J. Cornelius, P.J. Lea, R.M. Wallsgrove, B.J. Miflin, Photorespiratory Nitrogen Cycle, Nature 275 (1978) 741-743.

12. H. Bauwe, U. Kolukisaoglu, Genetic manipulation of glycine decarboxylation, J. Exp. Bot. 54 (2003) 1523-1535.

13. E.E. Dekker, U. Maitra, Conversion of Gamma-Hydroxyglutamate to Glyoxylate and Alanine – Purification and Properties of Enzyme System, J. Biol. Chem. 237 (1962) 2218-2227.

14. D.N. Kramer, N. Klein, R.A. Baselice, Quantitative Determination of Glyoxylic Acid, Anal. Chem. 31 (1959) 250-252.

15. M.W. Weatherburn, Phenol-Hypochlorite Reaction for Determination of Ammonia, Anal. Chem. 39 (1967) 971.

16. C.J. Patton, S.R. Crouch, Spectrophotometric and Kinetics Investigation of Berthelot Reaction for Determination of Ammonia, Anal. Chem. 49 (1977) 464-469.

17. A. Wingler, P.J. Lea, R.C. Leegood, Photorespiratory metabolism of glyoxylate and formate in glycine-accumulating mutants of barley and *Amaranthus edulis*, Planta 207 (1999) 518-526.

18. W. Weckwerth, K. Wenzel, O. Fiehn, Process for the integrated extraction, identification, and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks, Proteomics 4 (2004) 78-83.

19. K. Takeda, S. Tominaga, The Anthocyanin in Blue Flowers of *Centaurea cyanus*, Bot. Mag. Tokyo 96 (1983) 359-363.

20. S. Husted, C.A. Hebbern, M. Mattsson, J.K. Schjoerring, A critical experimental evaluation of methods for determination of NH_4^+ in plant tissue, xylem sap and apoplastic fluid, Physiol. Plant. 109 (2000) 167-179.

File List of Supplementary Data

Excel File containing two sheets as a supplemental file to chapter 3
Supplement to Chapter 3.xls 110KB

Supplemental table: Proteins identified with the three databases

Excel File containing three sheets as a supplemental file to chapter 4
Supplement to Chapter 4.xls 211KB

Supplementary table 1: Proteins identified in proplastid envelopes from cauliflower curd tissue
Supplementary table 2: Proteins identified in chloroplast envelopes from pea leaves (Bräutigam et al., 2008a) were extracted from PRIDE and reanalyzed
Supplementary table 3: Thylakoid proteins identified in envelopes of proplastids and chloroplasts

Excel File containing three sheets as a supplemental file to chapter 5
Supplement to Chapter 5.xls 224KB

Supplemental Table 5-S1. Proteins identified in C_4 PCA-type maize chloroplast envelope membranes. The table lists proteins identified, number of spectra mapping to each maize accession number, annotation, classification, number of membrane spanning domains, targeting prediction, and previous identifications in other proteomics studies.
Supplemental Table 5-S2. Proteins identified in C_3-type pea chloroplast envelope membranes. The table lists proteins identified, number of spectra mapping to each maize accession number, annotation, classification, number of membrane spanning domains, targeting prediction, and previous identifications in other proteomics studies.
Supplemental Table 5-S3. Percentage of total spectral counts for each protein identified in C_4 PCA-type and C_3-type chloroplasts of maize and pea, respectively.

Excel File containing three sheets as a supplemental file to chapter 7
Supplement to Chapter 7 217KB

Supplemental Table 7-1: All proteins which were identified in the proteome experiments
Supplemental Table 7-2: Proteins of the outer envelope fractions
Supplemental Table 7-3: Proteins of the plastid associated membrane fractions