Basil J. Nikolau
Eve Syrkin Wurtele
Editors

Concepts in Plant Metabolomics

Springer

CONCEPTS IN PLANT METABOLOMICS

Concepts in Plant Metabolomics

Edited by

BASIL J. NIKOLAU
Iowa State University, Ames, Iowa, U.S.A.

and

EVE SYRKIN WURTELE
Iowa State University, Ames, Iowa, U.S.A.

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-5607-9 (HB)
ISBN 978-1-4020-5608-6 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

PREFACE

Metabolomics is a word that progress in science forces linguists to invent in order to keep up with emerging technologies. The word is a hybridization of two words, metabolites and genomics, and it reflects a shift in biological research that is now possible in an era in which the entire genetic blueprint of an organism is available for scientific research. Although the concepts of metabolomics are in the scientific literature since the 1970s, the word "metabolomics" was first used in the title of a scientific publication in 2001. Since then, the field of metabolomics has expanded and is becoming an integral sector of post-genomic research in biology.

Analogous to genomics, which defines all genes in a genome irrespective of their functionality, metabolomics seeks to profile "all" metabolites in a biological sample irrespective of the chemical and physical properties of these molecules. Despite the fact that this is probably an unachievable goal, the ability to profile an ever-increasing proportion of the metabolome (the set of all metabolites of a sample) has many applications is solving biological problems. These range from the expansion of the tradition of natural products chemistry, to the finding of metabolic markers of disease states in humans and animals.

In the field of plant biology, metabolomics has a key role as a fundamental tool in basic research for elucidating gene functions that are currently undefined. Thus, metabolomics has the potential of defining cellular processes as it provides a measure of the ultimate phenotype of an organism, as defined by the collage of small molecules, whose levels of accumulation is altered in response to genetic and environmentally induced changes in gene expression.

As an emerging field of science, new developments will greatly change the practice of metabolomics; these will likely occur in the area of improvement in analytical technologies and computational integration and interpretation of data. We hope that this book will present a guide for new practitioners of metabolomics, providing insights as to its current use and applications. These chapters are derived from presentations made at the 3rd International Congress of Plant Metabolomics, which was held in 2004 at Iowa State University, Ames, Iowa. We are grateful to the National Science Foundation, the National Research Initiative program of the US Department of Agriculture, and the Office of Basic Science of the Department of Energy, for financial support of this meeting. Finally, we would like to acknowledge the contributors to this volume, for their patience and efforts to ensure a high scientific quality. Specifically, we acknowledge the professional editing provided by Ms. Julie Lelonek, her help was invaluable in getting this volume completed.

November 2006

Basil J. Nikolau
Eve Syrkin Wurtele

CONTENTS

Preface v

Chapter 1. Validated High Quality Automated Metabolome Analysis of *Arabidopsis Thaliana* Leaf Disks
Quality Control Charts and Standard Operating Procedures 1
Oliver Fiehn

Chapter 2. GC-MS Peak Labeling Under ArMet 19
Helen Jenkins, Manfred Beckmann, John Draper, and Nigel Hardy

Chapter 3. Metabolomics and Plant Quantitative Trait Locus Analysis – The Optimum Genetical Genomics Platform? 29
Daniel J. Kliebenstein

Chapter 4. Design of Metabolite Recovery by Variations of the Metabolite Profiling Protocol 45
Claudia Birkemeyer and Joachim Kopka

Chapter 5. Uncovering the Plant Metabolome: Current and Future Challenges 71
Ute Roessner-Tunali

Chapter 6. Lipidomics: ESI-MS/MS-Based Profiling to Determine the Function of Genes Involved in Metabolism of Complex Lipids 87
Ruth Welti, Mary R. Roth, Youping Deng, Jyoti Shah, and Xuemin Wang

Chapter 7. Time-Series Integrated Metabolomic and Transcriptional Profiling Analyses
Short-Term Response of Arabidopsis Thaliana Primary Metabolism to Elevated CO_2 - Case Study 93
H. Kanani, B. Dutta, J. Quackenbush, and M.I. Klapa

Chapter 8. Metabolomics of Cuticular Waxes: A System for Metabolomics Analysis of a Single Tissue-Type in a Multicellular Organism 111
M. Ann D.N. Perera and Basil J. Nikolau

Chapter 9. Metabolic Flux Maps of Central Carbon Metabolism in Plant Systems
Isotope Labeling Analysis 125
V.V. Iyer, G. Sriram, and J.V. Shanks

Chapter 10. MetNet: Systems Biology Tools for Arabidopsis 145
Eve Syrkin Wurtele, Ling Li, Dan Berleant, Dianne Cook, Julie A. Dickerson, Jing Ding, Heike Hofmann, Michael Lawrence, Eun-kyung Lee, Jie Li, Wieslawa Mentzen, Leslie Miller, Basil J. Nikolau, Nick Ransom, and Yingjun Wang

Chapter 11. Identification of Genes Involved in Anthocyanin Accumulation by Integrated Analysis of Metabolome and Transcriptome in *Pap1*-Overexpressing *Arabidopsis* Plants 159
Takayuki Tohge, Yasutaka Nishiyama, Masami Yokota Hirai, Mitsuru Yano, Jun-ichiro Nakajima, Motoko Awazuhara, Eri Inoue, Hideki Takahashi, Dayan B. Goodenowe, Masahiko Kitayama, Masaaki Noji, Mami Yamazaki, and Kazuki Saito

Chapter 12. Identifying Substrates and Products of Enzymes of Plant Volatile Biosynthesis with the Help of Metabolic Profiling 169
Dorothea Tholl, Feng Chen, Yoko Iijima, Eyal Fridman, David R. Gang, Efraim Lewinsohn, and Eran Pichersky

Chapter 13. Profiling Diurnal Changes in Metabolite and Transcript Levels in Potato Leaves 183
Ewa Urbanczyk-Wochniak, Charles Baxter, Lee J. Sweetlove, and Alisdair R. Fernie

Chapter 14. Gene Expression and Metabolic Analysis Reveal that the Phytotoxin Coronatine Impacts Multiple Phytohormone Pathways in Tomato 193
Srinivasa Rao Uppalapati and Carol L. Bender

Chapter 15. Profiling of Metabolites and Volatile Flavour Compounds from Solanum Species Using Gas Chromatography-Mass Spectrometry 209
Tom Shepherd, Gary Dobson, Rhoda Marshall, Susan R. Verrall, Sean Conner, D. Wynne Griffiths, Derek Stewart, and Howard V. Davies

Chapter 16. Metabolomic Analysis of Low Phytic Acid Maize Kernels 221
Jan Hazebroek, Teresa Harp, Jinrui Shi, and Hongyu Wang

Chapter 17. The Low Temperature Metabolome of *Arabidopsis* 239
Gordon R. Gray and Doug Heath

Chapter 18. Cloning, Expression and Characterization of a Putative Flavonoid Glucosyltransferase from Grapefruit (Citrus Paradisi) **Leaves** 247
Tapasree Roy Sarkar, Christy L. Strong, Mebrahtu B. Sibhatu, Lee M. Pike, and Cecilia A. McIntosh

Chapter 19. Application of Metabolite and Flavour Volatile Profiling to Studies of Biodiversity in Solanum Species 259
Gary Dobson, Tom Shepherd, Rhoda Marshall, Susan R. Verrall, Sean Conner, D. Wynne Griffiths, James W. McNicol, Derek Stewart, and Howard V. Davies

Chapter 20. Metabolic Profiling Horizontal Resistance in Potato Leaves (cvs. Caesar and AC Novachip) Against *Phytophthora Infestans* 269
Y. Abu-Nada, A.C. Kushalappa, W.D. Marshall, S.O. Prasher, and K. Al-Mughrabi

Chapter 21. *In Vivo* ^{15}N-Enrichment of Metabolites in *Arabidopsis* Cultured Cell T87 and Its Application to Metabolomics 287
Kazuo Harada, Ei-ichiro Fukusaki, Takeshi Bamba, and Akio Kobayashi

Chapter 1

VALIDATED HIGH QUALITY AUTOMATED METABOLOME ANALYSIS OF *ARABIDOPSIS THALIANA* LEAF DISKS

Quality Control Charts and Standard Operating Procedures

Oliver Fiehn
UC Davis Genome Center, Health Sci. Drive, Davis, CA 95616, USA

Abstract: Plants readily respond to changes in environmental conditions by alterations in metabolism. In addition, breeding processes as well as modern molecular tools often target at or result in constitutive changes in metabolite levels or metabolic pathways. These properties render metabolomics an ideal tool to characterize the degree of impact of genetic or environmental perturbation. In agronomic and agrobiotechnology, but also in some areas of fundamental plant biology research, this leads to experimental designs of *genotype × environment* (*G×E*) plots, which results in huge numbers of individual plants to be grown, harvested, processed, and analyzed. The benefit to add metabolomics is then to utilize analyses of metabolic events to better understand biochemical or regulatory mechanisms by which the plant responded to the *G×E* perturbations. However, technical challenges are still imminent regarding the complexity of plant metabolism and the need for high quality control in large projects. This chapter details how even larger projects with thousands of analyses can be managed in an academic laboratory while still keeping control over the total process by use of Standard Operating Procedures (SOP) and continuous Quality Control (QC) measures. This process is exemplified by SOP and QC implementations used for a larger study on effects of abiotic treatments on select Arabidopsis ecotype accessions.

1 INTRODUCTION

1.1 Theoretical considerations

Metabolomics aims at achieving qualitative and quantitative metabolite data from biological samples grown under a specific set of experimental conditions (Fiehn et al., 2000; Bino et al., 2004). In order to interpret and

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 1–18.

reuse data (*via* metabolomic databases), the sources of quantitative variability of data must be accurately described. Technical errors of the analytical process must be controlled and minimized in order to distinguish such variance in data (noise) from the inherent biological variability within and between the populations that are subjected to a certain experimental design. This chapter describes why and how Standard Operating Procedures and Quality Control charts are needed for larger metabolomic projects.

1.2 No data without metadata

The process of metabolomic analysis involves many steps from the actual experimental design of the biological trial to the conditions of plant growth and potential treatments by external factors such as changes in abiotic or biotic stressors, and following plant responses within temporal or spatial patterns, e.g., over plant organ development or within diurnal cycles. Reproducibility and reusability of metabolomic data sets therefore necessitate capturing this underlying information about the details of the total experimental design: without this, no data set can be understood and interpreted in a correct way. Such "data about the data" are called "metadata" in computing sciences and are at least in parts described and collated in the "materials and methods" sections of plant journals. However, in such sections plant biologists tend to focus on the novel parts of their experimental setup and do not give fully precise descriptions on more standard growth specifications. For example, unless researchers carry out specific light treatment studies, the light qualities (emission spectra) within green houses or climate chambers are usually not detailed out. The same is most often true for the type and dimensions of the climate chamber used, although it is known that each climate chamber has its own specifics with respect to air circulation conditions, which will ultimately affect water evaporation rates from the soil and by this, plant metabolic rates. One might argue that such description is overly detailed, but on the other hand, for each institution such information would only need be recorded once and then deposited as an object number in a database for future experiments.

The need to accompany metabolic data with exact experimental metadata is also given by the fact that each plant species and even each organ comprises a wealth of unannotated or unknown metabolites which will only reveal their specific importance when tracing back their relative levels under a multitude of conditions. Unlike other cellular components such as primary and secondary gene products (transcripts and proteins), most metabolites do not carry annotated biological functions which relate to well-described unique biological roles. For a few secondary metabolites like auxins or glucosinolates such roles are known for controlling plant growth or

herbivore defense but for most, especially for primary metabolites, multiple functions must be assumed. The complementary addition of (experimental) metadata is therefore a very important and necessary element when metabolomic data sets are to be stored in public repositories. One aspect of the results reported here is the development of required and optional metadata entries for typical plant metabolomic experiments, with the ultimate aim to enhance public access and reusability of such data sets.

1.3 Metabolomic methods require validation

The other aspect is to elaborate the details of the experimental procedures between the point of plant harvests and the result data output. Just like details in plant growth affect interpretation of results, so do also differences in the sample processing workflow impair comparisons between sets of data. Differences in equipment render it difficult to result in fully identical results between laboratories, but at least the process within a specific laboratory must be tightly defined and monitored to allow high reproducibility and re-usability of data. It is difficult to achieve long-term reproducibility due to a number of reasons: different staff may be responsible for sample work-up, each performing the dues somewhat differently, protocols may be understood and used in different ways (pointing to lack of training and supervision), suppliers of solvents, reagents, and consumables may lack tight specifications or change product characteristics without notification, the analytical instrumentation itself may be subjected to contamination or instability of sensitivity and selectivity (pointing to lack of ruggedness), and eventually data processing may be carried out in unexplained or varying ways, which may be the case for both, "relative" and "absolute" values. Most of these points are not relevant for small demonstration studies that only involve some 50 samples, since such projects will not take longer than a week and will be carried out by a single scientist. However, if hundreds of recombinant inbred lines or other genetic populations are to be compared and results are to be disseminated *via* public databases, far more rigid constraints have to be imposed. These constraints call for "validation" of the total process.

Validation in itself is a term that is often misunderstood. Krull and Swartz (1999) clarified that validation of (analytical) processes is needed not only for industry but equally important for the academic laboratory. The point is that validation means "valid for a purpose", and a valid method therefore needs first and for all a clear description of the purpose for which it is intended to be used. In many scientific papers, the difference is not made clear between method development and method validation: method development describes the steps which have been taken to evolve a process that led to a specific (analytical) result and ultimately to a protocol.

However, validation means more than that: a valid method would require that a certain result is always gained if defined processes are applied to a specified problem.

Similar to the difference between a developed and a validated method is the distinction between a laboratory protocol and so-called 'Standard Operating Procedures' (SOPs). Protocols may lack a number of details because the protocol developers deemed these to be common sense, or because they were unaware of their importance which is more often the case. For example, a protocol suggests adding 200 µL of a buffer solution to a tube. An SOP, however, would detail that these 200 µL need be taken by a pipette that has undergone a defined calibration check at regular intervals, and that signatures are required that these calibration checks have actually taken place. In fact, pipette volume accuracy is a critical factor that is often underestimated, and also the staff skills to routinely and correctly estimate such volumes. Even without going as far as an SOP, it is good laboratory practice to exercise calibrations (by weighing the volume of pure water at defined temperature) in regular terms by all staff members and for all pipettes. For the 200 µL buffer example given above, an SOP would further detail how the buffer solution was prepared: the water quality and its source, the manufacturers and brands of the buffer components, and the actual preparation procedure.

It is important to mention, however, that SOPs (a) must not be over-detailed and (b) that they must undergo regular inspection and checks against the real laboratory practices. This means that only the parts of the procedures are detailed that may actually hamper the overall results – which is determined through the "validation" process. Therefore, it is important to accurately check all aspects of the developed (analytical) method with respect to ruggedness, i.e., how smaller or larger deviations from the details of the procedure affect the result data. For example, a protocol may say a sample is shaken for 20 s by a vortexer. An SOP might even be less rigid by detailing that this mixing could take place between 10 s and 30 s (because the actual mixing time would cause no significant difference to the results), but it might add that the mixing would need be done at room temperature between 18°C and 28°C (and not, say, in the 4°C cold room). Furthermore, results may indicate after a certain time that an SOP needs revision, or the laboratory manager discovers that a certain laboratory practice has never been mentioned before. Then, a new SOP is written which supersedes the old one and which explains which parts have been altered.

1.4 Quality Control of data acquisition (QC)

Many metabolomic research papers emphasize the details of a specific instrument, e.g., the type of mass spectrometer or NMR instrument used for data acquisition. In fact, however, this should be less important than details

how the quality of measurements were ensured. For example, relative levels of amino acids can be analyzed by various means such as HPLC-fluorescence detection or ^{1}H-NMR or GC-MS, if the chosen data acquisition method suffices the specified sensitivity and selectivity. However, independent from the type of analytical instrument that is used are routines to ensure and prove long-term data precision. Some instruments may undergo a systematic drift in sensitivity over time; others may lose specificity for only certain compounds without affecting other metabolites, and certain instrumentation may simply lack robustness, producing highly oscillatory and hardly controllable measurements. All such errors cannot be evaluated in a single analytical sequence, be it 10 measurements or 100. Such trends will only be observable if identical samples are continuously subjected over long periods to (a) the total result of plant metabolome comparisons, (b) the total analytical method, or (c) the data acquisition process only. These three criteria are independent from each other. For example, an instrument may be perfectly in-control whereas sample preparation (grinding of leaf tissue followed by extraction, fractionation, drying, and derivatization) might cause high deviations. Going further, even the sample preparation step may be fully robust but plant growth conditions in the greenhouse or climate chamber might be altered by external factors (e.g., exchange of light bulbs, watering frequencies, temperature control, etc.). Depending on the purpose defined in the validation terms, different strategies may be adopted for maintaining robustness of the analytical results. In each case, the results must be monitored in so-called Quality Control charts that allow observing trends over time and deviations from upper and lower intervention limits: within these limits, the analytical process is in-control, but once these limits are crossed (out-of-control), results must be marked as unreliable and measures must immediately be taken (such as instrument maintenance or staff training) to bring the process back to control.

1.4.1 Quality Control calibration curves

Firstly, it must be ensured that the instrument itself does not cause systematic or large random deviations from routinely acquired data. In metabolomics, this can be ensured using compound mixtures covering all chemical classes of the typically analyzed metabolites (which were defined in the analytical scope before the validation process took place). If important classes of compounds cannot be analyzed within a given validated process, than the results by definition cannot be termed “metabolomics” but rather “metabolite profiling”. In addition to the requirement of selectivity of a metabolomic (or metabolite profiling) method, hence the ability to distinguish individual metabolites from other (matrix) components, it is equally important to control the analytical sensitivity, i.e., the increase in signal intensity upon increase in metabolite levels. The best way to perform

such quality control tests is by (daily) acquisitions of internal calibration curves, i.e., spiking a mixture of compound into a given matrix at concentrations spanning the whole range of typically detected levels in plants. Obviously, these compounds must not be present in the matrix before. This criterion can best be met using stable isotope labelled metabolites. If these are not available, alternatively external calibration curves may be employed by determining the instrument's sensitivity by analyzing varying QC mixture levels alone, without matrix. Such QC mixture analyses need be accompanied by reagent blank controls, i.e., measurements that only comprise the containers, reagents, or solvents used directly in conjunction with the analysis. Such reagent blank controls qualify which analytical signals may be assigned as "blank contaminants". Ideally, not a single peak should be detectable in such blanks. However, reality shows that this is generally not the case due to chemical impurities of reagents or ubiquitous (laboratory) contamination such as phthalate plastic additives originating, e.g., from pipette tips.

1.4.2 Plant sample "reference design"

The above-mentioned QC mix and reagent blank controls do not control for differences in sample preparation over time. Given the need to monitor this part of the process, a small set of identical plant specimen may be subjected to the homogenization, extraction, and fractionation procedure every day. However, as plant physiologists understand, there are no two fully identical plants. Even individual plants of homozygous lines grown under controlled standard growth conditions will show deviations from an assumed ideal mean (sometimes called "steady state"). Numerous factors account for this phenomenon, among which can be named small differences during germination (which may cause slight differences in growth rates and thus metabolism), subtle deviations of climate conditions (e.g., caused by position effects within the growth location) and, theoretically more fundamental, the basic properties of metabolic networks themselves which have been shown to amplify small oscillations of metabolic levels (e.g., external glucose fed into glycolysis) to larger metabolic deviations downstream along the pathways (Steuer et al., 2003; Weckwerth et al., 2004). For this reason, it is not easy to obtain identical plant specimen to control for potential sample preparation errors. One way to remedy this is by referring each metabolite value of experiments against data of a mixture of a control plant specimen that is always grown concomitant with the experiments, independent from the actual experiment design. The problem here is the large difference between (Arabidopsis) accessions. If, for example, C24 was chosen as "reference line", compounds that only occur in L*er* or Cvi could not be referred to. On the other hand, such a reference

design would not only account for differences in sample preparation but also for larger differences in plant growth conditions, and in this, be helpful for inter-laboratory comparisons.

1.4.3 Plant sample "master mix" design

An alternative is to utilize "master mixtures" of plant extracts that all originate from a specific (large) experiment. The idea would be here to aliquot a small fraction (e.g., 10 µL) of each plant extract into a larger container that would pool all processed samples of each day. A plant specimen randomization schema is then needed to dictate that each Arabidopsis ecotype accession (natural variant) included in the larger experiment would be included in this daily "master mix". Quality Control charts of these pools would then monitor deviations of the analytical results from these pool mixtures and compare them to the errors caused by the analytical instrument itself. The advantages of this procedure against a true "reference design" given above are threefold: (a) The total analytical error can be factored out into two distinct parts, the instrument error and the sample handling error. This facilitates decision making in where to put further emphasis in method refinements. (b) All metabolites (above a certain signal/noise threshold) would be monitored that are included in the Arabidopsis ecotype metabolomes under study. This alleviates the problem of using a specific accession that may not even be relevant for the experiment under study. (c) The largest advantage is that such a master mix design allows analysing plant samples that were grown elsewhere, i.e., where the (biological) design was not under control of the metabolomic laboratory. Such a situation is a daily reality in academia where colleagues ask the metabolomics specialist for collaboration in a certain project for which the growth is already complete or for which use of a specific reference accession would cause unacceptable complications.

Whatever chosen for a specific setting, i.e., either "reference design" or "master mix design", in any case method control blanks must be added. Such method control blanks are defined by utilizing all utilities, instruments, solvents, containers, and procedures in exactly the same way like the plant samples are treated, just without any sample in it. Method blank controls qualify which analytical signals may be assigned as "method blank contaminants" by comparison with the reagent blank controls. Ideally, not a single peak should be detectable in such blanks that are not also present in the reagent blank controls. However, reality shows that this is generally not the case due to the additive nature of laboratory contaminations (such as dishwasher detergents) and differences in lot qualities of solvents, tips, and glassware.

2 PRACTICAL IMPLEMENTATION OF QC AND SOP

2.1 ArMet database framework

Aforementioned considerations may be regarded as general recommendations for plant metabolomics experiments which may be subject to individual realizations depending on the biological focus or the techniques employed in a specific (academic) setting. Recently, an overall data model on how to capture the different components of such experiments has been suggested in order to enhance comparability and reusability of data, called ArMet (Architecture for Metabolomics). ArMet consists of nine basic modules (Jenkins et al., 2004):

1. ***Administration***: Informal experiment description and contact details.
2. ***Biological Source***: Genotype, provenance and identification data for items of biological source material.
3. ***Growth***: Description of the environments in which the biological material developed.
4. ***Collection***: Procedures followed for gathering samples from items of biological source material.
5. ***Sample Handling***: Handling and storage procedures following collection.
6. ***Sample Preparation***: Protocols for preparing samples for presentation to analytical instruments.
7. ***Analysis Specific Sample Preparation***: Protocols specific to particular analytical technologies.
8. ***Instrumental Analysis***: Process description of the chemical analysis of samples, including descriptions of analytical instruments and their operational parameters, quality control protocols, and references to archive copies of raw results.
9. ***Metabolome Estimate****:* The output from the analytical instruments after it has been processed from raw data to produce a metabolome description and metadata about its processing.

This model lays out the basic framework which general components need to be addressed and how ArMet compatible databases need to be structured. However, nothing has been standardized so far with respect to ontologies or controlled vocabularies to be used, and even less is agreed on which specific SOPs or QC measures need to be taken. In subsequent paragraphs, an example is given how this specific information content was implemented. All the metadata need to be acquired before the actual metabolomic analyses start.

2.1.1 Administration

Usually, more than one staff and often, more than one principal investigator (PI) participates in a larger biological experiment. The *Administration* component may include all investigators and staff involved, however, this is impractical in daily routines and may not even been fully known in all cases. The philosophy here is to enable back tracing the origins of samples and potential observations after data sets have been (statistically) evaluated. A useful implementation might therefore ask for the main biological "owner" of the experiment with full name, affiliation and email address and the responsible chief staff in the PI's laboratory who actually monitored plant growth and sample collection.

2.1.2 Biological source

In this component the genotype and pedigree of the plant sample is described. For the case of published mutants, database accessions must be given. For ecotype accessions (natural variants), reference to commonly used names are suitable. Further details of origins are required if the seeds were not garnered in the PI's growth locations but sent from elsewhere, e.g., germplasm seed stocks or collaborating institutions. For the case of mutants, the parental genetic background line must be given. For crosses and recombinant inbred lines, both parent lines need be named. For other lines (which may evolve in complex cultivar breeding programs), the closest isogenic relative(s) or cultivar identification codes must be given.

2.1.3 Growth

Plant growth is a complex, lengthy, and often variable process. It is almost impossible to capture all details along the development of individual plants. Furthermore, if all imaginable growth metadata were required to be collected and stored, collaborative efforts in large research consortia would likely be hampered. Ultimately, the details of the growth component often comprise a large part of the experimental design which could eventually involve many fragmented steps. Therefore, a single and inflexible metadata import schema is inadequate. A useful implementation may therefore require only very basic objects which should be relevant to >90% of typical plant biology experiments and ask for more details to be placed in string text.

(a) Name of growth location e.g. (InstituteGreenhouse#01).
(b) Sowing and transplanting date (mm-dd-yyyy).
(c) Standard growth conditions before treatment: Medium (GS standard soil or Agar, etc.), Temperature (20°C day, 18°C night), Light (16/8 H, 240 μMOL M^{-2} S^{-1}), and Humidity (80%).

(d) Treatment specifics, if applicable: when the treatment was performed (PLANT GROWTH STAGE, DATE and TIME), what kind of treatment was applied (string text such as "4°C cold acclimation") and how long the treatment was applied (DD-HH).

The current suggestion leaves the details of the overall experiment and the treatment specifics into (non-queryable) string text comment fields. These details may later be collated to higher levels of hierarchies (such as cold stress – abiotic perturbation, Cook et al., 2004) in database curation, but are not required before the metabolome experiment starts.

2.1.4 Collection

The collection component necessitates three entries:

a) harvest date and time (mm-dd-yyyy, hh).
b) harvested organ(s) and organ specifications (LEAF, ROSETTE).
c) harvesting procedure (string text such as "cork borer 4 mm id., immediate freezing in liquid nitrogen").

For the harvested organ descriptions, automatic spelling corrections, and public ontologies are needed (such as www.plantontology.org). The harvesting procedures may best be described using an SOP which was not employed for the example experiment.

2.1.5 Sample handling

The original metabolic composition must be ensured from the time of harvest to the actual analysis. Both biological and physicochemical factors may alter the metabolome during storage and handling:

(a) Chemists tend to underestimate the metabolic turnover rates of enzymes. If plant tissues are kept unfrozen, and even if they are partly thawed during sample preparation, enzymatic activity is rescued and metabolism starts. Therefore, samples need be kept deep-frozen at all times until extraction. No reports are published about how long Arabidopsis leaf metabolome integrity is preserved during storage. As general precautionary measure, samples should not be stored longer than 4 weeks at −80°C or longer than 2 weeks at −20°C.
(b) Biologists tend to underestimate the effects of oxidation and light treatment. Oxygen is a diradical that will act independently from enzymes and therefore also in the frozen state. Samples must thus be stored under argon or nitrogen in order to preserve internal redox state metabolite markers (such as cysteine/cystine, asorbate/dehydroascorbate, or glutathione/GSSG). Of minor importance is prevention from (excessive) light. Some molecules such as catecholamines or aromatic amines will undergo conformation changes or oligomerization when treated with light. Catecholamines are not found in

Arabidopsis leaves but in *Solanum tuberosum* leaves, however, in principle some Arabidopsis metabolites might also be affected.

2.1.6 Sample preparation

The details of sample preparation will undoubtedly affect metabolomic readouts. However, it should be kept in mind that there is no optimal or final single metabolomic method: metabolomic methods try to be as comprehensive as possible despite the vastly different chemical properties of primary and secondary metabolites. Therefore, the chosen sample preparation method can only be regarded as good compromise that fulfills its scope and which can therefore be validated for a given purpose. One of the most important criteria in the validation process is comprehensiveness (the breadth of different chemical compound classes being detected) and precision (the repeatability of analytical results), but not accuracy (correctness of a specific metabolite in absolute concentrations). Therefore, a sample preparation method may be valid despite its lack of recovery of, for example, a specific plant hormone. If this specific compound needs to be included, either a specific "metabolite target" method is developed or the existing metabolomic method is altered (which may eventually result in loss of other compounds).

For homogenization and extraction of Arabidopsis leaf metabolites, the following SOP is used at the metabolomics core of the UC Davis Genome Center. Note, that procedures for sample collection are documented but not solely allowed in connection with this SOP.

2.1.7 Analysis-specific sample preparation

In principle, sample extracts from the extraction SOP can be used for different analytical instruments such as NMR or GC-MS. Therefore, any further sample preparation steps must be regarded separate from the extraction. Subsequent steps such as "derivatization for GC-MS measurements" need again be detailed in an SOP. Many of the derivatization parameters do not have a dramatic influence on the overall result in GC-MS, if the conditions and parameters are hardly controlled. However, this is usually not the case. In all GC-MS instrument setups reported so far, samples were manually derivatized in batches and then placed on autosamplers prior to injection. The time between addition of GC-MS reagents and the actual measurements therefore remained uncontrolled, although it is known that certain compounds (especially amines and amino acids) undergo further reactions during this time. In a more tightly controlled SOP it is therefore reasonable to use a robotic system for automatic derivatization and injection. Such a system also allows automatic addition of

Table 1-1. Example SOP for 'extraction of Arabidopsis leaf tissue disks'

UC Davis Metabolomics Core	SOP Standard Operating Procedure	
date: 01/15/2004	Extraction of Arabidopsis leaf tissue disks	Code no.: 002_2005a
Issued: 01-15-2005 Valid from: 01-16-2005		Validity area: UC Davis - Metabolomics Core -
Responsible: Tobias Kind		
This SOP supersedes: SOP 002_2003a MPI Golm		Checked: Oliver Fiehn

Extraction of Arabidopsis leaf tissue disks

1. References: Weckwerth W., Wenzel K., Fiehn O. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks *Proteomics* 2004, *4,* 78–83

2. Starting material:
15-30 mg fresh weight, 1-3 mg dry weight Arabidopsis leaf disk, approx. 4 mm I.D.

Sample collection leaf tissue
Before sampling, take digital photo of plants to be harvested, indicating the target sample ID#. Take one or two disks of a fully mature rosette leaf from the leaf center by using the cork borer (diameter of cylinder will depend on the targeted amount of leaf material). Transfer the disk(s) immediately to an Eppendorf tube, equipped with a grinding metal ball and labeled by sample ID #. Close the Eppendof quickly and place it in liquid nitrogen.

Samples may be taken in other ways but procedures need be documented and accessible.

3. Equipment:
- Grinder (ball mill MM 200, Retsch corp.)
- Centrifuge Eppendorf 5417 C
- calibrated pipette 1000 µl (check conformity SOP 007_2005a)
- cork borer (diameter related to target disk weight. 4 mm I.D. appropriate for Arabidopsis leaves.)
- metal balls for grinder
- fine balance accuracy ± 0.1 mg
- Safe lock micro test tubes 2 mL, uncolored, (order no. Eppendorf corp.0030 120.094)
- Crimp V-vials glas 1ml, (order no. Fisher Scientific corp. 3102008)
- Julabo corp. cooling bath
- Vortex mixer/ stirrer, Scientific Industries corp.
- Thermo mixer HLC TM 130-6
- Speed vacuum concentration system, Heto corp.
- Large tweezers

4. Chemicals
- Methanol LC-MS Chromasolv, SAF order no. 34966
- Chloroform Chromasolv, SAF order no. 25685
- pure water "Purelab Plus" (Alternatively take LC-MS Chromasolv water SAF order no. 39253)

Table 1-1. (continued)

UC Davis Metabolomics Core	SOP Standard Operating Procedure	

- two dewar vessels filled with liquid nitrogen
- pH paper 1-14, Macherey-Nagel order no. 92110

5. Procedure

5.1. Preparation of extraction mix and material before experiment:
⇒ check pH of MeOH, CHCl3, and water (pH7) by adding one drop of pure water to pH paper, then one drop of solvent. Attention: this is not an accepted pH measurement by pH definition but suffices quality check here.
⇒ H2O, MeOH, CHCl3 is mixed in volumes in proportion 1 : 2,5 : 1.
⇒ rinse the extraction solution mix for 5 min with argon or gaseous nitrogen with small bubbles, for example using HPLC-solvent filters or pumice stones for home fish tanks.
⇒ switch on bath to pre-cool at –15°C to –20°C (validity temperature range).

5.2. Homogenization and extraction
Pre-chill two Eppendorf tube holders in liquid nitrogen for > 60 s. Then, take out six Eppendorf sample tubes from liquid nitrogen and place these into the Eppendorf-holder of the grinder, each three samples per tube holder. Take care to compensate for weight, maintaining equilibrium. Shake holders for 30 s with a frequency of 25 s^{-1}. Afterwards, immediately place Eppendorf tubes back into the second liquid nitrogen dewar. When all samples are ground and homogenized, take the tubes one by one out of the liquid nitrogen using the long tweezers and immediately add 1 ml of pre-chilled extraction solvent mixture (-15 to -20°C). Even partial thawing of samples must be prevented. Vortex for 10-20 s and shake for 4-6 min at 4°C (use thermo shaker in the 4°C cooling room). Samples may be stored on crushed ice (chilled at <0°C with NaCl) between vortexing and chilling for up to 10 min. After shaking, centrifuge samples for 2 min at 20,200 cfg at room temperature. Take out 800 µl supernatant into a labeled Eppendorf tube and vortex the tube for 5-10 s. (The remainder residue containing cell debris is discarded.) Take out 10 µl aliquot into the 'master mix pool'. Take out 400 µl into a round bottom 1 ml crimp glass vial. Store the residue aliquot under argon or nitrogen at -80°C. Dry the 250 µl sample aliquot in the speedvac concentrator to complete dryness. Once dry, store samples in darkness under argon or nitrogen at -20°C prior to analysis. Don't store longer than 6 weeks.

6. Problems
In order to prevent contamination, disposable material is used. Check pH of extraction solvent mix! Take care that Eppendorf tubes are completely closed before placing them in liquid nitrogen. Otherwise, liquid nitrogen will immerse into the tube and cause disruption once put back to room temperature.

7. Quality assurance
The method is invalid without at least one method blank control per 40 samples to which the total procedure was applied (i.e., employing all steps, materials and plastic ware), just leaving out the leaf tissue disk.

8. Waste disposal
Collect all chemicals in appropriate bottles and follow the disposal rules.

internal standards to control for injection errors and retention time shifts. A suitable SOP might use the following procedure: For derivatization and injection, use a robotic system with ovens, syringes, and manifolds to handle reagents and vials. Add 1 µL of a mixture of retention time standards and isotope labeled standards to the dried sample and immediately afterwards, 10 µL methoxyamine in pyridine (40 mg/mL). The solution is shaken at 28°C for 90 min before 90 µL MSTFA is added. The reaction solution is shaken at 37°C for 30 min and placed back to a waiting tray at room temperature. Each sample is injected exactly 3.5 h after addition of MSTFA. A macro program ensures that all steps are intimately linked.

2.1.8 Instrumental analysis

The ArMet database framework suggests distinguishing between the analysis specific sample preparation and the instrumental analysis. In the demonstration case presented here this is obviously not needed, but the different parts of the SOP can surely point to different database objects. The instrumental setup used here is specialized in that it also allows an automatic

Figure 1-1. Image of a robotic system for sample derivatization, liner exchange, and gas chromatography/time-of-flight mass spectrometry (GC-TOF).

liner exchange and cold injection into the GC-MS instrument. This ensures that each sample is subjected to a clean liner in order to avoid matrix carry over between samples. Especially for unsaturated free fatty acids but also for aromatics and phenolic compounds such a setup is needed for high-quality contamination-free analyses. Usual liners include some glass wool to prevent involatile material reach the chromatography column. The specialized liners used here include a microvial in which 1.5 µL of sample is injected at 40°C. Subsequent flash heating vaporizes all volatile components which are subsequently separated by gas chromatography, whereas all involatile components remain at the bottom of the microvial. The liner is then automatically exchanged against a new liner (with microvial) for the next injection. After use, microvials are discarded whereas liners are cleaned and reused.

The injection is the most critical part in GC-MS. Most other settings such as heat ramping rate, the actual separation column used, and even the type of mass spectrometer will usually not dramatically affect the overall metabolomics aims, i.e., comprehensiveness and precision. However, the injection process may do so as is outlined in the next section. The last ArMet component, the Metabolome Estimate, will also be discussed in brief in this section.

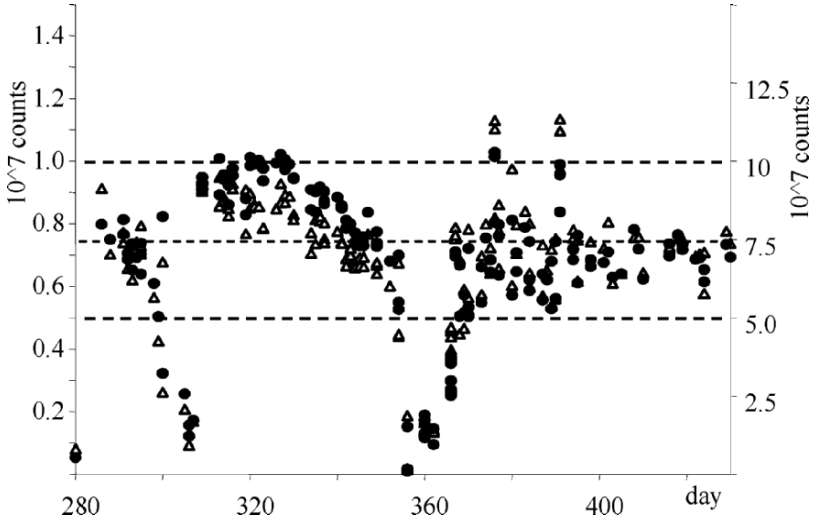


Figure 1-2. Long-time quality control chart for QC mix in GC-TOF mass spectrometry. Absolute values for ribitol (filled circles, left axis) and putrescine (open triangles, right axis) are notified in daily injections for absolute intensities at m/z 174+319. Dotted lines: upper intervention limit, mean, and lower intervention limit.

2.2 Quality control charts to monitor system suitability

The different SOPs outlined above are a result of monitoring the overall system performance over long times. The instrument Quality Control mixture contains 28 compounds (amines, amino acids, aromatics, hydroxy acids, mono- , di- and trisaccharides, sugar alcohols, and sterols) which are used for immediate recognition of some of the more frequent problems in high throughput GC-MS. This QC mix is daily analysed in six dilution steps.

(a) Detector response is monitored by absolute peak areas, e.g., ribitol and putrescine. These two chemically very different compounds elute in the midrange of the chromatogram and are not affected by injector boiling point discriminations. An example QC chart is given in Figure 1-2. Note that a lag time was used to get experience for suitable intervention limits. These limits were calculated by two-sigma standard deviations of the absolute peak areas during 100 subsequent days of operation, half a year after the instrument had been installed. Whenever the intervention limits were crossed, maintenance measures were taken to bring the process back into control, and no plant samples were run at that time.

(b) System selectivity is monitored by three different parameters: (i) ratios of putrescine to ribitol. Amine-silyl bonds are far weaker and decompose more easily in case any problem with injector quality occur (e.g., dirt, contamination). In dramatic cases, and especially at low levels, amines may completely get lost, but before, problems become imminent by altered ratios of amines to carbohydrates (e.g., putrescine to ribitol, fig. 1-3). (ii) peak height of maltotriose. High boiling point compounds are immediately affected by any kind of column contamination (fig. 1-4). A suitable maintenance measure is to cut the column by 10 cm. Therefore, 10 m empty guard columns are used in conjunction with the separation column. (iii) peak asymmetry of alanine. Low boiling point compounds are adversely affected by alterations in injector pressure regulation, especially when binary solvent systems are used like pyridine/MSTFA.

(c) System sensitivity is monitored for all compounds by alterations in slopes of the calibration curves. This may point to problems with liner lot quality delivered by the manufacturer.

3 CONCLUSIONS

Many research groups and initiatives have been started in the area of (plant) metabolomics but yet, comprehensive databases and metainformation are still to come. In this chapter, reasons have been outlined why this may be

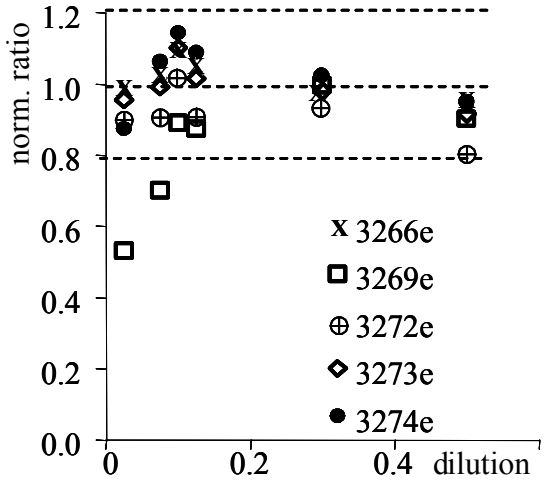


Figure 1-3. Quality control chart for liner discrimination of low abundant amines. Normalized relative ratios of putrescine/ribitol from injection sequence 3266–3274. Black dotted lines: upper intervention limit, mean, and lower intervention limit.

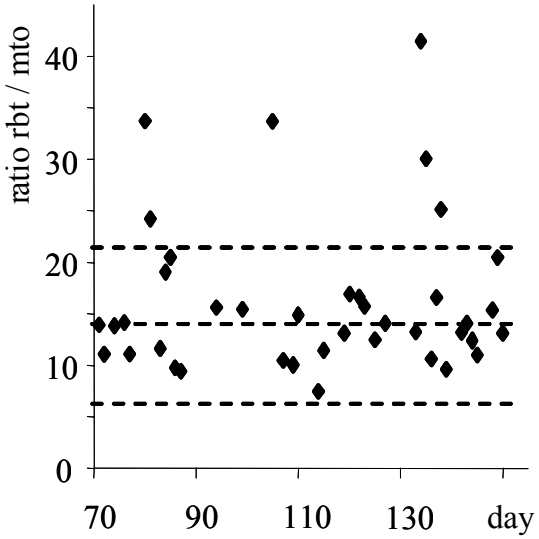


Figure 1-4. Quality control chart for discrimination of high boiling (maltotriose, mto) against mid boiling compounds (ribitol, rbt) caused by matrix depositions in injector, liner and/or column. Black dotted lines: upper intervention limit, mean, and lower intervention limit.

the case and which steps can be taken to circumvent, monitor, and control problems. As an example case, capture of metadata, and Arabidopsis leaf disk extraction prior to GC-MS analysis was taken, but similar long-time procedure checks and quality controls need be taken in other instrumental approaches. Eventually, databases giving access to experimental data will need to be accompanied by information on both biological and analytical metadata. Only by such measures can comparability and exchange of data be ensured, which is certainly true not only for metabolomics but also related areas such as proteomics and transcriptomics.

ACKNOWLEDGMENTS

I would like to acknowledge the assistance of Änne Eckardt who has worked as quality control manager in my laboratory from 2001 to 2004. Discussions with Helen Jenkins and Nigel Hardy, computer scientists at the University of Wales, Aberystwyth, have inspired this work, especially about the importance to capture metadata. I am thankful to Sandy Primrose who forced the implementation of SOPs into metabolic work throughout the G02 program on GM foods, funded by the Food Standards Agency, UK.

REFERENCES

Bino, R., Hall, R., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B., Mendes, P., Roessner-Tunali, U., Beale, M., Trethewey, R., Lange, B., Wurtele, E., and Sumner, L., 2004, Potential of metabolomics as a functional genomics tool, *Trends Plant Sci.* **9**:418–425.

Cook, D., Fowler, S., Fiehn, O., and Tomashow, M., 2004, Role of the CBF cold response pathway in configuring the low temperature metabolome of Arabidopsis, *Proc. Natl. Acad. Sci. USA* **101**:15243–15248.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L., 2000, Metabolite profiling for plant functional genomics, *Nat. Biotechnol.* **18**:1157–1161.

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A., Taylor, J., Fiehn, O., Goodacre, R., Bino, R., Hall, R., Kopka, J., Lane, G., Lange, B., Liu, J., Nikolau, B., Mendes, P., Oliver, S., Paton, N., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgard, J., Sumner, L., Wang, T., Walsh, S., Wurtele, E., and Kell, D.B., 2004, A proposed framework for the description of plant metabolomics experiments and their results, *Nat. Biotechnol.* **22**:1601–1605.

Krull, I.S. and Swartz, M., 1999, Analytical method development and validation for the academic researcher, *Anal. Lett*. **32**: 1067–1080.

Steuer, R., Kurth, K., Fiehn, O., and Weckwerth, W., 2003, Observing and interpreting correlations in metabolic networks, *Bioinformatics* **19**:1019–1026.

Weckwerth, W., Ehlers, M., Wenzel, K., and Fiehn, O., 2004, Metabolic networks unravel the effects of silent plant phenotypes, *Proc. Natl. Acad. Sci. USA* **101**:7809–7814.

Chapter 2

GC-MS PEAK LABELING UNDER ARMET

Helen Jenkins[1], Manfred Beckmann[2], John Draper[2], and Nigel Hardy[1]
[1]Department of Computer Science and [2]Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, Wales, UK, SY23 3DB

1 INTRODUCTION

ArMet (Jenkins and Hardy, 2004) is a data model for plant metabolomics. It provides a framework for representing the data associated with metabolomics research. A metabolome is the set of metabolites produced by an organism (Oliver and Winson, 1998), where metabolites are the end products of cellular regulatory processes (Fiehn, 2002). MIAMET (Bino and Hall, 2004) is a suggested checklist of the information necessary to provide context for metabolomics data but is not a formal description of a data model, such as is necessary to develop supportive data handling systems. ArMet is such a model and may be used in a variety of ways, including as a basis for the design of systems to store or transport data on metabolomics experiments. It may also serve to establish and promote standards for metabolomics experiment description. Similar initiatives in the microarray (Brazma and Hingamp, 2001) and proteomics (Orchard and Hermjakob, 2003; Taylor and Paton, 2003) (http://psidev.sourceforge.net/gps/) communities have already yielded some of these benefits. The ArMet designs and example implementations are freely available (http://www.armet.org/).

ArMet encompasses the timeline of metabolomics experiments from descriptions of the biological source material through to the results of chemical analyses. It describes not only the results of chemical analysis of a sample, but also data on the experimental context of those results (metadata). This metadata is an important part of experiment description as it enables correct interpretation of experimental results and meaningful comparison of experiments and their results across laboratories.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 19–28.

To provide a flexible and expandable data model ArMet has a component-based architecture. Each component describes part of the metabolomics process by way of a set of core data items. These are relevant to all experiments and serve as a minimal description. In addition to the core, each component may have zero or more alternative *sub-components* (detailed extensions), which support the core data plus additional data to describe particular methodologies and technologies in more detail. The opportunity to define sub-components means that the architecture may provide customized support to particular experiments or laboratories whilst maintaining the comparability of data sets from different origins through the core data. ArMet comprises nine components. Eight of these describe the context of experiments: Admin; Biological Source; Growth; Collection; Sample Handling; Sample Preparation; Analysis Specific Sample Preparation; Instrumental Analysis (Jenkins and Hardy, 2004) (http://www.armet.org for details). The ninth component, *Metabolome Estimate*, describes output from analytical instruments after raw data has been processed to produce metabolome descriptions appropriate for statistical analysis and data mining. While the core data for the eight components will stand alone as a minimal description the Metabolome Estimate component is abstract (in the computing sense) and requires sub-components to be meaningful; i.e. the core component is too general to represent any specific data in a useful way.

A range of different analytical instruments may be used to gain insight into the metabolomes of organisms. Gas Chromatography-Mass Spectrometry (GC-MS) is seen to have great potential in this area and there are many demonstrated examples of its use, (e.g., Fiehn and Kopka, 2000; Roessner and Wagner, 2000). Here we propose sub-components for the ArMet Metabolome Estimate component to support the results of certain types of metabolomics experiments carried out using GC-MS.

2 METABOLOME ANALYSIS

Before Metabolome Estimate subcomponents can be built an understanding of the types of metabolome analysis that may be carried out is required. Fiehn (2002) describes four analytical approaches:

- **Targeted analysis.** Detection and precise quantification of a single or small set of target compounds in a sample.
- **Metabolite profiling.** Detection and approximate quantification of a large set of target compounds in a sample. The target compounds will either have known chemical identities or will be reproducible between samples but be chemically unidentified.
- **Metabolomics.** Detection, approximate quantification and tentative identification of as many of the compounds in a sample as possible.

Metabolomics has the potential to discover previously undetected metabolites.

- **Fingerprinting.** Generation of a signature for a metabolome sample without regard for the individual compounds that it contains.

This paper considers data collected under the metabolite profiling and metabolomics approaches using GC-MS.

2.1 Metabolite profiling and metabolomics using GC-MS

The output of metabolite profiling or metabolomics is a quantified list of metabolites. The series of mass spectra output by GC-MS requires processing to produce such a list. Much work has been done to automate this processing. Many of the methods developed either implement or build on the work of Biller and Biemann (1974) or Dromey and Stefik (1976), whose methods are based on individual mass to charge ratio profiles across the scans. We identify the following stages of processing for both types of metabolome analysis:

- **Noise removal**: Removal of noise introduced by the analytical method
- **Peak detection**: Location of compound peaks in the elution profile
- **Peak deconvolution**: Data analysis to separate co-eluting compounds.

These stages establish the compound peaks. Two further stages of processing are required:

- **peak quantification;**
- **peak labeling.**

Quantification is typically based on calculation of the area under each peak. Effective noise removal, peak detection, and peak deconvolution to facilitate proper understanding of peak shapes are a prerequisite for this.

Labels may be chemical identities but, in profiling where a peak is reproducible between samples but has not been chemically identified or in metabolomics where there may be novel peaks, it must be some other identifier. Determination of a peak label may be performed on the basis of its mass spectrum. Each compound has a characteristic fragmentation pattern which may be sought in the reference list of target compounds (for profiling) or in mass spectrum libraries. Such lookup depends upon the ability to compare spectra. Common methods for this are the dot-product approach (Stein and Scott, 1994), probability-based matching (Pesyna and Venkataraghavan, 1976), and similarity indices (Hertz and Hites, 1971).

The retention time for a compound depends on its interaction with the stationary phase in the chromatograph column and is another attribute that is characteristic of chemical identity. Since the absolute retention time may be affected by analytical variability and instrument drift between runs, it is

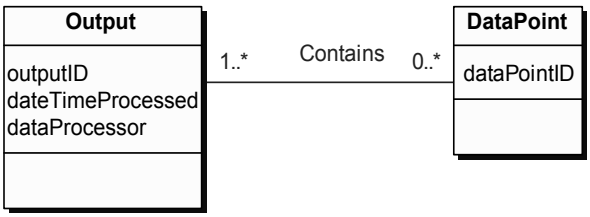


Figure 2-1. Metabolome Estimate component core.

common to convert it to a relative value or an index so that it may be used more reliably for peak labeling. Retention times may be expressed relative to times for internal or external standards. If temperature, stationary phase, and standard compounds are known relative retention times can be reproduced and used for identification (Willett, 1987). The retention index for an unknown is calculated by comparing its absolute retention time with those of a series of standard compounds. Retention indices are reproducible provided that the same temperature and stationary phase are used (Sewell and Clarke, 1987).

2.2 Understanding GC-MS metabolome descriptions

Understanding the types of metabolome analysis that may be carried out and how the resulting data sets may be produced from raw GC-MS output enables us to determine the data that should be supported by GC-MS subcomponents of the Metabolome Estimate component. The importance of contextual metadata, as outlined above should be emphasized.

All subcomponents of the Metabolome Estimate component must support its core data items. Figure 2-1 illustrates the core in a Unified Modeling Language (UML) (Booch and Rumbaugh, 1999) class diagram. UML enables us to model objects in an implementation-independent way. In the diagram, two classes represent the entities about which we wish to store data: `Output` models the data set as a whole and its processing to convert the raw instrument output to a metabolome description; `DataPoint` models a point in the processed results. The line between the two classes represents an association between them. The value "0..*" indicates that each `Output` is optionally associated with one or more `DataPoint`**.** The value "1..*" indicates that each `DataPoint` must be associated with one or more `Output`. The sparse nature of this model illustrates the abstract nature of the core data for the Metabolome Estimate component.

2.3 Modeling peak labels

The metabolome descriptions produced from GC-MS for both metabolite profiling and metabolomics comprise lists of labeled metabolite peaks.

For **metabolite profiling**, labeling is performed by comparison of the mass spectra for peaks found in a sample with those in a reference list of target metabolites. The sample peaks are labeled with either a chemical identity, where peaks have previously been identified, or a unique identifier where they have not. The complete peak label for a metabolite in a metabolite profiling experiment is, therefore, an identity and a quantity measurement that may be relative or absolute. Figure 2-2 depicts a data model, extending the core, to support this data. `DataPoint` has been extended to model an entry in a profiling data set as described above. Output supports two additional data items, `dataProcComments` for any notes that the data processor wishes to attach to the data set and `quantityType` which indicates whether the quantity measurements are relative or absolute. (`ProcessingProtocol` and associated classes are described in Section 2.4).

For **metabolomics**, peak labeling is performed by comparison of the peaks found in a sample with those in a library. Studies have shown (Stein and Scott, 1994; McLafferty and Zhang, 1998) that the best comparison algorithms may achieve only around 75% accuracy, so automatic lookup of chemical identity will provide only tentative identification. Therefore the model for metabolomics must support multiple candidate identities for each peak. Figure 2-3 depicts such a data model. Again it extends the core Metabolome Estimate data. `ChemicalID` supports the possible identities for a data point. To provide context for an association between a chemical identity and a data point the following metadata, which is represented in the model as attributes of the association, is required:

- The name and version number of the software/algorithms used to perform mass spectral lookup.
- The parameters to the software/algorithms.
- The name and version number of the mass spectral library used.
- A confidence value for the chemical identification.

The model allows zero or more chemical identities to be associated with each metabolite list entry (the 0..*:0..* label between `DataPoint` and `ChemicalID`). Therefore, the chemical identity cannot act as a unique label for a peak. As all chemical identities for metabolomics data points are only tentative a useful unique label would be one that allows third parties to perform their own tentative chemical identification. Mass spectral data and retention data are typically used for this purpose (see Sect. 2.2). Therefore, a metabolomics data point is modeled as:

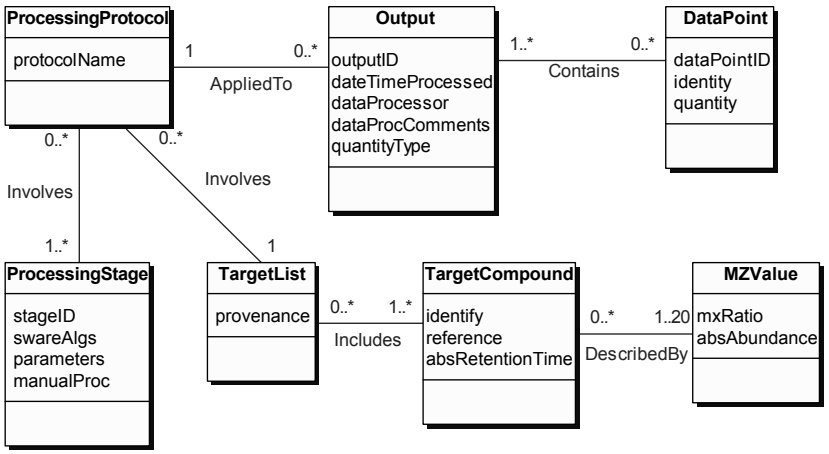


Figure 2-2. Metabolome Estimate subcomponent for metabolite profiling.

- A measurement of the quantity of the metabolite present in the sample; either absolute or relative.
- A retention value for the peak that represents the metabolite in the TIC chromatogram; either absolute retention time, relative retention time or a retention index.
- A maximum of 20 mass to charge ratio/ion abundance pairs from the mass spectrum for the peak that represents the metabolite in the TIC chromatogram.

These descriptions are supported by `DataPoint` and `MZValue`. `Output` supports (in addition to `dataProcComments`), `quantityType` and `retentionTimeType`, which indicate whether the quantity measurements are absolute or relative and whether the retention data is absolute, relative or indices. (`ProcessingProtocol` and associated classes are described in Sect. 2.4).

Figure 2-3 shows that the model supports a maximum of 20 peaks for each data point. McLafferty and Stauffer (1999) looked at the number of peaks required for successful comparison of mass spectra in the context of library lookup. The study used probability based matching and found that 15 peaks were 87% as effective as 150 peaks and 18 peaks were 97% as effective as 150 peaks. On this basis, and to yield the benefit of reduced data storage requirements in any implementation of the model, a representative mass spectrum with a maximum of 20 peaks is supported. Tong and Cheng (1999) compared three techniques for identifying the most significant peaks for spectral comparison and it is suggested that such approaches may be a basis for automatic selection of peaks for storage.

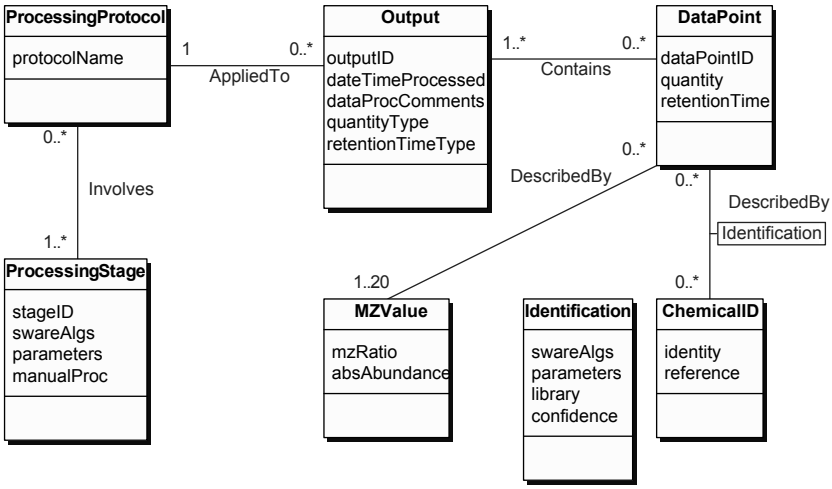


Figure 2-3. Metabolome Estimate sub-component for metabolomics.

2.4 Data processing metadata

Producing a metabolite list from raw GC-MS output involves the processes described in Sect. 2.1. For both approaches the first four processes (noise removal and peak detection, deconvolution and quantification) can be characterized by:

- The name and version number of the software/algorithms used to perform the process
- The parameters to the software/algorithms
- A description of any manual procedures employed to perform or adjust the output from the automated part of the process

In both Fig. 2-2 and Fig. 2-3 `ProcessingProtocol` and `ProcessingStage` provide support for these descriptions.

Metadata to describe the fifth process (peak labeling) for **metabolomics** data sets is included in the data point description. If relative retention times or retention time indices are used, it is necessary to support metadata to describe their calculation. This may be done in the same terms as noise removal, etc. and can be supported by the `ProcessingStage` entity. In addition, metadata to describe internal standards added to a sample to facilitate the calculation of approximate absolute quantities would be part of the sample preparation metadata supported by other ArMet components.

For **metabolite profiling** datasets the model should support additional metadata about the target compound reference list and the comparison

process used in peak labeling. While the comparison process may be described in the same terms as noise removal, etc. using `ProcessingStage`, the metadata to describe target compound reference lists should be sufficient to enable a third party to reproduce the results of an experiment, including details about the target compounds, i.e., mass spectral data and retention data. The data to describe a metabolite profiling reference list are characterized as a description of the provenance of the list (i.e., how it was compiled) and the following items for each target compound:

- Either a chemical identity or a unique identifier.
- Where the target compound has a chemical identity, a reference to further information on the compound in an external library or database.
- An expected absolute retention time for the compound.
- A maximum of 20 mass to charge ratio/ion abundance pairs from its mass spectrum.

This data will allow third parties to use the list with their own samples and provide input for chemical identification of unknown compounds. In Figure 2-2, a reference list is modelled by `TargetList` and associated classes.

3 CONCLUSIONS AND DISCUSSION

We propose a structure for the data required to describe GC-MS based metabolite profiling and metabolomics. It has been developed in the context of ArMet and consists of two subcomponents for its Metabolome Estimate component. Alternative models for the two analytical approaches and to support targeted analysis and fingerprinting are possible. Equally possible are combined models to support data collected under more than one approach (Jenkins and Hardy, 2004). Standard representations of GC-MS experiments for metabolome analysis may be used in a number of ways: to provide the basic design for systems to manage the output of such experiments; expressed in a data definition language such as XML and used to enforce data integrity when transporting data sets; as a standard for presentation and exchange of data sets, enabling cross-laboratory collaboration.

Metadata to provide experimental context for the data sets has been identified and is accommodated. The metadata required to describe the production of data sets from raw GC-MS output has also been identified and included in the models. This will enable data users to interpret the results of experiments, reproduce data sets, and identify data sets that may be meaningfully compared. We note that metadata and metabolome estimate data in an ArMet compliant system may be extracted and presented in the form suggested by Bino and Hall (2004) for “naming unknowns”.

These representations are not offered as an alternative to established and emerging standards for the representation of complete GC-MS data sets (Lampen and Hillig, 1994; Davies and Lampen, 2003) (http://animl.sourceforge.net/). Our intention is to model higher level metabolomics data for subsequent analysis and mining. We hope that our proposal will form the basis of discussions aimed at developing a standard for experiment descriptions in this area.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Oliver Fiehn of the University of California, Davis Genome Center for his input to this work and for his critical observations. The authors also gratefully acknowledge the United Kingdom Food Standards Agency (under the GO2006 project) for support of their work in metabolomics.

REFERENCES

Biller, J. E. and Biemann, K., 1974, Reconstruced mass spectra, a novel approach for the utilization of gas chromatograph-mass spectrometer data, *Analytical Letters* **7**(7):515-528.

Bino, R. J., and Hall, R. D., 2004, Potential of metabolomics as a functional genomics tool, *Trends in Plant Science* **9**(9):418-425.

Booch, G., and Rumbaugh, J., 1999, *The Unified Modeling Language User Guide*, Reading, Massachusetts, Addison-Wesley.

Brazma, A., and Hingamp, P., 2001, Minimum information about a microarray experiment (MIAME) - toward standards for microarray data, *Nature Genetics* **29**(4):365-371.

Davies, T., and Lampen, P., 2003, AnIMLs in the spectroscopic laboratory? *Spectroscopy Europe* **15**(5):25-28.

Dromey, R. G., and Stefik, M. J., 1976, Extraction of mass spectra free of background and neighbouring component contributions from gas chromatography/mass spectrometry data, *Analytical Chemistry* **48**(9):1368-1375.

Fiehn, O., 2002, Metabolomics - the link between genotypes and phenotypes, *Plant Molecular Biology* **48**(1-2):155-171.

Fiehn, O., and Kopka, J., 2000, Metabolite profiling for plant functional genomics, *Nature Biotechnology* **18**(11):1157-1161.

Hertz, H. S., and Hites, R. A., 1971, Identification of mass spectra by computer-searching a file of known spectra, *Analytical Chemistry* **43**(6):681-691.

Jenkins, H., and Hardy, N., 2004, A proposed framework for the description of plant metabolomics experiments and their results, *Nature Biotechnology* **22**(12):1601-1606.

Lampen, P., and Hillig, H., 1994, Jcamp-Dx for Mass-Spectrometry, *Applied Spectroscopy* **48**(12):1545-1552.

McLafferty, F. W., and Stauffer, D. A., 1999, Unknown identification using reference mass spectra. Quality evaluation of databases, *Journal of the American Society of Mass Spectrometry* **10**(2):1229-1240.

McLafferty, F. W., and Zhang, M.-Y., 1998, Comparison of algorithms and databases for matching unknown mass spectra, *Journal of the American Society for Mass Spectrometry* **9**(1):92-95.

Oliver, S. G., and Winson, M. K., 1998, Systematic functional analysis of the yeast genome, *Trends in Biotechnology* **16**(9):373-378.

Orchard, S., and Hermjakob, H., 2003, The proteomics standards initiative, *Proteomics* **3**(7):1374-1376.

Pesyna, G., and Venkataraghavan, M. R., 1976, Probability based matching using a large collection of reference mass spectra, *Analytical Chemistry* **48**(9):1362-1368.

Roessner, U., and Wagner, C., 2000, Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant Journal* **23**(1):131-142.

Sewell, P. A., and Clarke, B., 1987, *Chromatographic Separations*, Chichester, Wiley.

Stein, S. E., and Scott, D. R., 1994, Optimization and testing of mass spectral library search algorithms for compound identification, *Journal of the American Society for Mass Spectrometry* **5**(9):859-866.

Taylor, C. F., and Paton, N. W., 2003, A systematic approach to modeling, capturing, and disseminating proteomics experimental data, *Nature Biotechnology* **21**(3):247-254.

Tong, C. S. and Cheng, K. C., 1999, Mass spectral search method using the neural network approach, *Chemometrics and Intelligent Laboratory Systems* **49**(2):135-150.

Willett, J. E., 1987, *Gas Chromatography*, John Wiley & Sons.

Chapter 3

METABOLOMICS AND PLANT QUANTITATIVE TRAIT LOCUS ANALYSIS – THE OPTIMUM GENETICAL GENOMICS PLATFORM?

Daniel J. Kliebenstein
Department of Plant Sciences, Mail Stop 3, University of California, One Shields Ave, Davis, CA 95616, USA

1 INTRODUCTION

Biologists have long strived to understand what causes phenotypic differences between two individuals. This includes differences in morphology, disease susceptibility, and physiology as well as potential metabolic differences underlying these higher-order phenotypes. The diversity between individuals is partitioned into both environmental and genetic variation. Most genetic variation studied to date tends to be qualitative such that there are one or more distinct and non-overlapping phenotypic states. However, most phenotypic differences are quantitative such that there are numerous overlapping phenotypic states (Mackay, 2001; Flint et al., 2001; Lynch and Walsh, 1998; Mauricio, 2001). It has been known for nearly a century that the approximate genetic position of loci controlling these quantitative traits can be identified through associating marker and phenotype variation in a structured population (Sax, 1923). This association is the foundation for Quantitative Trait Locus (QTL) mapping experiments that attempt to identify the number, phenotypic impact and interaction of loci controlling a quantitative trait.

The latest incarnation of the QTL experiment is genetical genomics that phenotypes genetic mapping populations with genomics technology (Jansen and Nap, 2001). The goal is to merge the genomics technologies high-throughput and highly parallel phenotyping capacity, i.e., microarrays, proteomics, and metabolomics, with genetic segregation to test or generate specific hypothesis. The rationale is that a specific genes expression level is easier to quantify than the more complex developmental or physiological traits. Thus, by identifying loci controlling the differential gene expression

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 29–44.

patterns for all an organism's genes and comparing this to those loci controlling a specific physiological trait, the researcher could develop a systems biological understanding of more complex traits. Genetical genomics has predominantly utilized microarray analysis of stable mapping populations in a variety of species (Morley et al., 2004; Yvert et al., 2003; Brem and Kruglyak, 2005; Brem et al., 2002; Schadt et al., 2003).

Microarrays present a basic fiscal problem in that it is expensive to phenotype all lines in a mapping population much less replicate the phenotyping. Thus, microarrays are fiscally limited to small, highly defined mapping populations and replication limited to genes with highly heritable expression differences. This is a serious limitation, as most complex physiological traits are moderate-to-low heritability and controlled by numerous loci that require large populations with replicated experimental designs for reliable detection. Therefore, fiscal limitations alone will hinder microarray use in genetical genomics to all but the very largest or well funded of laboratories. Metabolomics platforms may provide a more widespread entry into genetical genomics. Metabolomics is much cheaper per sample than transcriptomics, enabling large populations to be studied with sufficient replication for moderate-to-low heritability traits. Additionally, most metabolomics platforms are higher-throughput than transcriptomics, allowing for rapid analysis (Fiehn, 2001; Fiehn et al., 2000; Hall et al., 2002).

Numerous studies have investigated QTL controlling plant metabolites but none with a metabolomics purview (Kliebenstein et al., 2002a; Kliebenstein et al., 2001a; Kliebenstein et al., 2002b; Monforte et al., 2001; Santos and Simon, 2002; Bushman et al., 2002; Thorup et al., 2000; McMullen et al., 1998; Byrne et al., 1996). This chapter's goal is to help provide guidance in developing, designing, and interpreting metabolomics genetical genomic experiments. I will focus on three questions that are frequently asked by individuals starting a metabolite QTL project: (1) How do I design the experiment? (2) What traits/variables do I measure? (3) What will I find? This will draw on literature both involved with the theory of QTL formation as well as experimental analysis of metabolite QTL detection and interpretation.

2 QTL QUESTIONS AND FINDINGS FOR METABOLOMICS

2.1 How do I design the experiment?

This question is best handled in three interrelated parts: population structure, population size, and replication. All three aspects are intertwined, such that population structure will influence the other two and vice versa, but I will deal with them separately. For more detailed information see the enclosed references (Mackay, 2001; Mauricio, 2001).

2.1.1 Which population do I chose?

For genetical genomics experiments, the optimal population structure is either Recombinant Inbred or Advanced Intercross lines. These populations allow for recombination and transgressive segregation similar to an F_2 population but are taken to homozygosity allowing independent replicated measurements of a given line. Homozygosity also increases the populations' power by forcing each genomic position to only have one of the two opposing haplotypes instead of the three possibilities in F_2 populations. Inbred line populations are not feasible in all systems due to generation time, inbreeding depression, or self-incompatability. In these species, the next best population structures are typically backcross populations as there are only two allelic classes at each locus, heterozygote, and one homozygote. Another factor that should be considered in determining the population is the availability of previously genotyped populations with phenotypic differences of interest. This is valuable as the majority of time and expense in any new population is not phenotyping but instead generating and genetically mapping the population. Thus, previously existing populations are highly desirable even if the structure is not optimal.

2.1.2 What population size do I use?

The next decision to resolve is the population's size. The general rule in determining the optimum population size is the larger the better. Ideally, populations should contain at least 300 individuals or lines. Larger populations provide several benefits. The first is that they have more recombination events increasing precision in measuring a QTLs position. Secondly, larger populations have more power to separate closely linked QTL due to the increased recombination. The increased line numbers also allow for better capacity to detect two- and three-way epistatic interactions because there are more lines in each combinatorial class. Finally, the larger population sizes allow for higher replication in terms of number of lines with Allele X at

position Y. Populations with less than 300 individuals can be utilized but will have limited power for traits with more than a couple QTL or moderate-to-low phenotypic effect QTL. In genetical genomics experiments, most traits may be controlled by numerous QTL with predominantly low-to-moderate phenotypic effect and thus small populations should be avoided (Mackay, 2001; Lynch and Walsh, 1998; Brem and Kruglyak, 2005).

2.1.3 How many replicates should I conduct?

Once the population is chosen, the next question is how many replicate measurements per line should be conducted and how these should be organized. The key to deciding these issues is to measure the experimental sources of variation. This involves designing an experiment whereby several samples are taken per plant with multiple independent plants per parental genotype per replicate. Multiple independent replicates are conducted and all samples independently analyzed *via* metabolomics. Analysis of variance for this experiment will allow the researcher to estimate the variation from spatial differences within a plant, from differences between plants, from differences between replicate experiments, and from differences between genotypes as well as any interactions between the different levels. The optimum result is that most of the variance is genetic with the rest of the error being split between plants within a replicate or between replicates. If this is the case, it is best to take one measurement per plant with each line being represented by two or more plants per replicate.

The analysis of parental variance also allows the researcher to obtain a very rough estimate of each traits heritability be estimating the variance due to genotype difference. There is a common perception that low heritability traits require high-replicate numbers to successfully map QTL. However, calculations show that even for traits with 30% heritability, only six replicates are required to diminish the error in the mean trait estimate to approximately 10% (Denby et al., 2004). Thus, it should be possible to identify QTL for most traits with less than 10 and as few as 6 replicates per line. Metabolomics platforms are probably the best current technology for fiscally achieving this replication in large populations. The analysis of parental variance will allow the researcher to identify the heritability distribution for the metabolites and make an informed decision on replication. Previous metabolite profiling projects have found heritabilities that range from 20% to 90% with most being in the range of 50–70% (Kliebenstein et al., 2002a; McMullen et al., 1998; Byrne et al., 1996; Kliebenstein et al., 2001b).

The ability to measure interactions in the above variance test is a key element of properly designing a QTL experiment. If there is a significant interaction between genotype and replicate, this suggests the presence of genotype × environment interactions. Previous metabolite profiling and

microarray QTL mapping projects have identified significant genotype × environment interactions (Brem and Kruglyak, 2005; Kliebenstein et al., 2002a). One option to minimize this is that each line should be repeated enough times per replicate to allow for QTL analysis within each replicate as there could be different QTL identified depending upon environmental fluctuation. Additionally, the researcher could attempt to better control the environmental variance by controlling the growth conditions between replicates to minimize this difficulty. Alternatively, the researcher may only be interested in QTL that impact the trait in all environments and would thus conduct the analysis in multiple environments.

The identification of a significant interaction of genotype with either within plant variance or between plant variance in the parental analysis suggests that there may be a developmental difference between the parents that is impacting the sampling. The best way to minimize this variance is to ensure that the same tissue at the same developmental stage is being sampled in all cases. A detailed analysis of the sources of variance before conducting a QTL mapping experiment will greatly enhance both the potential for success and the resulting QTL maps interpretability. This is especially important in a metabolomics genetical genomics experiment where thousands of traits will be analyzed simultaneously.

3 WHAT TRAITS/VARIABLES DO I MEASURE?

There are several aspects to this question. This includes what guidelines to use in deciding upon a metabolomics platform. Another important question to contemplate is which variables to use in the QTL mapping. Finally, should the data be altered to conform to the expectation of normality and what potential errors does this introduce? Each of these questions is dealt with below.

3.1 Which metabolomics platform to utilize?

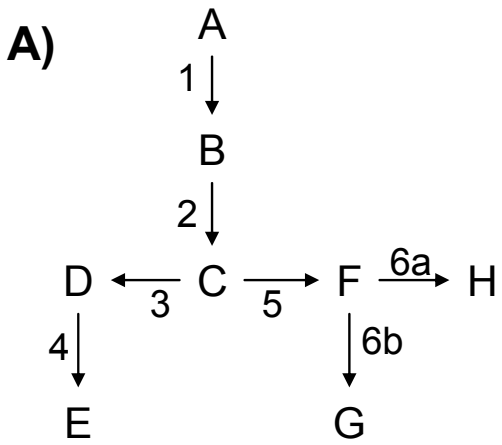
The first decision is which metabolomics platform should be utilized. This involves a compromise between the analytical speed and information content per analysis. The optimum platform should have significant high-throughput capacity to allow for the thousands of samples that are required for a statistically powerful QTL mapping experiment. In addition to high-throughput the best technology would individually quantify specific compounds and provide identification where possible and structural information for all compounds detected. This optimum requirement for individual quantification and identification provides the maximal power in the downstream QTL analysis. A number of high-throughput platforms like IR and NMR platforms are limited in providing specific compound

information and do not measure as many compounds per sample as other platforms. Thus, when QTL are identified the researcher will not be fully sure of the phenotypes identification. Thus, the researcher will be challenged to develop specific hypothesis about the locus's molecular function possibly even after cloning the underling gene. However, with platforms such as GC-MS and the rapidly gaining (LC-MS)/monolithic columns, it is possible to quantify and identify hundreds if not thousands of compounds (Bino et al., 2004; Tolstikov et al., 2003). Thus, when a QTL is found, the researcher will know the exact compound that locus is regulating and will be aided in developing specific hypothesis about the underlying molecular function.

3.2 What variables should I use for mapping?

The second part of this question is what aspects of the output are actually valid variables for QTL analysis. The most obvious variables are the actual amount of each individual compound. Considering that most metabolomics platforms can reliably detect hundreds of compounds, this creates a massive number of traits for QTL mapping. There are suggestions to decrease this dimensionality by using regression analysis to identify metabolite clusters, and then use an individual compound within each cluster to identify QTLs for that cluster. While this will decrease the computational power required, it will also decrease the experiments information content. The base assumption in regression clustering is that if two compounds are 80% correlated, that the other 20% is due to measuring error. However, it is equally likely that this 20% discrepancy is due to differences in the genetic control for the two compounds. Using a single compound per cluster would lose this genetic information. A better solution is to generate QTL software that can analyze 1000s of traits on the same population and present the results in a coherent manner. A challenge that is equally present for genetical genomics experiments using transcriptomics and proteomics.

Numerous variables/traits can be generated for QTL mapping using metabolomics data. The first is the absolute value of each variable (Figure 3-1B). Often times, there are known or predicted metabolic pathways providing relational context to the metabolites (Figure 3-1A). This relational context provides the ability to generate variables interrogating the interrelation between compounds (Figure 3-1C–E) (Weckworth et al., 2004; Steuer et al., 2003). These variables can either be the sum of specific groups of metabolites, the ratio between specific metabolites, or the ratio between different groups (Kliebenstein et al., 2001a). For instance, the equations in Figure 3-1C sequentially ask about the loci controlling the accumulation of the whole pathway (A–H), the accumulation of only those compounds on the right side (F–H), and the accumulation of those on the left side (D, E). These



B) A, B, C …

C) $\sum_{i=A}^{H} i \quad \sum_{i=F}^{H} i \quad \sum_{i=D}^{E} i$

D) $\frac{D}{E} \quad \frac{G}{H}$

E) $\frac{\sum_{i=D}^{E} i}{\sum_{i=C}^{H} i}$

Figure 3-1. Metabolomics Variables for QTL Mapping.

A. A hypothetical biosynthetic pathway is shown. The letters refer to the individual compounds. The numbers refer to the enzymes. Enzyme 6a and 6b are two different alleles of the same enzyme that lead to two different compounds. Arrows represent the direction of the biochemical reaction.
B. The first variable level is the individual compounds.
C. The second variable level is the broad summation meant to represent different branches of the pathway. i = the amount of specific compounds.
D. The third variable level is the ratio of two related compounds that may provide insight into particular enzymatic processes.
E. The final variable level is the ratio of different biosynthetic branches that may provide insight into more global regulation. i = the amount of specific compounds.

will identify a subset of common QTL as well as unique QTL. For example, it is possible to have a locus that has a 5% effect across the entire pathway. This

effect would most likely not be identified as QTL for any of the individual compounds but due to the smoothing impact of summing all of the compounds this effect may be seen at the pathway level. In addition to summations, ratios are other potential variables derived from a metabolomics data set (Figure 3-1D and E). These can allow the investigator to identify loci controlling regulation at specific branch points. For instance, the equation in Figure 3-1E measures the level of D and E with regards to all compounds produced from C. This would test for the presence of loci that impact the decision to go from compound C to either D or F. However, there are some statistical difficulties introduced in the use of ratio statistics that will be discussed later. When guided by known or predicted metabolic linkages, ratios, and summations provide powerful tools at querying the population for loci regulating whole branches or branch points in metabolic networks.

3.3 Should I worry about normality?

The final aspect before proceeding with the QTL analysis is data preparation. There is the underlying assumption that biological variables/traits should show a parametric distribution. This, however, presumes that the true biological distribution is in fact parametric and the skewing was technically introduced *via* the measurement. In metabolomics, this may not be the case especially in secondary metabolism (Kliebenstein et al., 2001a; McMullen et al., 1998; Byrne et al., 1996; Kliebenstein et al., 2001c; Yencho et al., 1998). In Figure 3-2A and B, the parents of a recombinant inbred mapping population differ in their capacity to make specific compounds due to enzymatic polymorphisms. Parent 1 contains a null allele of enzyme 5 but a hidden "a" allele at enzyme 6 and thereby does not accumulate compounds F, G, or H. Parent 2, however, contains a functional enzyme 5 allele but only the "b" allele of enzyme 6, leading to the accumulation of F and G (Figure 3-2A). When these two parents are mated, the recombinant inbred progeny will represent a mixture of parental genotypes and two recombinant genotypes, recombinant 1 will phenotypically look like parent 1 due to the enzyme 5 null allele while recombinant 2 will be a transgressive segregant producing F and H due to the "a" allele at enzyme 6 (Figure 3-2B). When the accumulation of either H or G in the progeny is plotted on a histogram, it will be a bimodal distribution due to the epistatic interaction between variation at enzymes 5 and 6 in controlling (Figure 3-2C). Normalization would destroy the information about both enzymes 5 and 6. The requirement for parametric distributions is a result of the QTL analysis algorithms. Most algorithms can handle skewed parametrics without normalization by using the bootstrapping methodology to empirically determine the significance threshold. Bimodal and true non-parametric distributions should instead be handled using non-parametric QTL

analysis techniques to obtain the maximal information (Diao et al., 2004; Kruglyak and Lander, 1995).

Metabolic pathway variation can generate other non-parametric distributions *via* transgressive segregation. The above epistasis example can generate non-parametric distributions if the enzyme 5 null allele hides functional enzymes such as 6b (Figure 3-2A). Another way for these non-parametric distributions to occur is when the compound is present in levels near the level of detection. QTL segregation can generate lines with undetectable levels while other lines are readily measurable. A common impulse in these situations is to take the undetectable lines and record them as no data/measurement when it is actually valid to assume that they are less than the other lines. By recording these lines as no measurement, the researcher is lowering the QTL detection power by diminishing the number of lines available for QTL analysis. A potential remedy is to give all undetectable lines a value equal to the detection threshold for the compound in question. This allows the researcher to include the fact that these lines are lower than the rest in the QTL analysis. However, if a significant number of lines are below the detection threshold, this may create a skewed parametric or non-parametric distribution. The skewed parametrics can be handled by the bootstrapping methodology as described. There are algorithms to handle the non-parametric distributions but they are not typically included into the common QTL mapping packages (Diao et al., 2004).

4 WHAT WILL I FIND IN THE QTLS?

Upon generating the metabolomics data and variables for QTL mapping there are numerous software options available to map QTL that are discussed elsewhere (Basten et al., 1999). These generally rely on the same composite or multiple interval mapping algorithms (Doerge and Churchill, 1996; Haley and Knott, 1992; Lander and Botstein, 1989; Zeng, 1994). Most programs, however, were not made to handle or present the massive number of traits generated in a standard genetical genomics experiment and thereby need to be modified to handle this data set. Once these hurdles are overcome and a QTL map is in hand for each trait, there are numerous questions to ask of the data. These include the size and number of QTLs for each trait, are the QTLs for different traits co-localized and is this because of a common polymorphism, as well as what is the level of epistasis and transgressive segregation in the population. I will briefly describe below what may be expected for each question.

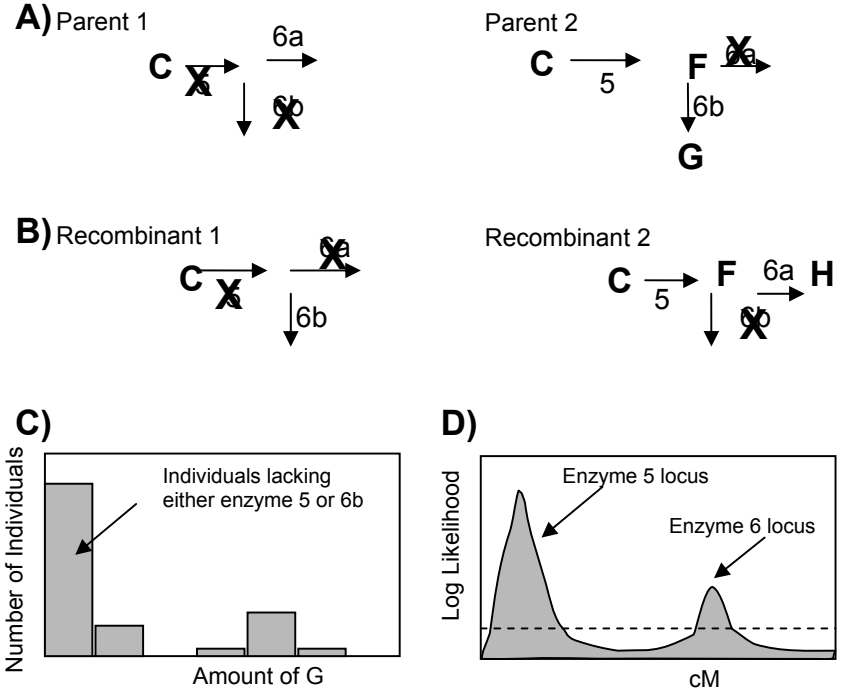


Figure 3-2. The Control of Epistasis and Transgressive Segregation by Enzymatic Variation.

A. The genotype and chemical composition of the parents is shown. The letters refer to the compounds present in each parent. The numbers refer to the enzymes. Enzyme 6a and 6b are two different alleles of the same enzyme that lead to two different compounds. Arrows represent the direction of the biochemical reaction. The X's indicate the presence of non-functional alleles for each enzyme.
B. The genotype and chemical composition of the recombinant individuals is shown. The letters refer to the compounds present in each genotypic class. The numbers refer to the enzymes. The X's indicate the presence of non-functional alleles for each enzyme.
C. The distribution of compound G's accumulation in the RIL population generated from crossing Parent 1 × Parent 2.
D. The QTL map generated for the accumulation of compound G in the RIL population generated from crossing Parent 1 × Parent 2.

4.1 QTL number and phenotypic effect

Recent analysis of a small yeast mapping population with 1 × replication *via* microarray has allowed a glimpse at what may be expected from a metabolomics genetical genomics experiment. This analysis found that most traits required at least 5 QTL's to partially explain the variation (Brem and Kruglyak, 2005). This experiment, however, was limited to a small number of lines with 1 × replication and as such, the analysis was limited to those

genes with at least 69% heritability. Nevertheless, it shows that most variable traits are under highly complex genetic regulation (Brem and Kruglyak, 2005). Metabolomics of higher-throughput and financial scale will allow for these experiments to be conducted with greater power, and therefore, to detect small to medium effect QTL. Thus, one could readily expect that the microarray indication is only the iceberg's tip.

One caveat should be made to the interpretation of both QTL number and phenotypic effect. Both the power to detect a QTL and the estimate of its phenotypic effect are dependent upon the populations' background variation. There could be epistatic interactions with other loci in one population that are not present in another. Alternatively, if the QTL is the only locus impacting the trait in one population, it will have a large phenotypic effect, whereas if the QTL is one of many in another population, it may have a smaller phenotypic effect. Thus, it should not be expected that a QTL or its phenotypic effect would be identical amongst all populations in which it is variable. An excellent example of this is shown in a paper investigating the quantitative inheritance of chlorogenic acid and flavones in three different maize populations. These populations were chosen as they are a pyramid such that all pair wise crosses of three different inbred lines were investigated (Bushman et al., 2002). While numerous QTLs were identified in more than one population, their significance and phenotypic effect were dependent upon the population studied. For example, one QTL was identified in two populations and controlled 34% of the trait variance in one population but only 8% in another (Bushman et al., 2002). Thus, the phenotypic effect and power to detect a QTL is relational and not absolute when comparing populations (Mackay, 2001; Lynch and Walsh, 1998).

4.2 QTL proximity – causality or proximity?

After the QTLs for each trait have been identified and surveyed, the next goal is to identify those QTL that pleiotropically impact different metabolites and as such may have global metabolic impacts. There are several major difficulties in this analysis. The first deals with validating if overlapping QTL for two different traits are caused by the same locus or closely linked loci. There are two possible techniques to try and differentiate between these two distinct possibilities. The first is a statistical approach to testing the possibility that the two QTL positions overlap by chance and hence are probably due to closely linked loci (Lebreton et al., 1998; Varona et al., 2004). It is possible to take the QTL models for each trait, fix the position and effect of the non-overlapping QTL as well as the effect of the overlapping QTL. Then randomly modify the position of the overlapping QTL from the same genetic position to gradually larger unlinked distances. At each step, use every lines genotype and the QTL model to predict all of

lines trait value. Then test the predictive strength of the model at each distance from identical to unlinked position and identify the genetic distance that maximizes the predictive power of both traits model (Lebreton et al., 1998; Varona et al., 2004). This would indicate whether the overlap is due to closely linked loci or a single locus.

Even if there is statistical support for a single locus, it is still possible that the overlap is due to two extremely tightly linked loci. The only way to validate the single pleiotropic QTL hypothesis is to clone the underlying molecular polymorphism and confirm that it impacts the expected traits. This requires fine-scale recombinational mapping in conjunction with some form of transgenic confirmation of the phenotypic effect (Mackay, 2001). Thus, once a pleiotropic region is identified, there still remains significant work to validate the pleiotropic QTL hypothesis. A further potential difficulty with highly pleiotropic QTL is to understand the mechanism by which it works and differentiate between direct and indirect effects. For example, in Arabidopsis, the ERECTA locus impacts leaf morphology, root architecture, floral shape and size, silique shape and size, pathogen resistance and numerous other traits (Qi et al., 2004; Xu et al., 2003; Godiard et al., 2003; Shpak et al., 2003; Douglas et al., 2002). However, even though the gene is a known receptor kinase with global impact, little is known about what the primary and secondary impacts are and how it controls these traits. Thus, highly pleiotropic QTL may not be easily interpreted panaceas of biological information.

4.3 Epistasis and transgressive segregation

Variation in biosynthetic pathways can easily form epistatic interactions measurable in metabolomics QTL mapping projects. One potential epistasis interaction is when variation at a preceding enzymatic step controls the production of another variable enzymes substrate (Figure 3-2A). In the example shown, functional variation at enzyme 5 determines whether the functional variation at enzyme 6a is seen. Thus, the accumulation of compound G in the population will form a bimodal distribution where the low accumulating lines have either a non-functional enzyme 5 or enzyme 6a (Figure 3-2C). Only those lines with functional enzymes 5 and 6a will accumulate compound G. When the level of compound G is used for QTL mapping it will identify at least two locations that epistatically interact in controlling the level of compound G, enzyme 5 and enzyme 6a, (Figure 3-2E). Other ways in which epistatic interactions may occur in biosynthetic pathways is if two proteins physically interact such that the variants from each parent prefer interacting with each other such as might occur in metabolic channeling. This will lead to any recombinant progeny between the loci having lower efficiencies and less compound accumulation. Taken

together, this suggests that epistatic interactions should be expected in metabolomics QTL mapping projects. It will be interesting to compare microarray and metabolomics estimates on epistatic interaction frequencies to see if metabolism is more prone to such interactions than gene expression.

In addition to epistasis, biosynthetic pathways readily generate transgressive segregation in both compound structure and amount. The easiest transgressive segregation to visualize is that impacting structure. In the example shown, the knockout in enzyme 5 hides the presence of the enzyme 6b allele. When recombination shuffles together a functional enzyme 5 and enzyme 6b, compound H is produced (Figure 3-2A and B). Compound H was not produced in either the parental genotype and is thus the product of transgressive segregation. This form of segregation has been identified in Plant Secondary Metabolite QTL projects (Kliebenstein et al., 2002a; Kliebenstein et al., 2001a). In addition to structural transgressive segregation, there is also the likelihood of transgressive segregation in compound amounts. If, for example, one parent has a bottleneck at one step in a biosynthetic pathway while the other is bottlenecked at a different step, segregation will produce lines that have both bottlenecks and compound levels lower than either parent while other lines will have neither bottleneck nor compound levels higher than both parents (Figure 3-3). This also illustrates how two parents can be indistinguishable *via* compound accumulation yet the progeny have highly variable levels. Both forms of transgressive segregation have been readily found in metabolite profiling experiments and should be expected in metabolomics QTL mapping (Kliebenstein et al., 2002a; Kliebenstein et al., 2001a). Further, over half of the highly heritable transcripts in yeast showed evidence of transgressive segregation, again supporting the idea that this will be a common hallmark of genetical genomics experiments (Brem and Kruglyak, 2005).

5 SUMMARY

Combining genomics technologies with segregating populations is becoming an area of increasing interest. The first experiments in this area were conducted with microarrays but it is likely that metabolomics due to cost and throughput advantages will become the “omics” platform of choice for genetical genomics. The hope in these experiments is to combine the massive parallel capacity albeit still reductionist “omics” technologies in a systems biological approach to understand why two organisms are different. In the process, fundamental aspects of biology may be illuminated. However, these same projects and data sets can be used to address base issues of quantitative genetics such as; How many loci control each trait?

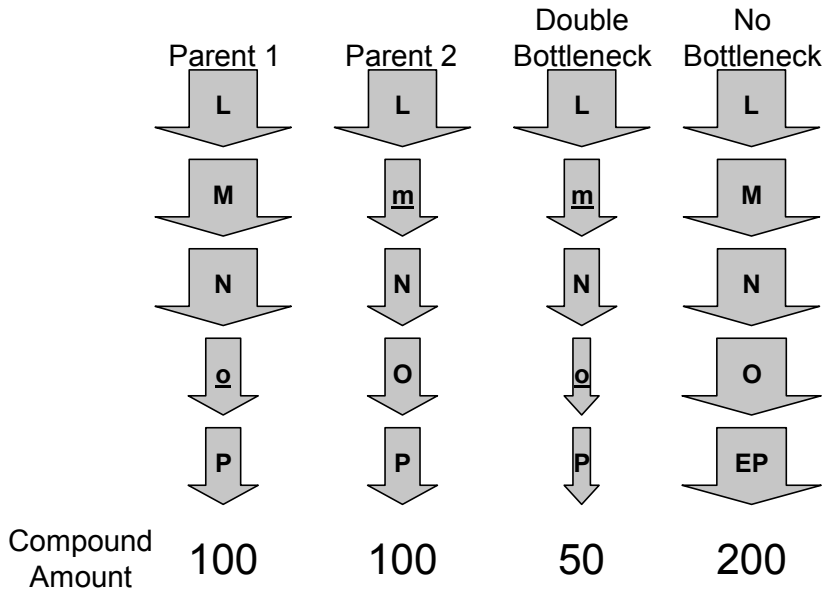


Figure 3-3. Transgressive Segregation Produced by Bottlenecks.
The enzymatic flux for two parents with different bottlenecks is shown. The specific bottlenecks are marked as small underlined letters with diminished arrow size representing the decreased flux. The two potential recombinant inbred line recombinant progeny are shown. The number at the bottom of each biosynthetic pathway shows the relative amount of the final product produced in each line.

What is the basis of the control, either additive or epistatic? Is there directionality in the parents with regards to the QTLs effect? Thus, it may be possible using QTL mapping and metabolomics to both understand key aspects of metabolic relationships as well as fundamental questions of quantitative genetics.

REFERENCES

Basten, C.J., Weir, B.S., and Zeng, Z.-B., 1999, *QTL Cartographer, Version 1.13*. Department of Statistics, North Carolina State University, Raleigh, NC.

Bino, R.J., et al., 2004, Potential of metabolomics as a functional genomics tool, *Trends In Plant Science* **9**(9):418-425.

Brem, R.B., and Kruglyak, L., 2005, The landscape of genetic complexity across 5,700 gene expression traits in yeast, *Proc. Natl. Acad. Sci. U S A*. **www.pnas.org/cgi/doi/10.1073/pnas.0408709102**(on-line).

Brem, R.B., et al., 2002, Genetic dissection of transcriptional regulation in budding yeast, *Science* **296**(5568):752-755.

Bushman, B.S., et al., 2002, Two loci exert major effects on chlorogenic acid synthesis in maize silks, *Crop Science* **42**(5):1669-1678.

Byrne, P.F., et al., 1996, Quantitative trait loci and metabolic pathways: Genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **93**(17):8820-8825.

Denby, K.J., Kumar, P., and Kliebenstein, D.J., 2004, Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana, *Plant Journal* **38**(3):473-486.

Diao, G.Q., Lin, D.Y, and Zou, F., 2004, Mapping quantitative trait loci with censored observations, *Genetics* **168**(3):1689-1698.

Doerge, R.W., and Churchill, G.A., 1996, Permutation tests for multiple loci affecting a quantitative character, *Genetics* **142**(1):285-94.

Douglas, S.J., et al., 2002, KNAT1 and ERECTA regulate inflorescence architecture in Arabidopsis, *Plant Cell* **14**(3): p. 547-558.

Fiehn, O., 2001, Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks, *Comparative and Functional Genomics* **2**(3):155-168.

Fiehn, O., et al., 2000, Metabolite profiling for plant functional genomics, *Nature Biotechnology* **18**(11):1157-1161.

Flint, J., and Mott, R., 2001, Finding the molecular basis of quantitative traits: Successes and pitfalls, *Nature Reviews Genetics* **2**(6):437-445.

Godiard, L., et al., 2003, ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt, *Plant Journal* **36**(3):353-365.

Hall, R., et al., 2002, Plant metabolomics: The missing link in functional genomics strategies, *Plant Cell* **14**(7):1437-1440.

Haley, C.S., and Knott, S.A., 1992, A simple regression method for mapping quantitative trait loci in line crosses using flanking markers, *Heredity* **69**:315-324.

Jansen, R.C., and Nap, J.P., 2001, Genetical genomics: the added value from segregation, *Trends In Genetics* **17**(7):388-391.

Kliebenstein, D.J., Figuth, A., and Mitchell-Olds, T., 2002a, Genetic architecture of plastic methyl jasmonate responses in Arabidopsis thaliana, *Genetics* **161**(4):1685-1696.

Kliebenstein, D.J., Pedersen, D., and Mitchell-Olds, T., 2002b, Comparative analysis of insect resistance QTL and QTL controlling the myrosinase/glucosinolate system in Arabidopsis thaliana, *Genetics* **161**(1):325-332.

Kliebenstein, D.J., Gershenzon, J. and Mitchell-Olds, T., 2001a, Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in Arabidopsis thaliana leaves and seeds*, Genetics* **159**(1):359-370.

Kliebenstein, D.J., et al., 2001b, Genetic control of natural variation in Arabidopsis thaliana glucosinolate accumulation, *Plant Phys.* **126**(2):811-825.

Kliebenstein, D., 2001c, et al., Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in Arabidopsis thaliana, *Plant Cell* **13**:681-693.

Kruglyak, L., and Lander, E.S, 1995, A Nonparametric Approach For Mapping Quantitative Trait Loci., *Genetics* **139**(3):1421-1428.

Lander, E.S., and Botstein, D., 1989, Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics* **121**: p. 185-199.

Lebreton, C.H., et al., 1998, A nonparametric bootstrap method for testing close linkage vs. pleiotrophy of coincident quantitative trait loci, *Genetics* **150**(2):931-943.

Lynch, M., and Walsh, B., 1998, *Genetics and analysis of quantitative traits,* Sinauer Associates, Inc., Sunderland, Massachusetts.

Mackay, T.F.C., 2001, The genetic architecture of quantitative traits, *Annual Review Of Genetics* **35**:303-339.

Mauricio, R., 2001, Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology, *Nature Reviews Genetics* **2**(5):370-381.

McMullen, M.D., et al., 1998, Quantitative trait loci and metabolic pathways, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **95**(5):1996-2000.

Monforte, A.J., et al., 2001, Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization, *Theoretical And Applied Genetics* **102**(4):572-590.

Morley, M., et al., 2004, Genetic analysis of genome-wide variation in human gene expression, *Nature* **430**(7001):743-747.

Qi, Y.P., et al., 2004, ERECTA is required for protection against heat-stress in the AS1/AS2 pathway to regulate adaxial-abaxial leaf polarity in Arabidopsis, *Planta* **219**(2):270-276.

Santos, C.A.F., and Simon, P.W., 2002, QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots, *Molecular Genetics and Genomics* **268**(1):122-129.

Sax, K., 1923, The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris, *Genetics* **8**:552-60.

Schadt, E.E., et al., 2003, Genetics of gene expression surveyed in maize, mouse and man, *Nature* **422**(6929):297-302.

Shpak, E.D., Lakeman, M.B., and Torii, K.U., 2003, Dominant-negative receptor uncovers redundancy in the Arabidopsis ERECTA leucine-rich repeat receptor-like kinase signaling pathway that regulates organ shape, *Plant Cell* **15**(5):1095-1110.

Steuer, R., et al., 2003, Interpreting correlations in metabolomic networks, *Biochemical Society Transactions* **31**:1476-1478.

Thorup, T.A., et al., 2000, Candidate gene analysis of organ pigmentation loci in the Solanaceae, *Proceedings of the National Academy of Sciences of the United States of America* **97**(21):11192-11197.

Tolstikov, V.V., 2003, et al., Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics, *Analytical Chemistry* **75**(23):6737-6740.

Varona, L., et al., 2004, Derivation of a Bayes factor to distinguish between linked or pleiotropic quantitative trait loci, *Genetics* **166**(2):1025-1035.

Weckwerth, W., et al., 2004, Differential metabolic networks unravel the effects of silent plant phenotypes, *Proceedings of the National Academy of Sciences of the United States of America* **101**(20):7809-7814.

Xu, L., et al., 2003, Novel as1 and as2 defects in leaf adaxial-abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity, *Development* **130**(17):4097-4107.

Yencho, G.C., et al., 1998, QTL mapping of foliar glycoalkaloid aglycones in Solanum tuberosum x S-berthaultii potato progenies: quantitative variation and plant secondary metabolism, *Theoretical And Applied Genetics* **97**(4):563-574.

Yvert, G., et al., 2003, Transacting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors, *Nature Genetics* **35**(1):57-64.

Zeng, Z.-B., 1994, Precision mapping of quantitative trait loci, *Genetics* **136**:1457-1468.

.

Chapter 4

DESIGN OF METABOLITE RECOVERY BY VARIATIONS OF THE METABOLITE PROFILING PROTOCOL

Claudia Birkemeyer and Joachim Kopka
Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany

Abstract: More than 670 GC-MS metabolite profiles were performed in the course of 3 years in an effort to probe robustness and reproducibility of metabolite profiling and to design metabolite recovery as well as the range of metabolite classes which are finally submitted to GC-MS based metabolite profiling. Experiments were performed with two important plant organs, namely root and leaf, using the model plant tobacco, *Nicotiana tabacum L.* var. Samsun NN (SNN). We investigated solvent composition, pH, and temperature during metabolite extraction and subsequent liquid partitioning of extracts. All permutations of the metabolite profiling protocol were directly compared to the initially published standard protocol. In agreement with the fundamental approach of profiling analyses, results were reported relative to this standard condition. Thus the consistency of results was maintained in the course of years. The resulting set of chromatograms was screened for mass spectral tags (MSTs), which represent identified as well as still unidentified metabolites. A non-supervised mass spectral and retention time index library (MSRI_NS) of these MSTs that was constructed for the future discovery of hitherto unidentified metabolic components will be made available at GMD (http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html). Cluster analyses and multivariate statistical techniques were applied to obtain insight into general trends of protocol variants. A choice of representative metabolites was analysed in depth for potential analytical improvement and final protocol optimization.

Key Words: gas chromatography mass-spectrometry (GC-MS); metabolite extraction; metabolite partitioning; metabolite profiling; protocol variation; recovery; retention time index; mass spectral library.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 45–69.

1 INTRODUCTION

The essence of metabolite profiling is the screening of biological samples for changes of metabolite levels relative to reference samples (Fiehn et al., 2000b; Sumner et al., 2003). The use of reference samples in biological experimentation and throughout relative quantitative analysis allows control of arbitrary changes in apparent metabolite levels. This approach of profiling analyses is essential because unwanted influences cannot always be avoided. For example, experimental investigations of biological systems often cannot exactly be reproduced in all aspects, and recovery artefacts of the chemical analysis are a common experience. In consequence, it can be argued that – provided the profiling approach is chosen – any type of comparative chemical analysis or experiment may be performed, if (i) reference samples are included and if (ii) these references and the samples under investigations are treated identically throughout the complete analytical process. Thus in-depth analyses aimed to optimize metabolite recovery may be deemed unnecessary for optimization of profiling analyses and thus only a small number of investigations have been performed so far (e.g. Fiehn et al., 2000b; Roessner et al., 2000; Roessner-Tunali et al., 2003; Gullberg et al., 2004). However, metabolite profiling may be described as the art of making as many metabolites as possible amenable to simultaneous analysis. This aspect is and will for a long time be one of the key aspects for improved metabolite profiling analyses. Two strategies allow extension and modulation of the scope of GC-MS based metabolite profiling: (i) the choice of chemical derivatization and (ii) the method of metabolite preparation.

We compared routine chemical derivatization by N-methyl-N-(trimethylsilyl)-trifluoroacetamide reagent (MSTFA) with tert.-butlydimethylsilylation using the N-methyl-N-(tert.-butyldimethylsilyl)-trifluoroacetamide reagent (MTBSTFA). The total data set comprises 1354 GC-MS experiments. Only half of the experiments were included in the present report: we focused on chemical derivatization by MSTFA only and, furthermore, kept subsequent GC-MS analysis invariable. As a consequence, the crucial factor for the modulation of chemical metabolite classes and the dynamic range of metabolite concentrations was the selected extraction and prefractionation protocol. To further our understanding of possibly more efficient variants of profiling analyses, we investigated temperature, pH, solvent composition and liquid partitioning. We first generated an inventory of identified metabolites which were present under the respective profiling regimes and constructed a mass spectral and retention time index (MSRI) library for future qualitative investigations of hitherto unidentified components (Wagner et al., 2003; Kopka et al., 2005; Schauer et al., 2005). We selected key metabolites, which represented the predominant metabolite classes covered by routine GC-MS based metabolite profiling and performed

exemplary analyses on metabolite recovery and robustness of analysis. In agreement with the general principle of metabolite profiling, all investigations were performed in direct comparison with standard, or in other words, reference samples. These reference samples were in this investigation defined to be samples, which were analysed in-parallel using the initially published protocol of GC-MS based metabolite profiling for plant material (Roessner et al., 2000; Fiehn et al., 2000b).

2 EXPERIMENTAL

2.1 Plant material

Nicotiana tabacum L. var. Samsun (SNN) plants were cultivated on quartz sand under 16/8 h long day conditions as described previously (Birkemeyer et al., 2003). Plant organs of 12 simultaneously grown plants were harvested 3 months after germination. All mature, non-senescent leafs were immediately shock-frozen in liquid nitrogen. The root systems were washed sand-free under tap water, shortly dried on filter paper, and subsequently frozen in liquid nitrogen. The complete root and leaf batches of 12 plants were homogenized separately in precooled mortars while permanently being kept under liquid nitrogen. Aliquots of 50 mg each were prepared and stored for up to 3 years at –80°C until further processing.

2.2 Extraction

The general extraction scheme started with ball-mill homogenization of the deep-frozen plant material in 2 mL microvials as described earlier (Fiehn et al., 2000a; Fiehn et al., 2000b; Roessner et al., 2000). Non-sample controls were entered into analysis at this step of the protocol. Extraction was performed in three steps: (i) the first extraction step was initiated by adding a 300 µL volume of pre-cooled polar water-miscible solvent, which contained the internal standard substances, ribitol, D(-)-isoascorbic acid for monitoring the recovery of GC-separated, oxidation-sensitive vitamin C, L(+)-ascorbic acid, and 2,3,3,3-D_4 alanine, to the deep-frozen powder without removal of the steel ball, (ii) after initial incubation a 200 µL volume of chloroform was added and shortly incubated 5 min at 37°C (iii) finally, polar and lipid phases were separated by adding 400 µL of bi-distilled water 1 min vigorous mixing and 10 min centrifugation at room temperature in a microvial centrifuge set to maximum speed. Cellular debris accumulated mainly below the bottom chloroform layer and to a small extent at the interphase boundary. Two 80, 120, or 160 µL aliquots were carefully drawn from the top (polar) and the bottom (lipid) layers of each experiment. One replicate was analysed by

trimethylsilylation the other using dimethyl-tert.-butyl-silylation. The aliquot volumes, which were dried by vacuum centrifugation and submitted to chemical derivatization, were adjusted to obtain an optimal number of analyte peaks. In general, we aimed to avoid overloading of the major components (Wagner et al., 2003). The representative total volumes of each liquid fraction from all protocol variations were determined.

Four protocol variations were performed. In the following we use a four character code to address respective variants and a "L_" or "R_" prefix to distinguish between leaf and root matrix, respectively. Each position of the code characterizes one altered parameter. The codes are compiled and briefly explained in Table 4-1. We modified (i) the major polar solvent of the first extraction step from pure methanol to acetone:water (2:1, v:v), first position code characters "m" and "a". (ii) The pH of the first solvent was either not adjusted, or acidified by 1.67% formic acid, or adjusted to basic pH by saturation with solid sodium carbonate, second position code characters "n", "a", and "b". (iii) Incubation of the first extraction step was either hot, i.e., 15 min at 70°C, or cold, over night at –20°C, third position code character, "h" or "c". (iv) Extraction was performed including a liquid partitioning into a polar and lipid phase, fourth position code characters "p" and "l", or phase separation was omitted, missing last code character. The preparations without liquid partitioning are in the following called "combined". The reference protocol was defined as the polar fraction of the hot methanol extraction without pH adjustment, protocol code mnhp.

2.3 Chemical derivatization

Dried extracts were derivatized in two consecutive steps as described earlier (Wagner et al., 2003). Carbonyl-moieties were converted into methoxyamine groups (MEOX) by 90 min incubation at 30°C with 40 µL freshly prepared methoxyamine hydrochloride (Sigma, Munich, Germany) which was dissolved at 20 mg mL^{-1} in pure pyridine (Merck, Darmstadt, Germany). Subsequently exchangeable protons were substituted by trimethylsilyl-groups (TMS) using N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA, Macherey & Nagel, Düren, Germany). Silylation was performed by adding a 70 µL volume of MSTFA followed by 30 min incubation at 37°C. Samples were subsequently kept at room temperature on the GC-MS injector tray for 2–12 h. Alternately the N-methyl-N-(tert.-butyldimethylsilyl)-trifluoroacetamide reagent (MTBSTFA, Macherey & Nagel, Düren, Germany) was used to derivatize a second equal aliquot of each extract preparation with 45 min incubation at 65°C (data not shown). Retention time index standard mixture was added in a 10 µL volume prior to silylation. *n*-Alkanes were dissolved in pyridine at a final concentration of 0.22 mg mL^{-1}

Table 1. Summary of protocol variations, protocol code, number of experiments, current entries into a non-supervized mass spectral and retention time index library and number of mass spectral tags (MSTs), which were generated by automated deconvolution using AMDIS software.

Protocol variants					*Nicotiana tabacum* L. SNN leaf			*Nicotiana tabacum* L. SNN root			*non-sample control*		
Code	Solvent	pH	Temperature	Fraction	Experiments	Library entries	MSTs/ entry	Experiments	Library entries	MSTs/ entry	Experiments	Library entries	MSTs/ entry
				∑:	309	16	9112	∑: 368	16	5796	∑: 48	8	1466
aacl	acetone	acidic	cold	lipid	7	-	-	7	-	-	-	-	-
aacp	acetone	acidic	cold	polar	10	-	-	10	-	-	-	-	-
abcl	acetone	basic	cold	lipid	7	2	496	7	2	247	4	1	152
abcp	acetone	basic	cold	polar	10	2	599	10	2	388	4	1	163
ancl	acetone	non-adjusted	cold	lipid	7	2	551	7	2	243	4	1	159
anc	acetone	non-adjusted	cold	mixed	8	-	-	10	-	-	-	-	-
ancp	acetone	non-adjusted	cold	polar	10	2	597	10	2	432	4	1	210
aahl	acetone	acidic	hot	lipid	5	-	-	13	-	-	-	-	-
aahp	acetone	acidic	hot	polar	10	-	-	20	-	-	-	-	-
abhl	acetone	basic	hot	lipid	4	-	-	17	-	-	-	-	-
abhp	acetone	basic	hot	polar	10	-	-	17	-	-	-	-	-
anhl	acetone	non-adjusted	hot	lipid	6	-	-	7	-	-	-	-	-
anh	acetone	non-adjusted	hot	mixed	8	-	-	10	-	-	-	-	-
anhp	acetone	non-adjusted	hot	polar	11	-	-	10	-	-	-	-	-
macl	methanol	acidic	cold	lipid	10	-	-	7	-	-	-	-	-
macp	methanol	acidic	cold	polar	10	-	-	10	-	-	-	-	-
mbcl	methanol	basic	cold	lipid	5	2	573	5	2	372	4	1	199
mbcp	methanol	basic	cold	polar	10	2	511	20	2	379	4	1	204
mncl	methanol	non-adjusted	cold	lipid	9	-	-	5	-	-	-	-	-
mnc	methanol	non-adjusted	cold	mixed	8	-	-	10	-	-	-	-	-
mncp	methanol	non-adjusted	cold	polar	11	-	-	19	-	-	-	-	-
mahl	methanol	acidic	hot	lipid	10	-	-	5	-	-	-	-	-
mahp	methanol	acidic	hot	polar	10	-	-	10	-	-	-	-	-
mbhl	methanol	basic	hot	lipid	5	-	-	5	-	-	-	-	-
mbhp	methanol	basic	hot	polar	10	-	-	10	-	-	4	1	167
mnhl	methanol	non-adjusted	hot	lipid	20	2	614	7	2	410	-	-	-
mnh	methanol	non-adjusted	hot	mixed	8	-	-	10	-	-	-	-	-
mnhp[a]	methanol	non-adjusted	hot	polar	70	2	617	90	2	429	20	1	212

[a] Standard protocol.

each, i.e., *n*-dodecane (RI 1200; CAS 112-40-3), *n*-pentadecane (RI 1500; CAS 629-62-9), *n*-octadecane (RI 1800; CAS 593-45-3), *n*-nonadecane (RI 1900; CAS 629-92-5), *n*-docosane (RI 2200; CAS 629-97-0), *n*-octacosane (RI 2800; CAS 630-02-4), *n*-dotriacontane (RI 3200; CAS 544-85-4), *n*-hexatriacontane (RI 3600; CAS 630-06-8). All above substances were obtained from Sigma, Munich, Germany, if not otherwise indicated.

2.4 GC-MS analysis

A MD 800 quadrupole GC-MS system (ThermoQuest, Manchester, UK) was equipped with a RTX-5Sil MS capillary column, 30 m length, 0.25 mm inner diameter, 0.25 µm film thickness and a 10 m IntegraGuard precolumn (Restek GmbH, Bad Homburg, Germany). The system was operated in constant flow mode with 1 mL min^{-1} Helium 5.0 carrier gas (Air Liquide, Magdeburg, Germany). GC-MS analysis was essentially as described earlier (Fiehn et al., 2000a; Fiehn et al., 2000b; Roessner et al., 2000; Roessner-Tunali et al., 2003; Colebatch et al., 2004). Injection was 1 µL in split-less mode at 230°C with a 2 min delay. The temperature program comprised 1 min isothermal time at 70°C, a 6 min ramp to 76°C, a 45 min ramp to 350°C, 1 min isothermal at 350°C, and further isothermal heating for 10 min at 330°C. The quadrupole mass selective detector was operated with electron impact ionization. The transfer line was set to 250°C, and the ion source operated at 200°C. Scan rate was 2 spectra s^{-1}, with the m/z range set to 40–600.

2.5 Mass spectral tags for analysis of metabolite recovery

In GC-MS metabolite profiling analyses, metabolites are represented by mass spectra of metabolite derivatives which occur in highly reproducible retention time index windows. Thus, in a more precise phrasing, an analyte, in other words a chemical derivative, is quantified by GC-MS and not the endogenous metabolic chemical. Exceptions are those metabolites which are not susceptible to derivatization and are volatile, for example nicotine. The majority of analytes comprising GC-MS profiles is still not identified. For these cases we created the expression “mass spectral tag” or MSTs (Colebatch et al., 2004). MSTs are defined by mass spectrum and retention behaviour and can thus be reproducibly deconvoluted and identified without knowledge about the chemical nature of the underlying metabolite. Each analyte or MST can be quantified by fragments which constitute respective full mass spectra of MSTs. In electron impact ionization analyses the relative intensities of fragments from a single compound are constant and independent

Table 2. List of 146 identified metabolites from leaf or root extracts of *Nicotiana tabacum L.* SNN. Identification was performed by automated comparison with the MSRI_ID library of identified mass spectra available at http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html using AMDIS software (MPIMP-ID; analyte identifier). Thresholds for accepting identifications were, signal to noise (S/N) > 20, reverse mass spectral match (Match) > 65 and retention time index (RI) deviation < 5.0. Relative quantitative analysis of changes in metabolite recovery (Recovery analysis) was performed on a subset of metabolites which were present under all tested protocol regimes. All suggested quantifying masses (QM) for quantitative metabolite analysis are indicated.

Analyte				RI		Match		S/N	QM	Recovery analysis
MPIMP-ID[b]	Metabolite	Derivative TMS	Derivative MEOX[b]	Expected	Deviation	Simple	Reverse		m/z	
176002-101	Aconitic acid, cis-	3		1762.8	-0.2	33	84	40	229|285|375|211|215	x
188005-101	Adenine	2		1872.4	-0.1	44	88	110	264|279|192|165|237	-
151006-101	Adipic acid[a]	2		1509.0	2.5	80	90	64	275|111|141|172|159	-
144001-101	Alanine, beta-	3		1431.4	0.5	86	93	131	248|290|174|160|100	x
138002-101	Alanine, DL-	3		1363.9	-0.2	96	99	52	188|262|290|100|114	x
138002-211	Alanine, DL-, 2,3,3,3-D_4-[c]	3		1359.9	0.6	96	99	440	192|266|294|104|117	-
167002-101	Arabinose	4	1	1675.3	0.7	93	99	185	307|217|160|103|189	x
168001-101	Asparagine, DL-	3		1683.3	0.9	88	97	176	116|188|231|258|159	-
152002-101	Aspartic acid, DL-	3		1525.0	1.4	76	99	255	232|218|306|202|334	x
128003-101	Benzoic acid[a]	1		1256.7	-1.6	65	95	109	179|105|135|77|194	x
164003-101	Benzoic acid, 4-hydroxy-	2		1639.5	-0.4	83	93	157	267|223|282|193|126	-
184001-101	Benzoic acid, p-amino-	2		1841.1	-0.4	40	87	26	266|281|222|192|126	-
117002-101	Butyric acid, 2-amino-, DL-	2		1169.9	2.5	67	89	28	130|204|218|232|142	-
153003-101	Butyric acid, 4-amino-	3		1530.7	0.4	98	100	397	174|304|216|246|100	x
126002-101	Butyric acid, 4-hydroxy-	2		1242.6	2.3	49	84	48	233|117|204|143|133	-
214001-101	Caffeic acid, trans-	3		2141.4	-0.8	81	95	157	396|381|219|307|205	x
311001-101	Caffeoylquinic acid, 3-trans-	6		3126.0	-2.3	76	98	1293	345|255|397|324|219	x
317001-101	Caffeoylquinic acid, 4-trans-	6		3188.8	-4.1	79	100	840	307|489|324|255|219	x
319001-101	Caffeoylquinic acid, 5-trans-	6		3209.8	-2.5	68	89	1407	307|447|345|255|219	x
329001-101	Campesterol	1		3262.1	0.4	97	99	506	472|382|343|367|129	-
319002-101	Cholesterol	1		3156.9	-1.6	97	98	415	458|368|353|329|129	-
148001-101	Citramalic acid, D(-)-	3		1473.8	-0.2	66	96	194	247|349|259|321|203	x
182004-101	Citric acid	4		1827.8	0.5	98	100	375	273|375|211|183|257	x
184011-101	Citric acid, 2-methyl-, DL-	4		1841.9	-1.6	43	67	66	287|479|389|197|225	-
194005-101	Coniferylalcohol, trans-	2		1945.7	0	84	97	64	324|293|235|309|219	-
195001-101	Coumaric acid, p-, trans-	2		1944.6	1.8	51	89	27	249|293|308|219|179	-
156002-101	Cysteine, DL-	3		1560.7	0.6	72	95	82	294|220|218|100|116	-
144002-101	Cysteine, S-methyl-, DL-	2		1427.3	0	54	77	36	162|218|236|100|115	-
147004-101	Decanoic acid, n-[a]	1		1462.1	2.3	67	89	81	229|244|117|201|145	-
185002-101	Dehydroascorbic acid, dimer[c]	+		1852.6	-1.1	79	86	121	316|173|157|245|231	x
166003-101	Dodecanoic acid, n-[a]	1		1662.5	2.2	90	99	96	257|272|117|201|145	-
245001-101	Eicosanoic acid, n-	1		2456.3	-0.2	72	96	186	369|384|117|201|145	-
150002-101	Erythritol	4		1510.2	1.8	91	98	221	217|293|307|205|320	x
154001-101	Erythronic acid	4		1548.7	1.6	97	100	605	292|220|117|319|205	-
146002-101	Erythrose	3	1	1459.1	2.5	28	75	24	205|117|161|233|262	-
128002-101	Ethanolamine	3		1269.1	0.7	70	97	52	174|86|100|188|262	-
210001-101	Ferulic acid, trans-	2		2098.6	2.4	33	89	29	338|249|323|293|308	-
187002-101	Fructose	5	1	1874.6	1	97	100	684	307|217|277|364|335	x
232002-101	Fructose-6-phosphate	6	1	2321.4	-0.6	80	99	161	459|315|357|217|	x
175001-101	Fucose	4	1	1746.5	-1.4	89	92	77	117|160|364|277|321	x
137001-101	Fumaric acid	2		1359.8	1.8	92	99	333	245|115|217|143|	x
299002-101	Galactinol	9		2993.5	-2.5	80	96	226	204|191|433|305|169	x
194001-101	Galactitol	6		1941.8	0.1	32	73	41	319|307|157|217|331	-
199002-101	Galactonic acid	6		1997.9	-0.3	95	99	404	333|292|319|305|157	x
189003-101	Galactono-1,4-lactone, DL-	4		1890.5	1.2	54	65	524	217|451|466|334|305	-
188001-101	Galactose	5	1	1892.3	0.1	96	99	534	160|319|229|343|305	-
194003-101	Galacturonic acid	5	1	1946.9	-2.5	34	77	37	333|160|423|292|364	x
283004-101	Gentiobiose	8		2828.3	-0.5	89	95	112	160|480|390|204|361	-
200001-101	Gluconic acid	6		2002.7	-1.5	88	99	334	333|292|319|305|157	-
189008-101	Gluconic acid-1,5-lactone	4		1887.9	1.2	68	86	68	220|229|319|451|129	-
190007-101	Glucopyranoside, 1-O-methyl-, beta-D-	4		1898.3	4.7	59	73	239	133|204|377|231|290	-
189002-101	Glucose	5	1	1897.3	0.4	96	100	594	160|319|229|343|305	x
172001-101	Glucose, 1,6-anhydro-, beta-D-	3		1715.1	0.3	96	99	164	204|217|333|243|317	-
233002-101	Glucose-6-phosphate	6	1	2334.5	-1.4	92	99	120	160|387|299|471|357	x
193004-101	Glucuronic acid	5	1	1937.4	-1.6	85	98	288	333|160|423|292|364	x
163001-101	Glutamic acid, DL-	3		1631.4	0.6	89	99	1183	246|363|128|348|156	x
178001-101	Glutamine, DL-	3		1785.1	-0.4	96	99	482	156|245|347|362|203	-
143001-101	Glutaric acid	2		1414.6	1.6	53	87	37	158|261|233|116|186	-
158004-101	Glutaric acid, 2-oxo-	2	1	1588.7	0.9	91	99	302	198|288|304|186|229	-
135003-101	Glyceric acid, DL-	3		1339.6	1.9	99	99	365	292|189|307|205|133	-
129003-101	Glycerol	3		1282.5	1.6	97	99	237	293|205|117|103|	x
174002-101	Glycerol-2-phosphate	4		1741.3	1.2	52	77	47	243|299|389|211|445	-
177002-101	Glycerol-3-phosphate, DL-	4		1775.1	1.1	86	100	621	357|445|299|315|211	-
133001-101	Glycine	3		1311.9	0	91	99	216	174|248|276|100|86	x
214002-101	Guanine	4		2132.7	2.7	71	97	56	352|367|264|202|99	x
278001-101	Guanosine	5		2781.8	-0.3	81	91	76	324|245|280|368|410	-
196001-101	Gulonic acid	6		1964.2	-2.2	79	95	70	333|292|423|433|319	-
205001-101	Hexadecanoic acid, n-	1		2050.2	0.1	98	100	1041	313|328|117|201|145	x
106001-101	Hexanoic acid, n-[a]	1		1064.0	-2.3	70	88	69	173|188|117|129|145	-
146001-101	Homoserine, DL-	3		1454.0	1.8	52	89	38	218|128|292|230|202	x
209002-101	Inositol, myo-	6		2091.9	-1.8	83	100	756	305|265|318|191|507	x
243003-101	Inositol-phosphate, myo-	7		2429.0	-2.2	69	99	199	299|318|387|315|217	x
182003-101	Isocitric acid	4		1831.6	-2.1	82	96	254	245|319|390|83|	x
132002-101	Isoleucine, DL-	2		1300.6	0.8	75	100	91	158|232|218|102|260	x
291002-101	Isomaltose	8	1	2907.0	-3.8	67	78	52	160|480|204|319|361	-
135004-101	Itaconic acid	2		1351.7	2	65	87	37	259|215|133|147|230	-
105001-101	Lactic acid, DL-[a]	2		1048.9	-0.2	94	99	173	219|117|191|133|234	-

Table 2. continued.

Analyte				RI		Match		S/N	QM	
MPIMP-ID[b]	Metabolite	Derivative TMS	Derivative MEOX[b]	Expected	Deviation	Simple	Reverse		m/z	
129002-101	Leucine, DL-	2		1278.8	0.9	54	100	294	158|232|102|260|	-
192003-101	Lysine, DL-	4		1920.8	-0.1	86	98	325	156|174|317|230|434	-
133003-101	Maleic acid[g]	2		1314.7	0.9	96	100	490	245|147|170|215|	x
137003-101	Maleic acid, 2-methyl-	2		1358.0	-2.5	58	83	31	259|184|122|157|231	-
149001-101	Malic acid, DL-	3		1492.3	1.1	97	99	495	233|245|335|307|217	x
122003-101	Malonic acid	2		1211.4	-0.1	96	99	333	233|248|147|133|109	-
284001-101	Maltitol	9		2839.2	-1.6	74	77	99	204|361|345|525|305	-
277002-101	Maltose	8	1	2768.4	-0.8	81	95	136	160|204|361|319|271	x
355003-101	Maltotriose	11	1	3550.2	-0.8	65	91	38	204|361|217|480|169	-
193002-101	Mannitol	6		1928.8	-2	94	99	161	319|307|157|217|331	x
189001-101	Mannose	5	1	1899.5	-2.2	43	79	1389	160|319|229|343|305	-
231001-101	Mannose-6-phosphate	6	1	2323.5	-0.4	35	76	24	160|471|387|357|299	x
346001-101	Melezitose	11		3475.7	-1.2	84	95	202	361|451|271|204|217	x
290002-101	Melibiose	8	1	2903.5	2.3	66	89	52	160|480|204|319|361	x
152001-101	Methionine, DL-	2		1521.2	0.6	95	99	128	176|128|250|293|202	-
259003-101	Nicotianamine	4		2606.3	-1.3	74	89	74	186|218|246|345|232	-
139002-101	Nicotine			1357.5	0.1	95	99	290	84|133|161|162|92	x
133004-101	Nicotinic acid	1		1303.7	1.7	83	95	85	180|136|106|78|195	-
138003-101	Nonanoic acid, n-[a]	1		1369.2	2.1	73	97	89	215|230|117|129|145	-
231003-101	Octadecadienoic acid	1		2315.1	0.4	78	85	108	337|352|262|129|117	-
221003-101	Octadecadienoic acid, 9,12-(Z,Z)-	1		2211.1	0.1	83	99	802	337|352|262|129|117	-
225002-101	Octadecanoic acid, n-	1		2247.0	-0.2	98	100	431	341|356|117|129|145	x
222003-101	Octadecatrienoic acid, 9,12,15-(Z,Z,Z)-	1		2219.1	-0.6	98	99	450	335|350|93|129|117	-
223003-101	Octadecenoic acid	1		2225.1	-0.4	56	65	108	339|354|117|129|145	-
222001-101	Octadecenoic acid, 9-(Z)-	1		2217.4	0.4	48	84	55	339|354|117|129|145	-
127006-101	Octanoic acid, n-[a]	1		1270.6	2.4	65	87	54	201|216|117|129|145	-
182002-101	Ornithine, DL-[d]	4		1821.9	-1.3	60	91	79	142|174|420|200|258	x
195004-101	Pentadecanoic acid, n-[a]	1		1949.0	0	48	95	92	299|314|117|201|145	-
164001-101	Phenylalanine, DL-	2		1635.4	-1.5	88	97	338	192|266|218|91|294	x
129001-101	Phosphoric acid	3		1281.9	0	95	100	665	314|299|211|283|225	x
348004-101	Phyllodihydroquinone	2		3487.2	-2.3	63	77	201	596|591|331|356|371	-
132003-101	Proline, DL-	2		1303.4	0	95	99	180	142|130|117|244	x
175002-101	Putrescine[c]	4		1741.6	-2	98	100	232	174|361|214|100|200	x
114003-101	Pyridine, 3-hydroxy-[a]	1		1137.0	-0.3	59	84	35	152|167|136|122|92	-
153002-101	Pyroglutamic acid[f]	2		1528.1	0.9	90	98	102	156|258|230|140|273	x
104002-101	Pyruvic acid	1	1	1036.6	2.5	78	76	65	174|189|115|89|158	-
185001-101	Quinic acid	5		1862.7	-1.1	86	100	1048	255|345|334|537|419	x
337002-101	Raffinose	11		3396.0	-4.7	94	99	110	437|451|361|217|204	-
172002-101	Rhamnose	4	1	1727.8	-0.4	87	93	113	117|160|364|277|321	-
173001-101	Ribitol[i]	5		1734.7	0.1	92	99	1359	319|307|422|217|205	-
168002-101	Ribose	4	1	1690.9	0	92	99	255	307|217|160|103|189	x
138001-101	Serine, DL-	3		1369.3	1.6	98	100	542	204|218|278|306|100	x
181002-101	Shikimic acid	4		1820.9	-0.8	84	98	492	204|462|372|255|357	x
209005-101	Sinapyl alcohol	2		2094.6	-1.2	80	97	119	354|234|323|339|293	-
338002-101	Sitosterol, beta-	1		3355.5	-1.6	79	91	810	396|357|486|471|129	-
193001-101	Sorbitol	6		1935.8	-1.1	80	91	102	319|307|157|217|331	x
226002-101	Spermidine	5		2252.6	-1.3	77	94	187	174|144|156|116|491	-
332001-101	Stigmasterol	1		3289.8	1.1	98	99	899	484|394|255|129|379	-
171007-101	Suberic acid	2		1710.8	0.7	44	79	34	187|303|169|117|129	-
134001-101	Succinic acid	2		1326.0	-0.2	97	100	360	247|172|147|262|129	x
264001-101	Sucrose	8		2653.4	0.3	90	89	1811	437|451|361|319|157	x
284005-101	Tetracosanoic acid	1		2838.7	0.7	70	90	211	425|440|117|132|145	-
185004-101	Tetradecanoic acid, n-[a]	1		1852.6	0.4	91	98	179	285|300|117|201|145	-
149002-101	Threitol	4		1501.7	2.5	77	94	130	217|293|307|205|320	-
156001-101	Threonic acid	4		1568.2	3.5	98	100	492	292|220|205|217|245	x
140005-101	Threonic acid-1,4-lactone	2		1382.5	2.2	96	99	166	247|147|262|217|101	x
140001-101	Threonine, DL-	3		1394.0	1.2	98	100	358	219|291|218|117|320	-
147008-101	Threose	3	1	1459.7	1.9	29	78	24	205|117|161|233|262	-
142006-101	Thymine[g]	2		1407.7	3.1	50	74	89	255|270|113|239|140	-
316001-101	Tocopherol, alpha-	1		3145.3	-0.9	88	99	759	502|503|236|237|277	-
300002-101	Tocopherol, gamma-	1		3002.9	1.7	42	95	110	488|489|222|223|263	-
274002-101	Trehalose	8		2749.1	0.3	65	87	151	191|169|361|243|331	x
223001-101	Tryptophan, DL-	3		2218.6	-1	89	99	452	202|291|218|303|130	-
191004-101	Tyramine	3		1910.4	-0.8	84	96	258	174|338|86|100|264	x
194002-101	Tyrosine, DL-	3		1941.4	-0.8	87	98	283	218|280|354|179|100	x
156004-101	Undecanoic acid, n-[a]	1		1559.9	2.8	33	68	22	243|117|129|132|145	-
136001-101	Uracil	2		1346.6	2.1	40	94	34	241|255|99|113|126	-
127002-101	Urea	2		1260.1	0.7	60	99	122	189|204|171|87|99	-
247002-101	Uridine	3		2468.0	0.7	46	77	47	217|259|243|230|169	-
122001-101	Valine, DL-	2		1220.2	-0.2	97	99	126	144|218|156|246|100	x
171001-101	Xylitol	5		1717.6	-0.9	78	75	157	307|319|332|217|205	-
166001-101	Xylose	4	1	1669.2	0.1	91	99	390	307|217|160|103|189	x

[a] May occur in non-sample controls.
[b] Methoxyamines form E- and Z- isomers in stable ratios; only the major isomer is reported.
[c] In the presence of ambient air dehydroascorbic acid dimer may be generated from ascorbic acid.
[d] Arginine and Citrulline may decompose and form ornithine.
[e] Agmatine may decompose and form putrescine.
[f] Glutamine and to a lesser extent glutamic acid may cyclize and form pyroglutamic acid.
[g] Thymidine readily decomposes and forms thymine.
[h] Detailed mass spectral infomation may be obtained by submitting the MPIMP-ID at http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html (Kopka et al. 2005).
The first six characters of MPIMP-ID identify the chemical compound, the last three characters the respective mass isotopomers.
[i] Internal standard compounds.

of concentration, except if mass detectors are operated beyond the linear range of detection. We analysed each analyte and MST using multiple fragment masses as noted in Table 4-2.

2.6 Response calculations

Peak areas, X, of selected ion traces were retrieved using the find algorithm of the MassLab software 1.4 (ThermoQuest, Manchester, UK). For each metabolite a single, specific and selective fragment mass was selected from a choice of analysed fragments (Table 4-2). Correct peak integration was monitored manually. Peak areas with low intensity were rejected. The resulting peak area values were defined to be what we call fragment responses (X_i of fragment i). Fragment responses were normalized to the fresh weight of the sample. In this investigation we did not use the internal standard substance, ribitol, for volume correction. Instead we determined the final volume of each extract variation and performed numerical correction to ensure analysis of equal fresh weight equivalents from all extraction regimes ($N_i = X_i$ * extract volume^{-1} * fresh weight^{-1}). In a further step, the relative response of a fragment, N_i, is divided by the averaged relative response of the same fragment, which was analysed in-parallel according to the initially developed standard protocol, mnhp ($R_i = N_i$ * $avgN_{i(mnhp)}{}^{-1}$). The resulting quotient is subsequently called response ratio R_i. R_i describes the x-fold change in metabolite recovery relative to the standard extraction protocol. Response ratios are calculated separately for root and leaf samples. As a consequence, leaf and root mnhp will exactly overlap in PCA analyses and indicate the origin of these plots (Figures 4-1 and 4-4).

Each protocol variation, p, was analysed in 5–20 fold replication (Table 4-1). In order to reduce the complexity of the data set we averaged the response ratios for the replicate analyses of each protocol, $avgR_{i,p}$. Fragments and underlying metabolites which exhibited high relative standard deviation (RSD), i.e., >35%, when applying the standard protocol to leaf samples, leaf mnhp, were excluded from further analysis. We observed increased RSD of all metabolites between early and late replicate experiments performed in the course of the 3-year storage period of our sample batches. For this reason, we had to accept a 35% threshold rather than the reported 10% average analytical RSD of all analytes, which was reported for shorter storage periods and GC-TOF-MS profiles (Gullberg et al., 2004; Weckwerth et al., 2004). Some fragments exhibited low RSD but were frequently missed by the MassLab find algorithm. For these cases we required successful peak finding in more than 50% of the replicate standard analyses.

Finally all $avgR_{i,p}$ were combined into a single matrix that described the complete set of changes in metabolite recoveries under the different extraction regimes employed in this experiment. The majority of metabolites which exclusively occurred in only one plant matrix or only under specific conditions, except for dehydroascorbic acid, were excluded from the present

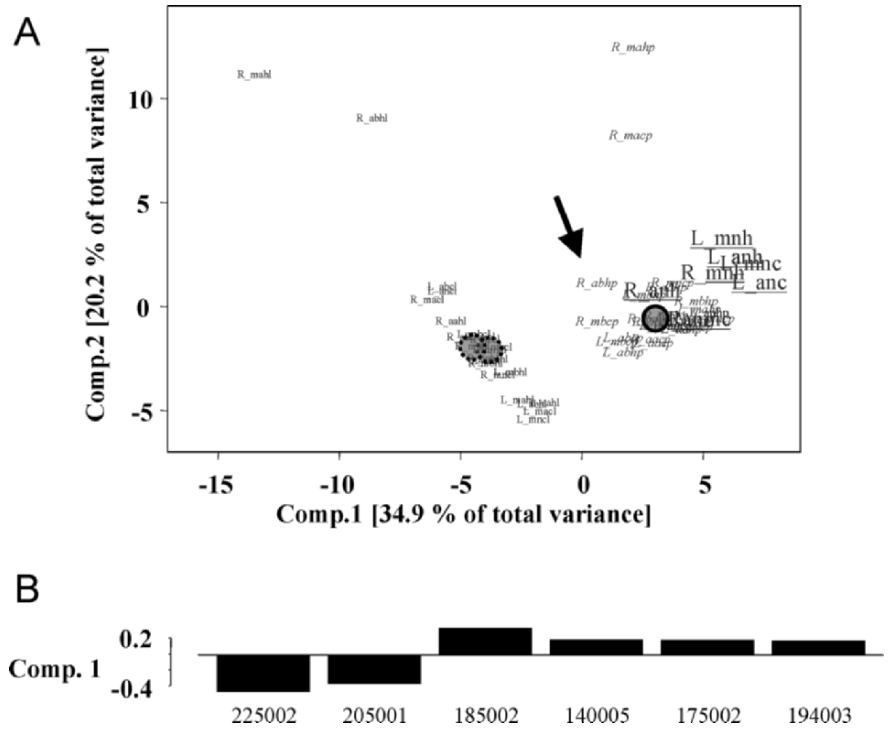


Figure 4-1. PCA analysis of the currently compiled data set of 64 analytes which represent those metabolites which were observed in GC-MS metabolite profiles of all tested protocol variations. Component scores are plotted in **A**. Ranked analyte loadings are shown in **B. Comp. 1:** 225002, octadecanoic acid, n-; 205001, hexadecanoic acid, n-; 185002, dehydroascorbic acid; 140005, threonic acid-1,4-lactone; 175002, putrescine and 194003, galacturonic acid. Protocol codes are as listed in Table 4-1 with L_- or R_-prefix to indicate leaf and root preparations. Circles indicate positions of the standard protocol mnhp (closed circle) and mnhl (dotted circle) of leaf and root. Lipid preparations (small, normal font), polar preparations (intermediate, italic font), and combined preparations (large, underlined font). The arrow indicates the position of basic pH protocols.

version of our comparative investigation. The currently available set of analysed metabolites is reported in Table 4-2.

Principal component analysis (PCA) using the covariance model (PCA) and hierarchical cluster analysis (HCA) using complete linkage of a Euclidian distance matrix (HCA) was performed without $\log_{10}$ transformation of the $avgR_{i(protocol)}$ matrix. The S-Plus 2000 software package standard edition release 3 (Insightful, Berlin Germany) was used for HCA, PCA, and visualization.

2.7 Mass spectral and retention time index libraries

Mass spectral metabolite tags (MSTs) were generated by automated deconvolution of GC-MS chromatograms using the publicly available deconvolution software AMDIS (http://chemdata.nist.gov/mass-spc/amdis/; National Institute of Standards and Technology, Gaithersburg, USA) (Stein, 1999). Mass spectra were collected with component width 20, adjacent peak subtraction set to 2, low resolution and shape requirement and medium sensitivity. The currently processed chromatograms and resulting number of MSTs are listed in Table 4-1. RIs of all MSTs were determined and thus annotated MSTs were uploaded into a non-supervised custom NIST02 mass spectral library (NIST02, mass spectral search program, http://chemdata.nist.gov/mass-spc/Srch_v1.7/index.html; National Institute of Standards and Technology, Gaithersburg, USA) (Ausloos et al., 1999). The approach of constructing non-supervised mass spectral libraries and applications of these libraries were described by Wagner et al. (2003).

2.8 Automated mass spectral identification

All chromatograms which were processed for MST library construction were also screened for already known analytes using AMDIS software and our Q_MSRI_ID library. This library contains repeatedly observed MSTs and those analytes which were identified by standard addition experiments. The Q_MSRI_ID mass spectral library is made available at http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html (Kopka et al., 2005). Identifications were automated using the AMDIS option to generate tab separated report files. Thresholds for acceptance of positive identifications were signal-to-noise (S/N) >20, reverse mass spectral match (Match) >65 and retention time index (RI) deviation <5.0 (Table 4-2).

3 RESULTS AND DISCUSSION

3.1 Inventory of analytes and MSTs

Automated and reliable mass spectral identification of deconvoluted mass spectra is one of the major challenges in GC-MS based metabolite profiling, not least because this process is prerequisite for the discovery of those MSTs and analytes which have hitherto not been found in a specific type of

biological sample or have not yet been archived in mass spectral and retention time index libraries (MSRI). We applied the concept of non-supervised mass spectral libraries (MSRI_NS) (Wagner et al., 2003) and our library of identified metabolites and frequently observed MSTs (MSRI_ID) (Kopka et al., 2005), which were developed in the course of recent years, to identify components found in different extracts of the model plant tobacco, *Nicotiana tabacum L.* var. Samsun NN (SNN). We manually selected 16 diverse and representative GC-MS chromatograms of polar and lipid fractions from non-pH-adjusted and basic extraction protocols including 8 complementary non-sample controls (Table 4-1). In total we obtained 16374 MSTs using AMDIS deconvolution. In agreement with earlier reports (Wagner et al., 2003, Gullberg et al., 2004) non-sample controls of our experiments had on average 183 MSTs per chromatogram, which comprised some laboratory contaminations, such as benzoic acid or lactic acid and a range of short-to-long chain fatty acids (refer to Table 4-2 for details) – correction by non-sample controls is required for the profiling of these compounds –, and unavoidable side products of the chemical reagents, which are used for GC-MS profiling. Besides immediate side products of the reagents, for example, hydroxylamine, the majority of contaminations, >50%, belonged to the class of linear and cyclic polysiloxanes. These compounds accumulate over time in MSTFA reagent, when exposed to traces of ambient air, but are also mobilized by MSTFA from GC-MS septum and crimp cap material. Capillary GC columns are an additional source of MSTFA mobilized polysiloxanes, especially toward end of column lifetime, and may contain mixed methyl-phenyl- or methyl-aryl-poly siloxanes.

When analysing the complexity of plant extracts, we found significant differences in retrieved numbers of MSTs. The following results were obtained after adjustment of major peaks to the upper detection limit. As a rule of thumb, these dominant components did not exhibit peak deformation due to chromatographic overloading, but may in cases exhibit slight mass spectral distortions due to saturation of the quadrupole mass detector used in this investigation. A few general trends were observed (Table 4-1). Profiles of root material had less components compared to leaf samples, a tendency which in general was more obvious for the lipid than for polar fractions. Acetone appeared to yield less lipid components as compared to methanol extracts. Other trends such as differences caused by pH adjustment or temperature may be revealed after further in-depth analysis.

The identification process resulted in 470 identified analytes and MSTs. We choose S/N threshold >20, a setting which drastically reduced false positives but unavoidably created a small fraction of false negative identifications. A RI deviation of ±5.0 units was applied based on the two observations that (i) the amount of an analyte may influence RI by approximately 2.5 units and (ii) the batch-to-batch reproducibility plus aging processes of the chosen capillary column had an approximately equal

contribution to the variability of absolute RI. Setting mass spectral match thresholds faced a fundamental problem which was caused by the observation that mass spectral deconvolution unavoidably results in increasing numbers of chimeric results when the complexity of samples increases. In addition GC-MS systems with low mass spectral scanning rates, such as quadrupole or ion-trap type of systems, are also prone to chimeric deconvolution, in other words a mixed mass spectra of a major compound and a co-eluting trace compound. In order to accommodate this inherent problem we choose the reverse match to be >65 on a scale of 100, because reverse matching allows identification of known mass spectra in chimeric MSTs. In addition, we report but do not threshold respective simple match values, to distinguish between hits based on perfect deconvolution and hits based on chimeric deconvolution (Table 4-2). As a final result of automated identification, we currently know 172 analytes and 298 MSTs from the extracts of, *Nicotiana tabacum L.* var. Samsun NN (SNN). In Table 4-2, we report only one major analyte for each of the identified 147 metabolites and two of 11 added internal standard compounds. Additional information on these compounds can be accessed using the MPIMP-ID identifiers or analyte names reported in Table 4-2 with our web query pages at the Golm metabolome data base (GMD; http://csbdb.mpimpgolm.mpg.de/csbdb/gmd/gmd.html) (Kopka et al., 2005).

For the subsequent investigations of this report we focused on metabolites which were – at least in detectable traces – present in all prepared fractions. We avoided those compounds which were close to detection limits in our reference protocol, mnhp, and those compounds which were frequently missed by the peak finding algorithm we used. Furthermore, most analytes were excluded which were specific for root, leaf, polar or lipid extracts under the conditions of our reference protocol, mnhp. We followed the reasoning of Roessner et al. (2001) for a generalized comparative analysis, which may be biased by condition specific components, and concentrated on common analytes for detecting general trends of protocol variants. Analysis and discovery of known metabolites and novel analytes which occur only under specific protocol regimes is of course a major aspect of this project and one of the prerequisites for the development and optimum design of extended and diverse complementary GC-MS profiles from single samples. Estimation and identification of novel analytes is an ongoing project in our laboratory (data not shown).

3.2 General trends caused by modifications of extraction

We tested five possible influences which may change qualitative and quantitative composition of metabolite profiles. For the protocol codes please refer to Table 4-1 and experimental Section 2.2. (i) The effect of

different biological matrices is well known in chemical analysis. Different types of biological samples may not only contain specific sets of metabolites, but in addition to this rather obvious fact, may influence the recovery of metabolites and thus introduce – without proper analytical standardization – apparent changes in metabolite levels, which are artefact and not due to *in vivo* effects. Here we compared root to leaf samples. (ii) The choice of solvent exerts effects on metabolite extraction and enzyme inactivation. We compared the polar-protic solvent methanol with polar-aprotic acetone, which is completely water-miscible and known to effectively precipitate proteins at high and low temperatures. (iii) pH is known to stabilise metabolites or induce acid and base catalysed hydrolysis of labile compounds. The initial protocol of metabolite profiling used non-pH-adjusted organic solvents. We tested acidified solvents similar to classical Bieleski mixtures (Bieleski, 1964) and carbonate saturated basic medium. (iv) Two general temperature regimes exist for the extraction of labile compounds. One regime uses hot short extraction for highly effective inactivation of enzymes in organic solvents and short exposure of temperature labile metabolites to heat. The second regime uses cold but long extraction to avoid loss of temperature labile compounds and reduce enzyme activity. Cold extraction has slow extraction kinetics and usually requires extended extraction times. This requirement, as a trade-off, allows for an influence of residual enzyme activity, an effect which has been described previously for sucrose cleavage in plant material possibly due to residual invertase activity (Gullberg et al., 2004; Weckwerth et al., 2004). We tested 15 min at 70°C vs –20°C over night. (v) The polarity of extraction and liquid partitioning allows differentiation into polar and lipophilic fractions (Fiehn et al., 2000b; Weckwerth et al., 2004). Combined analysis of both fractions should effectively increase the complexity of metabolite profiles (Gullberg et al., 2004) and avoid effects of irreproducible partitioning of amphipolar compounds. We tested separated lipid and polar liquid layers with combined extracts which can easily be obtained by omitting liquid partitioning from the standard protocol.

We currently have accessed data on 64 metabolites (Table 4-2). The derivatized GC-MS analytes of these metabolites were selected according to the criteria stated above (Section 2.6). PCA was performed on the complete data set and the two first principal components are shown for a first overview and general insight into the main variance within the data set (Figure 4-1). Figure 4-2 demonstrates the major profile classes observed with root and leaf extracts and the high similarity of classification between these two biological matrices. Separate PCA analyses of on one hand the lipid extracts (Figure 4-3) and on the other hand polar and combined extracts (Figure 4-4) not only demonstrated specific trends but also revealed the underlying analytes which contributed the major variance to each of the calculated principal components.

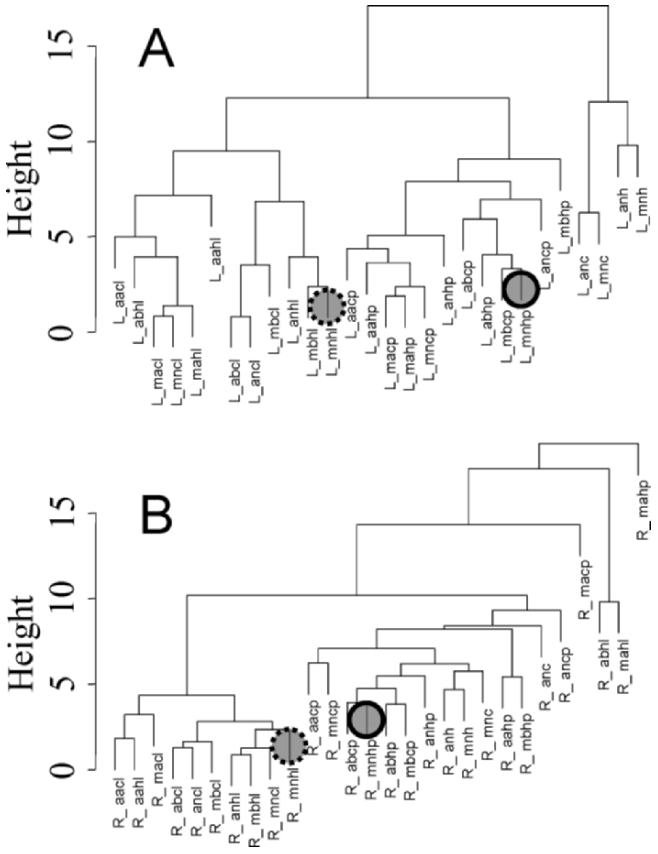


Figure 4-2. HCA analysis of the leaf subset of analyses **A** compared to the root subset **B**. Protocol codes are as listed in Table 4-1 with a L_- or R_-prefix to indicate leaf and root preparations. Circles indicate positions of the standard protocol mnhp (closed circle) and protocol mnhl (dotted circle) of leaf and root.

3.2.1 Influence of biological matrix

Influences of the biological matrix on metabolite profiles must be avoided (Kopka et al., 2004). In our analysis we used the leaf and root standard preparations for normalizing the analyte responses obtained from all other extracts (Section 2.6). Root and leaf standard preparations, R_mnhp and L_mnhp, therefore, co-localize and are centred to the origin within PCA analyses. The matrix effects on protocol variations can only be estimated relative to these preparations (Figures 4-1A, 4-3A, and 4-4A). The lipid fraction clearly

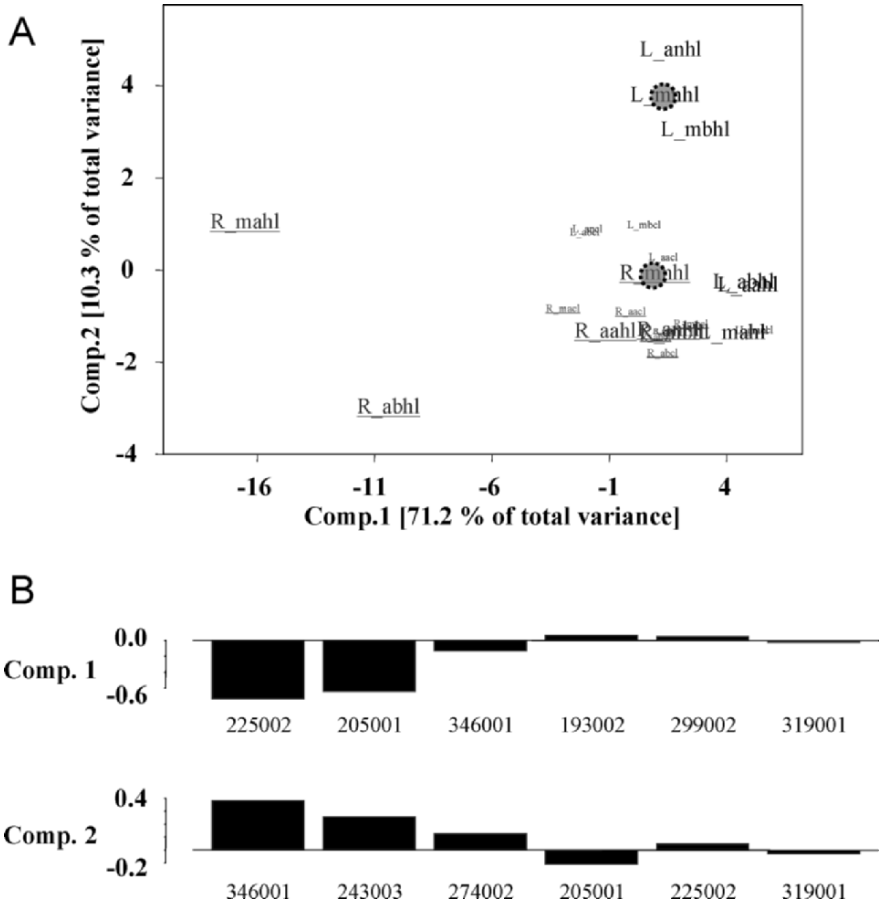


Figure 4-3. PCA analysis of the lipid subset of protocols. Component scores are plotted in **A**. Ranked analyte loadings are shown in **B. Comp. 1:** 225002, octadecanoic acid, n-; 205001, hexadecanoic acid, n-; 346001, melezitose; 193002 mannitol; 299002, galactinol and 319001, caffeoylquinic acid, 5-trans-; **Comp. 2:** 346001, melezitose; 243003 inositol-phosphate, myo-; 274002, trehalose; 225002, octadecanoic acid, n-; 205001, hexadecanoic acid, n- and 319001, caffeoylquinic acid, 5-trans-. Protocol codes are as listed in Table 4-1 with a L_- or R_- prefix to indicate leaf and root preparations. Circles indicate positions of the lipid fraction of leaf and root analysed with the standard protocol, mnhl (dotted circle). Cold preparations (small), hot preparations (large), leaf preparations (underlined font), and root preparations (normal font).

exhibited a systematic difference between root and leaf samples, for example, L_mnhl and R_mnhl (Figure 4-3A). Acidified cold and hot acetone extracts appeared to have smaller matrix effects on the lipid fractions.

In contrast, the majority of polar fractions did not separate, i.e., these fractions were subject to the same leaf and root matrix effects which occur under the conditions of our standard preparation. Caution is, however,

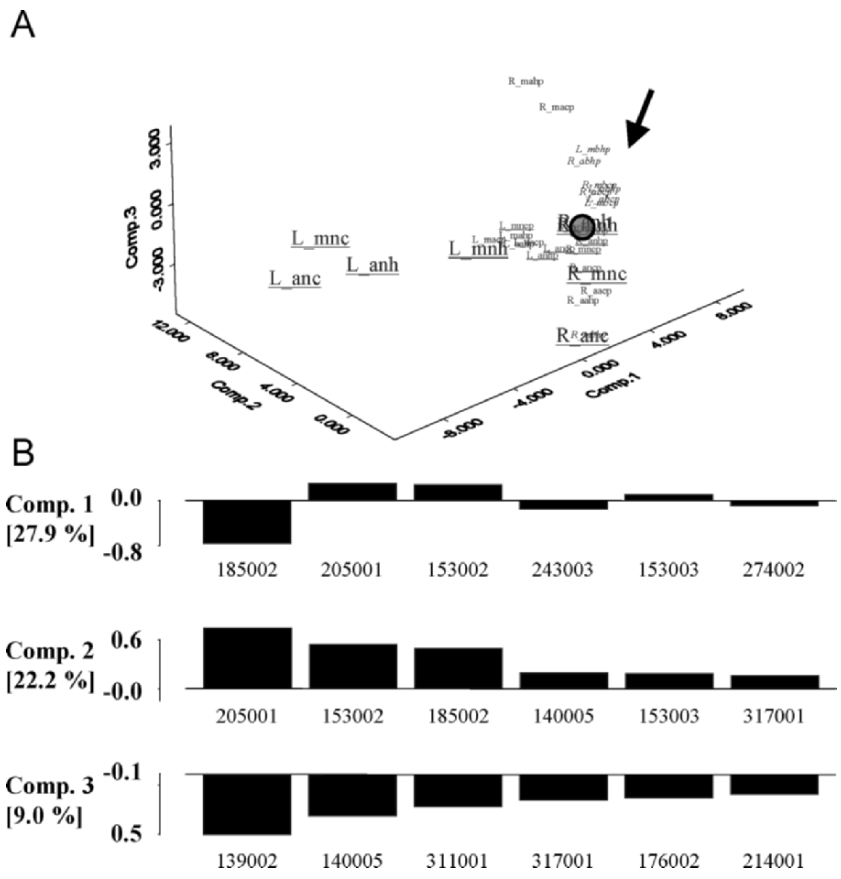


Figure 4-4. PCA analysis of the subset of polar and combined protocols. Component scores are plotted in **A**. Ranked analyte loadings are shown in **B. Comp. 1:** 185002, dehydroascorbic acid; 205001, hexadecanoic acid, n-; 153002, pyroglutamic acid; 243003, inositol-phosphate, myo-; 153003, butyric acid, 4-amino- and 274002, trehalose; **Comp. 2:** 205001, hexadecanoic acid, n-; 153002, pyroglutamic acid; 185002, dehydroascorbic acid; 140005, threonic acid-1,4-lactone; 153003, butyric acid, 4-amino- and 317001, caffeoylquinic acid, 4-trans-; **Comp. 3:** 139002, nicotine; 140005, threonic acid-1,4-lactone; 311001, caffeoylquinic acid, 3-trans-; 317001, caffeoylquinic acid, 4-trans-; 176002, aconitic acid, cis- and 214001, caffeic acid, trans-. Protocol codes are as listed in Table 4-1 with a L_- or R_-prefix to indicate leaf and root preparations. Circles indicate positions of the standard protocol mnhp (closed circle). Polar preparations (small), combined preparations (large), non-pH-adjusted preparations (underlined font), acidic preparations (normal font), and basic preparations (italic font). The arrow indicates the position of basic-pH protocols.

recommended when combined lipid and polar fractions of leaf and root are analysed. Cold extracts were most diverse when analysed in combined mode. In addition we found evidence for a clear matrix effect on dehydroascobic

acid. Dehydroascorbic acid was – with a few infrequent exceptions – absent from root extracts (data not shown). Further examples of matrix-specific metabolite recovery are to be expected from our future analyses.

3.2.2 Influence of extract polarity

Predominance of the influence of extract polarity was as expected independent of the biological matrix (Figure 4-2). Extract polarity was the major influence in our experiments (Figure 4-1A). Highly lipophilic compounds, such as octadecanoic and hexadecanoic acid, had the strongest influence on sample partitioning in PCA (Figure 4-1B). Interestingly, dehydroascorbic acid, threonic acid-1,4-lactone, and putrescine also ranked among those compounds which were sensitive to polarity changes of the extraction protocol. Combined analysis of polar and lipid fractions were similar to polar fractions, but unexpectedly did not result in an intermediate behaviour between lipid and polar fractions. For example, dehydroascorbic acid was strongly increased in combined analyses of leaf samples. Especially, combined analyses of leaf samples differed considerably as was demonstrated by Figure 4-4A. Component 1 and 2 (Figure 4-4) allowed clear differenttiation of combined analyses of leafs, L_anc, L_mnc, L_anh, and L_mnh, from respective root and non-combined preparations. Metabolite loadings indicated strong influences on the recovery of dehydroascorbic acid, hexadecanoic acid, pyroglutamic acid, inositol-phosphate, butyric acid, 4-amino-, and threonic acid-1,4-lactone (Figure 4-4B). Interestingly the combined but hot variations of the standard protocol, L_mnh and R_mnh, were least affected, and were in the case of root samples highly similar (Figure 4-4A).

3.2.3 Influence of pH

One of the controversial features of the metabolite profiling protocol may have been the absence of buffering during extraction. All tested buffer substances interfered with either chemical derivatization or chromatographic performance. Determination of the native pH of extracts demonstrated, however, an almost constant pH at approximately pH 6.3. Because slightly acidic pH is the most favourable condition for many pH labile compounds we deemed non-pH-adjusted metabolite profiling to be an acceptable procedure. Here, we show that only minimal differences are caused by acidifying extracts. In contrast, basic pH introduced a clear difference (Figures 4-1A and 4-4A). Substances which respond to basic conditions were revealed by component 3 (Figure 4-4B), i.e., nicotine, threonic acid-1,4-lactone, caffeoylquinic acids, aconitic acid, and caffeic acid. In lipid analyses acidic pH appeared to be even beneficial. In general acidification introduced less variance as compared to non-adjusted or basic pH (Figure 4-3A).

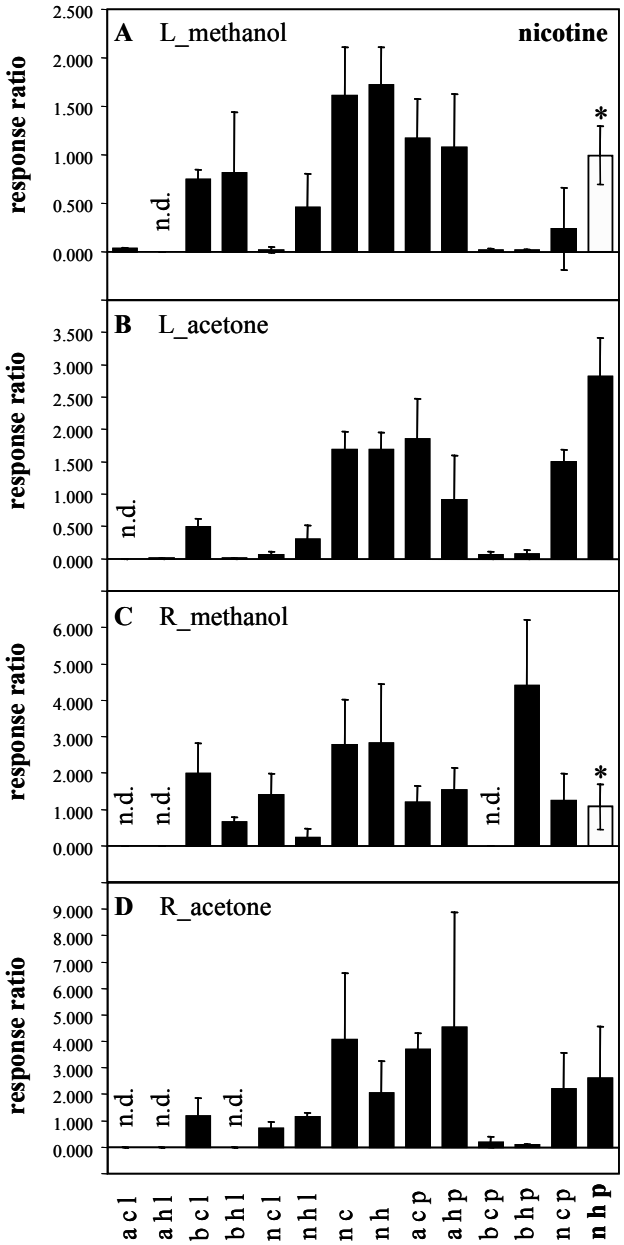


Figure 4-5. Recovery pattern of nicotine (MPIMP-ID: 139002-101). Leaf preparations (A, B), root preparations (C, D), methanol preparations (A, C), acetone preparations (B, D), n.d. (not detectable). Star indicates standard preparation set to response ratio 1.0.

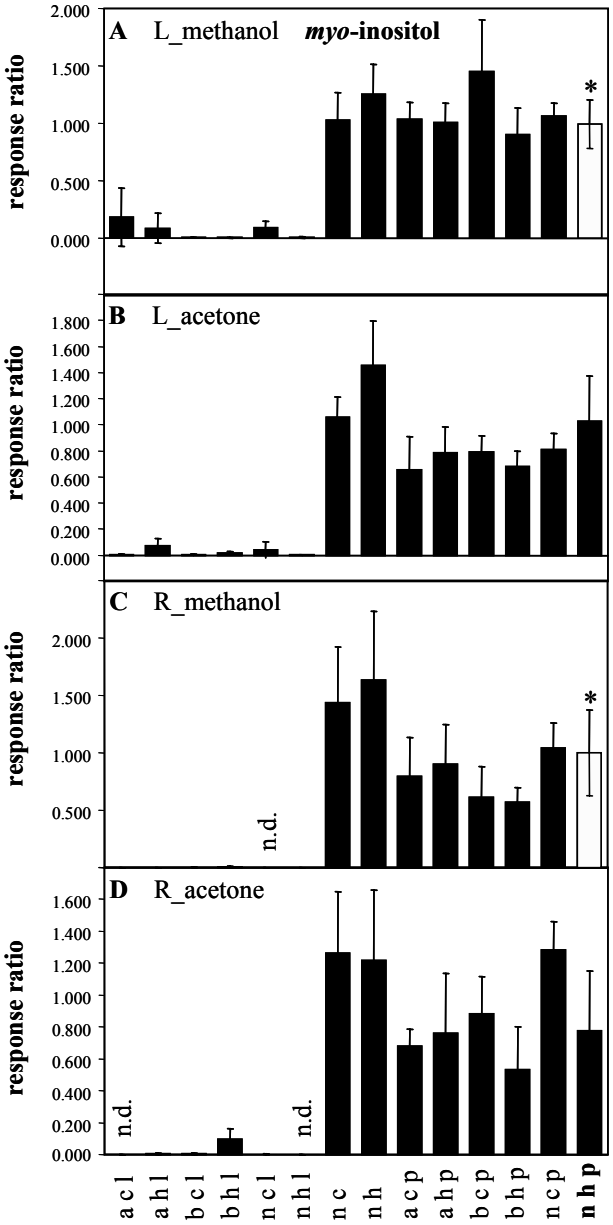


Figure 4-6. Recovery pattern of *myo*-inositol (MPIMP-ID: 209002-101). Leaf preparations (A, B), root preparations (C, D), methanol preparations (A, C), acetone preparations (B, D), n.d. (not detectable). Star indicates standard preparation set to response ratio 1.0.

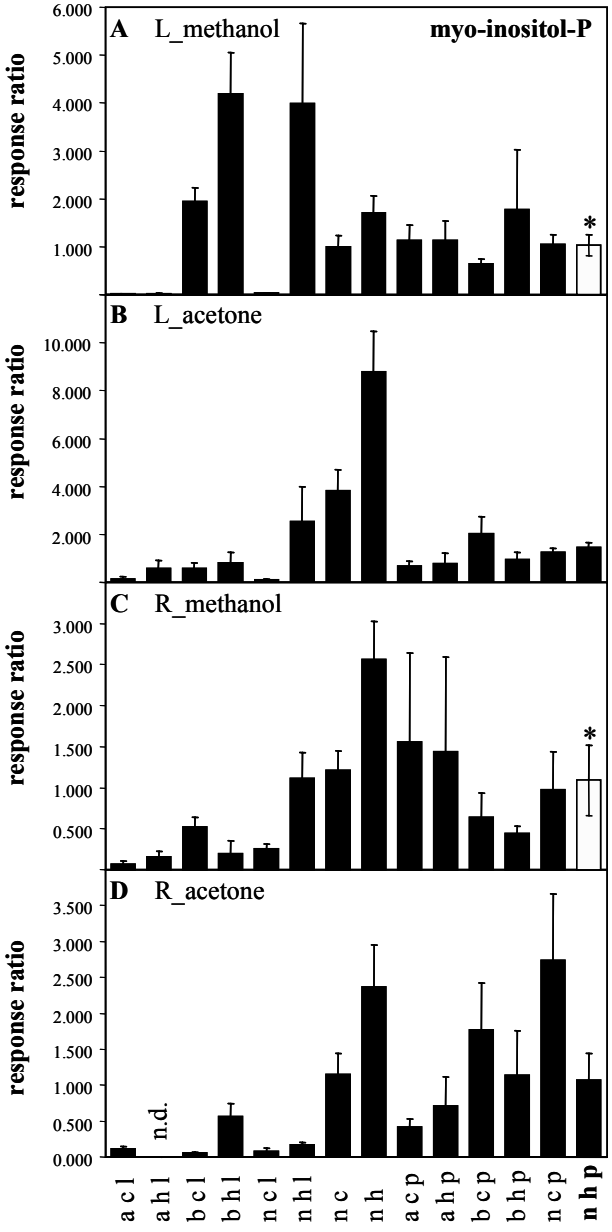


Figure 4-7. Recovery pattern of *myo*-inositol-phosphate (MPIMP-ID: 243003-101). Leaf preparations (A, B), root preparations (C, D), methanol preparations (A, C), acetone preparations (B, D), n.d. (not detectable). Star indicates standard preparation set to response ratio 1.0.

3.2.4 Influence of temperature

Temperature had only a minor influence on the pattern of metabolite recovery. Strong temperature effects were only observed in lipid profiles (Figure 4-3). In this case cold extractions were, as a rule of thumb, less variable than hot extractions. In our hands the previously described hydrolysis of sucrose during prolonged cold extraction in methanol (Weckwerth et al., 2004) was minor except for acidified acetone extractions (data not shown). This observation may be explained by a difference in the stability of invertase(s). *Arabidopsis thaliana* was used in the previous study, whereas this study inherently tested the homologous enzyme(s) from tobacco.

3.2.5 Influence of primary solvent

The choice of primary solvent, i.e., methanol or acetone, did not fundamentally affect metabolite recovery patterns (Figures 4-5, 4-6, and 4-7). More detailed analyses of single metabolites may, however, reveal better criteria for the choice of primary solvents.

4 ANALYSIS OF METABOLITE RECOVERY

We choose three metabolites, i.e., nicotine, *myo*-inositol, and *myo*-inositol-phosphate, which may serve as examples for the detailed analyses of metabolite recovery, which was made possible through our project.

The alkaloid, nicotine, while present in all polar fractions, was almost excluded from the basic polar fractions (Figure 4-5). Instead nicotine accumulated in the basic lipid fractions and was partially present in non-pH-adjusted lipid fractions. Cold extraction had the potential to improve nicotine partitioning into the lipid phase. Because partitioning into the lipid layer did rarely correlate with reduced presence in the polar layer, we concluded that basic pH acted on both extraction from the sample and liquid partitioning of nicotine. Increased nicotine recovery in R_mbhp (Figure 4-5C) was contrary to the generally observed trends. Thus a hitherto elusive factor was indicated, which may influence nicotine recovery.

myo-Inositol, as expected, exhibited clear partitioning into the polar fractions (Figure 4-6). Recovery was almost constant and robust. Temperature and pH effects were small. Improved recovery appeared to be possible by combined analysis of lipid and polar fractions.

myo-Inositol-phosphate exhibited unexpected partial or even dominant presence in lipid fractions (Figure 4-7A). Recovery of this compound consistently increased in combined hot analyses (Figure 4-7). We interpret this observation as possible breakdown of phosphatidylinositol lipids, especially under hot basic or non-pH-adjusted conditions.

5 CONCLUSION

Investigations of extraction conditions were clearly proven to be highly useful for obtaining detailed information on the analysis of metabolites in complex mixtures. Information on metabolite recovery, as was previously suggested and analysed for other plant matrices (Fiehn et al., 2000b; Roessner et al., 2000; Roessner-Tunali et al., 2003; Gullberg et al., 2004), is essential for metabolite profiling and a deeper understanding of designing increasingly robust and complex extracts for metabolite profiling. As cautionary remarks, we would like to stress that analysis of any novel biological matrix or change to a novel protocol has to be closely checked for possibly arising matrix effects (Kopka et al., 2004). Our investigation clearly demonstrated that these effects may unexpectedly occur and introduce major variances as shown for most of the analyses of lipid fractions (Figure 4-3A).

Unfortunately, we had at the beginning of this project only restricted means to test the matrix effect of root vs leaf material under standard conditions, for example, by adding stable isotope labelled internal standard compounds. We suggest improvement of the quantitative standardization of metabolites by internal standardization using complex mixtures of stable isotope-labelled metabolites. The use of synthetic internal standards was recommended earlier (Fiehn et al., 2000b; Gullberg et al., 2004). For discussion of an extended standardization concept and of the potential of in vivo stable isotope labelling, we would like the reader to refer Birkemeyer et al. (2005). The partial ^{13}C-labelling of yeast for use in yeast metabolomics studies (Mashego et al., 2004) and the *in vivo* labelling of metabolites by ^{15}N (Harada, et al., 2004) represent promising first applications.

Finally we expect that the subsequent in-depth analysis of our data set, especially the inventory and identification of novel and condition-specific analytes, will yield further valuable insight into improved range of metabolite classes, potential breakdown of complex metabolites, possible designs of GC-MS profiles with improved robustness, and the complementary design of multiple GC-MS profiles as compared to the standard protocols of polar and lipid GC-MS metabolite profiles, which are currently in use.

ACKNOWLEDGEMENTS

We would like to thank A.R. Fernie, A. Erban, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, and U. Roessner-Tunali, Australian Centre for Plant Functional Genomics, School of Botany, University of Melbourne, Australia, for critically reading and discussing our manuscript.

REFERENCES

Ausloos, P., Clifton, C.L., Lias, S.G., Mikaya, A.I., Stein, S.E., Tchekhovskoi, D.V., Sparkman, O.D., Zaikin, V., and Zhu, D., 1999, The critical evaluation of a comprehensive mass spectral library, *J. Am. Soc. Mass Spectrom.* **10**: 287–299.

Bieleski, R.L., 1964, Problem of halting enzyme action when extracting plant tissues, *Anal. Biochem.* **9**:431–442.

Birkemeyer, C., Kolasa, A., and Kopka, J., 2003, Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones, *J. Chromatogr. A* **99**3:89–102.

Birkemeyer, C., Luedemann, A., Wagner, C., Erban, A., and Kopka, J., 2005, Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling, *Trends Biotechnol.* **23:**28–33.

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Kloska, S., Kopka, J., and Udvardi, M.K., 2004, Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*, *Plant J.* **39**:487–512.

Fiehn, O., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000a, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry, *Anal. Chem.* **72**:3573–3580.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L., 2000b, Metabolite profiling for plant functional genomics, *Nat. Biotechnol.* **18**:1157–1161.

Gullberg, J., Jonsson, P., Nordstrom, A., Sjostrom, M., and Moritz, T., 2004, Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry, Anal. Biochem. **331**:283–295.

Harada, K., Fukusaki, E., Bamba, T., and Kobayashi, A. 2004, In-vivo ^{15}N-enrichment of metabolites in *Arabidopsis* cultured cell T87 and its application for metabolomics*, Third international congress on plant metabolomics*, #23.

Kopka, J., Fernie, A.F., Weckwerth, W., Gibon, Y., and Stitt, M., 2004, Metabolite profiling in Plant Biology: Platforms and Destinations, *Genome Biol.* **5(6):**109–117.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dörmann, P., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R., and Steinhauser, D., 2005, GMD@CSB.DB: The Golm Metabolome Database, *Bioinformatics* (doi:10.1093/bioinformatics/bti236).

Mashego, M.R., Wu, L., Van Dam, J.C., Ras, C., Vinke, J.L., Van Winden, W.A., Van Gulik, W.M., Heijnen, J.J., 2004, MIRACLE: mass isotopomer ratio analysis of U-^{13}C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites, *Biotech. Bioeng.* **85:**620–628.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000, Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.* **23**:131–142.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A.R., 2001, Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems, *Plant Cell* **13**:11–29.

Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D., and Fernie, A.R., 2003, Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development, *Plant Physiol.* **133**:84–99.

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgren, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., Fernie, A.R., and Kopka, J., 2005, GC-MS libraries for the rapid identification of metabolites in complex biological samples, *FEBS Letters* (in press).

Stein, S.E., 1999, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, *J. Am. Soc. Mass Spectrom.* **10**:770–781.

Sumner, L.W., Mendes, P., and Dixon, R.A., 2003, Plant metabolomics: large-scale phytochemistry in the functional genomics era, *Phytochemistry* **62**:817–836.

Wagner, C., Sefkow, M., and Kopka, J., 2003, Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles, *Phytochemistry* **62:**887–900.

Weckwerth, W., Tolstikov, V., and Fiehn, O., 2001, Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, ASMS, Chicago, pp. 1–2.

Weckwerth, W., Wenzel, K., and Fiehn, O., 2004, Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks, *Proteomics* **4**:78–83.

Chapter 5

UNCOVERING THE PLANT METABOLOME: CURRENT AND FUTURE CHALLENGES

Ute Roessner
Australian Centre for Plant Functional Genomics, School of Botany, University of Melbourne, Victoria 3010, Australia

Abstract: Within the plant kingdom, it has been estimated that several hundred thousand different metabolic components may be produced, with abundances varying through of six orders of magnitude. The goal of metabolomics is a comprehensive and non-targeted analysis of metabolites in a biological system. Any valid metabolomic approach must be able to unbiasedly extract, separate, detect and accurately quantitate this enormous diversity of chemical compounds. These requirements dictate the challenges that are continually addressed in the field of plant metabolomics. To date, both gas- and liquid-based chromatography systems, in combination with various MS detection technologies, have been employed to analyse complex mixtures of extracted metabolites. In addition, nuclear magnetic resonance spectroscopy has been used to fingerprint plant systems, but will be not discussed in this context and has been reviewed elsewhere.

Although the technologies employed in metabolomic analyses are uncovering a huge amount of new knowledge in biology, a range of challenges are still to be faced. One bottleneck in metabolomic analysis is the identification of novel compounds. Additionally, in order to allow greatest spatial resolution, the sensitivity and selectivity of currently available technologies has to be increased. Multi-parallel and high-throughput analyses result in large data sets which need be evaluated, extracted, and interpreted. As a result, automated algorithms have to be developed. One of the major future challenges in the metabolomics field will be the integration of metabolic data with genomic and proteomic data sets. The ultimate goal is to comprehensively describe complex biological systems and as such, metabolomics has become an important player in systems biology. In the following text each of these challenges concurrently being connected with metabolomic analyses will be discussed.

Key Words: plant metabolome; chromatography; mass spectrometry; data analysis; data interpretation.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 71–85.

1 INTRODUCTION

The development of tools to characterize genetic diversity in plant systems has made enormous progress over the last few years. Transgenic knockout populations, transposon insertions, and highly efficient ways to genotype single nucleotide polymorphisms within large populations have paved the way to a much broader base of diversity than imagined a few years ago. Furthermore, these developments have occurred in tandem with the elucidation of complete genomes and the rapid development of multiparallel technologies designed to access and describe the properties of biological systems (Celis et al., 2000). Most prominent amongst these new technologies has been the establishment of protocols for the determination of the expression levels of many thousands of genes in parallel (for review see Hardiman, 2004) and the detection, identification, and quantification of the protein complement (for review see Heazlewood and Millar, 2003). The logical progression from the large-scale analysis of transcripts to proteins is the determination of metabolite profiles in cells, tissues, and organisms. Importantly, the improvement of analytical instrumentations, such as mass spectrometry, has opened up the possibilities of determining and identifying a large number of metabolic compounds in parallel and in a high-throughput manner. The term metabolomics describes the comprehensive, non-targeted detection, and quantification of all compounds derived from a biological system.

2 ANALYTICAL TECHNOLOGIES FOR METABOLITE ANALYSES IN PLANT TISSUES

2.1 GC-MS

To date, both gas- and liquid-based chromatographic systems in combination with various mass spectrometry (MS) detection technologies, as well as nuclear magnetic resonance spectroscopy (NMR), have been employyed to analyse complex mixtures of extracted metabolites. Due to its overall robustness, gas chromatography coupled to electron impact ionization mass spectrometry (GC-EI-MS) has played a major role in high-throughput metabolite analyses (for review see Roessner et al., 2002). The use of GC allows separation of mixtures of compounds with high separation efficiency and sensitivity. In combination with MS, it also provides very accurate, sensitive and selective identification and quantification of separated compounds by their specific mass spectrum. Moreover, MS analysis further increases the resolution of the chromatography used as two co-eluting substances can be separated by their fragmentation pattern. Off-the-shelf

instruments are now able to rapidly and quantitatively detect up to 500 compounds simultaneously in crude plant extracts, depending on tissue and extraction procedure. In the past, GC-MS technology has been applied and optimized for simultaneous analyses of metabolites in many different plant species, such as *Arabidopsis thaliana* (Fiehn et al., 2000), *Solanum tuberosum* (Roessner et al., 2000), *Medicago truncatula* (Duran et al., 2003), *Lycopersicon esculentum* (Roessner-Tunali et al., 2003), *Saccharum officinarum* (S. Bosch, personal communication), *Lotus japonicus* (Colebatch et al., 2004), and *Cucubita maxima* (Fiehn, 2003).

In many of these detailed characterizations, it was shown that a one-dimensional GC separation approach does not resolve all compounds in high-complex extracts of plants. Recently a new approach has been taken, in which a second dimension of GC is applied to further separate the mixtures. GC × GC-TOF-MS has been already successfully applied to highly resolve volatile compounds of roasted coffee beans (Ryan et al., 2004). In the future, this technology will allow a more complete definition of the chemical composition of plants.

Despite the many advantages that GC-EI-MS has in metabolomics applications, there are also limitations of this technology. One of these is that GC can only be used for low molecular weight (<1000 Da) compounds, which are either volatile at relatively low temperatures, or which can be chemically transformed into volatile derivates. Thus, for a comprehensive analysis of a greater range of plant metabolites, complementary techniques have to be established (Kopka et al., 2004).

2.2 LC-MS

One complementary approach to GC-EI-MS in metabolite analyses is the application of liquid chromatography coupled to electrospray ionization mass spectrometry (LC-ESI-MS). The main advantages of LC-ESI-MS are twofold. Firstly, compounds do not have to be chemically altered prior to analysis and secondly, highly polar, thermo-unstable and high-molecular weight compounds, such as oligosaccharides or lipids, are able to be separated and quantified. LC in combination with an ultraviolet or visible light (UV/VIS) or diode-array detection (DAD) has been applied for many years in plant metabolite analyses. An enormous range of different columns and elution procedures exist for the separation and detection of many different classes of compounds. When coupled to MS, these provide further selectivity, unbiased detection, and most importantly, information about the structure of detected compounds. This multidimensional approach has been successfully applied for the analysis of a wide range of primary and secondary metabolites in plant tissues (Tolsitkov and Fiehn, 2002; Huhmann and Sumner, 2002). Recently, the use of a monolithic column enabled the separation of several hundred chromatographic peaks derived from extracts

of *Arabidopsis* (Tolstikov et al., 2003). Another research group has reported the detection of 1,400 components (based on mass-to-charge ratios) by direct injection of *Arabidopsis* extracts into a quadrupole time-of-flight (QTOF) hybrid mass spectrometer (von Roepenack-Lahaye et al., 2004). The resolution and selectivity of mass detection can be dramatically increased to up to 5,000 signals from a single plant extract by application of Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) as shown by Aharoni et al. (2002). In the future, this technique will play an increasing role in metabolic fingerprinting approaches where large mutant collections are screened for metabolic alterations following random mutations.

2.3 Increasing sensitivity

An additional challenge in metabolite analyses is the development of technologies for the isolation and detection of metabolites from very small samples sizes in order to increase spatial resolution in single cell or tissue-specific investigations. These techniques have to be designed to combine high sensitivity with selectivity. First remarkable reports have been given on the determination of the distribution of IAA in *Arabidopsis* plants (Muller et al., 2002) or even the distribution of ATP in *Vicia faba* embryos (Borisjuk et al., 2003). Future research has now to face multiparallel analyses of metabolites on a cell and organ level. One attractive technology to increase sensitivity is capillary electrophoretic separation techniques in combination with laser-induced fluorescence (CE-LIF) or mass spectrometric detection (CE-MS), which has been already proven to give promising results. For example, CE-LIF allowed the separation and quantification of a large range of amino acids and sugars in approximately 50 pL of phloem sap or in five pooled mesophyll cells of *Cucurbita maxima* (Arlt et al., 2001; S. Brandt, 2004, personal communication). By using CE-MS, more than 80 main metabolites belonging to glycolysis, photorespiration or the oxidative pentose phosphate pathway could be analysed in rice leaf extracts (Sato et al., 2004). It has to be especially noted that in this study, the ability to analyse many unstable substances in parallel, which only occur in low concentrations in *planta,* such as fructose-1,6-bisphosphate or ribulose-1,5-bisphosphate, was presented.

In the past decade many new technologies have been established which are currently used in novel biological information discovery in plant physiology and functional genomics. In summary, if the working definition for metabolomics means the analysis of all metabolites in a biological system, it requires a platform of complementary analytical technologies for comprehensive selectivity and sensitivity.

3 IDENTIFICATION OF UNKNOWN COMPOUNDS

Non-targeted metabolite detection in plant tissue results in a large number of chromatographic peaks and mass spectra, which cannot be identified easily with respect to the chemical nature of the compound. It has been shown in many metabolomic approaches that, for example, up to 70% of all peaks in a typical GC-MS chromatogram of a plant extract still remain unidentified. The interpretation of mass spectra following GC-EI-MS analysis is very difficult for two reasons. Firstly, derivatization dramatically alters the chemical structure of the compounds. Secondly, the use of electron impact (EI) to ionize the compounds is a very harsh method that leads to complex fragmentation patterns. As a result, two strategies are used to identify the chemical nature of as many peaks as possible. Firstly, the spectra of all resolved peaks are compared to commercially available EI mass spectrum libraries such as NIST (http://www.nist.gov/: National Institute of Standards and Technology, Gaithersburg, USA). However, although these libraries contain over 350,000 entries, the majority of these are non-biological compounds. In a second approach, commercial standard compounds, that are assumed to be present at detectable levels within plant tissues, are analysed. A reference library containing both the retention time of these compounds (as determined under the same conditions) and the corresponding mass spectrum can be created (Wagner et al., 2003). Identification by retention time is verified by co-chromatography of each standard substance with substances obtained in the plant extract. A major problem of this approach is that most plant compounds are not commercially available, especially the enormous number of secondary metabolites. Recently the publication of the first "biological" public domain GC-MS mass spectra library (MSRI; http://csbdb.mpimp-golm.mpg.de/gmd.html) was described (Kopka et al., 2005; Schauer et al., 2005). This library contains a large number of identified and unknown, but repeatedly observed EI-mass spectra of many different plant species and organs. A feature of this library is its compatibility with the NIST software and GC-MS evaluation software packages, such as automated mass spectral deconvolution and identification system (AMDIS) (see below). Further references to this mass spectral and retention time index library and its applications may be found in Chapter 4X by C. Birkemeyer and J. Kopka.

For LC-MS signal identification the situation is much more difficult. Mass spectra generated by LC-MS are typically instrument dependent and therefore, standard reference LC-MS spectral libraries are of limited use. The minimum information acceptable for the identification of novel organic compounds or metabolites has been traditionally defined by the scientific

literature criteria and often includes elemental analysis, NMR and MS spectral data for the isolated compound. One method for preliminary identification of unknowns appears to be the use of multidimensional instrumental techniques (based on combinations of GC-MS, LC-MS, MS/MS, or MS/NMR), which enable both comparative profiling and structural elucidation. For example, LC-QTOF-MS/MS (liquid chromatographic quadrupole tandem time-of-flight mass spectroscopy) has the potential to provide accurate mass and product-ion information of chromatographically separated metabolites. Experimental mass data can then be used for the calculation of an elemental composition and be compared with available mass information in, e.g., the NIST or KEGG database for possible structure suggestions. Further stepwise fragmentation by tandem MS (MS^n) leads to product-ion information, which can be used to determine/confirm structure. Although this gives much information about the potential structure of the compound, the final confirmation of the identity of the compound has to be done by either analysis of an authentic standards substance or by analysis of the purified sample using NMR.

The method of choice for unambiguous peak identification is NMR, which offers high chemical selectivity. In combination with LC and MS (LC-MS-NMR), it represents the ultimate technology for peak identification and structure elucidation (Wolfender et al., 2003) although the in-line version of this combination to date is still highly limited by the low sensitivity of the NMR instrument.

4 AUTOMATION OF DATA EVALUATION

Once an analytical platform is established a large number of samples can be analysed very quickly. This makes it an impractical and tedious task to manually extract information of each single chromatogram. One challenge of multitargeted compound analysis is the development of automated chromatogram evaluation. Many software packages delivered with the GC- or LC-MS system (Xcalibur, ThermoElectron, Austin, USA or HP Chemstation, Agilent, Palo Alto, USA) are able to use either self-created or commercial mass spectra libraries for peak detection, identification, and integration. The limitation of these software packages are that they search and integrate only targets, which the researcher has to know and enter into the search lists. This situation as been improved recently with the development of novel software packages for untargeted chromatogram evaluation based on mass spectral deconvolution. Deconvolution means the separation of corresponding fragments to one mass spectrum and thus for a single compound. This can be either achieved in an automated fashion by the software packages provided

with the GC-MS instrument (Pegasus, Leco, St. Jospehs, USA) or separate software can be applied, such as AMDIS (http://chemdata.nist.gov/mass-spc/amdis/; National Institute of Standards and Technology, Gaithersburg, USA). Recently other helpful commercial and free software packages have become available. Examples include MSFacts for GC-MS (Duran et al., 2003) or MetAlign for GC- and LC-MS (www.metalign.nl), which automatically import, reformat, align, correct the baseline and export large chromatographic data sets to allow more rapid visualization and interrogation of metabolomic data. To date, these software packages are indispensable for unambiguous data extraction. Very recently a novel software package named AnalyzerPro (www.spectralworks.com; Runcorn, Cheshire, UK) has been made available which meets the high requirements of an automatic GC-MS and also LC-MS^n chromatogram evaluation. In addition to signal deconvolution, mass spectra library matching and quantification, the implementation of retention time indices (RI) for improved signal identification are beneficial. Retention times of eluted substances following chromatographic separation do change dramatically over time. Retention time indices include for their calculation a range of added time references (e.g., long-chain alkanes) and therefore provide a better prediction of the absolute retention time of the analytes. In addition, retention time indices are very stable both within and between systems, allowing valid system to system comparisons, provided that injection, separation and ionization parameters are kept similar (Schauer et al., 2005).

5 DATA INTERPRETATION AND VISUALIZATION

As mentioned above, high-throughput analysis of a collection of samples results in large data sets, which have to be interpreted in a biological context. To date, statistical tools for pattern-recognition, such as hierarchical clustering (HCA) or principle component analysis (PCA), are routinely used for ease of comparison, and visualization of similarities and differences between data sets by definition of clusters (Fiehn et al., 2000; Roessner et al., 2001a, 2001b). Another approach is to detect dependencies and connections between metabolites and more recently, between genes, proteins, and metabolites by using pair-wise analysis of linear correlations (Urbanczyk-Wochniak et al., 2003; Steuer et al., 2003). Interestingly, when significant correlations are connected, the construction of regulatory networks becomes possible. The comparison of network connectivity between different genotypes allows not only the identification of novel pathways, it also represents a way of uncovering “silent” mutations, which do not show any obvious phenotype in any of the parameters under analysis (Weckwerth et al., 2004).

6 COMBINATION OF STEADY-STATE METABOLOMICS WITH METABOLIC FLUX ANALYSIS

The measurement of steady-state levels of metabolites, as described in this review, gives new insights into metabolic networks at a given time. But the real behaviour of plant metabolism can be only understood by determination of the dynamics of metabolism. The basis of metabolic flux analysis (MFA) is a combination of stable isotope labelling under steady-state conditions and NMR or MS-based detection systems to follow the distribution of label. This technique has been applied in detail in micro organism research but will play an increasingly important role in plant research (for review see Schwender et al., 2004). The application of a multiparallel detection method such as GC-or LC-MS allows determination of isotope label in very many metabolites in one experiment and therefore gives the opportunity to calculate metabolic fluxes of many different pathways simultaneously (Schwender et al., 2003; Roessner-Tunali et al., 2004). The power of this method becomes striking when it is incorporated with steady-state metabolite level determinations. This has been demonstrated by Foerster et al. (2002), showing *in silico* pathway analysis using stoichiometric models in yeast, which were constructed from knowledge of biochemical reaction networks in the cells. By further implementation of available genomic, biochemical, and physiological information, these authors reported the reconstruction of a genome-scale metabolic network from *S. cerevisiae* (Famili et al., 2003), which produced computed predictions for phenotypes following *in silico* mutation and therefore allowed gene function identification. In conclusion, a metabolomics approach in combination with stable isotope metabolic flux analysis will provide important insights in plant functional genomics studies. Another obvious use of this information will be in more rational approaches in metabolic engineering of novel, valuable biotech-crops (Sweetlove et al., 2003).

7 DEVELOPMENT OF DATABASES FOR METABOLOMICS-DERIVED DATA

In the past it has been noted by several scientists, that the large data sets generated by post-genomics technologies have to be transmitted, stored safely and be made available in convenient and accessible formats (Goodarce et al., 2004). The implementation of relational databases for data storage requires well-designed data standards. The DNA microarray community has agreed on the development of a minimum information about a microarray experiment

(MIAME, Brazma et al., 2001) and its structure has been widely accepted. Similar initiatives are underway for the proteomics community (PEDRo, Taylor et al., 2003). Whilst metabolic databases such as the KEGG system (Goto et al., 2002) or MetaCyc (Krieger et al., 2004) provide detailed information about metabolic pathways and enzymes of a variety of organisms, the development of a data standard equivalent to MIAME and PEDRo describing metabolomics data in their experimental context has been proposed only very recently (MIAMET, Bino et al., 2004; ArMet, Jenkins et al., 2004). On the other hand it will be not only important to store metabolic profiling data but to also integrate these data with metabolic pathway information which will be the future source of knowledge discovery. Recently, a database has been developed, which assembles information about different *Arabidopsis* metabolic pathways (AraCyc) and provides diagrams showing metabolites and genes encoding the enzymes in each pathway (Mueller et al., 2003). For a holistic integration of numerous multiparallel genomic, proteomic, metabolomic, and metabolic flux analysis data sets with metabolic pathway information the "Pathway Tools Omics Viewer" (http://www.arabidopsis.org:1555/expression.html) has been enabled, which in an easy and powerful manner paints experimental data onto the biochemical pathway map. Another example for such "mapping" tool is MapMan (Thimm et al., 2004), which allows users to visualize comparative metabolic and also transcriptional profiling data sets on existing metabolic templates and design their own templates. For a holistic integration of numeric multiparallel genomic, proteomic and metabolomic data sets a data managing system for editing and visualization of biological pathways was developed, which on a publicly available domain will be very important for data-mining in the functional genomics field (MetNetDB, Syrkin Wurtele et al., 2003; PaVESy, Luedemann et al., 2004). These software tools henceforth will become important to map novel findings onto metabolic pathways and fully understand the function of each gene, encoded protein and metabolite.

8 FROM TECHNOLOGY TO BIOLOGY

Once a robust metabolite analysis platform has been established and reliable data can be produced, the range of plant research applications is enormous. These can vary from answering simple biological questions, i.e., what are the metabolic differences between two cultivars, to investigations regarding complex metabolic networks. For example, a metabolomics approach can be used to determine the influence of transgenic and environmental manipulations on the metabolite profile as demonstrated by a detailed characterization of the metabolic complement of a number of transgenic potato tubers altered in their starch biosynthetic pathway and wild

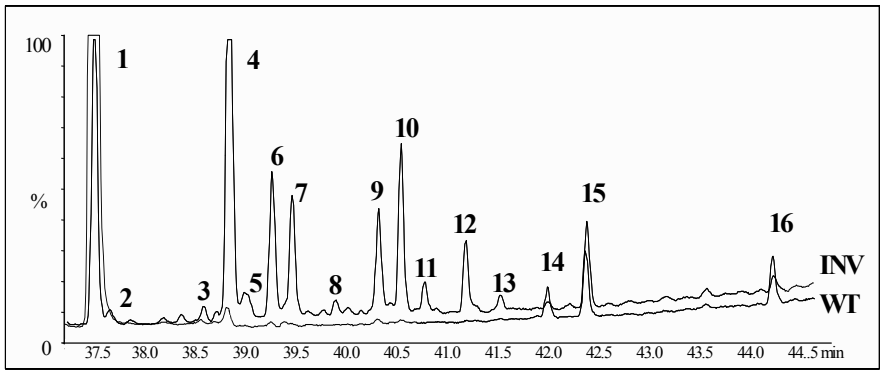


Figure 5-1. Comparison of a specific region of a GC-MS chromatogram of wild type potato tuber (WT, lower line) compared to tubers expressing a yeast invertase in the cytosol (INV, upper line). 1: sucrose; 3: maltose TMS; 4: maltose MEOX1; 5: trehalose TMS; 6: maltose MEOX2; 7: maltitol TMS; 12: isomaltose MEOX1; 13: isomaltose MEOX2, 2, 8, 9, 10,11, 14, 15 and 16 are not identified, mass spectra suggest they are sugars or sugar derivatives.

type tubers incubated in different sugars using GC-MS (Roessner et al., 2001a, 2001b). Due to this non-targeted approach, many unintended differences of transgenic tubers compared to wild type were detected (Roessner et al., 2001a; Figure 5-1). This study showed that using a metabolomic approach, it is possible to easily phenotype genetically and environmentally diverse plant systems.

Another useful application of metabolomics is in the field of functional genomics, which aims to identify of gene functions using high-throughput phenotyping technologies, as for example in investigations of responsible genes and their products in plant adaptations to different abiotic stresses. Often the role of certain metabolites in stress response could be assigned, as for example proline plays a major role in salt stress adjustments in rice (Garcia et al., 1997). The detailed characterization of metabolic adaptations to low and high temperatures in *Arabidopsis thaliana* has already demonstrated the power of this approach (Kaplan et al., 2004; Cook et al., 2004). Interestingly, it could be shown that low temperatures have more profound effects than heat, and novel findings of metabolic adaptation to temperature stress were identified (Kaplan et al., 2004). Another important report on using metabolomics as a tool in investigating metabolic responses of *Medicago truncatula* cell cultures to biotic and abiotic elicitors has revealed both elicitor-specific responses of metabolite levels as well as more generic responses in which similar metabolites responded independently of the type of stress (Broeckling et al., 2004). Nutrient deficiencies and

toxicities represent another example of common stress situations, e.g., it has been already demonstrated that the availability of inorganic nitrogen can reprogram carbohydrate metabolism (Stitt et al., 2002). This has been recently verified in more detail by a metabolomic investigation of the effects on tomato leaf metabolism grown in saturated, replete, and deficient nitrogen supplement conditions (Urbanczyk-Wochniak et al., 2005), showing the impact of nitrogen levels in the growth solutions on a wide range of metabolites. Similar striking effects on metabolite levels have been found when barley plants were grown in conditions where other inorganic nutrients are unavailable, e.g., phosphate or zinc (Roessner-Tunali, unpublished results). In the future, this approach will lead to the determination of the role of both metabolites and genes in stress tolerance and thus provide new ideas for genetic engineering of novel stress-resistant crops.

The next step of interpretation of metabolomic data sets can be achieved when they are integrated with other "omics" data such as transcriptomic or proteomic data. First attempts to face this challenge have been presented by Urbanczyk-Wochniak and co-workers who combined data obtained from microarray analysis and metabolite profiling of the same sample (Urbanczyk-Wochniak et al., 2003). A co-response analysis of both data sets has resulted in a large number of significant correlations between mRNA transcripts and metabolites. Some of these could be explained easily with existing biochemical knowledge but most were found to be novel, and thus highlighted the power of these integrated approaches for gene and metabolite function identifications. A similar investigation simultaneously analysed transcripts and metabolite levels in *Lotus japonicus* nodules to study symbiotic nitrogen fixation in detail (Colebatch et al., 2004). This report has shown clear interrelationships between transcript and metabolite responses dependent on a physiological event.

Last but not least, it has to be noted that a detailed characterization of the metabolome of a biological organism plays an integral role in a systems biology approach (Weckwerth, 2003). The aim of the emerging area of systems biology is to investigate the dynamics of all genetic, regulatory and metabolic processes in a cell and to understand the complexity of cellular networks (Kitano, 2002). Further this will give the opportunity to investigate the behaviour of biological systems with respect to the environment.

ACKNOWLEDGEMENTS

I would like to thank Dr. J. Patterson and Dr. Siria Natera for proof-reading the manuscript. My special thanks to Dr. Joachim Kopka for helpful discussions and to the Australian Centre for Plant Functional Genomics for funding.

REFERENCES

Aharoni, A., Ric de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., and Goodenowe, D.B., 2002, Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry, *OMICS* **6**:217–234.

Arlt, K., Brandt, S., and Kehr, J., 2001, Amino acid analysis in five pooled single plant cell samples using capillary electrophoresis coupled to laser-induced fluorescence detection, *J. Chrom. A* **926**:319–325.

Bino, R.J., Hall, R.H., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B., Mendes, P., Roessner-Tunali, U., Beale, M., Trethewey, R.N., Lange, B.M., Syrkin Wurtele, E., and Sumner, L., 2004, Opinion: Potential of Metabolomics as a Functional Genomics Tool, *Trends Plant Sci.* **9**:418–425.

Borisjuk, L., Rolletschek, H., Walenta, S., Panitz, R., Wobus, U., and Weber, H., 2003, Energy status and its control on embryogenesis of legumes: ATP distribution within Vicia faba embryos is developmentally regulated and correlated with photosynthetic capacity, *Plant J.* **36**:318–329.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M., 2001, Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nat. Genet.* **29**:365–371.

Broeckling, C.D., Huhman, D.V., Farag, M.A., Smith, J.T., May, G.D., Mendes, P., Dixon, R.A., and Sumner, L.W., 2005, Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism, *J. Exp. Bot.* **56**: 323–336.

Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T., Gromov, P., Yu, J., Palsdottir, H., Magnusson, N., and Ornoft, T.F., 2000, Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics, *FEBS Lett.* **480**:2–16.

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., and Udvardi, M.K., 2004, Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*, *Plant J.* **39**:487–512.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F., 2004, A prominent role for the CBF cold response pathway in configuring the low-temperature metabolomie of *Arabidopsis*, *Proc. Natl. Acad. Sci. USA* **101**:15243–15248.

Duran, A.L., Yang, J., Wang, L., and Sumner, L.W., 2003, Metabolomics spectral formatting, alignment and conversion tools (MSFACTs), *Bioinformatics* **19**:2283–2293.

Famili, I., Foerster, J., Nielsen, J., and Palsson, B.O., 2003, *Saccharomyces cerevisiae* phenotypes can be predicted by using contraint-based analysis of a genome-scale reconstructed metabolic network, *Proc. Natl. Acad. Sci. USA* **100**:13134–13139.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L., 2000, Metabolite profiling for plant functional genomics, *Nat. Biotechnol.* **18**:1157–1161.

Fiehn, O., 2003, Metabolic networks of *Cucurbita maxima* phloem, *Phytochem* **62**:875–86.

Foerster, J., Gombert, A.K., and Nielsen, J., 2002, A functional genomics approach using metabolomics and in silico pathway analysis, *Biotechnol. Bioengineering* **79**:703–712.

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., and Kell, D.B., 2004, Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends Biotechnol.* **22**:245–252.

Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M., 2002, LIGANS: database of chemical compounds and reactions in biological pathways, *Nucleic Acid Res.* **30**:402–404.

Garcia, A.B., Engler, J., Iyer, S., Gerats, T., Van Montagu, M., and Caplan, A.B., 1997, Effects of Osmoprotectants upon NaCl Stress in Rice, *Plant Physiol.* **115**:159–169.

Hardiman, G., 2004, Microarray platforms – comparisons and contrasts, *Pharmacogenomics* **5**:487–502.

Heazlewood, J.L. and Millar, A.H., 2003, Integrated plant proteomics – putting the green genomes to work, *Funct. Plant Biol.* **30**:471–482.

Huhman, D.V. and Sumner, L.W., 2002, Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer, *Phytochemistry* **59**:347–360.

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R., Kopka, J., Lane, G.A., Lange, B.M., Liu, J.R., Mendes, P., Nikolau, B.J., Oliver, S.G., Paton, N.W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, J., Sumner, L.W., Wang, T., Walsh, S., Syrkin Wurtele, E., and Kell, D.B., 2004, A proposed framework for the description of plant metabolomics experiments and their results, *Nat Biotechnol* **22**:1601–1606.

Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y., and Guy, C.L., 2004, Exploring the temperature-stress metabolomie of *Arabidopsis*, *Plant Physiol.* **136**:4159–4168.

Kitano, H., 2002, Systems Biology: A Brief Overview, *Science* **295**:1662–1664.

Kopka, J., Fernie, A.R., Weckwerth, W., Gibon, Y., and Stitt, M., 2004, Metabolite profiling in plant biology: Platforms and destinations, *Genome Biol.* **5**:109–117.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dörmann, P., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R., and Steinhauser, D., 2005, GMD@CSB.DB: The Golm Metabolome Database*, Bioinformatics* **21**:1635–1638.

Krieger, C.J., Zhang, P., Mueller, L.A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S.Y., and Karp, P.D., 2004, MetaCyc: a multiorganism database of metabolic pathways and enzymes, *Nucleic Acid Res* **32**(Database issue):D438–442.

Luedemann, A., Weicht, D., Selbig, J., and Kopka, J., 2004, PaVESy: pathway visualization and editing system, *Bioinformatics* **20**:2841–2844.

Mueller, L.A., Zhang, P., and Rhee, S.Y., 2003, AraCyc: a biochemical pathway database for Arabidopsis, *Plant Physiol.* **132**:453–460.

Muller, A., Duchting, P., and Weiler, E.W., 2002, A multiplex GC-MS/MS technique for the sensitive and quantitative single-run analysis of acidic phytohormones and related compounds, and its application to Arabidopsis thaliana, *Planta* **216**:44–56.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000, Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.* **23**:131–142.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A.R., 2001a, Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems, *Plant Cell* **13**:11–29.

Roessner, U., Willmitzer, L., and Fernie, A. R., 2001b, High-resolution metabolic phenotyping of genetically and environmentally diverse plant systems – identification of phenocopies, *Plant Physiol.* **127**:749–764.

Roessner, U., Willmitzer, L., and Fernie, A.R., 2002, Metabolic profiling and biochemical phenotyping of plant systems, *Plant Cell Rep.* **21**:189–196.

Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D., and Fernie, A.R., 2003, Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development, *Plant Physiol.* **133**:84–99.

Roessner-Tunali, U., Lui, J., Leisse, A., Balbo, I., Perez-Melis, A., Willmitzer, L., and Fernie, A.R., 2004, Flux analysis of organic and amino acid metabolism in potato tubers by gas chromatography-mass spectrometry following incubation in ^{13}C labelled isotopes, *Plant J.* **39**:668–679.

Ryan, D., Shellie, R., Tranchida, P., Casilli, A., Mondello, L., and Marriott, P., 2004, Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, *J. Chrom. A* **1054**:57–65.

Sato, S., Soga, T., Nishioka, T., and Tomita, M., 2004, Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection, *Plant J.* **40**:151–163.

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgen, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., Fernie, A.R., and Kopka, J., 2005, GC-MS libraries for the rapid identification of metabolites in complex biological samples, *FEBS Lett.* **579**: 1332–1337.

Schwender, J., Ohlrogge, J.B., and Shachar-Hill, Y., 2003, A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos, *J. Biol. Chem.* **278**:29442–29453.

Schwender, J., Ohlrogge, J., and Shachar-Hill, Y., 2004, Understanding flux in plant metabolic networks, *Curr. Opin. Plant Biol.* **7**:309–317.

Stitt, M., Muller, C., Matt, P., Gibon, Y., Carillo, P., Morcuende, R., Scheible, W.R., and Krapp, A., 2002, Steps toward an integrated view of nitrogen metabolism, *J. Exp. Bot.* **53**:959–570.

Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W., 2003, Observing and interpreting correlations in metabolomic networks, *Bioinformatics* **19**:1019–1026.

Sweetlove, L.J., Last, R.L., and Fernie, A.R., 2003, Predictive metabolic engineering: A goal for systems biology, *Plant Physiol.* **132**:420–425.

Syrkin Wurtele, E., Li, J., Diao, L., Zhang, H., Foster, C.M., Fatland, B., Dickerson, J., Brown, A., Cox, Z., Cook, D., Lee, E-K. and Hofmann, H., 2003, MetNet: software to build and model the biogenetic lattice of *Arabidopsis*, *Comp. Funct. Genom.* **4**:239–245.

Taylor, C.F., Paton, N.W., Garwood, K.L., Kirby, P.D., Stead, D.A., Yin, Z., Deutsch, E.W., Selway, L., Walker, J., Riba-Garcia, I., Mohammed, S., Deery, M.J., Howard, J.A., Dunkley, T., Aebersold, R., Kell, D.B., Lilley, K.S., Roepstorff, P., Yates, J.R. 3rd, Brass, A., Brown, A.J., Cash, P., Gaskell, S.J., Hubbard, S.J. and Oliver, S.G., 2003, A systematic approach to modeling, capturing, and disseminating proteomics experimental data, *Nat. Biotechnol.* **21**:247–254.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M., 2004, MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes, *Plant J.* **37**:914–939.

Tolstikov, V.V., and Fiehn, O., 2002, Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion mass trap spectrometry, *Anal. Biochem.* **301**:298–307.

Tolstikov, V.V., Lommen, A., Nakanishi, K., Tanaka, N., and Fiehn, O., 2003, Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics, *Anal. Chem.* **75**:6737–6740.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., and Fernie, A.R., 2003, Parallel analysis of transcript and metabolic profiles: a new approach in systems biology, *EMBO Rep.* **4**:989–992.

Urbanczyk-Wochniak, E. and Fernie, A.R., 2005, Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (*Solanum lycopersicum*) plants, *J. Exp. Bot.* **56**:309–321.

von Roepenack-Lahaye, E., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., Schmidt, J., Scheel, D., and Clemens, S., 2004, Profiling of arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry, *Plant Physiol.* **134**:548–559.

Wagner, C., Sefkow, M., and Kopka, J., 2003, Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles, *Phytochemistry* **62**:887–900.

Weckwerth, W., 2003, Metabolomics in systems biology, *Annu Rev. Plant Biol.* **54**:669–689.

Weckwerth, W., Loureiro, M.E., Wenzel, K., and Fiehn, O., 2004, Differential metabolic networks unravel the effects of silent plant phenotypes, *Proc. Natl. Acad. Sci. USA* **18**:7809–7814.

Wolfender, J.L., Ndjoko, K., and Hostettmann, K., 2003, Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites, *J. Chromatogr. A* **1000**:437–455.

Chapter 6

LIPIDOMICS: ESI-MS/MS-BASED PROFILING TO DETERMINE THE FUNCTION OF GENES INVOLVED IN METABOLISM OF COMPLEX LIPIDS

Ruth Welti[1], Mary R. Roth[1], Youping Deng[1,3], Jyoti Shah[1], and Xuemin Wang[2,4]

[1]Division of Biology, Ackert Hall, Kansas State University, Manhattan, KS 66506; [2]Department of Biochemistry, Willard Hall, Kansas State University, Manhattan, KS 66506; [3]Current address: Department of Biological Sciences, Johnson Science Tower 1009, University of Southern Mississippi, Hattiesburg, MS 39406-0001; and [4]Current addresses: Department of Biology, R223 Research Building, 1 University Boulevard, University of Missouri at St. Louis, St. Louis, MO 63121 and Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132

Abstract: Electrospray ionization tandem mass spectrometry in the precursor and neutral loss scanning modes is utilized to obtain profiles of the complex, polar lipids of plants. This method is rapid, accurate, and sensitive. The technique is being used to determine the metabolic functions of known genes, to implicate new metabolic pathways, and to help identify mutant genes from their functions.

Key Words: lipidomics; lipid profiling; electrospray ionization; mass spectrometry; phospholipids; galactolipids.

1 INTRODUCTION

Metabolomics may be viewed as a comprehensive strategy to study the function and levels of metabolites in relation to the function of genes and their proteins. In this context, lipidomics can be considered the branch of metabolomics in which non-water-soluble metabolites are studied.

The aims of our group's lipidomic strategies are to determine the role of gene products involved in lipid metabolism and to determine the importance of specific genes and specific lipid compositional changes in plant responses to stress and hormones. To determine the role. of gene products, mutants

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 87–92.

that lack the function of genes encoding lipid metabolic enzymes and putative lipid metabolic enzymes are being examined. Comparison of the metabolic responses of these mutants with the responses of wild-type plants allows identification of gene function, including identification of *in vivo* substrates and products of the gene products. Such comparisons can be made at the levels of specific tissues, cell types, or subcellular fractions. To understand the role of particular lipid compositional changes that may be brought about by particular genes, the lipid composition of mutant plants can be correlated with physiological responses to stress or hormones.

2 LIPID PROFILING METHODOLOGY

Traditional, chromatographic analysis of polar, complex lipids is a time-consuming process that involves separation of the lipids into classes, derivatization, and analysis of the fatty acyl chains. In contrast, mass spectrometry (MS)-based lipid analysis is a rapid process that produces a detailed profile of lipid molecular species. We are currently offering mass-spectrometry-based lipid analysis through the Kansas Lipidomics Research Center, which is described on the web at www.k-state.edu/lipid/lipidomics/. The lipid profiling process involves (1) solvent extraction of tissues from wild-type and/or mutant organisms, (2) addition of a mixture of phospho-lipid and/or galactolipid internal standards and appropriate solvents, and (3) analysis by electrospray ionization tandem mass spectrometry (ESI-MS/MS) (Brügger et al., 1997; Welti et al., 2002, 2003). This process produces spectra from which hundreds of lipid molecular species can be identified by head group and mass. The mass can be interpreted as the total number of acyl carbons and total number of acyl double bonds. A lipid profile can be obtained from a small amount of material, such as a few percent of an Arabidopsis leaf.

The lipid profiling methodology utilizes a tandem mass spectrometer ("triple quad" or MS/MS) with a collision cell, where fragmentation occurs, between the two mass spectrometers and a detector after the second mass spectrometer in the ion path. Our sample introduction is electrospray ionization (ESI). No preseparation is used. The sample is introduced by continuous infusion in solvent into the ESI source. Lipid molecular ions are produced from the lipid molecules. Phosphatidylcholine (PC) and phos-phatidylethanolamine (PE) are analyzed as singly charged positive $[M+H]^+$ ions, phosphatidylglycerol (PG), phosphatidylinositol (PI), phosphatidic acid (PA), and phosphatidylserine (PS) are analyzed as singly charged negative $[M–H]^–$ ions, and monogalactosyldiacylglycerol (MGDG) and digalacto-syldiacylglycerol (DGDG) are analyzed as singly charged positive $[M+Na]^+$ ions.

The lipidomic technology utilizes precursor and neutral loss scanning to allow detection of individual lipid molecular species in extracted biological samples that contain complex mixtures of non-water-soluble compounds that produce molecular ions with essentially every unit mass up to over 1,000 amu. Precursor and neutral loss scanning modes allow the user to obtain a separate spectrum for each class of polar lipids, while many other components are being simultaneously infused and ionized. To perform "precursor" scanning, the second mass spectrometer is set to allow only ions with a mass that corresponds to a charged fragment, characteristic of a particular head group class, to move to the detector. When scanning occurs in the first mass spectrometer, the second mass spectrometer effectively acts as a "filter", so that signal is recorded at the detector only when a molecular ion from the first mass spectrometer produces the characteristic head group fragment. Thus, the spectrum (mass of molecular ions scanned by the first mass spectrometer vs signal detected after the second mass spectrometer) that is collected shows only the lipid molecular ions of those species that can produce the head group fragment. Usually this corresponds to the lipids in one head group class. A neutral loss spectrum also depicts the lipids in a single class; neutral loss scanning is performed when the charge does not localize to the lipid head group fragment after fragmentation. This is the case with PE and PS. In these lipids, the fragment ions containing the two acyl species vary in mass as a function of the molecular ion acyl composition, but the *difference* in mass between the molecular ion and the charged diacyl-containing fragment is constant (corresponding to the mass of the neutral head group fragment). Thus, when the second mass spectrometer scans in synchrony with the first mass spectrometer with an offset that corresponds to the mass of the neutral head group fragment, signal at the detector is again observed only when the first mass spectrometer is at the mass of a molecular ion that generates the characteristic neutral loss of the head group.

In a lipid profiling experiment, a series of precursor and neutral loss scans are executed sequentially (Table 6-1). The signal for each molecular species is corrected for isotopic overlap of the lipid species with other species and then compared in magnitude with the signals of the internal standards (Welti et al., 2002). Currently, this methodology allows routine analysis of 144 polar plant lipid molecular species in eight head group classes. As mentioned, these species are identified in terms of total carbon number and total double bonds. More detailed information about the acyl species can be determined separately *via* product ion analysis of the molecular ions in the negative mode (Welti et al., 2002).

Table 6-1. Precursor and neutral loss scans for analysis of lipid species from plants

Time (min)	Scan mode	Fragment detected	Classes analyzed
3[a]	+	Precursor of 184^+	LysoPC/PC
2	+	Neutral loss of 141	LysoPE/PE
4	-	Precursor of 153^-	LysoPG/PG/PA
2	-	Precursor of 241^-	PI
3	-	Neutral loss of 87	PS
5	+	Precursor of 243^+	MGDG
5	+	Precursor of 243^+	DGDG

The first five scans are performed on an aliquot of extract dissolved in chloroform/ methanol/water (300:665:35) containing 10.5 mM ammonium acetate, while the MGDG and DGDG scans are performed on a second aliquot of extract that is dissolved in chloroform/methanol/water (300:665:35) containing 1.75 mM sodium acetate. Spectra are acquired sequentially by scanning in the listed modes for the indicated time period.

3 USES OF LIPID PROFILING TECHNOLOGY

Lipid profiling technology has been utilized to determine the metabolic functions of genes involved in lipid metabolism (Welti et al., 2002; Nandi et al., 2003, 2004; Abbadi et al., 2004; Li et al., 2004), to examine lipid changes during developmental processes (Fauconnier et al., 2003), to implicate new metabolic pathways (Welti et al., 2002), and to help identify mutant genes from their functions (Nandi et al., 2003). One example of determination of the metabolic function of a gene was for phospholipase Dα1, one of the 12 Arabidopsis gene products that encode phospholipase Ds. Lipid profiles of rosettes sampled before and after freezing stress from wild-type Arabidopsis were compared with similarly treated samples from Arabidopsis deficient in phospholipase Dα1. This comparison showed that phospholipase Dα1 accounts for about half of the PA formed upon freezing. These lipid profiles also showed that phospholipase Dα1 acts on PC, rather than PE or PG.

Lipid profiles of wild-type Arabidopsis during freezing suggested the existence of a previously undescribed pathway leading from MGDG to PA. During freezing, a molecular species of PA that is not detectable before freezing is formed. This species, 34:6 PA, is likely to be derived from MGDG, the only diacyl lipid class that contains large amounts of 34:6 diacylglycerol. The metabolic steps and the gene products involved in this pathway await elucidation.

An example of how lipid profiling expedited the identification and cloning of genes affecting biochemical processes *via* a candidate gene approach is provided by *SFD1*. Lipid profiling of the *ssi2 sfd1* mutant plants suggested

that the *SFD1* was involved in plastid lipid biosynthesis (Nandi et al., 2003). The profile suggested a lipid composition similar to that described by Miquel et al. (1998) for a mutant involved in glycerol phosphate metabolism. A survey of genes near the map location of *SFD1* identified a gene (At2g40690) that putatively encoded a DHAP reductase. Sequencing of *sfd1* mutant alleles confirmed that there were mutations in At2g40690. Finally, genetic complementation and studies of the SFD1 gene expressed in *Escherichia coli* confirmed the identity of *SFD1*.

4 LONG-TERM GOALS

The long-term goals of our group are to determine the roles of genes and enzymes that are involved and potentially involved in lipid metabolism in generating lipid compositional changes during plant stress. In addition, through the Kansas Lipidomics Research Center, we plan to continue to provide high-throughput, sensitive, and accurate lipid profiling and analysis. Finally, we will continue to develop mass-spectrometry-based lipid profiling strategies.

ACKNOWLEDGMENTS

Grant support from National Science Foundation (MCB 0110979 and MCB 0416839) and support of the Kansas Lipidomics Research Center Analytical Laboratory from National Science Foundation's EPSCoR program, under grant EPS-0236913 with matching support from the State of Kansas through Kansas Technology Enterprise Corporation and Kansas State University are gratefully acknowledged, as is Core Facility Support from Kansas Biomedical Infrastructure Network, under grant P20 RR016475.

REFERENCES

Abbadi, A., Domergue, F., Bauer, J., Napier, J.A., Welti, R., Zahringer, U., Cirpus, P., and Heinz, E., 2004, Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation, *Plant Cell* **16**:2734–2748.

Brügger, B., Erben, G., Sandhoff, R., Wieland, F.T., and Lehmann, W.D., 1997, Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry, *Proc. Natl. Acad. Sci. USA* **94**:2339–2344.

Fauconnier, M.-L., Welti, R., Blée, E., and Marlier, M., 2003, Lipid and oxylipin profiles during aging and sprout development in potato tubers (*Solanum tuberosum* L.), *Biochim. Biophys. Acta* **1633**:118–126.

Li, W., Li, M., Zhang, W., Welti, R., and Wang, X., 2004, The plasma membrane-bound phospholipase Dδ enhances freezing tolerance in Arabidopsis, *Nature Biotech.* **22**:427–433.

Miquel, M., Cassagne, C., and Browse J., 1998, A new class of Arabidopsis mutants with reduced hexadecatrienoic acid fatty acid levels, *Plant Physiol.* **117**:923–930.

Nandi, A., Krothapalli, K., Buseman, C.M., Li, M., Welti, R., Enyedi, A., and Shah, J., 2003, The *Arabidopsis thaliana sfd* mutants affect plastidic lipid composition and suppress dwarfing, cell death and the enhanced disease resistance phenotypes resulting from the deficiency of a fatty acid desaturase, *Plant Cell* **15**:2383–2398.

Nandi, A., Welti, R., and Shah, J., 2004, The *Arabidopsis thaliana* dihydroxyacetone phosphate reductase gene supressor of fatty acid desaturase deficiency1 is required for glycerolipid metabolism and for the activation of systemic acquired resistance, *Plant Cell* **16**:465–477.

Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H.-E., Rajashekar, C.B., Williams, T.D., and Wang, X., 2002, Profiling membrane lipids in plant stress responses: role of phospholipase Dα in freezing-induced lipid changes in Arabidopsis, *J. Biol. Chem.* **277**:31994–32002.

Welti, R., Wang, X., and Williams, T.D., 2003, Electrospray ionization tandem mass spectrometry scan modes for plant chloroplast lipids, *Anal. Biochem*. **314**:149–152.

Chapter 7

TIME-SERIES INTEGRATED METABOLOMIC AND TRANSCRIPTIONAL PROFILING ANALYSES

Short-Term Response of Arabidopsis thaliana Primary Metabolism to Elevated CO_2 – Case Study

H. Kanani[1], B. Dutta[1], J. Quackenbush[2,3,+], and M.I. Klapa[1,4,*]
*[1]Department of Chemical Engineering, University of Maryland, College Park, MD 20742, USA; [2]The Institute for Genomic Research (TIGR), Rockville, MD 20850, USA; [3]Department of Biochemistry, The George Washington University, Washington DC 20052, USA; [4]Institute of Chemical Engineering and High Temperature Chemical Processes, Foundation of Research and Technology-Hellas (FORTH/ICE-HT), Patra, Greece [+]Current Address: Dana Farber Cancer Research Institute and Harvard School of Public Health, Boston, MA 02115, USA.*To whom correspondence should be addressed (mklapa@eng.umd.edu)*

1 INTRODUCTION

In the conventional analysis of a biological system, its response to a particular perturbation was usually monitored from the change in macroscopic physiological properties and, at the molecular level, from the expression of few genes and/or the concentration of few proteins or metabolites. Relying on a small number of markers was imposed primarily by limitations in the available measurement techniques. Therefore, the conventional analysis had to heavily count on an initial hypothesis based on which the macro- and microscopic markers had to be selected. Any conclusions or models derived from such analysis depended on how sensitive sensors of the examined process the selected measurements were. Moreover, any simultaneously occurring phenomena that could not be observed from the selected markers, risked being missed.

Advances in the computational and robotic techniques, along with better understanding of biological processes, allowed for the development of the high-throughput (OMICS) techniques. These techniques enabled researchers to obtain detailed and comprehensive information about the state of a biological system at the molecular level. In contrast to the conventional

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 93–110.

analysis, high-throughput relies less on an initial hypothesis. Moreover, different phenomena can be observed simultaneously and thereby connected in the context of the system's physiology. Hence, high-throughput techniques can significantly upgrade the information, which is obtained about a biological system.

Transcriptional profiling using cDNA microarrays (Schena et al., 1995) or Genechip® (Pease et al., 1994) has been the most widely used high-throughput analysis in the post-genomic era. However, it is becoming increasingly clear that comprehensive analysis of a complex biological system requires the quantitative integration of all cellular fingerprints: genome sequence, maps of gene and protein expression, metabolic output, and *in vivo* enzymatic activity (Ideker et al., 2001). For a systematically perturbed cellular system, such integration could provide insight about the function of unknown genes, the metabolic regulation and even the reconstruction of the gene regulation network (Klapa and Quackenbush, 2003). It is, therefore, very important to carefully design experiments that can provide comparable gene expression and proteomic/metabolomic measurements that can lead to useful results (Ideker et al., 2001).

Before, however, such integrated analysis can be carried out, the challenges of quantitative high-throughput analysis at each individual level of cellular function need to be resolved. These range from limitations in the available experimental protocols to lack of data normalization and analysis techniques for upgrading the information content of the acquired measurements. These challenges at each of the genomic and metabolic levels and in their integration, along with suggestions for their resolution are discussed in the next sections. All issues are presented in the context of the time-series integrated metabolomic and transcriptomic analysis of the short-term response of the *Arabidopsis thaliana* primary metabolism to elevated levels of CO_2 in its growth environment. The need for integrated molecular analyses becomes apparent through the discussion of the obtained results.

2 EXPERIMENTAL DESIGN AND DATA NORMALIZATION

2.1 Selection of the system/perturbations

Always, the selection of the system depends on the specific aims of the study. In the case of integrated genomic and metabolic studies, which are still at their infancy, we believe that appropriate model systems and easily observable physiological changes should be, respectively, used and targeted, as a means of justifying this new systems biology approach and validating algorithms developed for the combination of the data (Klapa and Quackenbush,

2003). Moreover, the experiments should be designed in such a way that the observed changes in the molecular profiles could be attributed only to the applied perturbations (Klapa and Quackenbush, 2003).

In this context, the selection of *A. thaliana* liquid culture as the system, and elevated CO_2 levels in its growth environment as the perturbation under investigation in the case study, was based on the following reasons:

- Plants are complex eukaryotic organisms; besides then any functional insight that might be gained through the study, it is anticipated that analysis of multitissue organisms will also contribute to the development of systems biology principles that will have broader applicability than studies in yeast or prokaryotic systems.
- *A. thaliana* is considered the model system of plant physiology, because of its short growth cycle and a small genome of five chromosomes, which has been fully sequenced (*Arabidopsis* Genome Initiative, 2000).
- Liquid compared to soil cultures provide a controllable growth environment, ensuring that all plants in all experiments receive the same nutrients.
- In liquid cultures, any signaling molecule/growth hormone added to the media is uniformly distributed.
- Each liquid culture comprises of 50–80 plants. Therefore, a large number of replicates are harvested at the same time, increasing, thereby, the confidence in the statistical significance of the acquired measurements. In the case of time-series analysis, this might help in partially overcoming the trade-off between number of replicates and number of timepoints (see further discussion in following section).
- CO_2 is the main carbon source for plants; thereby, any change in its concentration in the plants' growth environment is expected to have direct implications in the activity of its central carbon metabolism and amino acid biosynthesis. The latter are considered a good model framework for integrated metabolic and genomic analyses, because:
    - There exists extensive information about their function both at the metabolic and genomic levels.
    - The majority of the involved metabolic pathways have been well characterized in plants.
    - The regulation of these pathways has been extensively investigated, at least in prokaryotic systems.
    - At the genomic level, the majority of the genes related to these pathways have already been identified and annotated.

2.2 Metabolomic analysis

The metabolomic profile of a biological system – referring to the concentration profile of all its free small metabolite pools (Roessner et al., 2000; Fiehn et al., 2000) – provides a phenotypic equivalent of the high-throughput transcriptional and proteomic profiles (Fiehn et al., 2000). To date, metabolomic profiling in plants has been mainly used to differentiate between metabolic states (Kanani, 2004; Cook et al., 2004; Broeckling et al., 2005; Noguchi et al., 2003; Sakai et al., 2004; Hirai et al., 2004) and/or identify an environmental or genetic phenotype (Fiehn et al., 2000; Roessner et al., 2001; Taylor et al., 2002; Weckwerth et al., 2004). Identification and quantification of small metabolites is possible by Gas (or Liquid) Chromatography - Mass Spectrometry (GC/LC-MS) or Nuclear Magnetic Resonance (NMR) Spectroscopy. Among these, GC-MS has been the technique of choice for most quantitative high-throughput metabolomic profiling analyses, of polar metabolites, in particular (Roessner et al., 2000; Kanani, 2004; Cook et al., 2004; Broeckling et al., 2005; Noguchi et al., 2003; Sakai et al., 2004; Roessner et al., 2001; Taylor et al., 2002; Weckwerth et al., 2004). More detailed comparison of the 3 techniques and their specific advantages/disadvantages in the context of metabolomic analysis can be found in Fiehn (2001) and Kopka (2004). Taking into consideration this information, GC-MS was selected for the acquisition of the metabolomic profiles in the case study. The main objective of the latter was the analysis of the primary metabolism of the plant liquid cultures, which comprises mainly polar metabolites.

A typical metabolomic analysis using GC-MS consists of the following four distinct stages:

1. *Metabolite extraction*: In the case study, part of the homogenized ground sample was used in the metabolomic analysis, while the rest in the trascriptomic analysis. Polar metabolites are obtained from the homogenized ground plant sample through methanol-water extraction (Roessner et al., 2000; Kanani, 2004), while nonpolar through chloroform extraction (Fiehn et al., 2000; Weckwerth et al., 2004). If used in combination, the entire metabolome could be obtained.
2. *Metabolite derivatization*: Derivatization is imperative for the conversion of the small metabolites to volatile, nonpolar, and stable derivatives through their reaction with a particular derivatization agent. The most commonly used derivatization method in metabolomics analysis involves the original metabolites' conversion into their trimethylsilyl (TMS) and Methoxime (MEOX) derivative(s) (Roessner et al., 2000). To ensure accuracy of the metabolomic analysis, the derivatization time should be optimized (see Kanani and Klapa, 2007, for the optimization strategy).

3. *GC-(electron ionization (EI)) MS analysis*: Gas chromatography enables the separation of the metabolites, while their identification and quantification is based on the acquired mass spectra. In the case study, GC-(ion trap) MS (GCQ, Thermo-Finnigan) was used because of the significant reported advantages of ion-trap MS (Klapa et al., 2003). Details concerning the actual GC-MS operating conditions are provided in Kanani (2004).
4. *Metabolite identification and quantification*: First, the acquired mass spectra undergo peak de-convolution (Stein, 1999). The identification of (known or putative) metabolite peaks is based on the retention time and mass spectra standards available in public (http://www.mpimpgolm.mpg.de/mms-library/details-e.html), commercial (Ausloos et al., 1999) or prepared in each laboratory libraries (Kanani, 2004; Kanani and Klapa 2007). Each metabolite is quantified from the peak area of its marker ion/s (Roessner et al., 2000; Kanani and Klapa 2007).

Because of the derivatization step, in metabolomic analysis using GC-MS the actually measured metabolomic profile is the derivative profile. In this case, metabolomic analysis is based on the assumption that the concentration of each metabolite in the original sample is in one-to-one directly proportional relationship with the peak area of its marker ion (or the sum of the peak areas of its marker ion(s)). Biases, however, introduced at each of the four steps of the GC-MS data acquisition process might affect this proportionality, hindering the comparison between data from different experiments/batches. In this case, appropriate normalization is required before any data analysis is attempted. The potential biases in GC-MS metabolomic analysis can be divided into three categories, for each of which a specific normalization strategy is suggested:

Errors that affect all metabolites equally: These biases, e.g., unequal division of a sample into replicates, injection errors, different split ratios, are expected to change the proportionality ratio between a metabolite's original concentration and the peak area of its marker ion to the same fold-extent for all metabolites. Therefore, barring any other type of biases, the relative composition of the measured derivative metabolomic profile should be the same as that of the original sample. To account for this bias and render the results from different experiments/batches comparable ***Internal Standard Normalization*** is required. The selected internal standard should not be produced – at least not to the extent that it distors the acquired data – by the biological system (ribitol or isotopes of known metabolites have been the most commonly used, Roessner et al., 2000; Fiehn et al., 2000). It is added just before the initiation of the four-step process described above. Each metabolite is then quantitatively characterized by the ratio of the peak area of its marker ion(s) to the peak area of the marker ion(s) of the internal standard. Detailed explanation of internal standard normalization is provided

in Kanani (2004). The peak area ratio thus obtained is referred to as "relative peak area" of the metabolite.

Errors that affect specific metabolites: These biases are expected to change the proportionality ratio between a metabolite's original concentration and the peak area of its marker ion to a different fold-extent for the various metabolites in the sample. They concern the derivatization process and time, e.g., incomplete derivatization of a metabolite, formation of multiple derivatives or changes in GC-MS conditions that lead to variations in a metabolite's fragmentation pattern (Kanani and Klapa, 2007). The extent of this type of bias introduced in a particular metabolite's measurement depends on the molecular structure and/or concentration of the metabolite. These errors should be identified in the measured profile and properly accounted for, because if not, they could change the relative composition of the measured derivative metabolomic profile with respect to that of the original sample. In this case, changes in the profile that are due only on chemical and/or setup reasons could be attributed biological significance leading to erroneous conclusions (Kanani and Klapa, 2007). Kanani and Klapa (2007) propose a normalization strategy for this type of biases.

Process/*s*etup or Biological Outliers: To potentially enable identification of these outliers through clustering analysis (Kanani, 2004), at least three biological (if allowed from the experimental setup/resources) and experimental (i.e., parts of the same sample or different injections of the same sample) replicates should be acquired. The identified outliers should be removed from the rest of the analysis as not representing the true metabolic state of the plant sample to avoid the distortion of the attained results/conclusions.

These three data normalization steps are necessary in any metabolomic analysis using GC-MS. In addition, in the case when different experiments/perturbations of the same biological system/setup are conducted on different days, the potential change in the initial/control conditions (e.g., ambient air composition, different batch of seeds, and/or media) between the various experiments should also be taken into consideration for the experiments to be comparable. In the case of a time-series analysis at which time zero represents the initiation of each perturbation, the change in control conditions between the experiments is represented by the difference in their metabolomic profiles at time zero. To account for this variation and to scale the metabolomic data around the value of 1 ($\log_2[1] = 0$), the metabolomic profiles of all timepoints of an experiment could be normalized with respect to the metabolomic profile of its time zero. Then, any identified difference between the metabolomic profiles of the experiments is due only to the applied perturbation(s). Similar normalization strategy has also been used in snapshot analysis involving comparison between different genotypes, grown on different days (Roessner et al., 2001). The metabolomic profiles as

obtained after this four-step data normalization and validation procedure could now be used in further data analysis.

2.3 Transcriptional profiling analysis

High-throughput transcriptional profiling analysis using DNA microarrays is based on the principle of the highly specific affinity between complementary DNA/RNA strands. There exist two widely used microarray platoforms: the Affymetrix GeneChip® is prepared using photolithographic technology (Pease et al., 1994), while the spotted array using robotic printing technology (Schena et al., 1995). Selection of either platform has been based mainly on the available resources. To-date, spotted array has been the technique of choice of most academic laboratories.

A typical transcriptomic analysis using spotted microarray consists of three distinct stages:

Slide printing: In the described case study, full-genome *A. thaliana* spotted arrays printed at The Institute for Genomics Research (TIGR) were used. Each of these arrays comprised 27,648 spots, each corresponding to a particular gene or EST. The array design and fabrication protocol can be obtained from Hegde (2000).

Total RNA and mRNA extraction: Total RNA can be extracted from the ground plant samples using trizol (*http://atarrays.tigr.org/arabprotocols.html*). mRNA is then obtained from the total RNA *via* the amplification of RNAs with a polyA tail through reverse transcription (*http://atarrays.tigr.org/ arabprotocols. html*).

Hybridization and scanning: mRNA from two different samples (query and reference) is reverse transcribed to produce cDNA, which is attached to two different Cy3 and Cy5 dyes by biochemical reactions. The samples are then hybridized on a microarray slide followed by scanning (http://atarrays. tigr.org/PDF/Probehyb.pdf). The relative intensities of the two different dyes provide a measure of the relative amount of mRNA present in the query vs the reference sample.

During a typical transcriptomic analysis, errors in the acquired DNA microarray data could be mainly originated from:

- Unequal quantities of starting mRNA in the query and reference samples.
- Difference in the labeling efficiencies of the Cy3 and Cy5 dyes.
- Difference in the sensitivity of the scanner for the two dyes.
- Variation of intensity between the spots due to variation between the slide printing pins.

To eliminate these errors, various normalization methods have been proposed, among which total intensity normalization (Quackenbush, 2002), standard deviation regularization (Yang et al., 2002a), lowess (Yang et al., 2002a; Yang et al., 2002b) and flip-dye analysis (Quackenbush, 2002) were

used in this sequence in the case study. Previous studies in the group of Dr. Quackenbush have validated this normalization sequence as generally valid for the normalization of spotted arrays. After this normalization, outlier analysis and, in the case of time-series analyses, time zero normalization should be carried out in a similar manner and for the same reasons as described in metabolomic analysis.

2.4 Time-series vs snapshot analysis

The response of a complex biological system, such as plants, to perturbations in its environment is inherently a temporal process. When cells are exposed to a new condition (treatment or stress), they respond to the situation by modifying their gene expression and/or altering protein activity. Usually, the cascade of molecular events is initiated through the activation of transcription factor(s), which, in turn, induce(s) the expression of genes encoding proteins necessary to respond to the new conditions. The proteins are the catalysts and regulators of (almost) all cellular functions, including signaling and metabolic reaction networks. In the latter, metabolic regulatory mechanisms might alter the enzymatic activity leading to a certain redistribution of metabolic fluxes. If a snapshot of the new conditions is compared with the old one, a set of variables (genes/proteins/metabolites) that undergo change in expression can be found. However, in order to determine the "trend" of change, it is preferable to obtain expression data over a certain period of time. In this way, knowledge not only about the difference between two states, but also about the pathways involved in the physiological change, could be obtained (Joseph, 2004). One additional argument in favor of time-series analysis is that, in the case of integrated genomic and metabolic analyses, it is the metabolic flux redistribution that is directly integratable with the gene expression data (Klapa and Quackenbush, 2003), because it characterizes the change in the degree of enzymatic engagement in a conversion process (Klapa and Quackenbush, 2003). However, metabolic flux analysis requires extensive knowledge of the biochemical reaction network and is to date limited to steady-state conditions (Klapa and Quackenbush, 2003). In this case, the time-series metabolomic profiles, which inherently contain information about the flux distribution, could provide the metabolic fingerprint of a biological system in high-throughput integrated analyses when steady- or pseudo-steady state conditions are a risky assumption to make.

To design a time-series integrated metabolomic and transcriptomic analysis, the following two issues should be addressed:

Number of timepoints Vs number of bioreplicates: For a given experimental setup and set of resources, the experimental design should optimize between the number of timepoints at which the biological system

should be sampled and the number of biological replicates at each timepoint. Both are desired to increase the information content and statistical significance of the analysis.

Selection of timepoints/duration of sampling time: If there is large time difference between samples, intermediate key events of shorter timescale might be missed. However, for a given number of sampling times due to the available resources as described above, decreasing the time difference between consecutive samples will result in shorter duration of the experiment. This might lead to missing important physiological events that are activated at a later stage after the initiation of the perturbation. Thus, there exists a trade-off between the frequency of sampling and the duration of the experiment. Moreover, in the case of integrated analyses, the selection of sampling times should take into consideration the difference in the timescale of response between the transcriptional and the metabolic processes. In this context, adapting a particular time-course experiment from one organism to another might prove cumbersome, since the rates at which similar biological processes take place differ across organisms and environmental and/or genetic conditions (Joseph, 2004; Spellman et al., 1998).

2.5 Case study

Taking into consideration all the issues regarding the design of a time-series integrated high-throughput metabolomic and transcriptomic study as described above, the short-term effect of elevated CO_2 on the physiology of *A. thaliana* was studied based on the following experiment:

Two sets of 19 and 20, respectively, *A. thaliana* (Columbia strain) liquid shake cultures were grown under constant light (80–100 μmole/cm^2 s), relative humidity (60%) and temperature (23°C) at ambient air conditions for 12 days; of note, each set was grown on different days (see Kanani (2004) for detailed description of the experimental setup). At the beginning of the 13^{th} day, 3 and 4 cultures from each set, respectively, were harvested to account for the reference physiological state (time 0 h). Immediately afterwards, each set of the 16 remaining plants was fed continuously (i.e., at constant rate) for 23 h with air of ambient composition and 1% CO_2, respectively. Two cultures from each set were harvested at each of the 0.5 h, 1 h, 1.5 h, 3 h, 6 h, 12 h, and 23 h after initiation of the perturbation. The harvested samples were frozen using liquid nitrogen and stored at –80°C for further analysis. Each frozen plant was subsequently ground, homogenized and separated into two sections used in each of the analyses. The data normalization strategy followed for each of the obtained metabolomic and transcriptomic data sets are shown in parallel in Figure 7-1.

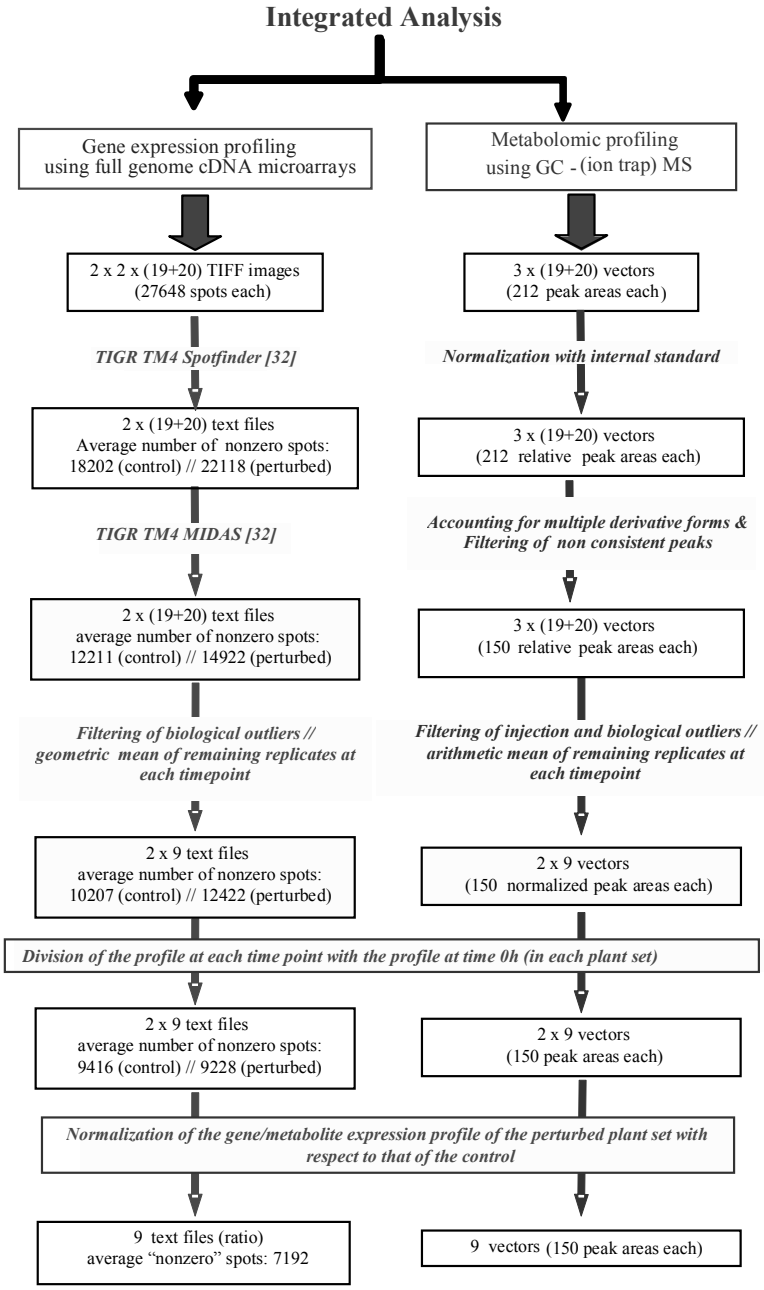


Figure 7-1. The steps of the data normalization of the transcriptional and metabolomic profiles; the specific numbers of files and vector sizes refer to the described case study.

3 MULTIVARIATE DATA ANALYSIS

3.1 Identification of different physiological states

Clustering techniques, like Principal Component Analysis (PCA) and Hierarchical Clustering (HCL) have been used in metabolomic and genomic analysis to identify different physiological states, representing genetic mutation(s), environmental perturbation(s), or external treatment(s). Typically, such analysis is carried out for the comparison between snapshots (Fiehn et al., 2000; Hirai et al., 2004; Roessner et al., 2001; Taylor et al., 2002), where presence of clusters represents different states. Figure 7-2 shows the results of applying PCA analysis (Saeed et al., 2003) in the time-series metabolomic and transcriptional profiling data of the case study. They indicated significant difference between the control (ambient CO_2 concentration) and the perturbed (1% CO_2) plant sets at both the metabolic (Kanani, 2004) and transcriptional (Dutta, 2004) levels even for short CO_2 treatment. This difference validated the choice of the system and perturbation during the experimental design. However, the clustering pattern, reflecting the shape of the majority of the metabolites' and genes' expression profiles over time, indicated difference in the plant's response between the shorter vs the longer CO_2 treatment. The large fluctuations observed over the first 2 hours of treatment in the metabolomic and transcriptional profile individually, even in the control plant sets, have to be attributed to factors other then the applied perturbation. Taking into consideration (a) the high sensitivity of the plants to slight changes in their environment, and (b) the experimental setup that was causing changes in the ambient pressure of the plants during each harvesting, the initial timepoints (0.5h–2 h) were removed from further data analysis as not directly reflecting the physiological consequences of the applied treatment. Moreover, the experimental setup was accordingly modified to avoid this problem in following experiments (Dutta et al., 2006). This is an example of how the results of one experiment could contribute in further optimising the design of future studies.

3.2 Identification of significant genes and metabolites

One of the main objectives of high-throughput analysis is the identification of these biological variables that characterize the difference between physiological states. In most of the reported snapshot "-omic" studies (Fiehn et al., 2000; Cook et al., 2004; Roessner et al., 2001) this has been achieved through t-test or fold change (FD) analysis. These methods, however, do not include any distinct threshold characterizing significance.

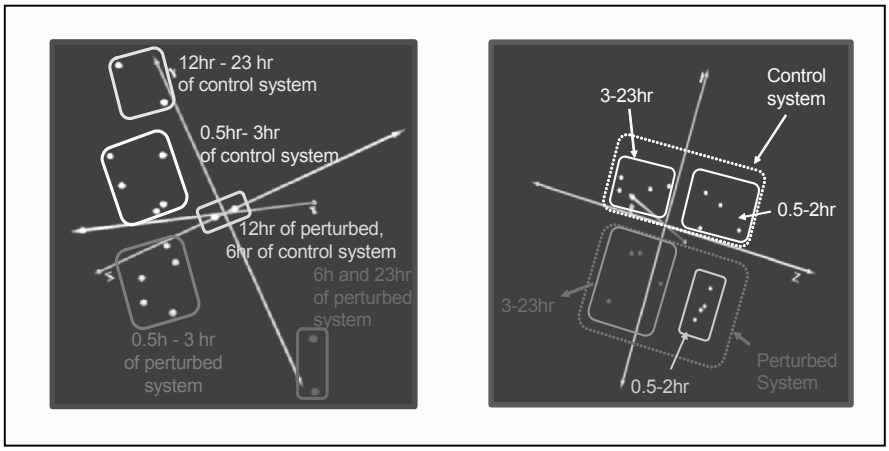


Figure 7-2. Identifying the short-term effect of elevated CO_2 on the transcriptional and metabolic response of the A. thaliana liquid cultures (case study) using PCA analysis on the (A) Transcriptional (Euclidean distance) and (B) Metabolomic (Pearson correlation) profiles. Only the first 3 principal components are shown, encountering, respectively, for 75% and 65% of the variation in the data. (Kanani, 2004; Dutta, 2004).

To overcome this problem in DNA microarray analysis, a methodology called Significance Analysis of Microarrays (SAM) was recently developed and used for the comparison of different experimental conditions (Hirai et al., 2004; Tusher et al., 2001; Xiang et al., 2004). SAM was further modified to allow one-to-one pairing between corresponding samples in each compared class and this method is known as two-class paired-SAM to be differentiated from the original unpaired. In both paired and unpaired SAM analyses, the probability of genes being falsely assigned significance (referred to as False Detection Rate (FDR)) is calculated at each significance level (referred to as delta value). Thus, when comparing multiple data sets, instead of using a fixed limit of significance as in traditional analysis (p-value in t-test, FC value in fold change), SAM allows comparison between the same FDR values, which is determined based on the overall variation in each data set.

In the CO_2 case study, two-class paired (between the corresponding timepoints) SAM analysis was used on both the transcriptional and metabolomic profiles to identify genes and metabolites, respectively, that exhibited significant change in their expression between the control and the perturbed plant sets. The analysis identified ~900 significant out of ~9000 genes (Dutta, 2004) and ~45 significant out of 150 metabolites (Kanani, 2004) (see Figure 7-1).

Taking into consideration that the identification of the significant biological variables in time-series analysis is based on multiple snapshots of the perturbed vs the control state compared to the single "picture" of the system in snapshot analysis, the former could be considered more "robust" than the latter. However, the available and employed statistical tests (SAM, t-test, FC) compare the profiles at the various timepoints as simply different snapshots of the physiological state, without taking into consideration their time history, in terms of their specific order and the time difference between them. Hence, there exists currently a need for extension of these analytical techniques into time-series analysis. Until, however, such method(s) become(s) available, the experimental design should opt for equal time intervals between consecutive samplings.

3.3 Data interpretation in biological context

The identified significant genes and metabolites have to be discussed in the context of the existing knowledge about the physiology and regulation of the particular biological system. This is necessary in order to (a) validate the statistical results and (b) determine new interactions, relations, and phenomena concerning the system and the applied perturbation. In the CO_2 case study, such analysis indicated following physiological changes.

3.3.1 Transcriptional profiling

- Increase in the expression of genes involved in cell division and cell/plant wall synthesis (Dutta, 2004);
- Increase in the expression of multiple genes involved in the Calvin Cycle (carbon fixation) & the chlorophyll production pathway (Dutta, 2004);
- Decrease in the expression of the gene encoding the nitrate reductase enzyme; the latter catalyzes the first step of the nitrogen assimilation pathway (Dutta, 2004).

3.3.2 Metabolomic profiling

- Significant increase in the concentration of metabolic intermediates required for the synthesis of structural carbohydrates (Kanani, 2004);
- Significant decrease in the concentration of the nitrogen storage and transport amino acids (Kanani, 2004);
- Significant decrease in the concentration of metabolites involved and usually characterizing the activity of the photorespiration pathway (Kanani, 2004).

In combination, the obtained transcriptional and metabolic measurements suggested, (1) inhibition of the inorganic nitrogen assimilation; this result agreed with previous studies (Smart et al., 1998; Bloom et al., 2002),

(2) increase in the photosynthesis and the CO_2 fixation rate and (3) increase in the production of structural carbohydrates required for the formation of cell/plant wall to accommodate the increased cell division. Of note, this is the first time that the CO_2 effect on plant physiology has been associated with increase in the concentration of structural vs nonstructural carbohydrates (Idso and Idso, 2001; Paul and Foyer, 2001; Hui et al., 2002). Taking into consideration that most of the to date reported studies referred to long-term CO_2 effect, it is speculated that the first response of young, healthy plants to elevated CO_2 would be to grow instead of storing it in the form of nonstructural carbohydrates.

These results illustrate the advantages of time-series integrated high-throughput transcriptional and metabolomic profiling analyses. It becomes clear that the latter enhance the quality and quantity of the information obtained about a biological system. Moreover, the "concrete" picture that was obtained about the physiology of *A. thaliana* provided strong support to the selected experimental design, data normalization, and analysis strategy as were described throughout this chapter. Specifically:

- The selected duration of the treatment and the integrated high-throughput nature of the analysis allowed for the short-term effect of elevated CO_2 on young, healthy plants, i.e., their growth, to be observed. Such effect of the CO_2 treatment had never been reported in earlier studies, because they referred to longer treatments. In addition, the former studies not being high-throughput had to focus on the measurement of specific class(es) of molecules and/or genes based on particular hypotheses. Therefore, phenomena, e.g., increase in the concentration of structural carbohydrates that might have been taking place at the same time, but had not been associated with the considered hypotheses, could not have been captured by the acquired data.
- If only one of the two analyses had been employed to describe the physiological state of the system, some of the currently derived conclusions would have remained at the level of speculation, if even observed. For example, most metabolites in the Calvin cycle and chlorophyll production are not identified by GC-MS; therefore conclusion (2) above is primarily based on the genomic data, and only by its association to the nitrogen assimilation inhibition, on the metabolomic data. Similarly, the inhibition of the photorespiration pathway, which competes with the Calvin cycle (Coshigano et al., 1998; Siedow and Day, 2001), was observed primarily at the metabolomic level; the competition between the two pathways originates in the enzyme kinetics level, due to their common reaction sites in RubisCO.

4 SUMMARY AND FUTURE CHALLENGES

Time-series integrated metabolomic and transcriptional profiling analyses can provide a comprehensive picture of the examined biological system. Therefore, they should be preferred over the studies that are based on only one level of cellular function. However, there are still several challenges associated with this type of analyses in general and in the study of plants in particular that need to be addressed.

Systematic perturbations/data submission protocols: The detailed analysis of the molecular mechanisms that determine the physiological state of a biological system and the development of theoretical models that describe and predict cellular function must be based on integrated data from a large number of experiments/systematic perturbations. As the microarray community has come to realize, this will require the development of protocols (see MIAME protocol, Brazma et al., 2001) for describing experimental conditions and for submitting (any type of biolgical) data to public databases.

Data normalization: A systematic and widely accepted strategy (Kanani and Klapa, 2007) is required for the normalization of the high-throughput metabolomic data. Software tools similar to those existing for the normalization of transcriptional profiling data (e.g., TIGR MIDAS, Saeed et al., 2003) should be developed for the metabolomic measurements as well as for the parallel processing of multiple data types.

Data analysis: Extension of the currently available for the analysis of high-throughput data sets statistical hypothesis testing methods to account for the time history in time-course measurements is required.

Integrated data visualization and interpretation: There is a clear need for development of integrated data visualization and mining software tools that can be used to infer the relationships that exist between the various data sets in the context of the known and expected cellular physiology.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge the financial support of the US National Science Foundation (Award No: QSB-0331312), the University of Maryland Minta Martin Foundation through a fellowship to Dr. Klapa and the Department of Chemical Engineering through its junior faculty startup fund. Further, we would like to acknowledge contribution of Dr. Tara Vantoai, Ms. Linda Moy, Ms. Lara Linford, and Mr. Jeremy Hasseman in the CO_2 case study experiment.

REFERENCES

Arabidopsis Genome Initiative, 2000, Analysis of the genome sequence of the flowering plant, *Arabidopsis thaliana*, *Nature* **408**:796–815.

Ausloos, P., Clifton, C.L., Lias, S.G., Mikaya, A.I., Stein, S.E., Tchekhovskoi, D.V., Sparkman, O.D., Zaikin, V., and Zhu, D., 1999, The critical evaluation of a comprehensive mass spectral library, *J. Am. Soc. Mass Spectrom.* **10**:287–299.

Bloom, A.J., Smart, D., Nguyen, D., and Searls, P., 2002, Nitrogen assimilation and growth of wheat under elevated carbon dioxide, *Proc. Natl. Acad. Sci. USA* **99**:1730–1735.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M., 2001, Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nat. Genet*. **29**:365–371.

Broeckling, C.D., Huhman, D.V., Farag, M.A., Smith, J.T., May, G.D., Mendes, P., Dixon, R.A., and Sumner, L.W., 2005, Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism, *J. Exp. Bot.* **56**:323–336.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F., 2004, *Proc. Natl. Acad. Sci. USA*. **101**:10583–89.

Coshigano, K., Schultz, C., Melo-Oliveria, R., Lim, J., and Coruzzi, G., 1998, Arabidopsis *gls* mutants and distinct Fd-GOGAT genes: implications for photorespiration and primary nitrogen assimilation, *Plant Cell.* **10**:741–752.

Dutta, B., 2004, Time series transcriptional profiling analysis of the *Arabidopsis thaliana* response using full genome microarray and metabolic information, Master's thesis, University of Maryland, College Park, MD.

Dutta, B., Kanani, H.H., Quackenbush, J., and Klapa, M.I., 2006, Dynamic transcriptomic and metabolomic short-term response to elevated CO_2 stress in *Arabidopsis thaliana*: A plant systems biology case, *submitted*

Fiehn, O., 2001, Integrated studies in plant biology using multi parallel techniques, *Current Opinions in Biotechnology*. **12**:82–86.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L., 2000, Metabolite profiling for plant functional genomics, *Nature Biotech*. **18**:1157–1168.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N., and Quackenbush, J., 2000, A concise guide to cDNA microarray analysis, *Biotechniques*. **29** (3): 548–556.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., and Saito, K., 2004, Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA* **101**:10205–10210.

Hui, D., Sims, D., Johnson, D., Cheng, W., and Luo, Y., 2002, Effects of gradual versus step increase in carbon dioxide on *Plantago* photosynthesis and growth in a microcosom study, *Environmental and Experimental Botany*. **47**:51–66.

Ideker, T. V., Thorsson, J.A., Ranish, R., Christmas, J., Buhler, J.K., Eng, R., Bumgarner, R., D.R., Goodlett, D.R., R., Aebersold, R., and Hood, L., 2001, Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science.* ***2*92**:929–934.

Idso, S., and Idso, K., 2001, Effect of atmospheric CO_2 enrichment on plant constituents related to animal and human health, *Environmental and Experimental Botany.* **45**:179–199.

Joseph, Z.B., 2004, Analyzing time series gene expression data, *Bioinformatics*. **20**:2493–2503.

Kanani, H., 2004, Time series metabolic profiling analysis of the short term *Arabidopsis thaliana* response to elevated CO_2 using gas chromatography mass spectrometry, Master's thesis, University of Maryland, College Park, MD.

Kanani, H., and Klapa, M.I., 2007, Data normalization strategy for quantitative high-throughput metabolomic profiling analysis using Gas Chromatography – Mass Spectrometry, *Metabolic Engineering*. **9**: 39–51.

Klapa, M.I., Aon, J.C., and Stephanopoulos, G., 2003, Ion-trap mass spectrometry used in combination with gas chromatography for high-resolution metabolic flux determination, *Biotechniques.* **34**(4):832-6.

Klapa, M., Quackenbush, J., 2003, The Quest for the Mechanisms of Life, *Biotech. & Bioeng.* **84**:739–742.

Kopka, J., Fernie, A., Weckwerth, W., Gibson, Y., and Stitt, M., 2004, Metabolic Profiling in plant biology: platforms and destinations, *Genome Biology* **5**:109.

Noguchi, Y., Sakai, R., and Kimura, T., 2003, Metabolomics and its potential for assessment of adequacy and safety of amino acid intake, *J. Nutr.* **133**:2097S–2100S.

Paul, M., and Foyer, C., 2001, Sink Regulation of photosynthesis, *J. Exp. Bot.* **52**:1383-1400.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P., 1994, Light-generated oligonucleotide arrays for rapid DNA sequence analysis, *Proc. Nat. Acad. Sci. USA* **91**:5022–5026.

Quackenbush, J., 2002, Microarray data normalization and transformation, *Nat. Genet.* **32**(S2):496–501.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A., 2001, Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems, *Plant Cell* **13**:11–29.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R., and Willmitzer, L., 2000, Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.* **23**:131–142.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J., 2003, TM4: a free, open-source system for microarray data management and analysis, *Biotechniques*. **34**(2):374–378.

Sakai, R., Miura, M., Amao, M., Kodama, R., Toue, S., Noguchi, Y., and Kimura, T., 2004, Potential approaches to the assessment of amino acid adequacy in rats: a progress report, *J. Nutr.* **134**:1651S–1655S.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O., 1995, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**:467–470.

Siedow, J., and Day, D.A., 2001, Respiration and Photorespiration, in: *Biochemistry & Molecular Biology of Plants*, B. Buchanan, W. Gruissem, and R. Jones eds., American Society of Plant Physiologists, Rockville, Maryland.

Smart, D., Ritchie, K., Bloom, A., and Bugbee, B., 1998, Nitrogen balance for wheat canopies (*Triticum aestivum cv. Veery* 10) grown under elevated and ambient CO_2 concentrations, *Plant Cell Environ.* **21**:753–763.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Ansers, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B., 1998, Comprehensive identification of cell-cycle regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of Cell.* **9**:327–397.

Stein, S.E., 1999, An integrated method for spectrum extraction and compound identification from gas chromatography / mass spectrometry data, *J. Am. Soc. Of Mass Spectrometry.* **10**:770–781.

Taylor, J., King, R.D., Altmann, T., and Fiehn, O., 2002, Application of metabolomics to plant genotype discrimination using statistics and machine learning, *Bioinformatics* Suppl. **2**:S241–248.

Tusher, G.V., Tibshirani, R., and Chu, G., 2001, Significance analysis of microarray applied to the ionizing radiation response, *Proc. Natl. Acad. Sci. USA* **98**:5116–5121.

Weckwerth, W., Loureiro, M.E., Wenzel, K., and Fiehn, O., 2004, Differential metabolic networks unravel the effects of silent plant phenotypes, *Proc. Natl. Acad. Sci. USA* **101**:7809–7814.

Weckwerth, W., Wenzel, K., and Fiehn, O., 2004, Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks, *Proteomics* **4**:78–83.

Xiang, W., Windl, O., Wünsch, G., Dugas, W., Kohlmann, A., Dierkes, N., Westner, I.M., and Kretzschmar, H.A., 2004, Identification of Differentially Expressed Genes in Scrapie-Infected Mouse Brains by Using Global Gene Expression Technology, *J. Virol.* **78**(20): 1105160.

Yang, I.V., Chen, E., Hasseman, J.P., Liang, W., Frank, B.C., Wang, S., Sharov, V., Saeed, A.I., White, J., Li, J., Lee, N.H., Yeatman, T.J., Quackenbush, J., 2002a, Within the fold: assessing differential expression measures and reproducibility in microarray assays, *Genome Biology*. **3**:research0062.1– 0062.12.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P., 2002b, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Research.* **30**:e15.

Chapter 8

METABOLOMICS OF CUTICULAR WAXES: A SYSTEM FOR METABOLOMICS ANALYSIS OF A SINGLE TISSUE-TYPE IN A MULTICELLULAR ORGANISM

M. Ann D.N. Perera and Basil J. Nikolau
W.M. Keck Metabolomics Research Laboratory, Iowa State University, Ames, IA 50011, USA

Abstract: Electrospray ionization tandem mass spectrometry in the precursor and neutral loss scanning modes is utilized to obtain profiles of the complex, polar lipids of plants. This method is rapid, accurate, and sensitive. The technique is being used to determine the metabolic functions of known genes, to implicate new metabolic pathways, and to help identify mutant genes from their functions.

Key Words: lipidomics; lipid profiling; electrospray ionization; mass spectrometry; phospholipids; galactolipids.

1 INTRODUCTION

The emerging field of metabolomics seeks to globally identify the low molecular weight (<1,000 Da) biochemical constituents of biological materials (Hall et al., 2002; Bino et al., 2004). These molecules are primarily either metabolites of intermediary metabolism or end products of metabolism. These molecules therefore represent the final level at which most genes express their functionality (Fiehn et al., 2000; Fiehn, 2002; Weckwerth and Fiehn, 2002). Hence, one of the many potential utilities of metabolomics data is in the field of functional genomics, which seeks to identify the biochemical and physiological function of all genes in a genome (Weckwerth and Fiehn, 2002; Bino et al., 2004). The application of metabolomics data to functional genomics faces a number of inherent limitations. One of these is the fact that many individual metabolites are common to different metabolic processes.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 111–123.

Thus, unlike proteomics and transcriptomics, where there is a one-to-one correspondence between individual molecules (proteins and mRNAs) and individual genes, no such correlation exists for individual metabolite molecules (Oliver et al., 2002). Another limitation of metabolomics, at least as currently practiced for eukaryotic multicellular organisms, is the lack of discrimination of metabolites from different cellular and subcellular compartments. Namely, because metabolomic analyses are usually conducted on metabolite extracts made by the homogenization of a number of different tissue types and subcellular compartments, data concerning the spatial arrangement of metabolites is lost. This is of particular significance in the case of metabolites that are common to different metabolic processes that occur in distinct cellular and subcellular compartments (e.g., acetyl-CoA; (Ke et al., 2000; Fatland et al., 2005)).

Cuticular waxes are constituents of the cuticle that coat all aerial organs of terrestrial plants (Martin and Juniper, 1970; Post-Beittenmiller, 1996; Kunst and Samuels, 2003). Because cuticular waxes are products of the metabolism of a single cell layer of plants (i.e., the epidermis), their metabolomic analysis offers a convenient system for evaluating the utility of metabolomics in functional genomics in the absence of the complexity associated with cellular compartmentalization of metabolites. Furthermore, because cuticular waxes are extracellular and they are not covalently bound to the organism, they are a readily extracted and analyzed. Thus, the "exometabolome" of the aerial organs of terrestrial plants has the potential of assessing the metabolic status of the epidermis.

The biological function of the cuticle is complex and not precisely defined (Martin and Juniper, 1970). It has been implicated as having a role in plant–water relationships, and in responses of plants to biotic and abiotic stimuli. Although biochemical studies have provided the skeleton of the metabolic processes that underlie the biosynthesis of this lipid mixture, the genetic and molecular regulation of this process is poorly understood (Post-Beittenmiller, 1996; Kunst and Samuels, 2003). A number of mutant collections that affect the normal accumulation of cuticular waxes are available for elucidating the molecular genetics of cuticular wax biosynthesis. These include an extensive mutant collection in Arabidopsis (the cer mutants; (Jenks et al., 1995; Jenks et al., 1996), in barley (the cer-que mutants; von-Weittstein-Knowles, 1986), and in maize (the glossy mutants; (Schnable et al., 1994)). Additional, but less extensive collections have also been generated in cabbage (Eigenbrode et al., 1995), pea (Macey and Barber, 1970), and sorghum (Jenks et al., 2000). Each of these collections offers unique opportunities for combined biochemical and genetic studies that should reveal different aspects of a very complex metabolic process. This chapter presents the procedures that have been developed for the metabolomics analysis of the cuticular waxes of maize and Arabidopsis.

2 MATERIALS AND METHODS

2.1 Plant materials

The maize (*Zea mays*) inbred line B73 was used in all the studies presented herein. Seeds were germinated in a sand-bench maintained in a greenhouse whose temperature was maintained at 25°C. Plants were illuminated for 16 hours per day with natural sunlight, supplemented with artificial lighting, at a level of 210 μmol m^{-2} s^{-1}. Seedlings at the 2-leaf stage (7-9 days old) were harvested at between 4- and 6-hours after the start of the illumination period, by cutting seedlings at ground level. The coleoptile was removed from the harvested seedling and cuticular waxes were immediately extracted.

The Columbia ecotype of Arabidopsis thaliana was used in all the studies reported herein. Seeds were germinated in soil (professional growing mix Sun Gro LC1) and plants were grown in a growth-chamber, which was maintained at constant illumination level of 60-80 μmol m^{-2} s^{-1}, at a temperature of 22°C, and 75% relative humidity. Waxes were extracted from the bolt, when it was approximately 15 cm tall.

2.2 Extraction of cuticular waxes

The harvested plant material was transiently immersed in chloroform for 60 seconds. The chloroform extract was filtered through a plug of glass wool and the filtrate was dried under reduced pressure in a rotary evaporator at 30°C. The dried wax sample was dissolved in a small volume (250 mL) of chloroform and analyzed *via* HPLC or GC-MS.

2.3 HPLC separation of cuticular waxes

Cuticular wax extracts were separated into their chemical classes by reverse phase HPLC. Chromatography was conducted with a 53 mm x 7 mm (3 μm particle size) Adsorbosphere C18 (12% C) column (Altech, Deerfiled, IL), using a Beckman 110B HPLC system. The flow rate was at 1.0 mL/min. Elution was monitored with an evaporative light scattering detector (ELSD 11A, Varex, Maryland). The nebulizer and the evaporator of the detector were set at 70°C. For maize cuticular waxes the HPLC solvent gradient system was: 0-10 min, 100% THF; 10-20 min, linear gradient to heptane:THF (70:30); 20-25 min at THF:heptane (70:30); 25-36 min, 100% heptane; 36-40 min, 100% THF. For Arabidopsis cuticular waxes the HPLC solvent gradient system was: 0–7 min, 100% THF; 7–17 min, linear gradient to THF:heptane:hexane) (70:15:15); 17–25 min, pentane:hexane (50:50), 25–31 min, linear gradient to 100% THF. Fractions containing constituents of

different chemical classes (i.e., alkanes, alkenes, alcohols, aldehydes, esters, fatty acids, and ketones) were collected using a fraction collector. All chemical standards used in these studies were purchased from Altech (Deerfield, IL, USA).

2.4 Gas chromatography-mass spectrometric analysis

Chromatographic analysis was performed with a Model 6890 series gas chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a Model 5973 mass detector (Agilent Technologies, Palo Alto, CA, USA). Chromatography was conducted with a 30 m length, 0.25 mm I.D. HP-5MS cross-linked (5%)-diphenyl-(95%)-dimethyl polysiloxane column, using helium as the carrier gas. The injection temperature was at 300°C. The column oven temperature was initially at 80°C and was increased to 260°C at a rate of 5°C/min. After holding this temperature for 10 minutes, it was ramped to 320°C at a rate of 5°C/min and held at this temperature for 30 min, and finally cooled to the starting temperature (80°C) over a 5-minute interval. Using the HP enhanced chemical analysis software G1701BA version B.01.00 with Windows NT^{TM} operating system facilitated peak identification.

2.5 Analysis of unsaturated metabolites

The position of carbon–carbon (C–C) double bonds in unsaturated components was determined by the GC-MS analysis of dimethyl disulfide adducts. Isolated cuticular waxes (~1 mg) were dissolved in 20 mL of heptane, and incubated overnight at 40°C°C with 50 mL dimethyl disulfide, and 5 mL 0.06% (w/v) I_2 in diethyl ether. The reaction was stopped by the addition 50 mL heptane, and 25 mL aqueous solution of (5% w/v) sodium thiosulfate. The organic phase was recovered and concentrated prior to GC-MS analysis.

3 RESULTS AND DISCUSSION

To facilitate the complete identification of cuticular wax constituents, extracted cuticular waxes were separated to chemical class components *via* HPLC. Figure 8-1A shows the fractionation of maize waxes into the five major chemical class constituents (aldehydes, alcohols, ketones, esters, and alkanes), and Figure 8-1B shows the similar fractionation of Arabidopsis waxes. The identity of each peak was based on the co-elution with known standard mixtures for each chemical class. These standards were n-alkanes

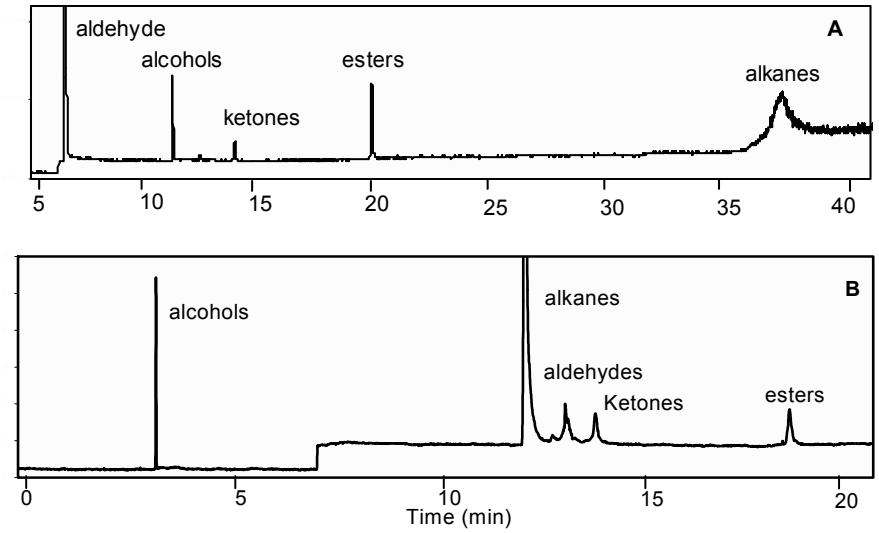


Figure 8-1. The HPLC fractionation of cuticular waxes. Cuticular waxes were extracted from 9-day-old maize seedling leaves (A) and rosette leaves of 21-days old Arabidopsis seedling (B), and fractionated by reverse phase HPLC coupled to an evaporative light scattering detector.

(between 12- and 26-carbons), alcohols (1-octacosanol, 1-octadecanol and 1-docosanol), ketones (2-heptadecanone, 14-heptacosanone, and 6-tricosanone), esters (docosanyl eicosanoate, docosanyl docosanoate, and docosanyl hexacosanoate), aldehydes (octadecanal, decanal, and dodecanal) and n-fatty acid acids (mixture of 16–20 carbon chain lengths).

The fractionated chemical classes were collected, and each fraction was then analyzed *via* GC-MS. By comparing the total ion chromatographic (TIC) profile of the unfractionated cuticular wax extract with that of the TIC of each fraction, it was possible to classify each individual cuticular wax constituent to a chemical class. Thus, this HPLC pre-fractionation, simplified the identification of the cuticular wax constituents prior to GC-MS analysis. Figures 8-2A–C illustrates the application of this strategy for identifying the alcohol and aldehyde constituents of maize cuticular waxes. These analyses identified aldehydes of between 16 and 32 carbon chain lengths (Figure 8-2A), and alcohols of similar chain length distribution (Figure 8-2C). Figure 8-3 illustrates the identification of the cuticular wax components isolated from bolts and siliques of Arabidopsis.

Verification of the chemical identity of each metabolite was achieved by the interpretation of the mass spectra obtained from the electron-impact (EI) ionization/fragmentation of each metabolite. The interpretation of these mass-spectra is illustrated with an example of a metabolite for each chemical class. Figure 8-4A presents the mass spectrum of the 29-carbon n-alkane.

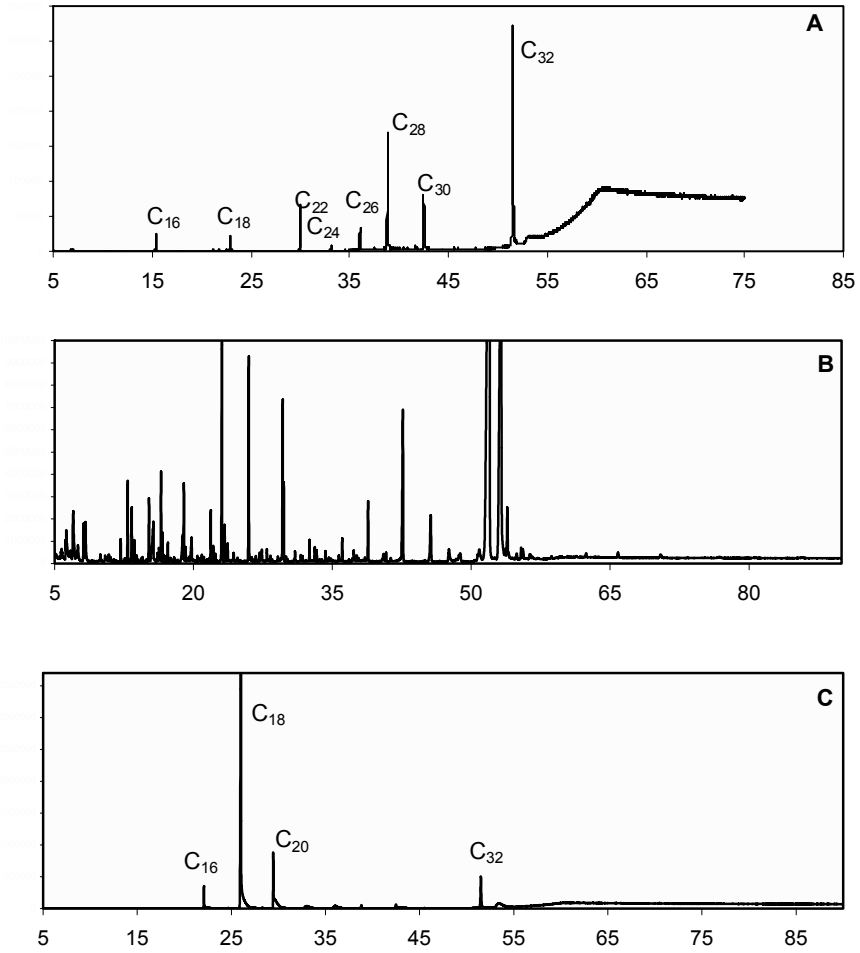


Figure 8-2. Identification of cuticular wax components by combined HPLC and GC fractionation. Isolated maize cuticular waxes were separated into an alcohol and aldehyde fractions by HPLC. The purified alcohol (A) and aldehyde fractions (C) were analyzed by GC, and the resultant chromatograms are compared to the chromatograms of the isolated cuticular waxes (B).

Typical of linear hydrocarbons the spectrum is composed of clusters of fragmentation products that differ from each other by 14 *m/z* mass units, which represents loss of $(CH_2)_nCH_3$ groups from the molecular ion. The *m/z* value of the molecular ion (408 units), which confirms the identity of this alkane as an alkane of 29-carbons. Primary alcohols are identified by the signature loss of a water molecule from the molecular ion, which leads to the increased abundance, relative to that of the molecular ion, of an ion of 18

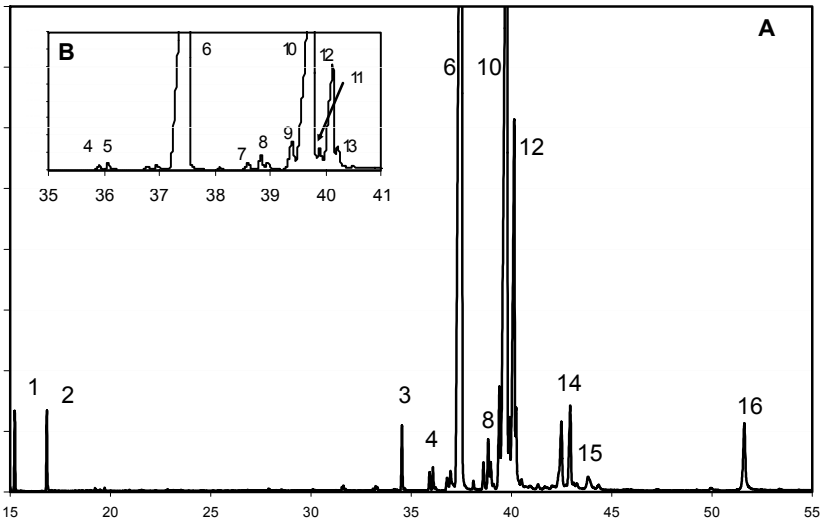


Figure 8-3. GC chromatography of Arabidopsis cuticular waxes. Insert B, is an expanded view of the chromatogram between 35 and 41 minutes of elution time. Peaks were identified as: 1, hexadecanoic acid; 2,octadecanoic acid; 3, 1-tetracosanol; 4, heptacosane; 5, 13-heptacosanone; 6, nonacosane; 7, hexacosanoic acid; 8, secondary alcohol of hexacosanol; 9, 1-octacosanol; 10, 15-nonacosanone; 11, octacosanal; 12, 1- triacontanol; 13, hentriacosane; 14, triacontanal; 15, amyrin; 16, C_{44} ester.

m/z units less than the molecular ion. Thus, as illustrated in Figure 8-4B, the molecular ion of 466 *m/z* units, in combination with the water-loss fragment of 448 *m/z* units, identifies this metabolite as a primary alcohol of 32 carbon chain length.

The signature base-ion facilitates the mass-spectroscopic identification of aldehydes (Figure 8-4C). This fragmentation ion is due to the loss of a CHO group resulting in an ion that is 29 *m/z* units smaller than the molecular ion. However, identification of aldehydes based solely on such a fragmentation pattern is complicated by the fact that alkanes also generate such a fragmentation pattern by the loss of a CH_2–CH_3 group. This complication was clarified by the fact that we had pre-fractionated the alkanes and aldehydes *via* HPLC (Figures 8-1A and B), and thus could independently identify these two classes of metabolites.

Esters are the only molecules present in cuticular waxes that are "hybrid" molecules, being composed of an alcohol and acid moieties. Thus, their characterization required the identification of both moieties. Depending on the combination of these two moieties, each ester at a defined carbon chain length could consist of several isomers, i.e., C_{44} esters could be isomers of C_{20} acid + C_{24} alcohol, C_{22} acid + C_{22} alcohol, C_{24} acid + C_{20} alcohol etc. Our strategy of pre-fractionating the cuticular wax extract *via* HPLC,

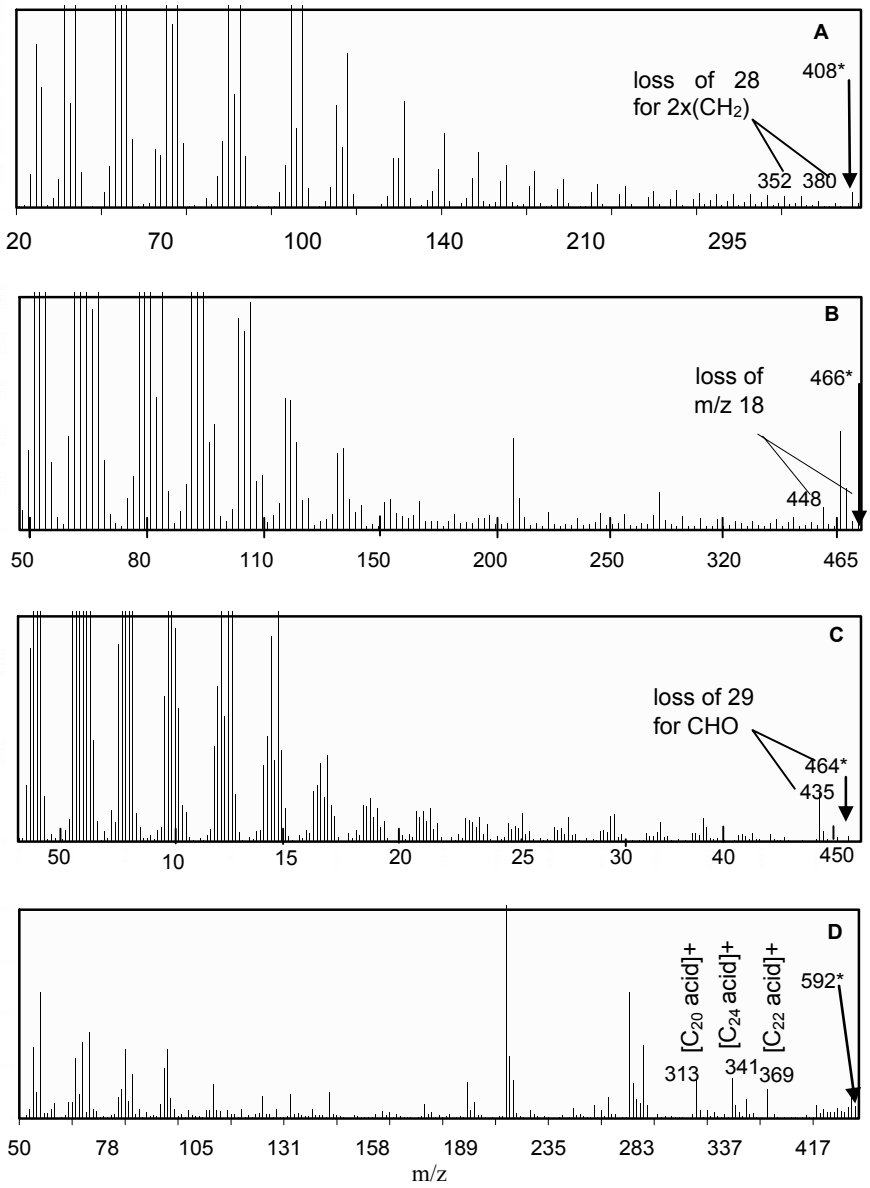


Figure 8-4. Mass-spectra of cuticular wax components. Characteristics of the mass spectra that lead to the identification of alkanes (A), alcohols (B), aldehydes (C) and esters (D) are illustrated with nonacosane, dotriacontanol, dotriacontanal, and C40 ester, respectively.

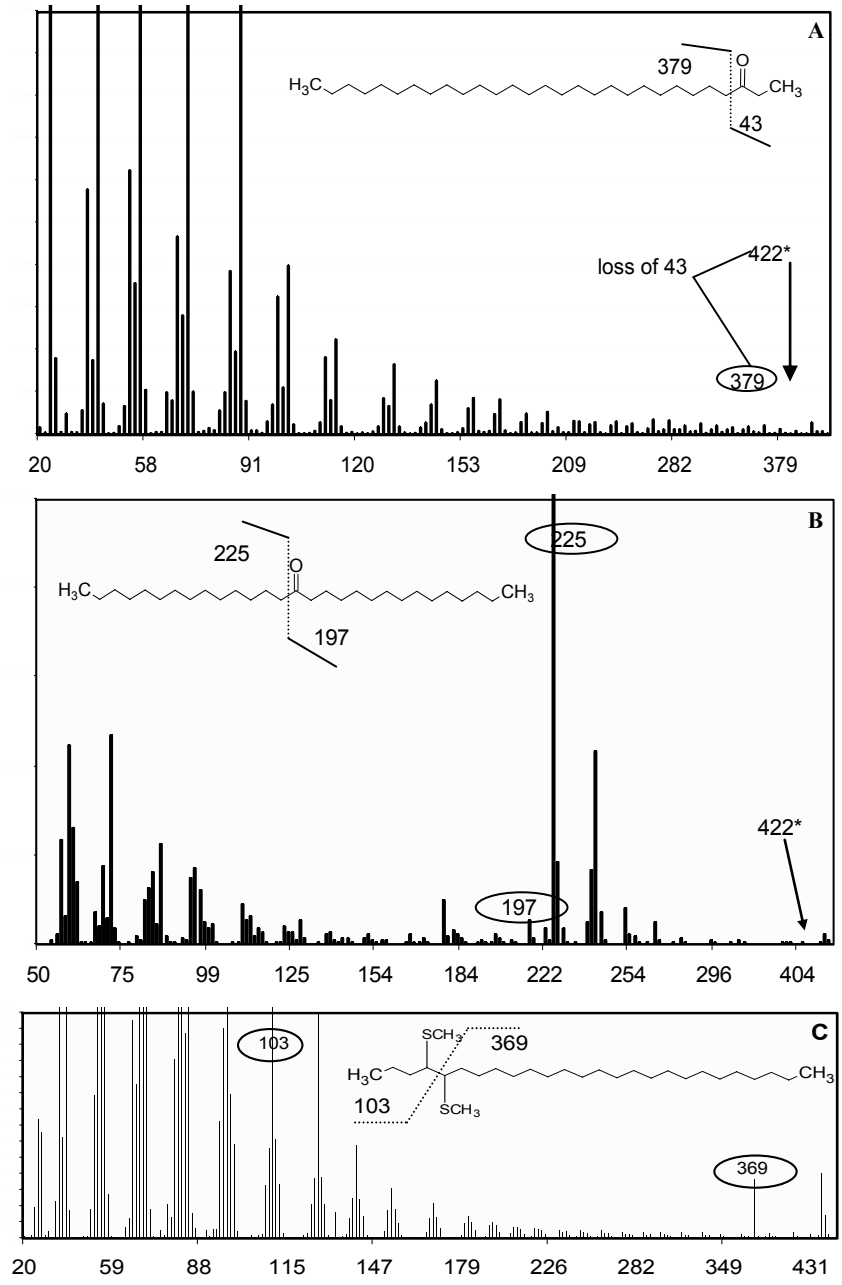


Figure 8-5. Mass-spectra of cuticular wax components. Characteristics of the mass spectra that lead to the identification of methylketones (A), symmetrical ketone (B), and alkenes (C) are illustrated with 2-nonacosanone, 15-nonacosanone, and dimethyl disulfide adduct of 4-heptacosene, respectively.

enabled us to identify these metabolites, however, the fractionation of these esters *via* GC fractionates these metabolites only on the basis of their total carbon number. Ester isomers with the same total carbon number, but differing in their acid and alcohol moieties were identified *via* the characteristic protonated acid fragmentation ions (Reiter et al., 1999). Figure 8-4D, illustrates such an analysis of the C40 ester (identified as such by the *m/z* value of the molecular ion), which is a mixture of three isomers that are each composed of C_{20} acid + C_{20} alcohol, C_{22} acid + C_{18} alcohol, and C_{24} acid + C_{16} alcohol.

Once the identity of the ketones was established by the HPLC fractionation, GC-MS analyses identified the carbon chain length of these molecules and the position of the carbonyl group. The former could be calculated from the *m/z* value of the molecular ion; for example, Figure 8-5 illustrates two isomers of 29-carbon ketones, both of which display a molecular ion of 422 *m/z* units. However, these isomers display distinct fragmentation patterns that reveal the different position of the carbonyl group. EI-induced fragmentation generates ions from the cleavage of the C–C bonds adjacent to the carbonyl group. Thus, in maize the carbonyl group is at the C-2 position (i.e., they are methyl ketones) generating a characteristic ion that is 43 *m/z* units less than the molecular ion due to the loss of a $CO–CH_3$ group (Figure 8-5A). In contrast, the Arabidopsis ketones have a centrally located carbonyl group, which fragment to generate stable base-ions as illustrated in Figure 8-5B. Thus, the Arabidopsis cuticular waxes contain symmetric ketones.

The cuticular waxes of some maize organs, particularly silk and pollen contain unsaturated components. These were identified as alkenes, dienes, aldehydes, and methyl ketones. The position(s) of the carbon-carbon double bonds on these metabolites was identified by the GC-MS analysis of dimethyl disulfide adducts. Upon MS analysis, such adducts preferentially fragment at the bond between the carbon atoms that have been derivatized, yielding two substantial fragment ions that identify the positions of the carbon atoms involved in the carbon-carbon double bond. Figure 8-5C, illustrates the mass spectrum of the dimethyl disulfide adduct of 4-heptacosene (a 27-carbon alkene, with the double bond at the 4^{th} position), which upon fragmentation generates substantial fragment ions of 103 and 369 *m/z* units.

In total these analyses identified 232 metabolites from the cuticular waxes of maize and Arabidopsis (Table 8-1). Associated with each metabolite is a high-quality mass-spectrum, measured on a double-focusing sector field spectrometer (70 eV EI). In addition, each metabolite is identified *via* retention indices from a nonpolar stationary phase column (HP5, Agilent Technologies, Palo Alto, CA, USA). The resulting database has been useful in the characterization of plants that carry cuticular wax mutations (Nikolau et al., 2002, ; Nikolau et al., 2003; Dietrich et al., 2005;

Perera et al., 2005). Moreover, this database has been used to discover new genes in new biosynthetic pathways (Perera et al., 2005).

Table 8-1. Cuticular wax constituents

Chemical class	Carbon chain lengths[a]	
	Maize	Arabidopsis
Saturated aldehydes	C_{22}, C_{24}, C_{26}, C_{28}, C_{30}, **C_{32}**	C_{26}, C_{28}, **C_{30}**
Unsaturated aldehydes	C_{26}, **C_{28}**, C_{30}	none
Primary alcohols	C_{16}, C_{18}, C_{20}, C_{24}, C_{26}, C_{30}, **C_{32}**	C_{24}, **C_{28}**
Secondary alcohols	none	C_{27}, **C_{29}**
Alkanes	C_{15}, C_{19}, C_{23}, C_{29}, **C_{31}**	C_{27}, **C_{29}**, C_{31}
Alkenes	C_{19}, C_{23}, C_{25}, C_{27}, **C_{29}**, C_{31}	none
Dienes	C_{25}, C_{27}, **C_{29}**, C_{31}	none
Methyl ketones	C_{17}, C_{23}, C_{25}, C_{27}, **C_{31}**	C_{25}, C_{27}, **C_{29}**
Symmetric ketones	none	C_{25}, C_{27}, **C_{29}**
Unsaturated ketones	C_{21}, C_{25}, C_{27}, **C_{29}**, C_{31}	none
Esters	C_{40}, C_{42}, C_{44}, C_{46}, C_{48}, **C_{52}**	C_{42}, **C_{44}**
Esterified alcohols	C_{16}, C_{18}, C_{20}, C_{24}, **C_{26}**, C_{28}, C_{30}, C_{32}	C_{16}, C_{18}, **C_{20}**
Esterified acids	C_{16}, C_{18}, C_{20}, **C_{24}**, C_{26}	C_{18}, C_{20}, **C_{24}**
Free fatty acids	Trace	**C_{16}**, C_{18}, C_{24}, C_{30}

[a]In each chemical class, the carbon chain length of the most abundant metabolite is identified in bold-text.

ACKNOWLEDGEMENTS

This research was supported in part by grants from the National Science Foundation (IBN-9808559 and IOB-0344852) and by Hatch Act and State of Iowa funds.

REFERENCES

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S., and Sumner, L.W., 2004, Potential of metabolomics as a functional genomics tool, *Trends Plant Sci.* **9**:418–425.

Dietrich, C.R., Perera, M.A., M. DY-N, M., Meeley, R.B., Nikolau, B.J., and Schnable, P.S., 2005, Characterization of two GL8 paralogs reveals that the 3-ketoacyl reductase component of fatty acid elongase is essential for maize (Zea mays L.) development, *Plant* **42**:844–861.

Eigenbrode, S.D., Moodie, S., Castagnola, T., 1995, Predators mediate host plant resistance to a phytophagous pest in cabbage with glossy leaf wax. Entomologia Experimentalis et Applicata **77**: 335–342.

Fatland, B., Nikolau, B.J., and Wurtele, E.S., 2005, Reverse Genetic Characterization of Cytosolic Acetyl-CoA Generation by ATP-citrate lyase in Arabidopsis, *Plant Cell* **17**:182–203.

Fiehn, O., 2002 Metabolomics – the link between genotypes and phenotypes, *Plant Mol. Biol.* **48**:155–171.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L., 2000, Metabolite profiling for plant functional genomics, *Nat. Biotechnol.* **18**:1157–1161.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., and Bino, R., 2002, Plant metabolomics: the missing link in functional genomics strategies, *Plant Cell* **14**:1437–1440.

Jenks, M.A., Rashotte, A.M., Tuttle, H.A., and Feldmann, K.A., 1996, Mutants in Arabidopsis thaliana Altered in Epicuticular Wax and Leaf Morphology, *Plant Physiol.* **110**:377–385.

Jenks, M.A., Rich, P.J., Rhodes, D., Ashwort, E.N., Axtell, J.D., and Din, C.K., 2000, Leaf sheath cuticular waxes on bloomless and sparse-bloom mutants of Sorghum bicolor, *Phytochemistry* **54**:577–584.

Jenks, M.A., Tuttle, H.A., Eigenbrode, S.D., and Feldmann, K.A., 1995, Leaf Epicuticular Waxes of the Eceriferum Mutants in Arabidopsis, *Plant Physiol* **108**:369–377.

Ke, J., Wen, T.N., Nikolau, B.J., and Wurtele, E.S., 2000, Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the heteromeric acetyl-coenzyme A carboxylase, *Plant Physiol.* **122**:1057–1071.

Kunst, L. and Samuels, A.L., 2003, Biosynthesis and secretion of plant cuticular wax, *Prog. Lipid Res.* **42**:51–80.

Macey, M.J.K. and Barber, H.N., 1970, Chemical genetics of wax formation on leaves of Pisum sativum-D, *Phytochemistry* 9.

Martin, J.T., and Juniper, B.E., 1970, *The Cuticle of Plants*, Edward Arnold Ltd, Edinburgh.

Nikolau, B.J., Perera, M.A., Dietrich, C.R., and Schnable, P.S., 2002, High-throughput metabolomic analysis of an extensive new collection of maize glossy mutants and of

maize inbreds reveals novel genetic regulation of epicticular wax biosynthesis, *In*: 1st International Congress on Plant Metabolomics, Wageningen, The Netherlands.

Nikolau, B.J., Perera, M.A.D.N., Dietrich, C.R., and Schanble, P.S., 2003, Metabolomic analysis of epicuticular wax biosynthesis in maize using a collection of maize glossy mutants, *In*: 2nd International Congress on Plant Metabolomics, Potsdam, Germany.

Oliver, D.J., Nikolau, B., and Wurtele, E.S., 2002, Functional genomics: high-throughput mRNA, protein, and metabolite analyses, *Metab. Eng*. **4**:98–106.

Perera, A., Dietrich, C.R., Schanble, P.S., and Nikolau, B.J., 2005, Metabolomic analysis of a collection of maize glossy mutants reveals novel aspects of cuticular wax biosynthesis, In preparation.

Post-Beittenmiller, D., 1996, Biochemistry and molecular biology of wax production in plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**:405–430.

Reiter, B., Marion, L., Lorbeer, E., Aichholz, R., 1999, Isolation and characterization of wax esters in fennel and caraway seed oils by SPE-GC, *J. High Resol.Chromatogr.* **22**:514–520.

Schnable, P.S., Stinard, P.S., Wen, T.-J., Heinen, S., Weber, D., Zhang, L., Hansen, J.D., and Nikolau, B.J., 1994, The genetics of cuticular wax biosynthesis, *Maydica* **39**:279–287.

Von Wettstein-Knowles, P., 1986, Role of *cer-cqu* in epicuticular wax biosynthesis, *Biochemical Society Transactions* **14**:576–579.

Weckwerth, W., and Fiehn, O., 2002, Can we discover novel pathways using metabolomic analysis? *Curr. Opin. Biotechnol.* **13**:156–160.

Chapter 9

METABOLIC FLUX MAPS OF CENTRAL CARBON METABOLISM IN PLANT SYSTEMS

Isotope Labeling Analysis

V. V. Iyer[1], G. Sriram[2] and J. V. Shanks[1]
[1]*Department of Chemical and Biological Engineering, Iowa State University, Ames, IA 50011, USA;* [2]*Department of Human Genetics and Department of Chemical and Biomolecular Engineering, UCLA, Los Angeles, CA 90095, USA*

Abstract: Metabolic flux analysis (MFA) quantifies carbon flow in a biological system, which is an important characteristic reflective of physiology. Nodal rigidity of the metabolic network at branchpoints can be assessed from flux ratios to compare genetic and environmental variants and identify targets for potential genetic manipulations. MFA coupled with systems-wide tools such as transcriptomics and metabolomics have significant potential for building predictive models of plant metabolism. This chapter aims to explain the methodology behind MFA using carbon labeling experiments (CLE), nuclear magnetic resonance spectroscopy and a comprehensive mathematical framework (NMR2Flux) for a better understanding of central carbon metabolism in plants.

1 INTRODUCTION

Genetic engineering marked the advent of modifying specific enzymatic reactions using recombinant DNA technology. Early genetic engineering manipulations showed that a single gene transformation can result in unexpected changes in the metabolic pathways and phenotypic behavior and gave credence to a systems-level understanding of physiology. Consequently, the field of metabolic engineering emerged, which dealt with a systematic approach towards pathway modification to understand the underlying physiology (Stephanopoulos et al., 1998). The significance of metabolic engineering lay in the fact that the metabolic network was considered in its entirety as opposed to a single reaction.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 125–144.

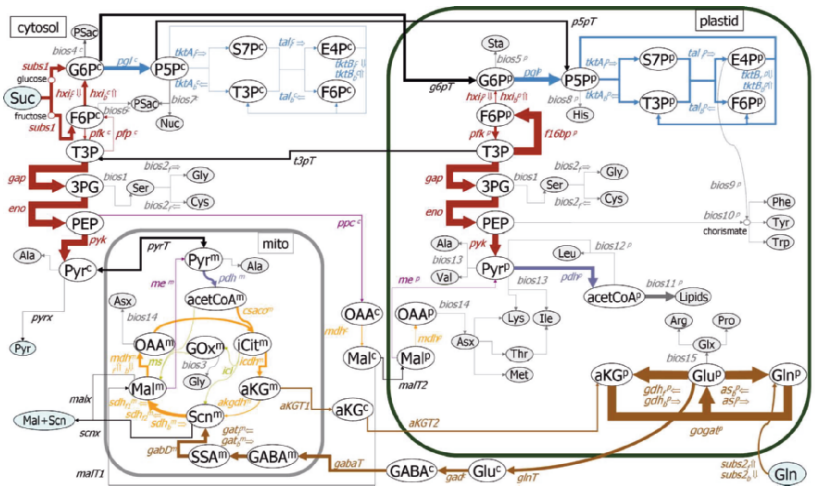


Figure 9-1. Metabolic network for central carbon metabolism in embryos of soybean (Glycine max). Parallel pathways for glycolysis and the pentose phosphate pathway exist in the cytoplasm and the plastid and communication between them occur through three transporters: glucose 6-phosphate (g6pT), pentose 5-phosphate (p5pT), and triose 3-phosphate (t3pT). The thickness of the arrows is directly proportional to the flux values. Reprinted from Sriram et al. (2004) with permission of the American Society of Plant Biologists.

The importance of metabolic fluxes as a fundamental determinant of cell physiology was promoted by metabolic engineering (Stephanopoulos, 1999). Metabolic flux is defined as the net rate of conversion of a precursor metabolite to a product in a metabolic pathway. The quantification of intracellular metabolite fluxes in a network of metabolic pathways is termed as metabolic flux analysis (MFA). In particular, MFA has been applied to network models of central carbon metabolism due to it importance in cellular physiology (Stephanopoulos, 1999). Central carbon MFA calculates steady-state intracellular fluxes using a stoichiometric model supplemented with extracellular measurements such as substrate intake and effluxes of metabolites. In larger network models for which additional measurements are required, constraints in the form of labeling data from Nuclear Magnetic Resonance (NMR) spectroscopy or Mass Spectroscopy (MS) can be applied (Section 2 of this chapter). The result of MFA is a metabolic flux map (Figure 9-1) which indicates the steady-state fluxes through various reactions of the metabolic pathway. Such metabolic flux maps can be effectively used for comparing flux differences in genetic or environmental variants. Subsequently, once the effect of a genetic or environmental manipulation is analyzed, further hypotheses are developed and tested (genetic modification followed by analysis) in an interactive cycle to further characterize the cellular physiology (Nielsen, 1998).

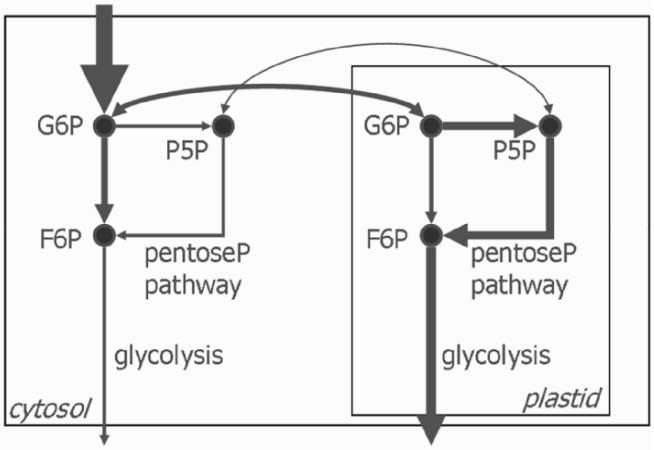


Figure 9-2. Parallel glycolytic and pentose phosphate pathways in cytosol and plastid. Transporters facilitate plastidic and cytosolic interactions.

Although the application of MFA in central carbon metabolism in microbes has been appreciable, unfortunately the same cannot be said in plants. Metabolic networks are more complex to analyze in plants than in microbes. One of the major factors contributing to the complexity of plant networks is compartmentation. In plants, the same reaction pathway may occur in more than one compartment as shown in Figure 9-2. Transporters facilitate the exchange of metabolites between compartments thus making intracellular transport processes important. Hence, the quantification of fluxes in parallel compartments becomes vital (Shanks, 2000). Additionally, higher plants are separated on various levels such as the tissue (roots, stems, and leaves) and cellular levels within a tissue. Furthermore, the topology of plant networks is often incomplete.

As a result of aforementioned complexity of plant metabolic networks, the few "flux" labeling studies in plants have focused on either the identification of metabolic network topology (Glawschnig et al., 2002; Krook et al., 1998; Schwender et al., 2004) or flux quantification using analytical or a highly simplified ^{13}C NMR constrained analysis (Dieuaide-Noubhani et al., 1995; Rontein et al., 2002; Schwender et al., 2003). Plant systems biology has reemphasized the importance of fluxes (Girke et al., 2003; Stitt and Fernie, 2003; Sweetlove et al., 2003) in achieving the "in silico" plant (Minorsky, 2003). Thus, it has become more essential that application of MFA to different plant systems be promoted. Toward this goal, a comprehensive flux analysis tool for central carbon metabolism, NMR2Flux, was developed using recent mathematical advances from our research group (Sriram et al., 2004). This chapter aims at explaining the theoretical background and the methodology that NMR2Flux employs in the

evaluation of intracellular fluxes using the example of the developing soybean embryo.

2 METABOLIC FLUX ANALYSIS

Metabolic flux analysis (MFA) involves the quantification of intracellular steady-state fluxes in the cell, using metabolite balances and extracellular measurements.

2.1 Stoichiometric flux analysis

Metabolic flux analysis relies on the principle of conservation of mass: mass cannot be created or destroyed. Stoichiometric MFA is the most basic approach of metabolic flux analysis and requires details of the reaction stoichiometry. Consequently, mass balances around each intracellular metabolite are written to generate a system of linear equations (Stephanopoulos et al., 1998; Varma and Palsson, 1994). Thus in vector notation,

$$d\mathbf{X}/dt = \mathbf{r} - \mu\mathbf{X} \tag{1}$$

where, **X** represents the concentration of the metabolite under consideration; **r**, the rate of formation of the metabolite and μ is the biomass growth rate. Assuming a pseudo-steady-state, where the rate of turnover of X (left-side of equation (1)) is smaller than the sum of the rate of metabolite formation and dilution due to cell growth (right-hand side of equation (1)), we have,

$$\mathbf{r} - \mu\mathbf{X} = 0 \tag{2}$$

The dilution due to biomass growth is generally small and the second term can be neglected and we have,

$$\mathbf{r} = \mathbf{G}^{\mathrm{T}}.\ \mathbf{v} = 0 \tag{3}$$

where, **v** is the vector containing the fluxes and **G** is the stoichiometric matrix. If the network model has J reactions and K internal metabolites, the degrees of freedom F, is represented by,

$$F = J - K \tag{4}$$

Hence, to solve for the intracellular fluxes, some measurements such as substrate consumption, metabolite effluxes etc. have to be supplied. Thus, the measured and calculated fluxes can be partitioned into $\mathbf{v}_m$ and $\mathbf{v}_c$,

respectively. Correspondingly, the stoichiometric matrix can be partitioned into $\mathbf{G}_m$ and $\mathbf{G}_c$. Thus, knowing $\mathbf{v}_m$ and $\mathbf{G}$, we can calculate $\mathbf{v}_c$, the set of unmeasured intracellular fluxes. If the number of supplied measurements is same as F, it is an exactly determined system; if greater than F, an overdetermined system; and if less than F, an underdetermined system. The exactly determined and overdetermined systems will have a unique solution for the distribution of fluxes through the metabolic network. In addition, for an overdetermined system, the extra measurements can be used to check the validity of the metabolic network.

On the other hand, to solve an underdetermined system, cofactor balances (NADPH/NADH) may need to be supplemented as additional constraints (Varma and Palsson, 1994). However, the NADPH, NADH and ATP balances are not closed in reality due to futile cycles and incomplete pathway knowledge. Stoichiometric MFA also fails in certain cases of parallel pathways and metabolic cycles (Wiechert, 2001). It is hence essential to provide further information and also elucidate flux distribution at branchpoints. For larger networks with an increase in the number of reactions, flux analysis becomes more difficult as the number of measurements required correspondingly increases. Consequently, ^{13}C labeling experiments can be used to complement stoichiometric balancing and extracellular measurements, thereby providing a rigorous alternative to traditional flux analysis.

2.2 ^{13}C metabolic flux analysis

Carbon labeling experiments (CLE) involve feeding a combination of labeled (^{13}C or ^{14}C) substrates along with ^{12}C substrates such as glucose or sucrose to the biological system of interest. The label gets distributed throughout the network when the substrate is assimilated into metabolites. The labeling pattern of various metabolites depends on the network topology and the intracellular fluxes. The labeling patterns can be detected by Nuclear Magnetic Resonance (NMR) spectroscopy (Marx et al., 1996; Szyperski, 1995) or Mass Spectroscopy (MS) (Christensen and Nielsen, 1999) or a combination of the two (Klapa et al., 1999). The labeling data can be translated into flux information, using the concept of isotopomers as explained in Section 3.3 (Klapa et al., 1999; Schmidt et al., 1997).

Plant systems exemplify complex metabolic networks due to compartmentation issues, futile cycling, and anaplerotic reactions. Consequently, additional measurements are required in plant systems and the number of isotopomer balances increases, further increasing the computational burden. Due to the mathematical burden required for quantification of flux, most papers that have reported labeling studies in plants have focused on the identification of metabolic network topology

rather than flux quantification (Glawschnig et al., 2002; Krook et al., 1998). In an elegant example of the use of labeling for network topology, Schwender et al. demonstrated the use of labeling studies to identify a new pathway in *Brassica napus* embryos (Schwender et al., 2004). They characterized the role of Rubisco in the absence of Calvin cycle in improving the efficiency of carbon utilization during oil synthesis. Earlier studies of flux quantification in plants have used analytical or a highly-simplified ^{13}C NMR constrained analysis (Dieuaide-Noubhani et al., 1995; Rontein et al., 2002; Schwender et al., 2003). These simplified analyses may lead to erroneous fluxes – a comprehensive analysis from an abundance of data is needed to verify the assumptions (Sauer, 2004). Recently, we have been able to execute comprehensive flux analysis of central carbon metabolism in plant tissues (Sriram et al., 2004). Section 3 of this chapter describes our analysis methodology in detail.

3 FLUX EVALUATION METHODOLOGY

Fluxes in a biological system can be quantified from isotopomer abundances, extracellular measurements, and biomass accumulation data coupled with a mathematical framework, using the evaluation methodology as explained below.

3.1 Experimental design

The selection of the type of labeled substrate, i.e., selective or uniformly labeled is a fundamental component of experimental design (Schmidt et al., 1999). In addition, it is essential that the relative extents of the labeled and unlabeled substrate be decided *a priori* to get adequate information from the NMR data (Stephanopoulos et al., 1998; Szyperski, 1995). In the case of selectively labeled substrates, a large percentage of labeling (as high as 90%) has to be used to obtain meaningful data (Park et al., 1999). On the other hand, when a mixture of uniformly labeled and unlabeled substrates is used, carbon bond-bond connectivities are traced as opposed to fractional enrichments. Hence, percentages of uniformly labeled substrate required are much lower (approximately 10%) for adequate NMR data (Szyperski, 1995).

Once the type of labeled substrate and their extents are decided, the cells are cultured with the mixture of labeled and unlabeled substrate. The experiment is carried out at metabolic (the rate of change of intracellular metabolite concentrations is much less than that of fluxes in and out of the metabolite) and isotopic (labeling patterns of the metabolites do not change with time) steady-states. Finally, the biomass from the plant tissue is extracted and broken down into its corresponding components. Depending on the

network topology and intracellular fluxes, different labeling patterns of the metabolites will be reflected from the biomass components (e.g., protein or starch sample), which can be detected using NMR or MS (Christensen and Nielsen, 1999; Klapa et al., 1999; Szyperski, 1995). The conversion of NMR data to fluxes using both type of substrates involve the concept of isotopomers as explained in section 3.3. Details of the experimental setup for labeling studies in developing soybean embryos and NMR sample preparation have been discussed in our recent work (Sriram et al., 2004).

3.2 NMR spectroscopy

NMR spectroscopy has proved to be an efficient analytical technique to gain significant insights into plant metabolism (Ratcliffe and Shachar-Hill, 2001). In an NMR measurement, the spin of the ^{13}C nucleus is detected and gives rise to a signal. The signal to noise ratio (S/N) in an NMR experiment depends directly on the concentration of the nuclei (C) and the number of scans (N_s).

$$\mathrm{S/N} \propto C * (N_s)^{0.5} \tag{5}$$

The accumulation time (T_a) for the signal depends on N_s and the pulse interval T_p (Shanks, 2000),

$$T_a = T_p * N_s \tag{6}$$

T_p is given by the following equation,

$$T_p = t_p + T_{acq} + T_{rd} \tag{7}$$

where, t_p is the length of radiofrequency pulse, T_{acq} is the acquisition time and T_{rd} is the relaxation delay (Shanks, 2000). Hence, for example, to double S/N, the number of scans needs to be increased four times. Also, T_a depends directly on N_s and will increase proportionately. Thus, it is essential to balance the parameters T_a, N_s and S/N to keep the NMR analysis cost at a reasonable limit without compromising on the S/N ratio. Using the developing soybean embryo system as an example, some of the key parameters for the NMR analysis have been discussed below and two dimensional (2D) experiments for detection of labeling patterns have been suggested.

In the soybean *in vitro* experiment, only 10% uniformly labeled sucrose was fed to the soybean cotyledons. Assuming this 10% labeling randomly distributes through the network, the probability that two adjacent atoms are

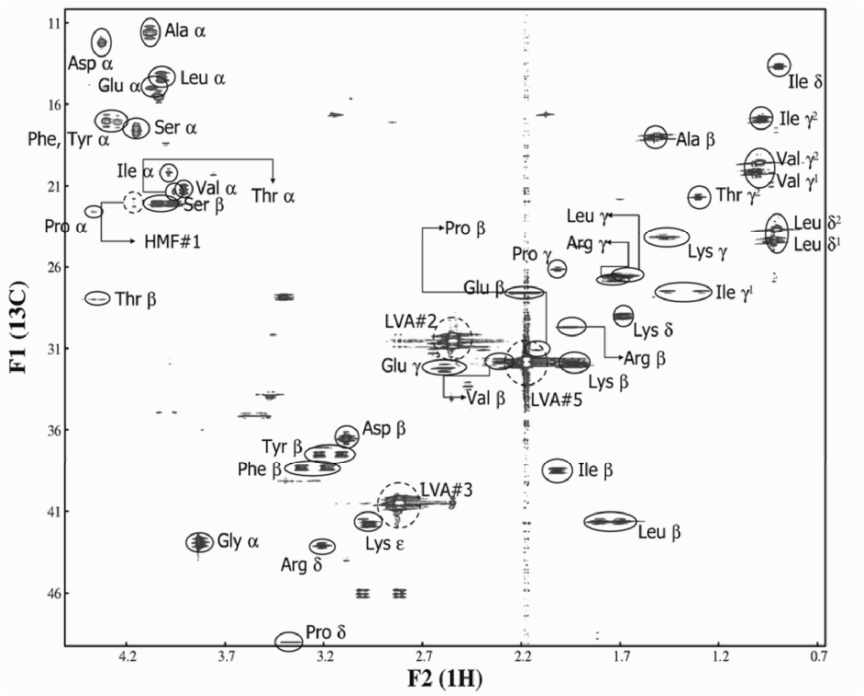


Figure 9-3. 2D [^{13}C, ^{1}H] HSQC spectrum of protein hydrolysate from soybean cotyledons cultured on sucrose (10% w/w U-^{13}C) and glutamine. Cross peaks represent carbon atoms of hydrolysate constituents (proteinogenic amino acids and hydrolysis products of sugars from glycosylated proteins – 5-hydroxymethyl furfural (HMF) and levulinic acid (LVA)). Reprinted from Sriram et al. (2004) with permission of the American Society of Plant Biologists.

labeled, is about 1% and the probability that two adjacent atoms originating from the same metabolite are labeled is 10% (Szyperski, 1995). From the specifications of the 500 MHz spectrometer, the minimum concentration for a 2D NMR analysis was determined to be 1 mM. The amino acid with the lowest concentration in the soybean protein hydrolysate was methionine (2 mol%). Hence, for a 20–22 hour [^{1}H, ^{13}C] Heteronuclear Single Quantum Correlation (HSQC) experiment, taking the aforementioned parameters into consideration, the minimum amount of soybean protein required for an adequate S/N ratio was 20 mg.

Two experiments, the HSQC and [^{1}H, ^{1}H] Total Correlation Spectroscopy (TOCSY) were performed on a Bruker Avance DRX 500 MHz spectrometer at 298 K on the soybean protein hydrolysate. For more details on the parameters of the NMR experiment, the reader is referred to our previous work (Sriram et al., 2004). The HSQC analysis determines the labeling pattern between the adjacent carbon atoms (Szyperski, 1998). Also, since we have unlabeled glutamine as a carbon source in addition to the

labeled sucrose, there is a dilution of ^{13}C in the system. The 2D TOCSY analysis detects the protons attached to ^{12}C and ^{13}C, thus providing the enrichment of each carbon atom of amino acids.

NMR spectra were acquired and processed using the Xwinnmr (Bruker) software. Peak assignments were verified using 2D [^{1}H, ^{1}H] TOCSY and 3D [^{13}C, ^{1}H, ^{1}H] TOCSY (Braunschweiler and Ernst, 1983) experiments on 100% labeled protein sample, with a pH of 1.0. Hence, the pH of the soybean sample was also adjusted to 1.0 to avoid a change in the chemical shift data caused by the environmental variation. The 2D HSQC spectrum of the hydrolyzed soybean protein with peak assignments of the carbon atoms of amino acids is shown in Figure 9-3. The HSQC and TOCSY spectra were analyzed using the free software NMRview (Johnson and Blevins, 1994). The deconvolution of the multiplet peaks was carried out using software based on the spectral processing proposed by van Winden et al. (van Winden et al., 2001). The isotopomer theory employed to convert the NMR data to intracellular fluxes using the software NMR2Flux (Sriram et al., 2004) is explained below.

3.3 Isotopomer theory

The concept of isotopomers arises from a combination of the terms isotope and isomers, which represent various labeling patterns of a given metabolite. For example, for a three carbon metabolite, there are $2^3 = 8$ isotopomers possible (Figure 9-4). Hence for a metabolite with n carbons, there are 2^n labeling patterns possible. As mentioned before, 2D HSQC detects the labeling patterns of adjacent carbon atoms. The peak fine structure obtained from the HSQC experiment shows multiplet patterns proportional to the isotopomer abundances (Figure 9-5). The labeling data is converted to flux data by comparing the experimental data with simulated

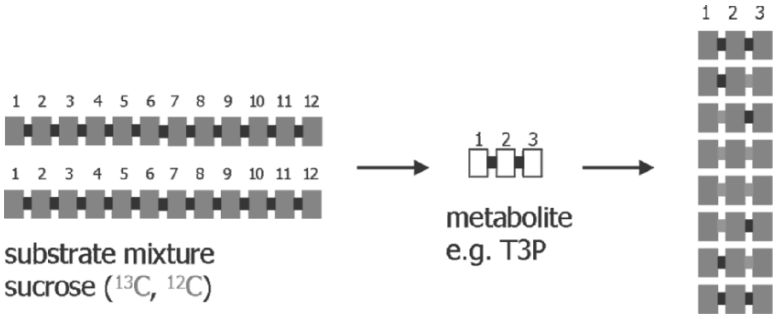


Figure 9-4. Isotopomers of a three carbon metabolite from a mixture of uniformly labeled and unlabeled sucrose.

isotopomer abundances generated using isotopomer mapping matrices (IMM). Assuming a set of intracellular fluxes, IMM uses the concept of isotopomer distribution vectors (IDV) and reaction stoichiometry to generate the simulated isotopomer abundances (Schmidt et al., 1997). The "best" set of intracellular fluxes satisfies the reaction stoichiometry and also shows the least mean square error between experimental and simulated isotopomer abundances. IMM are analogous to the Atom Mapping Matrices (AMM) used to calculate the TCA flux ratios in a hybridoma cell line (Zupke and Stephanopoulos, 1994).

3.4 Additional measurements

In addition to NMR labeling data, extracellular measurements such as substrate intake, product effluxes, and fluxes contributing to biomass accumulation are essential inputs to the model. The substrate intake and product effluxes can be measured by carrying out a high performance liquid chromatography (HPLC) of the culture media. It is also required that the biomass composition be known so that the carbon balance of the system is adequately accounted for. For example, in the soybean embryo culture, in addition to protein, lipids, and starch, a major constituent of the biomass were seed coat carbohydrates. The sugars that contributed to the carbohydrates were estimated from literature values (Mullin and Xu, 2000). The dry weight of the embryo and the fractions of protein, lipids, and starch were measured using standard protocols (Sriram et al., 2004). The fatty acid composition of the lipid fraction was estimated from literature (Dey and Harborne, 1997). When the molecular formula of the biomass is known, it can be used to close the carbon balance more efficiently.

The external fluxes contributing to protein were determined from the amino acid HPLC analysis, coupled with the precursor-amino acid stoichiometry (Szyperski, 1995). To elucidate, let us consider the synthesis of the amino acid phenylalanine (Phe) from erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP),

$$1 \text{ mole } E4P + 2 \text{ moles } PEP \rightarrow 1 \text{ mole } Phe \tag{8}$$

Thus, from protein data and HPLC analysis of the amino acids, the total number moles of Phe in the sample are known. Consequently, from equation (8) total moles of the precursor metabolite PEP required for synthesis of Phe can be calculated. Additionally, we know that tyrosine (Tyr) and tryptophan (Trp) are the other amino acids synthesized from PEP. Hence, the total external flux from PEP can be calculated from the sum of the moles of PEP required for synthesis of the corresponding three amino acids (Tyr, Trp, and Phe). Similar analysis can be carried out for remaining precursor metabolites

(for example, Pyr, OAA, P5P, etc.) associated with the synthesis of amino acids.

3.5 Metabolic reaction network

It is essential that a metabolic network mimicking the underlying physiology be proposed to convert the labeling data to intracellular fluxes. Figure 9-1 shows a metabolic network that describes sucrose metabolism in developing soybean embryos. The fluxes in a reaction network are stoichiometrically related to each other and can be expressed in terms of flux parameters. The selection of flux parameters is important to solve the metabolic network (Sriram et al., 2004). Some potential candidates for flux parameters are the independent reactions of the network (Stephanopoulos et al., 1998), reversibilities of key reactions and also scrambling extents of parallel reactions (Szyperski, 1995). Product effluxes, substrate consumption, and biosynthetic reactions are additional essential extracellular measurements required for accounting for complete carbon balance, thereby providing a better estimate of the intracellular fluxes.

Further, labeling data which give key information about branchpoints are additional important inputs. In the event that the labeling data does not satisfy the proposed reaction network, certain reactions may need to be added or removed from the proposed network to satisfy the labeling data. Also, sometimes the error in the experimental NMR data can translate to a very high probability distribution of the flux, leading to "identifiability" problems (Wiechert et al., 2001). For example, flux analysis of the

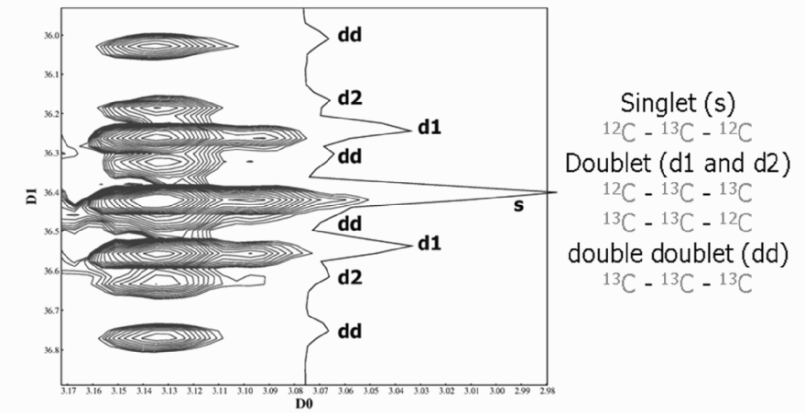


Figure 9-5. Peak fine structure of Aspβ. The multiplet intensities are proportional to the isotopomer abundances. Reprinted from Sriram et al. (2004) with permission of the American Society of Plant Biologists.

reversibilities of the transketolase and transaldolase reactions in the pentose phosphate pathway depend primarily on the labeling information obtained from the PEP family of the amino acids and histidine. If the NMR data is not sufficient to estimate these fluxes or if the error in the measurements is large, then the fluxes become "structurally unidentifiable". The problem of structural identifiability can be solved by increasing the number of external measurements pertaining to that particular part of the metabolic network or providing low error NMR data. However, in some cases, the relationship between the NMR measurements and the fluxes are highly nonlinear. In such cases, the fluxes become "statistically unidentifiable" and a very low noise level can translate into large probability distributions of the corresponding fluxes. Thus, in the case of a statistically unidentifiable flux, the model cannot estimate the flux irrespective of redundant measurements pertaining to that flux. Such issues need to be studied in detail in the course of developing the experimental design of the biological system.

3.6 Mathematical modelling of the reaction network

The metabolite balances from the metabolic network coupled with the carbon skeleton rearrangements are fundamental in enumerating the isotopomers of the metabolites in the network. Both analytical approaches (Klapa et al., 1999; Park et al., 1999; Rontein et al., 2002) and numerical solutions (Schmidt et al., 1999; Wiechert and De Graaf, 1997a; Wiechert et al., 1999; Zupke and Stephanopoulos, 1994) have been used to solve isotopomer abundances for calculating intracellular fluxes. A generic software using the concept of isotopomer balancing for flux analysis is also available (Wiechert et al., 2001).

More recently, a generic tool NMR2Flux (Figure 9-6) has been developed in our lab by employing recent mathematical advances, that can be extended to complex plant systems (Sriram et al., 2004). The tool chooses an initial set of flux parameters (independent net fluxes, reversibilities, and scrambling extents) that are stoichiometrically feasible (Sriram et al., 2004; Stephanopoulos et al., 1998). From the feasible set of flux parameters, the remaining fluxes can be calculated. These fluxes are converted to isotopomer distributions using a recently developed efficient Boolean function mapping method (Figure 9-7), coupled with explicit solution methods (Wiechert and Wurzel, 2001). Boolean function mapping is a novel method of simulating isotopomer distributions. Carbon skeletal rearrangement steps are modeled as Boolean or arithmetic operations on the decimal representation of an isotopomer. The Boolean function mapping method is based upon the fact that all reactions in a metabolic network can be represented as occurring between two reactants (R_1, R_2) to give two

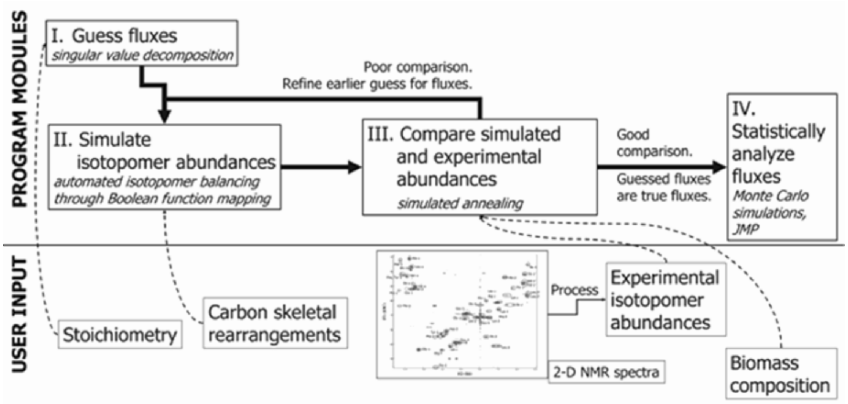


Figure 9-6. Flux evaluation methodology.

products (P_1, P_2), i.e., they can be represented as "bi–bi" reactions (Wiechert and Wurzel, 2001). Reactions steps in this schema can be described as a function of four different "moves": fragmentation, reversal, transposition, and condensation. The simulated and experimental (from NMR data) isotopomer abundances are compared and the error between them is minimized using a global optimization routine (employing simulated annealing).

The reduction in computation time achieved using the Boolean function mapping method allows additional statistical analysis of fluxes. The errors in the NMR input intensities are used to perform multiple Monte Carlo estimation of fluxes (Press et al., 1992); thereby generating probability distribution of the intracellular fluxes (Sriram et al., 2004). For further information on the tool and comprehensive explanation of the mathematical details, refer our recent work (Sriram et al., 2004).

Recently, a new concept, bondomer was introduced which can be used in case of single carbon substrate experiments only. Bondomers are similar to isotopomers except that the bonds instead of the carbon atoms are being followed (Sriram and Shanks, 2002; van Winden et al., 2002). Bondomers are molecules of the same metabolite, which have different bond integrities for different carbon–carbon bonds (Sriram and Shanks, 2004). Bondomer analysis is advantageous in plant tissue cultures that require only sucrose or glucose as a carbon source.

4 INSIGHTS FROM MFA INTO PLANT METABOLISM

Metabolic fluxes form the most important link in translating transcript and metabolite information to the existing physiology (Sauer, 2004). Flux

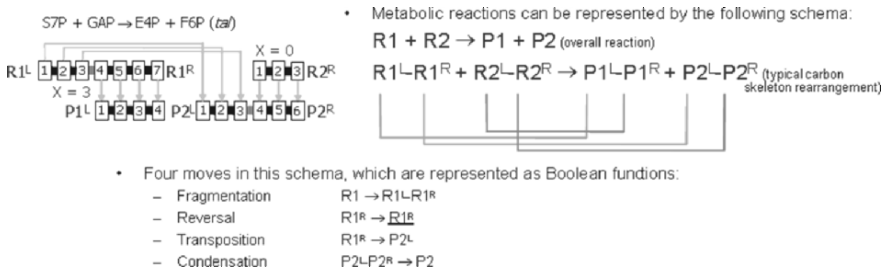


Figure 9-7. Boolean function mapping.

ratios can be used to analyze different nodes in the reaction network where there is a partitioning of the flux into multiple branches. The node under consideration can be either rigid or flexible. A node is said to be "flexible" if the ratio of the carbon flow into multiple reactions changes with a change in the incoming flux. In the case of flexible nodes, the distribution of the precursor metabolite can be modified inherently by the system without the need of any major genetic modification. For a "rigid" node, the ratio of the carbon flow into multiple branches remains the same irrespective of changes to the total flux coming into the node. Subsequently, genetic modifications will prove more effective in altering the metabolic flow in the desired direction in the case of rigid nodes (Stephanopoulos and Vallino, 1991). Network topology is less understood in plants as compared to microbes and the application of MFA can help elucidate plant reaction pathways. Examples of application of flux analysis in revealing network topology has been discussed below.

4.1 Segregation of pathways

As mentioned before, plant metabolism is compartmented, and features multiple copies of the same reaction of a pathway in different subcellular compartments. A classic example is the glycolysis/pentose phosphate pathway subnetwork, which exists in both the cytosol and the plastid in plant cells. In our flux analysis, it was therefore, critical to determine if these pathways were in equilibrium (i.e., they exchanged metabolites so rapidly that for all practical purposes, they could be considered one consolidated pathway) or were segregated (the fluxes through the cytosolic and plastidic pathways are significantly different, and the pathways did not rapidly exchange metabolites).

The segregation or equilibration of cytosolic and plastidic pathways can be ascertained by examining isotopomer abundances or ^{13}C enrichments of

metabolites synthesized in those compartments. Previously, Krook et al. have also reported significantly different ^{13}C enrichments of cytosolic and plastidic hexose pools in *Daucus carota* cells (Krook et al., 1998), which showed that cytosolic and plastidic pathways were segregated. However, hexose phosphate pools were found to be in equilibrium in tomato cells (Rontein et al., 2002) and *B. napus* embryos (Schwender et al., 2003), which showed that the cytosolic and plastidic pathways were in equilibrium.

In our work, on comparing the isotopomer abundances of the carbon atoms of glucosyl units from protein hydrolysate (which are derived from the cytosolic glucose-6-phosphate pool ($G6P^c$) and starch hydrolysate (which are derived from the plastidic glucose-6-phosphate pool, ($G6P^p$) from soybean embryos, we found that they were significantly different, and not in equilibrium (Sriram et al., 2004). In contrast, we did find that the isotopomer abundances of Ala α and Phe α in soybean embryos were similar. Phe α is obtained from the second carbon atom of PEP and is exclusively synthesized in the plastid; whereas Ala α is synthesized in all three compartments, cytosol, plastid, and mitochondrion, respectively from the second carbon atom of pyruvate. From biochemistry, we know that the three carbon atoms of PEP translate to the three carbon atoms of pyruvate without any rearrangement. This result indicated that the T3P pools in the plastid and cytosol were exchanging rapidly between the two compartments (Sriram et al., 2004).

To account for the above observations, we developed a compartmented model of the metabolic network, with separate glycolysis and pentose phosphate pathways in the cytosol and plastid. This model, when used in conjunction with NMR2Flux, was able to explain the observed isotopomer abundances well (see Sriram et al., 2004). Additionally, a fructose-1,6-bisphosphatase reaction had to be included in the plastid to fully account for the experimental isotopomer abundances (see below). Thus, labeling-based flux analysis is competent in segregating pathways in multiple compartments thereby accounting for complex compartmentation inherent in plant systems.

4.2 Identification of new pathways

In the case of the pyruvate family of amino acids the δ^1 carbon of Leu, the β carbon of alanine and γ^1 carbon of Val reflect the same carbon atom of pyruvate respectively (Szyperski, 1995). Hence, the multiplet intensities should be similar for these carbon atoms in the above-mentioned amino acids. However, our recent soybean work (Sriram et al., 2004) indicated that the δ^1 carbon of Leu shows a 30% difference from Ala and Val. This disparity in the isotopomer abundances of the Pyr family of amino acids has been observed in our labeling experiments on another plant system, *Catharanthus roseus* hairy roots as well (Sriram, G., and Shanks J. V.,

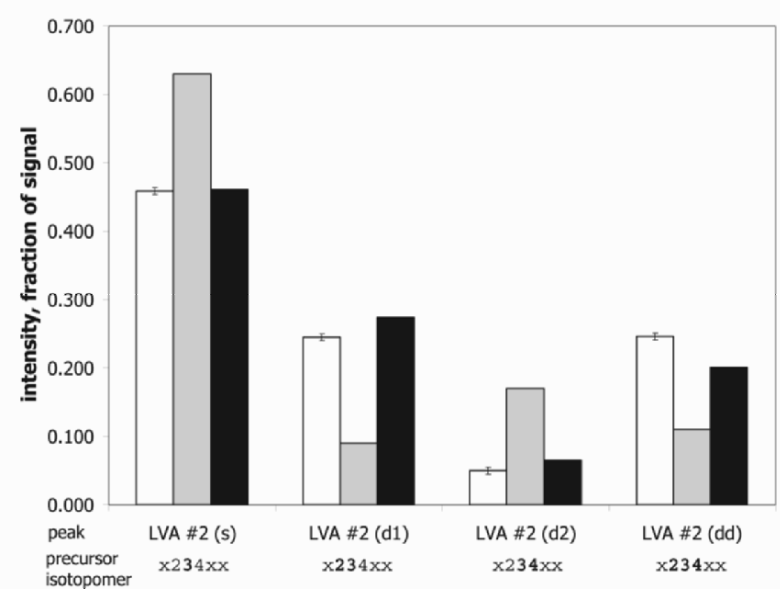


Figure 9-8. Identification of fructose-1, 6-bisphosphatase in the plastid. White bars are the experimental isotopomer abundances of levulinic acid atom 2 (LVA #2) from starch hydrolysate. This atom reflects the isotopomer abundances around carbon #3 of plastidic glucose-6-phosphate. Grey bars are simulated isotopomer abundances from a compartmented model with glycolysis and pentose phosphate pathways in the cytosol and plastid that included no plastidic fructose-1, 6-bisphosphatase. Black bars are from a similar compartmented model that included plastidic fructose-1, 6-bisphosphatase.

unpublished data). Its cause still remains a mystery and we believe that it may involve a currently unknown reaction or pathway related to Leu metabolism.

In addition, we identified the fructose-1,6-bisphosphatase (F16BP) reaction, which converts T3P to F6P, in the plastid. Although acompartmented model with separate glycolysis and pentose phosphate pathways in the cytosol and plastid accounted for most of the isotopomer abundances of the glucosyl units from protein and starch hydrolysates, we found that the isotopomer abundances around levulinic acid atom 2 (LVA #2) were not accounted for (compare white and grey bars in Figure 9-8). LVA #2 is derived from atom 3 of plastidic glucose-6-phosphate pool ($G6P^p$), and the above observation hinted that some pathway or reaction that significantly affects atom 3 of $G6P^p$ was absent in our initial compartmented model. This led us to hypothesize that a significant flux from T3P to F6P may be present in our system. Since such a reaction would cause two three-carbon T3P molecules to form a six-carbon F6P (and eventually a G6P) molecule in the plastid, it may result in isotopomer patterns different from those resulting

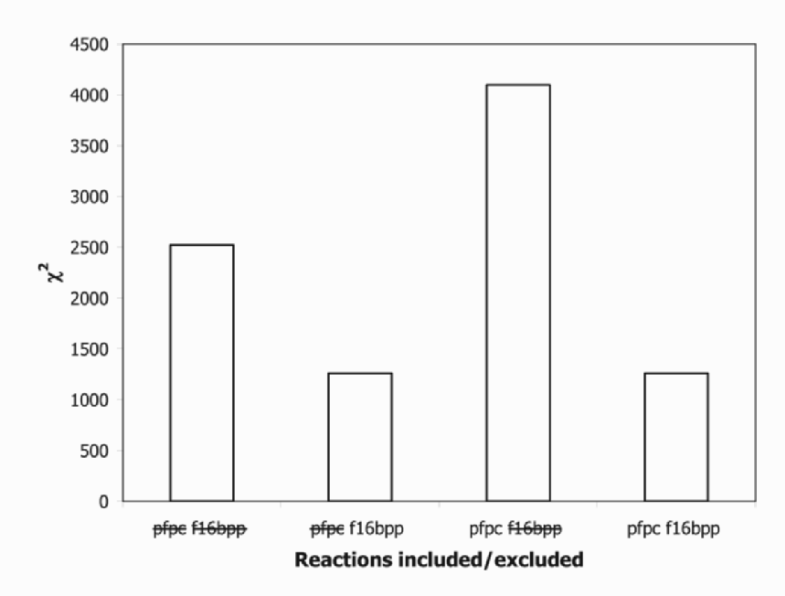


Figure 9-9. Identification of fructose-1,6-bisphosphatase in the plastid: Effect of the inclusion of plastidic fructose-1,6-bisphosphatase (f16bpp, catalysing F6P → T3P in the plastid), and corresponding cytosolic enzyme, pyrophosphatase (pfpc, catalysing F6P → T3P in the cytosol), on the χ^2 error.

due to the presence of the pentose phosphate pathway alone. Such a reaction is usually not present in nonphotosynthetic plant tissues. However, on including this reaction into our compartmented model, we found that observed isotopomer abundances for LVA #2 were well-accounted for (compare white and black bars in Figure 9-8).

Figure 9-9 depicts the improvement in the χ^2 error between experimental and simulated isotopomer abundances, due to the inclusion of plastidic F16BP and the corresponding cytosolic reaction, pyrophosphatase (pfpc). Only the plastidic conversion of T3P is evident in our system, and the cytosolic flux may be small or negligible as it does not significantly improve the χ^2 error. The example of fructose-1,6-bisphosphatase illustrates a systematic approach to pathway identifiability. More recently, Schwender et al. characterized the role of Rubisco in the absence of Calvin cycle (Schwender et al., 2004). They found that Rubisco improves the carbon efficiency during the formation of oil synthesis in *Brassica napus* embryos using an alternative pathway.

5 SUMMARY

The flux evaluation methodology described in this chapter is a promising powerful tool for understanding plant physiology. We expect that the generic computer program NMR2Flux (Sriram et al., 2004) available for calculating fluxes from the labeling data encourages the applicability of flux analysis in plants. Furthermore, once the methodology is established for a particular plant system, the tool can be used to compare the plants environmental and genetic variants. Currently, flux analysis of both environmental and genetic variants of plants is in progress in our laboratory. The ability of the labeling method to establish key regulatory nodes of metabolism thereby enabling identification of potential targets for genetic manipulations makes MFA important from a metabolic engineering perspective (Stephanopoulos and Vallino, 1991). Quantification of fluxes thus is an important tool, which when complemented with metabolite, transcript, and genomic data can contribute toward an overall correct picture of plant physiology (Sanford et al., 2002; Sauer, 2004; Schwender et al., 2004).

REFERENCES

Braunschweiler, L., and Ernst, R. R., 1983, Correlation transfer by isotropic mixing: Application to proton correlation spectroscopy, *Journal of Magnetic Resonance,* **53**: 521-528.

Christensen, B., and Nielsen, J., 1999, Isotopomer analysis using GC-MS, *Metabolic Engineering* **1**:282-290.

Dey, P.M. and Harborne, J.B., 1997, *Plant Biochemistry*, Academic Press, San Diego, CA.

Dieuaide-Noubhani, M., Raffard, G., Canioni, P., Pradet, A., and Raymond, P., 1995, Quantification of compartmented metabolic fluxes in maize root tips using isotope distribution from ^{13}C or ^{14}C-labeled glucose, *Journal of Biological Chemistry* **270**:13147-13159.

Girke, T., Ozkan, M., Carter, D., and Raikhel, N. V., 2003, Towards a modeling infrastructure for studying plant cells, *Plant Physiology* **132**:410-414.

Glawschnig, E., Gierl, A., Tomas, A., Bacher, A., and Eisenreich, W., 2002, Starch biosynthesis and intermediary metabolism in maize kernels. Quantitative analysis of metabolite flux by nuclear magnetic resonance, *Plant Physiology* **130**:1717-1727.

Johnson, B. A., and Blevins, R. A., 1994, NMRView: a computer program for the visualization and analysis of NMR data, *Journal of Biomolecular NMR* **4**:603-614.

Klapa, M. I., Park, S. M., Sinskey, A. J., and Stephanopoulos, G., 1999, Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. Theory, *Biotechnology and Bioengineering* **62**(4):375-391.

Krook, J., Vreugdenhill, D., Dijkema, C., and van der Plas, L. H. W., 1998, Sucrose and starch metabolism in carrot (*Daucus carota*) cell suspensions analyzed by ^{13}C labeling: indications for a cytosol and a plastid-localized oxidative pentose phosphate pathway, *Journal of Experimental Botany* **49**:1917-1924.

Marx, A., De Graaf, A. A., Wolfgang, W., Eggeling, L., and Sahm, H., 1996, Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear

magnetic resonance spectroscopy combined with metabolite balancing, *Biotechnology and Bioengineering* **49**:111-129.

Minorsky, P. V., 2003, Achieving the in silico plant. Systems biology and the future of plant biological research, *Plant Physiology* **132**:404-409.

Mullin, W. J., and Xu, W., 2000, A study of the intervarietal differences of cotyledon and seed coat carbohydrates in soybeans, *Food Research International* **33**:883-891.

Nielsen, J., 1998, Metabolic engineering: Techniques for analysis of targets for genetic manipulations, *Biotechnology and Bioengineering* **58**:125-132.

Park, S. M., Klapa, M. I., Sinskey, A. J., and Stephanopoulos, G., 1999, Metabolite and isotopomer balancing in the analysis of metabolic cycles: II. Applications, *Biotechnology and Bioengineering* **62**(4):392-401.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vellerling, W. T., 1992, *Numerical recipes in C* (Second edn.): Cambridge University Press.

Ratcliffe, R. G., and Shachar-Hill, Y., 2001, Probing plant metabolism with NMR, *Annual Review of Plant Physiology and Plant Molecular Biology* **52**:499-526.

Rontein, D., Dieuaide-Noubhani, M., Dufourc, E. J., Raymond, P., and Rolin, D., 2002, The metabolic architecture of plant cells, *Journal of Biological Chemistry* **277**(46):43948–43960.

Sanford, K., Soucaille, P., Whited, G., and Chotani, G., 2002, Genomics to fluxomics and physiomics – pathway engineering, *Current Opinion in Microbiology* **5**(3):318-322.

Sauer, U., 2004, High-throughput phenomics: Experimental methods for mapping fluxomes, *Current Opinion in Biotechnology* **15**:58-63.

Schmidt, K., Carlsen, M., Nielsen, J., and Villadsen, J. (1997). Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnology and Bioengineering,* **55**(6), 831-840.

Schmidt, K., Norregaard, L. C., Pedersen, B., Meisnner, A., Duus, J. O., Nielsen, J., et al. 1999, Quantification of intracellular metabolic fluxes from fractional enrichment ^{13}C-^{13}C coupling constraints on the isotopomer distribution in labeled biomass components, *Metabolic Engineering* **1**:166-179.

Schwender, J., Goffman, F., Ohlrogge, J., and Shachar-Hill, Y., 2004, Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds, *Nature* **432**:779-782.

Schwender, J., Ohlrogge, J., and Shachar-Hill, Y., 2003, A flux model of glycolysis and oxidative pentose phosphate pathway in developing *Brassica napus* embryos, *Journal of Biological Chemistry* **278**(32):29442-29453.

Schwender, J., Ohlrogge, J., and Shachar-Hill, Y., 2004, Understanding flux in plant metabolic networks, *Current Opinion in Plant Biology* **7**(3):309-317.

Shanks, J. V., 2000, *In Situ* NMR systems. in: *NMR in microbiology: Theory and applications,* J. N. Barbotin & J. C. Portais eds., Horizon scientific press, pp. 49-75.

Sriram, G., Fulton, D. B., Iyer, V. V., Peterson, J. M., Zhou, R., Westgate, M. E., et al., 2004, Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing, *Plant Physiology* **136**:3043-3057.

Sriram, G. and Shanks, J.V., 2002, *A mathematical model for carbon bond labeling experiments: Analytical solutions and sensitivity analysis for the effect of reaction reversibilities on estimated fluxes.* Paper presented at the Biochemical Engineering Symposium, Kansas city.

Sriram, G. and Shanks, J.V., 2004, Improvements in metabolic flux analysis using carbon labeling experiments: bondomer balancing and boolean function mapping, *Metabolic Engineering* **6:**116-132.

Stephanopoulos, G., 1999, Metabolic fluxes and metabolic engineering, *Metabolic Engineering* **1**:1-11.

Stephanopoulos, G., Nielsen, J., and Aristidou, A. A., 1998, *Metabolic engineering: Principles and methodologies* (First ed.): Elsevier Science & Technology Books.

Stephanopoulos, G., Vallino, J. J., 1991, Network rigidity and metabolic engineering in metabolite overproduction, *Science* **252**(5013):1675-1681.

Stitt, M., and Fernie, A. R., 2003, From measurements of metabolites to metabolomics: An 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions, *Current Opinion in Biotechnology* **14**:136-144.

Sweetlove, L. J., Last, R. L., and Fernie, A. R., 2003, Predictive metabolic engineering: A goal for systems biology, *Plant Physiology* **132**:420-425.

Szyperski, T., 1995, Biosynthetically directed fractional ^{13}C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism, *European Journal of Biochemistry* **232**:433-448.

Szyperski, T., 1998, ^{13}C-NMR, MS and metabolic flux balancing in biotechnology research, *Quarterly Reviews of Biophysics* **31**(1):41-106.

van Winden, W., Heijnen, J., and Verheijen, P., 2002, Cumulative bondomers: A new concept in flux analysis from 2D ^{13}C, ^{1}H COSY NMR data, *Biotechnology and Bioengineering* **80**:731-745.

van Winden, W., Schipper, D., Verheijen, P., and Heijnen, J., 2001, Innovations in generation and analysis of 2D [^{13}C, ^{1}H] COSY NMR spectra for metabolic flux analysis purposes, *Metabolic Engineering* **3**:322-343.

Varma, A., and Palsson, B. O., 1994, Metabolic flux balancing: Basic concepts, scientific and practical use, *Bio/technology* **12**:994-998.

Wiechert, W., 2001, Minireview: ^{13}C Metabolic flux analysis, *Metabolic Engineering* **3**:195-206.

Wiechert, W., and De Graaf, A. A., 1997a, Bidirectional reaction steps in metabolic networks: I Modeling and simulation of carbon isotope labeling experiments, *Biotechnology and Bioengineering* **55**:101-117.

Wiechert, W., Mollney, M., Isermann, N., Wurzel, M., and De Graaf, A. A., 1999, Bidirectional reaction steps in metabolic networks:III Explicit solution and analysis of isotopomer labeling systems, *Biotechnology and Bioengineering* **66**:69-85.

Wiechert, W., Mollney, M., Petersen, S., and De Graaf, A. A., 2001, A universal framework for ^{13}C metabolic flux analysis, *Metabolic Engineering* **3**:265-283.

Wiechert, W., and Wurzel, M., 2001, Metabolic isotopomer labeling systems. Part I: Global dynamic behavior, *Mathematical Biosciences* **169**:173-205.

Zupke, C., and Stephanopoulos, G., 1994, Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices, *Biotechnology Progress* **10:**489-498.

Chapter 10

METNET: SYSTEMS BIOLOGY TOOLS FOR ARABIDOPSIS

Eve Syrkin Wurtele[1], Ling Li[1], Dan Berleant[3], Dianne Cook[2], Julie A. Dickerson[3], Jing Ding[3], Heike Hofmann[2], Michael Lawrence[2], Eun-kyung Lee[2], Jie Li[1], Wieslawa Mentzen[1], Leslie Miller[4], Basil J. Nikolau[5], Nick Ransom[1], and Yingjun Wang[1]

[1]Departments of Genetics, Development and Cell Biology; [2]Statistics; [3]Electrical and Computer Engineering; [4]Computer Science; [5]Biochemistry Biophysics and Molecular Biology, Iowa State University, Ames, IA, USA

Abstract: MetNet (http://metnetdb.org) is an emerging open-source software platform for exploration of disparate experimental data types and regulatory and metabolic networks in the context of Arabidopsis systems biology. The MetNet platform features graph visualization, interactive displays, graph theoretic computations for determining biological distances, a unique multivariate display and statistical analysis tool, graph modeling using the open source statistical analysis language, R, and versatile text mining. The use of these tools is illustrated with data from the *bio1* mutant of Arabidopsis.

1 INTRODUCTION

Plant composition, form, and function are the ultimate consequences of gene expression. High-throughput detection and measurement of changes in the accumulation of tens of thousands of cellular components – RNAs, proteins, and metabolites, and metabolic flux information, lead to complex, valuable data sets (Oliver et al., 2002; Sriram et al., 2004; Fernie et al., 2005; Nikiforova et al., 2005). Each data set has the potential to contribute to our understanding of cellular function, and combined experimental data sets impart an added potential to understand and predict the behaviour of a cell. Comparative analysis of mRNA and proteins can provide insights into the processes that affect mRNA accumulation (gene transcription and/or mRNA stability) and protein accumulation (mRNA translation and/or protein stability), but do not give direct information on metabolism. Metabolite profiling gives information about the accumulation of metabolites, but does

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 145–157.

not reveal which pathways produced those metabolites; however, in combination with microarray and proteomics pathways may be surmised. Techniques for metabolomic flux analysis in plants are becoming more sophisticated (Sriram et al., 2004; Ratcliffe and Shachar-Hill, 2005), and these data can contribute information on the flow through specific metabolic pathways and when combined with "omics" data can provide clues about regulatory mechanisms. Other data sets for plants that could provide additional information for analysis of cellular systems, such as protein-protein or protein-DNA interactions, are on the horizon.

Due to the complexity of each data set, a human mind cannot comprehend data of a single type, let alone the data sets *en toto*. Also, the data sets are flawed. Even for the model plant species Arabidopsis, the majority of genes are not yet well annotated, and current technologies to identify metabolites and proteins yield incomplete data sets. Furthermore, most interactions between the biomolecules, as well as most of the kinetics of the established interactions, are not yet known. Even given the availability of comprehensive "omics" data sets, and a full understanding of the interactions and kinetics of a cell, there are not yet modeling methods capable of predicting the behaviour of such a complex system (Du et al., 2005; Ma'ayan et al., 2005; Lee et al., 2005; Xiong et al., 2004).

Thus, the challenge in prediction of a biological network is complex, and requires consideration of a variety of factors: (1) How to represent a biological network; (2) How to evaluate data sets that have only part of their constituents determined and a subset of the possible interactions elucidated; (3) How to model processes that have wide-ranging kinetics parameters, most of which are not yet determined.

MetNet is being designed to provide an integrated, open-source platform to develop hypotheses about which genes and proteins might be involved in a process, which pathways and interactions might be important under particular conditions, and ultimately how the biological system functions. We discuss MetNet, and illustrate its use with data from an experiment designed to analyse the biotin metabolic network. Biotin is required as a cofactor by all living organisms. It is synthesised almost exclusively by photosynthetic organisms, is an essential cofactor for several key enzymes in plants (Nikolau et al., 2003). It is also a potential metabolic regulator (Che et al., 2002, 2003). Understanding the multiple functions of this metabolite presents a formidable challenge in systems biology.

2 RESULTS

2.1 MetNetDB contains an integrated metabolic and regulatory map of Arabidopsis interactions

The MetNetDB database contains a repository of curated expert-created regulatory and metabolic pathways, as well as processed information from repositories of metabolic-only pathways for Arabidopsis: AraCyc (Mueller et al., 2003), and in the near future, BioPathAt (Lange and Ghassemian, 2005), and MapMan (Thimm et al., 2004). Expansion of the MetNetDB database is ongoing. Biomolecules that can be represented in MetNetDB include metabolites, genes, RNAs, polypeptides, and protein complexes; interactions that can be represented include catalysis, conversion, transport, and a wide variety of regulatory interactions (e.g., allosteric inhibition, transcriptional inhibition, and covalent modification). Because the concentration of each biomolecule, as well as the interactions it is able to participate in, vary across subcellular compartments, MetNetDB includes subcellular location information. Thus, multiple entries are permitted for each biomolecule (e.g., a metabolite can participate in more than one reaction, and can be located in more than one subcellular compartment). The MetNetDB curator interface is designed for curation of biomolecules, interactions, and associated information about subcellular location, synonyms, and references. The interface includes a simple graphic representation of the pathways in which biological interactions and complexes can be viewed, created, or modified.

The network is stored in a MYSQL (www.mysql.org) relational database. We have constructed an XML file format that accurately encodes the network topology information from MetNetDB. The network itself is designed for analysis with experimental data, using tools such as MetNet in Cytoscape and ExploRase, which currently receive network information in XML format. A versatile XML file-builder (http://metnetdb.gdcb.iastate.-edu/MetNet/MapBuilder.html) can be used to export current data from the MetNetDB database network.

2.2 Statistical and visualization software tools

ExploRase provides a multivariate approach to detect patterns in gene expression, and to explore connections between “omics” data sets and the known and hypothesized regulatory and metabolic network of Arabidopsis. ExploRase is built on the open-source statistical analysis software R (*http://*www.R-project.org), and the open-source data visualization software, GGobi (*http://www.ggobi.org*), and includes a user-friendly interface for both. ExploRase also adds a spreadsheet with TAIR annotations about each

gene, links to literature, menus of analysis and visualization options, and an interface to lists of genes in MetNetDB pathways. Common statistical analyses are provided through GUIs. Alternatively, code for new functionality can be written using R commands. Thus, the GUIs in ExploRase make the R functionality transparent for the novice, but allow a more advanced user to do more sophisticated analysis.

ExploRase has a highly interactive graphics system, designed specifically for exploratory mining of high-dimensional data. It has multivariate graphics including parallel coordinate plots and tours (rotations of high-dimensional scatterclouds). Users can label elements of the plots by clicking on genes, proteins, and/or metabolites of interest. Metabolic and regulatory networks can be displayed using the add-on package GGVis. Users can layout a network (in 2, 3 or higher dimensions), or read in a layout from another package such as MetNet in Cytoscape.

To elucidate the biotin network of plants from a systems biology viewpoint, we have been analysing mutants blocked, overexpressed, or underexpressed in steps of this network. One such step is encoded by the *BIO1* gene, which encodes 7,8-diaminopelargonic acid aminotransferase, the third step in the synthesis of biotin from pimelic acid (Patton et al., 1996). A homozygous mutation in *BIO1* is lethal without addition of exogenous biotin; however, the seedlings appear normal for several days, due to a residue of biotin originally supplied to the parent plants (Weaver et al., 1996; Patton et al., 1996; Che et al., 2003).

One aspect of ExploRase is illustrated with an example of microarray data from a portion of a larger experiment (Figure 10-1). In this experiment, seeds of homozygous mutants for the *bio1* gene are grown in medium with and without biotin. The upper part of Figure 10-1A shows a dialog window with information on each mRNA. This includes Affy8k ID, Locus ID, TAIR annotations and other descriptions. On the left, is a list of available chips (eg., biotin.WT.02, biotin.WT.B1, biotin.WT.B2). Prior to this visualization, the chips were normalized using a quantiles' normalization and a robust median is used for the expression value. Below the dialogs are two plots: a scatterplot and a parallel coordinate plot. The scatter plot shows a comparison of transcript accumulation between two replicates of WT seeding grown without biotin. The two RNAs marked by the user in yellow, both chloroplast encoded transcripts, are seen to accumulate at a much higher level in the second replication than in the first. The yellow highlight marking is automatically shown on the parallel coordinate plot on the right, and on the annotation list. Both genes exhibit a similar pattern: they are expressed at a high but fairly stable level for all of the chips except for biotin.WT.02 biotin.WT.02, indicating, that a data error might have occurred on the chip biotin.WT.02.

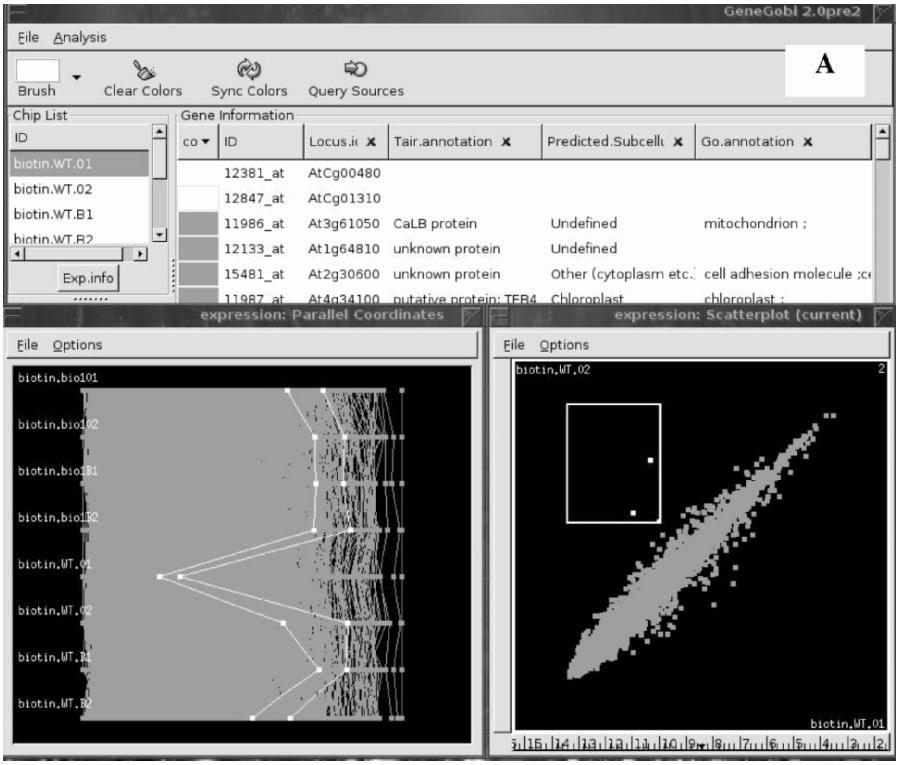


Figure 10-1. Visualization of microarray data by ExploRase. (A) Examining data quality. A scatter plot analysis of data from microarray replicates of two biological samples quickly reveals a problem with values for two plastid-encoded genes inherent in one of the replicates; these RNAs are simultaneously highlighted in a parallel coordinate plot view of the data. The raw data had previously been normalized, using R functions in ExploRase (not shown). (B) Differentially expressed genes. Scatterplots show data from the *bio1* mutant with added biotin compared to without added biotin. Several genes appear differentially expressed in the scatter plot, and were selected by the user (blue or pink highlights); the corresponding parts of the parallel coordinate plot are simultaneously highlighted; the annotation for these genes is also displayed. Results of statistical analyses can also be superimposed on the visualization results (not shown). The y-axis of the parallel coordinate plot has a log scale.

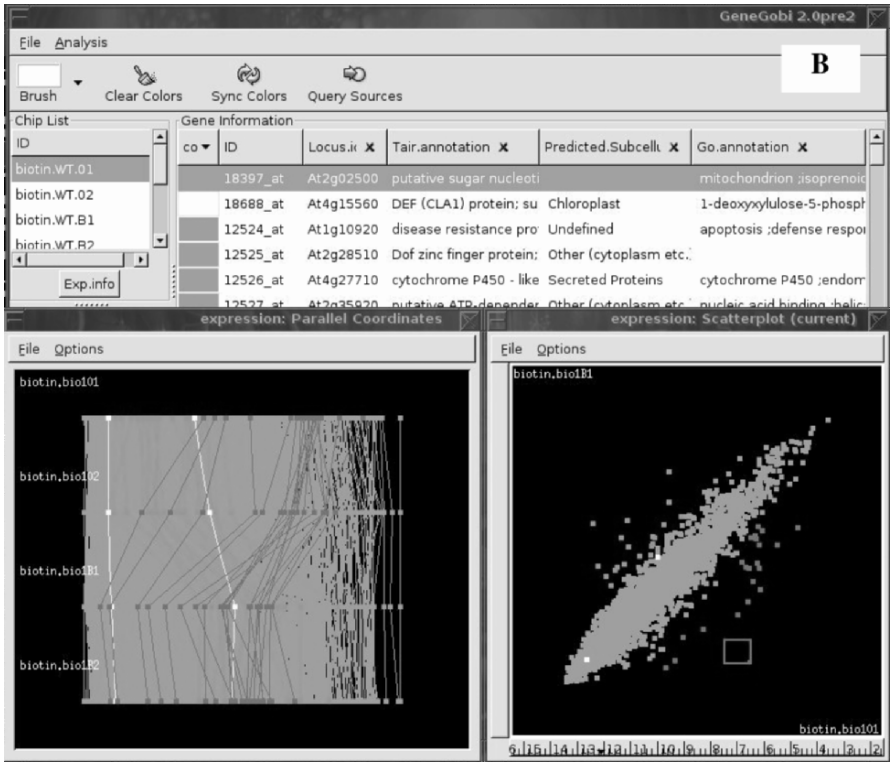


Figure 10-1(continued).

To identify patterns of co-accumulation of RNAs associated with biotin, the user selected a parallel coordinate plot, gene annotation list, and a scatter plot, and displayed data from the *bio1* genotype grown with and without biotin (Figure 10-1B). There are numerous RNAs, visible in the scatter plot, that accumulate to similar levels. Clicking on outliers in the scatter plot (those accumulating at higher levels in the *bio1* mutant with added biotin were colored in blue; those with decreased accumulation when biotin is absent are colored orange) links to the same genes in the parallel coordinate plot and the highlighted genes are also displayed in the gene annotation list. For comparison, two genes that are not outliers in the scatter plot were highlighted in yellow. Using this approach identified At2g02500 (encoding 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (ISPD)) and At4g15560 (encoding putative 1-deoxy-D-xylulose 5-phosphate synthase (DXPS), both genes of isoprenoid synthesis. At2g02500 and At4g15560 were upregulated 7 and 2-fold, respectively in the *bio1* mutant plus biotin as compared to the *bio1* mutant without biotin.

2.3 Metabolic network display and modeling (MetNet in Cytoscape)

MetNet in Cytoscape (Figure 10-2A) uses the Cytoscape (http://www.-cytoscape.org/) a Java program, together with plug-ins specialized for MetNet and its database, to dynamically display complex biological networks and analyse their structure (Wurtele et al., 2003; Du et al., 2005). Data from experiments (i.e., microarray, proteomics, or metabolomics) can be directly overlayed on the network. An interface to R allows the user to analyse "omics" data in R, cluster biomolecules that behave similarly, search for biomolecules with significant changes, and to custom-write R scripts and apply them to experimental data.

MetNet in Cytoscape uses graph theoretic methods to display and analyse biological networks, such as those in MetNetDB. Graphs can be visualized by employing the P-neighborhood function around nodes or reactions of interest; in this mode, the user selects any group of biomolecules or pathways in the MetNetDB network, and extends the network in all directions by a user-designated number of steps. Graphs also can be dynamically displayed as pathways and cycles. (A simple cycle could include a gene transcribed to a protein which, when that protein was over-accumulated, would inhibit the gene's transcription.) For example, a user could display the network that includes all genes that are differentially expressed between *bio1* seedlings grown with and without biotin, and find pathways in that network. Different pathways might indicate multiple mechanisms for control of a process. Common steps among pathways may reflect critical paths in the network.

By displaying all pathways containing genes identified as differentially expressed using ExploRase, the user obtained a very complex network (the insert at upper left of in Figure 10-2A shows a portion of this network); the network was pared down so that only steps connecting biotin with the At2g02500 and At4g15560 proteins remained (Figure 10-2A). Both these encode enzymes that are early in the plastidic methylerythritol 4-phosphate (MEP) isoprenoid pathway. Pyruvate is a common substrate for both plastidic fatty acid synthesis and the MEP pathway. Biotin is required for acetyl-CoA carboxylase activity; therefore the carbon needed for the formation of plastidic malonyl-CoA (indirectly from plastidic pyruvate *via* acetyl-CoA) could be limited in the *bio1* mutant when grown without added biotin. A decrease in the flux through the fatty acid biosynthetic pathway due to decreased acetyl-CoA carboxylase activity might influence the flux of pyruvate towards MEP, and provide a signal that alters gene expression in the MEP pathway. This potential interconnection between the fatty acid and MEP pathway could be explored further by experimentation and by modeling (e.g., Du et al., 2005).

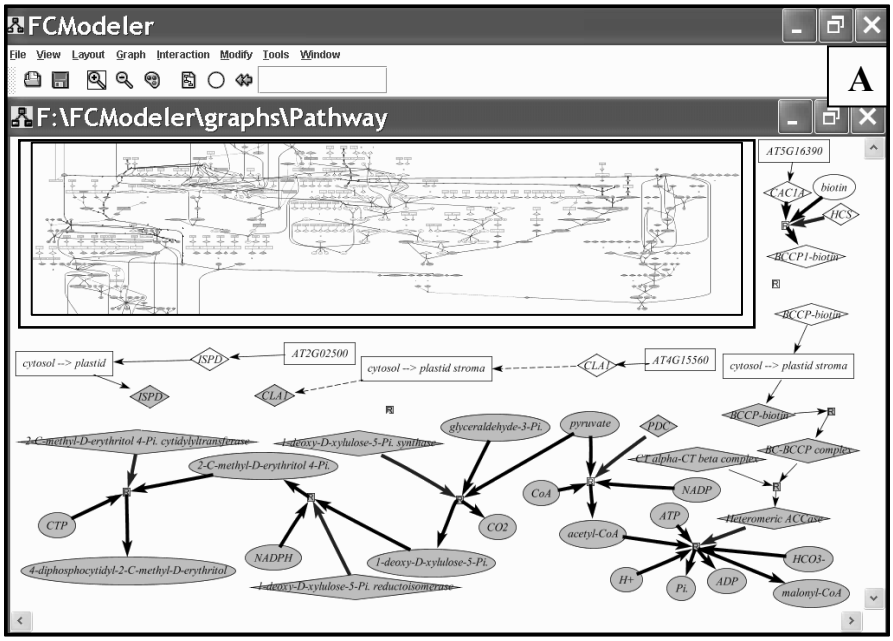


Figure 10-2. Exploration of genes whose expression increases in response to biotin in the *biol* mutants. (A) MetNet in Cytoscape (previously FCM) was used to explore possible interrelationships between a suboptimal level of biotin and changes in gene expression as revealed by global microarray analysis. This example focuses on the increase in accumulation of two RNAs in the methylerythritol phosphate (MEP) pathway of plastidic isoprenoid synthesis. A graph containing pathways of central metabolism, including isoprenoid metabolism and starch metabolism, and a subset of the upregulated genes was selected (insert box in upper left). Clicking within this graph identified a subgraph that includes both biotin and the MEP differentially expressed genes. (B) MetaOmGraph was used to determine the Pearson correlations across 1000 Arabidopsis ATH1 chips in the NASCArrays database, comparing the expression pattern of At2g02500 to that of the other 22,746 genes on the chip. The genes most similar to At2g02500 are At5g45930 and At1g32990 (87% and 86% correlation, respectively). Both of these genes are involved in plastid function. At5g45930 encodes a magnesium-chelatase subunit, ChlI, which is required for biosynthesis of the isoprenoid-porferin hybrid molecule, chlorophyll. (C) PathBinder was used to explore interconnections in the literature between biotin and isoprenoids.

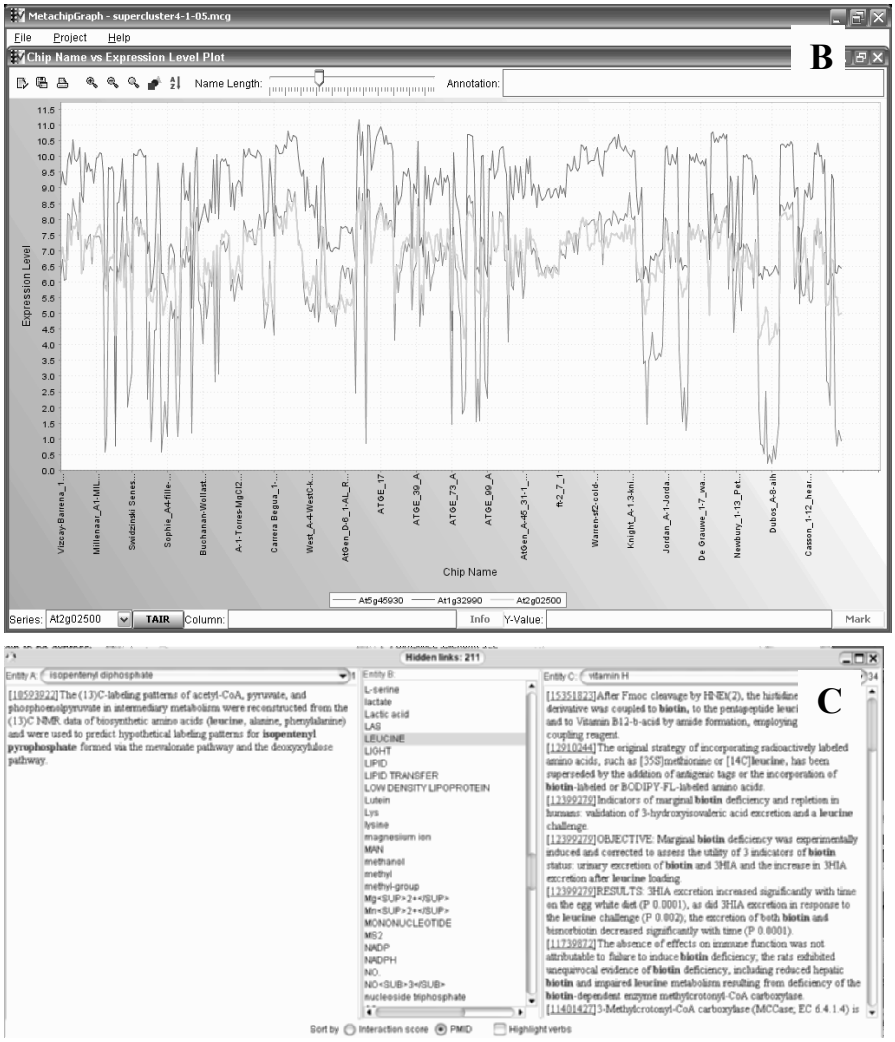


Figure 10-2 (continued).

2.4 MetaOmGraph

MetaOmGraph is a JAVA program designed to analyse co-expressed genes across large data sets. Unlike other analysis programs, which store all the data in memory all the time, MetaOmGraph uses the RandomAccessFile

class to read and store data only as it's needed. This allows the program to work with extremely large sets of data while requiring relatively little memory. The program comes with a set of Arabidopsis data (both experimental data and metadata) from NASCArrays (http://arabidopsis.info/) that we have selected as being high quality and normalized. It is also simple to analyse a microarray data set from any species (or indeed any other type of data set) using MetaOmGraph. Graphs can be sorted according to the expression value and metadata information.

MetaOmGraph was used to determine the Pearson correlations of the differentially expressed genes At2g02500 and At4g15560, across 1000 chips from the NASCArrays database. At2g02500 and At4g15560 have a 63% correlation with each other across all the chips (not shown). This corresponds to a p-value below 1.4e-45. Among the 22,746 genes on the Affymetrix ATH1chip, the most similar expression profiles to that of At2g02500 are those of At5g45930 and At1g32990 (87% and 86% correlation respectively) (Figure 10-2B). At5g45930 encodes a magnesium-chelatase subunit, ChlI, which is required for chlorophyll biosynthesis. At1g32990 encodes plastidic ribosomal protein L11. These results suggest a possible relationship between the plastidic biosynthetic processes of isoprenoid synthesis and photosynthesis.

2.5 Text mining

PathBinder(http://www.public.iastate.edu/~mash/MetNet/MetNet_PathBinder.htm) is a text-mining tool designed specifically to explore metabolic and regulatory interactions in plants. The tool queries the MEDLINE database and retrieves sentences that contain two terms of interest; each sentence is a clickable link pointing to the original online PubMed citation. PathBinder contains an extensive set of synonyms from MetNetDB, many tailored to plant biology, which are co-searched when a term is selected. An API (applications programming interface) is provided so that the PathBinder text-mining tool can be integrated into other analysis tools. The API has been used to incorporate PathBinder into the MetNetDB database, and in the future could automatically extract references for interactions, which could then be manually curated. We have created a novel "hidden links" tool to identify and explore potential intermediate links in networks. Given biomolecules A and C that do not co-occur in any sentence, the tool will find biomolecules B that co-occur in sentences both with A and with C. In Figure 10-2C, the user explored possible literature connections between isoprenoids and biotin. These two terms did not co-occur in any sentence. The user chose isopentenyl diphosphate as biomolecule A, and biotin [an automatically-selected synonym was vitamin H] as biomolecule C, and selected the Hidden Links algorithm. Two hundred and eleven biomolecule B terms

co-occurred independently in sentences with both isopentenyl diphosphate and with biotin. The user clicked on LEUCINE; a single sentence containing isopentenyl diphosphate and LEUCINE, and 34 sentences containing biotin and LEUCINE were retrieved. Here, a second possible connection between biotin and isoprenoids is suggested, as leucine catabolism requires the biotin-enzyme methylcrotonyl-CoA carboxylase.

2.6 Major venues for improvement and expansion of the MetNet platform

Expansion of the MetNet platform is in progress. The software tools will be further integrated such that the users will be able to access all tools from a single platform. A node- and edge-labeled graph model for the database will be implemented. This model will address a major database challenge: tracking changes in biological network data, as such data are being continuously revised and expanded. By broadening the current MetNetDB relational database to a node- and edge-labeled graph model format, information about the date and person (or web source) for each data entry must be captured, to track data revisions and new biological knowledge, as well as to provide automated methods for addition of large-scale data-dumps from online resources such as AraCyc. This model would enable addition of several features not present in other databases. In particular it would provide a flexible method for tracking changes. Such a model would also enable researchers to create their own version of networks to test, model and compare with other networks. In addition, the database would be able to model, as well as store, the data.

MetNet can be modified for analysis of species other than Arabidopsis; in particular, we are beginning to expand the MetNet platform to soybean.

3 CONCLUSION

The MetNet platform is designed for the exploration of diverse data sets, and formulation of hypotheses based on this data in the context of known Arabidopsis regulatory and metabolic interactions.

REFERENCES

Che, P., Weaver, L.M., Wurtele, E.S., and Nikolau, B.J., 2003, The role of biotin in regulating 3-methylcrotonyl-CoA carboxylase expression in Arabidopsis, *Plant Physiol*. **131**:1479–1486.

Che, P., Wurtele, E.S., and Nikolau, B.J., 2002, Metabolic and Environmental Regulation of 3-Methylcrotonyl-CoA Carboxylase Expression in Arabidopsis, *Plant Physiol.* **129**:625-637.

Cook, D., Hofmann, H., Lee, E-K., Yang, H., Nikolau, B., Wurtele, E.S., 2004, Visual Methods for Data from Two Factor Single Replicate Gene Expression Studies, Technical Report, Department of Statistics, Iowa State University 04-06: http://www.stat.iastate.edu/preprint/articles/2004-04.pdf.

Dickerson, J.A., Berleant, D.A., Du, P., Ding, J., Foster, C.M., Ling, L., and Wurtele, E.S., 2005, Creating modeling and visualizing metabolic networks, in: *Medical Informatics*, H. Chen, H., S.S. Fuller, S.S., C. Friedman, C., and W. Hirsch, W., eds., Springer Verlag, pp. 491-518.

Ding, J., Berleant, D., Nettleton, D.E., and Wurtele, E.S., 2002, Mining medline: abstracts, sentences, or phrases? *Pacific Symposium on Biocomputing*, 1-12.

Du, P., Gong, J., Wurtele, E.S., Dickerson, J.A., 2005, Modeling Gene Expression Networks using Fuzzy Logic, *IEEE Trans. On Systems, Man and Cybernetics*, Part B **35**(6): 1351-1359.

Fernie, A.R., Geigenberger, P., and Stitt, M., 2005, Flux an important, but neglected component of functional genomics, *Current Opinion in Plant Biology* **8**:174-182.

Lange, B.M., and Ghassemian, M., 2005, Comprehensive post-genomic data analysis approaches integrating biochemical pathway map, *Phytochemistry* **66**:413-451.

Lee, E-K., Cook, D., Hofmann, H., Wurtele, E., Kim, D., Kim, J., and An, H., 2004, Genegobi: visual data analysis tools for microarray data compstat' 2004 Symposium c Physica-Verlag/Springer.

Lee, S.G., Kim, C.M., Hwang, K.S., 2005, Development of a software tool for in silico simulation of Escherichia coli using a visual programming environment*, J. Biotechnol*. (Epub).

Ma'ayan, A., Blitzer, R.D., and Iyengar, R., 2005, Toward predictive models of mammalian cells, *Annu. Rev. Biophys. Biomol. Struct*. **34**:319-349.

Mueller, L.A., Zhang, P., and Rhee, S.Y., 2003, AraCyc: A Biochemical Pathway Database for Arabidopsis, *Plant Physiology* **132**:453-460.

Nikiforova, V.J., Daub, C.O., Hesse, H., Willmitzer, L., and Hoefgen, R., 2005. Integrative gene-metabolite network with implemented causality deciphers informational fluxes of sulphur stress response, *J. Exp. Bot.* **56**:1887-1896.

Nikolau, B.J., Ohlrogge, J.B., and Wurtele, E.S., 2003, Plant Biotin-Containing Carboxylases, *Arch. Biochem. Biophys*. **414**:211-222.

Oliver, D.J., Nikolau, B., and Wurtele, E.S., 2002, Functional genomics: high-throughput mRNA, protein, and metabolite analyses, *Metab. Eng*. **4**:98106.

Patton, D.A., Volrath, S., and Ward, E.R., 1996, Complementation of an Arabidopsis thaliana biotin auxotroph with an Escherichia coli biotin biosynthetic gene, *Molecular Genetics and Genomics.* **251**:261-266.

Ratcliffe, R.G., and Shachar-Hill, Y., 2005, Revealing metabolic phenotypes in plants: inputs from NMR analysis, *Biol. Rev. Camb. Philos. Soc*. **80**:27-43.

Sriram, G., Fulton, D.B., Iyer, V.V., Peterson, J.M., Zhou, R., Westgate, M.E., Spalding, M.H., and Shanks, J.V., 2004, Quantification of Compartmented Metabolic Fluxes in Developing Soybean Embryos by Employing Biosynthetically Directed Fractional 13C Labeling, Two-Dimensional [13C, 1H] Nuclear Magnetic Resonance, and Comprehensive Isotopomer Balancing, *Plant Physiol*. **136**:3043-3057.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A, Rhee, S.Y., and Stitt, M., 2004, Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes, *The Plant Journal* **37**:914-939.

Weaver, L.M., Yu, F., Wurtele, E.S., and Nikolau, B.J., 1996, Characterization of the cDNA and gene coding for the biotin synthase of Arabidopsis thaliana, *Plant Physiol.* **110**:1021-1028.

Wurtele, E.S., Dickerson, J.D., Cook, D., Hofmann, H., Li, J., and Diao, L.. 2003, MetNet: software to build and model the biogenetic lattice of Arabidopsis, *Comp. Funct. Genomics* **4**:239-245.

Xiong, M., Zhao, J., and Xiong, H., 2004, Network-based regulatory pathways analysis, *Bioinformatics* **20**:2056-2066.

Chapter 11

IDENTIFICATION OF GENES INVOLVED IN ANTHOCYANIN ACCUMULATION BY INTEGRATED ANALYSIS OF METABOLOME AND TRANSCRIPTOME IN *PAP1*-OVEREXPRESSING *ARABIDOPSIS* PLANTS

Takayuki Tohge[1,2], Yasutaka Nishiyama[1,3], Masami Yokota Hirai[1,4], Mitsuru Yano[1], Jun-ichiro Nakajima[1], Motoko Awazuhara[1], Eri Inoue[2], Hideki Takahashi[2], Dayan B. Goodenowe[5], Masahiko Kitayama[3], Masaaki Noji[1], Mami Yamazaki[1] and Kazuki Saito[1,2,4]

[1]Department of Molecular Biology and Biotechnology, Graduate School of Pharmaceutical Sciences, Chiba University; [2]RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan;[3]Institute of Life Science, Ehime Women's College, 421 Ibuki-cho Baba, Uwajima-shi, Ehime, 798-0025, Japan; [4]CREST, JST (Japan Science and Technology Agency), Yayoi-cho 1-33, Inage-ku, Chiba-shi, Chiba 263-8522, Japan; [5]Phenomenome Discoveries Inc., 204-407 Downey Road, Saskatoon, SK S7N 4L8, Canada

Abstract: The *PAP1* gene, which encodes an MYB transcriptional factor, up-regulates the flavonoid biosynthetic gene expression. We studied an integrated analysis of metabolomics and transcriptomics with wild-type, *pap1-D* mutant, and *PAP1*-overexpressing transgenic plant, to elucidate a detailed anthocyanin accumulation mechanism and to identify the novel gene functions involved in flavonoid biosynthesis. The flavonoid-targeted analysis by high-performance liquid chromatography-mass spectrometry, indicated the specific over-accumulation of cyanidin derivatives and quercetin glycosides in *PAP1*-overexpressing leaves. The transcriptome analysis on a DNA microarray revealed the upregulation of 38 genes by ectopic *PAP1*-overexpression. In addition to well-known genes involved in anthocyanin production, several genes with unidentified functions or annotated with putative functions have been upregulated. From the enzymatic activity of their recombinant proteins *in vitro* and the analysis of anthocyanins in the respective T-DNA-inserted mutants, two putative glycosyltransferase genes (At5g17050 and At4g14090) induced by *PAP1*-overexpression were confirmed to encode flavonoid 3-*O*-glucosytransferase and anthocyanin 5-*O*-glucosyltransferase, respectively. Our approach by integration of transcriptomics and metabolomics provides an innovative way for comprehensive identification of genes involved in plant metabolism.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 159–168.

Key Words: metabolome; transcriptome; MYB factor; anthocyanins; glycosyltransferase.

1 INTRODUCTION

Plants have always been the most important resource for novel medicines, flavours, and industrial materials as alternatives for fossil resources. Plants are considered to produce ~200,000 natural products, involved in their wide chemical diversity (Dixon and Strack, 2003). After the determination of the whole genome sequence of *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), it is now possible to elucidate gene-to-metabolite correlation through the comprehensive analysis of gene expression (transcriptomics) and metabolite accumulation (metabolomics) (Fiehn, 2002; Sumner et al., 2003; Weckwerth, 2003; Bino et al., 2004; Kopka et al., 2004). For non-targeted metabolome analysis, it is necessary to combine several different analytical technologies, particularly those based on mass spectrometry such as gas chromatography-mass spectrometry (GC-MS) (Fiehn et al., 2000; Weckwerth et al., 2004), high-performance liquid chromatography-mass spectrometry (LC-MS) (Yamazaki et al., 2003; Roepenack-Lahaye et al., 2004), and Fourier-transform ion-cyclotron mass spectrometry (FT-MS) (Aharoni et al., 2002). The integration of the transcriptomics and metabolomics or detailed targeted chemical analysis would be a breakthrough in identifying the function of unknown genes and determining all gene-to-metabolite correlations in the plant cells. Only a limited number of reports, however, have been available on successful identification of novel gene functions by this approach (Aharoni et al., 2002; Guterman et al., 2002; Goossens et al., 2003; Mathews et al., 2003; Mercke et al., 2004; Hirai et al., 2004).

The *pap1-D* mutant is a T-DNA activation-tagged line overproducing anthocyanins by the ectopic overexpression of the *PAP1* gene. The *PAP1* gene encodes a MYB transcriptional factor by the action of an enhancer from the promoter of the cauliflower mosaic virus 35S transcript in the inserted T-DNA (Borevitz et al., 2000). In the *pap1-D* mutant, some structural genes for anthocyanin biosynthesis are expressed constitutively, and the accumulation of some phenylpropanoid derivatives such as anthocyanins is significantly enhanced (Borevitz et al., 2000). However, the transcriptome and metabolome have not been extensively characterized in this mutant, and elucidated the whole cellular process. The *PAP1*-overexpressing plants are an ideal model system for elucidating the whole cellular mechanisms at both transcriptome and metabolome levels under the expression of a single transcriptional factor.

The structures of flavonoids and their biosynthetic genes in *A. thaliana* have not yet been completely elucidated. Recently, the structures of several anthocyanin derivatives (Bloor et al., 2002) and flavonol glycosides (Veit

et al., 1999; Graham, 1998) have been reported. However, no genes encoding glycosyltransferase and acyltransferase for the modification of anthocyanin aglycones have been identified yet. For the identification of such genes involved in the production and modification of terminal metabolites in biosynthetic pathways, the combined analysis of transcripts and metabolites is a powerful technology (Jones et al., 2003). Here, we studied metabolomics by LC-MS for the targeted metabolite analysis of ~17 compounds combined with FT-MS for the non-targeted metabolite profiling of ~1,800 putative metabolites, and transcriptomics using the DNA microarrays covering 22,810 genes of the Arabidopsis genome. We could show that a set of genes involved in anthocyanin accumulation were upregulated together with the production of cyanidin derivatives and quercetin glycosides; thus we determined induced gene functions in production of these compounds. Subsequently, two genes coding for flavonoid glucosyltransferases were identified by *in vitro* study using recombinant proteins and by the anthocyanin analysis of T-DNA-inserted mutants. The present study shows a novel means of studying functional genomics through the integral analyses of the transcriptome and metabolome in plants.

2 COMBINED METABOLOME ANALYSIS

Metabolome analysis was carried-out by combination of flavonoid-targeted analysis by LC-MS, and non-targeted large-scale metabolite analysis by FT-MS.

2.1 Flavonoid-targeted analysis by LC-MS

Flavonoid accumulation profiles were analysed by HPLC/photodiode array/detection/electrospray ionization mass-spectrometry (HPLC/PDA/ESI-MS). The metabolites were identified by their UV-visible absorption spectra and mass fragmentation pattern of tandem MS spectroscopy in comparison with the authentic compounds in laboratory stock, and reported data (Bloor et al., 2002; Veit et al., 1999; Graham, 1998). In total, 17 peaks (A1-A11 and F1-F6) were identified in the leaves and roots (Figure 11-1).

In the *PAP1*-overexpressing lines (*pap1-D* mutant and *PAP1* cDNA-overexpressing transgenic plant), 11 anthocyanin pigments (A1-A11) and 3 quercetin glycosides (F4-F6) were accumulated in leaves. In leaves, the total anthocyanin accumulation in *pap1-D* mutant in leaves is 50 times (grown on agar plate) as high as in wild-type plant. The major anthocyanin in leaves, A11, is the most highly modified anthocyanin with 4 glycosides and 3 acyl-moieties attached in its molecule. In roots, on the other hand, 5 anthocyanins (A1, A2, A3, A5 and A8) were accumulated in the roots grown on the agar

plate. The total anthocyanin in the roots of the *pap1-D* mutant (grown on agar plate) is 14 times as high as that in the roots of the wild-type plant. The anthocyanins attached with sinapoyl moiety (A4, A7, A9, A10 and A11) were not detected in roots. In contrast to leaves, no significant differences with the amounts of quercetin glycosides were observed in roots of wild-type plant and *pap1-D* mutant.

2.2 Non-targeted analysis by FT-MS

Non-targeted FT-MS metabolite analysis was conducted on the leaf and root samples of the wild-type plant, *pap1-D* mutant, and *PAP1* cDNA-overexpressing transgenic plant grown on either agar or vermiculite. To decide the key determinant factors of the metabolome, principal component analysis (PCA) was conducted with ~1800 peaks of non-targeted FT-MS analysis (data not shown).

The first component of the PCA results (76% variance) predominantly reflects the difference in the type of organ (leaf or root), and the second component (9% variance) primarily indicates a difference in growth conditions (agar or vermiculite) as well as a secondary reflection of the total anthocyanin content (wild or *pap1-D*). Two major clusters (leaf on vermiculite and root on agar) formed two separate groups each reflecting two different genotypes (wild and *pap1-D*). This is presumably due to the small but significant difference in total anthocyanin content between the wild-type and *pap1-D* plants as detected by FT-MS, supporting the results of LC-MS analysis. Altogether, these results suggest that the major determinant

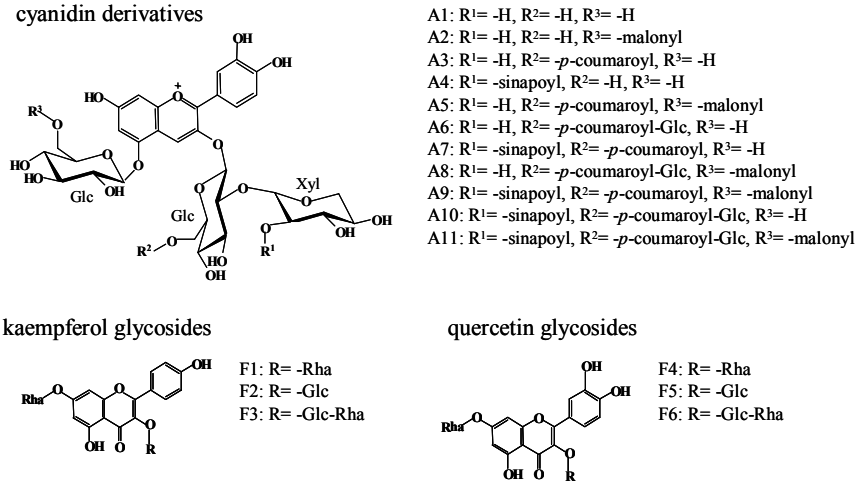


Figure 11-1. Cyanidin derivatives and flavonol glycosides accumulated in *PAP1* overexpressing *Arabidopsis*.

factors of the metabolome were the type of organ (leaf or root) and growth condition (agar or vermiculite). This implies that the global metabolome profiles of *PAP1*-overexpressing lines are relatively similar to those of wild-type plants despite the marked difference in total anthocyanin observed.

The PCA results of the anthocyanin-targeted analysis (data not shown) indicate that the major determinant factor of anthocyanin patterns is the genotype of plants reflected to the first component (68% variance). The *PAP1*-overexpressing lines form three distinct clusters: (1) root on agar; (2) leaf on agar; and (3) leaf on vermiculite. In contrast, the wild-type plants form a single cluster regardless of the type of organ and growth condition, exhibiting only slightly affected anthocyanin patterns. These results suggest that the *PAP1* gene regulates anthocyanin accumulation in a relatively specific manner, causing only a small change in the metabolome.

3 TRANSCRIPTOME ANALYSIS

3.1 Upregulated expression of novel genes by *PAP1*

The transcript levels of 22,810 genes on Affymetrix Arabidopsis Genome ATH1 GeneChip array were determined. Four different sets of comparisons were made to sort out the candidate genes responsible for anthocyanin accumulation in *PAP1*-overexpresing lines. To identify the genes exhibiting the reproducible changes of expression, the genes expressing more than 1.5-fold in the comparisons of *pap1-D* leaf exp.1 *vs* wild-type leaf exp.1, *PAP1,* cDNA-overexpressing plant exp.1 *vs* wild-type leaf exp.1, and *pap1-D* leaf exp.2 *vs* wild-type leaf exp.2 were selected as induced genes. Thirty-nine upregulated genes in leaf including *PAP1* (Table 11-1) were resulted. Eight iIn the 39 upregulated genes in leaves 8 were annotated as encoding well-known anthocyanin biosynthetic enzymes or regulatory proteins characterized previously; *TT3*, *TT4*, *TT5*, *TT7*, *PAP1*, *TT8*, *TTG2* and *TT19*. Besides these well-known anthocyanin biosynthetic genes, those that are putatively annotated to anthocyanin biosynthetic genes such as At5g05270 (*CHI* homologue), At4g22870 (*ANS* homologue), and At1g20490 (4CL1 homologue) were also upregulated. These paralogous genes are presumably involved also in anthocyanin biosynthesis in addition to the previously characterized genes. In addition, the comparison of *pap1-D* root *vs* wild-type root was conducted. Out of 39 genes upregulated in leaves, 17 genes encoding well-known anthocyanin biosynthetic enzymes, glycosyltransferases, and acyltransferases were also induced in roots.

Combing with the results of metabolite profiling, these results suggest that *PAP1* gene induces specifically the gene expression of involved in

anthocyanin production or in accumulation leading to specific accumulation of anthocyanins.

4 INTEGRATION OF METABOLOMICS AND TRANSCRIPTOMICS

4.1 Global changes of metabolome and transcriptome incited by the expression of *PAP1*

With metabolome analysis, it was observed that the ectopic *PAP1*-overexpression resulted in remarkable over-accumulation of cyanidin-type anthocyanins and quercetin-type flavonols. Metabolic profiling indicated that the alteration in metabolite patterns is specific to flavonoids. With transcriptome analysis, being coordinated with those metabolome changes, *PAP1*-overexpression resulted in upregulation of almost all genes encoding anthocyanin biosynthesis enzymes (Table 11-1). All these metabolome and transcriptome data suggested that PAP1 regulates specifically flavonoid biosynthetic genes causing specific accumulation of cyanidin- and quercetin-type flavonoids.

From the results of metabolome and transcriptome analysis, we could putatively assign the function of upregulated genes. Besides known anthocyanin biosynthetic genes as indicated above, several genes of unconfirmed in particular gene families were upregulated: two acyltransferase-family genes (At1g03940 and At3g29590), three glycosyltransferase-family genes (At5g54060, At4g14090 and At5g17050), and two glutathione *S*-transferase-family genes (At1g02930 and At1g02940). Considering accumulation of specific molecular species of anthocyanins in *PAP1*-overexpresing plants, the functions of those up regulated genes can be

Table 11-1. Upregulated genes by ectopic *PAP1*-overexpression

	Gene family	Gene name	Number
			>1.5-folds
Upregulated	Flavonoid pathway	*TT3, TT4, TT5, TT7*	4
	Putative flavonoid pathway		3
	Glycosyltransferase		4
	Acyltransferase		2
	Glutathione *S*-transferase	*TT19*	3
	Transcription factor	*PAP1, TT8, TTG2*	4
	Sugar transporter protein		2
	Ca^{2+} binding protein		2
	Others		7
	Unknown		8
	Total		39

putatively assigned to be related to the production of specific anthocyanin derivatives for their modification and transport.

4.2 Flavonoid acyltransferases

In the Arabidopsis genome, ca.~70 genes related to acyl-CoA dependent acyltransferase are contained (Dudareva et al., 2000). Two putative acyltransferase genes, At1g03940 and At3g29590, were upregulated by the *PAP1* expression. The most extensively modified anthocyanin A11 contains three acyl groups, *i.e.,* sinapoyl, *p*-coumaroyl and malonyl. Taking into accounts the distinct expression patterns of the two genes and the accumulation of anthocyanin molecules in leaves and roots, At1g03940 and At3g29590 would be the candidates of either malonyltransferase or *p*-coumaroyltransferase.

4.3 Networks of transcription factors

Recently, a network model of TTG1-dependent transcriptional pathway was proposed including anthocyanin accumulation, seed coat pigmentation, and initiation of trichomes (Zhang et al., 2003). In the present study of *PAP*1-overexpressing plants, three transcription factor genes, *TT8* (bHLH protein), *TTG2* (WRKY protein) and At5g61600 (an *AP2* domain factor), were upregulated in addition to *PAP1*. The other well-known transcription-factor genes were not changed. These results demonstrated that *PAP1* was responsible for anthocyanin-specific downstream of the transcription network.

4.4 Functional identification of two flavonoid glycosyltransferases

Three glycosyltransferase genes, At5g54060, At4g14090, and At5g17050, were induced in *PAP1*-overexpression plants, suggesting the involvement of these three proteins in modification of sugar moieties of anthocyanins. Besides, in the Arabidopsis genome, 107 UDP-sugar-dependent glycosyltransferase genes are present (Bowles, 2002). In Arabidopsis 107 UDP-glycosyltransferase gene, three induced glycosyl-transferase genes (Table 11-1) in our present investigation, one other gene, At3g21560, was induced in *PAP1*-overexpressing leaves. It is suggesting possible participation of this protein in the production of accumulated anthocyanins in *PAP1*-overexpressing plants. Due to weak induction of At3g21560 by PAP1, this gene is not listed in Table 11-1; however, the induction in *pap1-D* was reproducible. Figure 11-2 shows the molecular phylogenetic tree of the amino acid sequences of the flavonoid

glycosyltransferases. Since the most extensively modified anthocyanin molecule A11 possesses, in addition to 3-*O*- and 5-*O*-glucose, a xylose residue attached to C2-position of 3-*O*-glucoside and a glucose residue attached to *p*-position of coumaroyl moiety. The phylogenetic tree shows that At5g17050 belongs to the subfamily of 3GT, At4g14090 belongs to the subfamily of 5GT, At5g54060 suited Petunia A3G-2''RT, and At3g21560 does not belong to these subfamilies.

Considering the different patterns of anthocyanin accumulation and of gene expression profiles between leaves and roots, At3g21560 was suggested the *p*-coumaroyl glucosyltransferase. The clustering in phylogenic tree of glycosyltransferase family is also consistent with these assumptions.

Two of them, At5g17050 and At4g14090, were functionally identified as coding for flavonoid 3-*O*-glucosyltransferase (3GT) and anthocyanin 5-*O*-glucosyltransferase (5GT), respectively. In the gene knockout mutant of At5g17050, the levels of 4 flavonol glycosides (F2, F3, F5 and F6) with glucose attached at 3-position were reduced. In contrast, the slight increases were observed with the levels of 2 flavonol glycosides (F1 and F4) with rhamnose residue attached at 3-position. These results indicate that At5g17050 protein is responsible for glucosylation of 3-positions of both anthocyanins and flavonols. Moreover, with recombinant protein of At5g17050, flavonoid 3-*O*-glucosyltransferase activity was detected. Three anthocyanidins (cyanidin, pelargonidin, and delphinidin) and three flavonols (kaempferol, quercetin, and myricetin) were tested for substrates of the

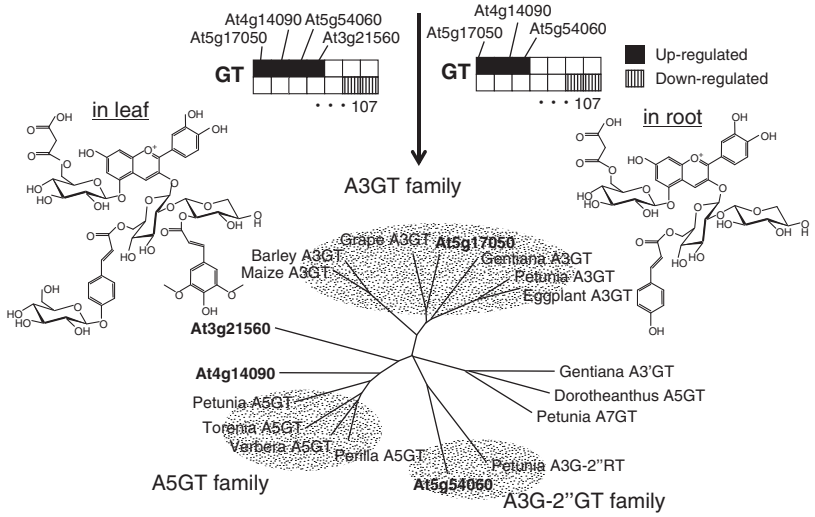


Figure 11-2. Cyanidin derivatives accumulated in *Arabidopsis* leaf and root, and Molecular phylogenetic tree of the amino acid sequences of the flavonoid glycosyltransferases.

reaction by the recombinant protein of At5g17050. All these results indicate that the protein of At5g17050 catalyses glucosylation at 3-position of both cyanidins and flavonols *in planta* as UDP-glucose; flavonoid 3-*O*-glucosyltransferase.

The gene knockout mutant of At4g14090 exhibited the altered anthocyanin pattern, accumulating six new anthocyanins, which are not produced in the wild-type plant. Detailed investigation of mass spectra using MS^2 analysis indicated that six cyanidin derivatives are de-glucosylated anthocyanins at 5-position of A1, A5, A4, A8, A7 and A11, respectively, suggesting complete lack of 5-glucosylation activity of anthocyanin in this mutant. These results clearly indicated the protein of At4g14090 is a functional UDP-glucose: anthocyanin 5-*O*-glucosyltransferase.

ACKNOWLEDGEMENTS

We thank Dr. Richard A. Dixon (Samuel Roberts Noble Foundation, Admore, OK, USA) for providing the *pap1-D* mutant. We also thank the Salk Institute Genomic Analysis Laboratory for providing the sequence-indexed *A. thaliana* T-DNA insertion mutants, and the RIKEN BioResource Center for providing the full-length cDNA. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology (Japan; Grants-in-Aid for Scientific Research), by CREST of Japan Science and Technology Agency (JST), and by Research for the Future Program (grant no. 00L01605; "Molecular Mechanisms on Regulation of Morphogenesis and Metabolism Leading to Increased Plant Productivity").

REFERENCES

Aharoni, A., Ric De Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., and Goodenowe, D.B., 2002, Non-targeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry, *OMICS* **6**:217–243.

Bino, R.J, Hall, R.D, Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S., and Sumner, L.W., 2004, Potential of metabolomics as a functional genomics tool, *Trends Plant Sci*. **9**:418–425.

Bloor, S.J. and Abrahams, S., 2002, The structure of the major anthocyanin in *Arabidopsis thaliana*, *Phytochemistry* **59**:343–346.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C., 2000, Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis, *Plant Cell* **12**:2383–2393.

Bowles, D., 2002, A multigene family of glycosyltransferases in a model plant *Arabidopsis thaliana*, *Biochem. Soc. Trans*. **30**:301–306.

Dixon, R.A. and Strack, D., 2003, Phytochemistry meets genome analysis and beyond, *Phytochemistry* **62**:815–816.

Dudareva, N. and Pichersky, E., 2000, Biochemical and molecular genetic aspects of floral scents, *Plant Physiol.* **122**:627–633.

Fiehn, O., 2002, Metabolomics – the link between genotypes and phenotypes, *Plant. Mol. Biol.* **48**:155–171.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., and Willmitzer L., 2000, Metabolite profiling for plant functional genomics, *Nat. Biotechnol.* **18**:1157–1161.

Goossens, A., Hakkinen, S.T., Laakso, I., Seppanen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A.M., Soderlund, H., Zabeau, M., Inze, D., and Oksman-Caldentey, K.M., 2003, A functional genomics approach toward the understanding of secondary metabolism in plant cells, *Proc. Natl. Acad. Sci. USA* **100**:8595–8600.

Graham, T.L., 1998, Flavonoid and flavonol glycoside metabolism in Arabidopsis, *Plant Physiol. Biochem.* **36**:135–144.

Guterman, I., Shalit, M., Menda, N., Piestun, D., Dafny-Yelin, M., Shalev, G,. Bar, E., Davydov, O., Ovadis, M., Emanuel, M., Wang, J., Adam, Z., Pichersky, E., Lewinsohn, E., Zamir, D., Vainstein, A., and Weiss, D., 2002, Rose scent: genomics approach to discovering novel floral fragrance-related genes, *Plant Cell* **14**:2325–2338.

Hirai, M.Y., Yano, M., Goodenowe, B.G., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., and Saito K., 2004, Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stress in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA* **101**:10205–10210.

Jones, P., Messner, B., Nakajima, J., Schäffner, A.R., and Saito K., 2003, UGT73C6 and UGT78D1 glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*, *J. Biol. Chem.* **45**:43910–43918.

Kopka, J., Fernie, A., Weckwerth W., Gibon, Y., and Stitt, M., 2004, Metabolite profiling in plant biology: platforms and desitinations, *Genome Biol.* **5**:109.

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J., and Wagner, D.R., 2003, Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport, *Plant Cell* **15**:1689–1703.

Mercke, P., Kappers, I.F., Verstappen, F.W., Vorst, O., Dicke, M., and Bouwmeester, H.J., 2004, Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants, *Plant Physiol.***135**:2012–2124.

Roepenack-Lahaye, E., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., Schmidt, J., Scheel, D., and Clemens, S., 2004, Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry, *Plant Physiol.* **134**:548–559.

Sumner, L.W., Mendes, P., and Dixon, R.A., 2003, Plant metabolomics: large-scale phytochemistry in the functional genomics era, *Phytochemistry* **62**:817–836.

Veit M., Pauli G.F., 1999, Major flavonoids from *Arabidopsis thaliana* leaves, *J. Nad. Prod*. **62**:1301–1303.

Weckwerth, W., 2003, Metabolomics in systems biology, *Annu. Rev. Plant Biol.* **54**:669–689.

Weckwerth, W., Loureiro, M.E., Wenzel, K., and Fiehn, O., 2004, Differential metabolic networks unravel the effects of silent plant phenotypes, *Proc. Natl. Acad. Sci. USA* **101**:7809–7814.

Yamazaki, M., Nakajima, J., Yamanashi, M., Sugiyama, M., Makita, Y., Springob, K., Awazuhara, M., and Saito K., 2003, Metabolomics and differential gene expression in anthocyanin chemo-varietal forms of *Perilla frutescens*, *Phytochemistry* **62**:987–995.

Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A., 2003, A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*, *Development* **130**:4859–4869.

Chapter 12

IDENTIFYING SUBSTRATES AND PRODUCTS OF ENZYMES OF PLANT VOLATILE BIOSYNTHESIS WITH THE HELP OF METABOLIC PROFILING

Dorothea Tholl[1,2], Feng Chen[1], Yoko Iijima[1], Eyal Fridman[1], David R. Gang[3], Efraim Lewinsohn[4], and Eran Pichersky[1]

[1]Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109-1048 USA; [2]Max Planck Institute for Chemical Ecology, Hans Knoell Strasse 8, D-07745, Jena, Germany; [3]Department of Plant Sciences and Institute for Biomedical Science and Biotechnology, University of Arizona, Tucson, Arizona 85721-0036 USA; [4]Department of Vegetable Crops, Newe Ya'ar Research Center, Agricultural Research Organization, P.O. Box 1021, Ramat Yishay, 30095, Israel

Abstract: Ongoing efforts in metabolic profiling of both cultivated and wild plants continue to identify new plant compounds, many of them unique to a single species or found only in closely related species. Such compounds are defined as specialized, or secondary, metabolites and they play many physiological and ecological roles, including in plant-insect and plant-pathogen interactions. To date, only a few of the enzymatic reactions leading to the synthesis of such compounds have been elucidated and the enzymes responsible identified. Our group has concentrated on the biosynthesis of plant volatiles. We present several examples in which metabolic profiling together with gene expression profiling and biochemical methods have led to the identification of enzymes responsible for the synthesis of volatile terpenes in Arabidopsis flowers, benzenoid esters in Arabidopsis leaves, and terpenes and methylated phenylpropenes in glands of sweet basil.

Key Words: Secondary metabolites; gene expression profiling; bacterial expression system; biochemical assays; methylation; terpenes; esters; phenylpropenes.

1 INTRODUCTION

Achieving the goal of cataloguing all the components of the cell – genes, enzymes (surely the majority of the cellular proteins), and other types of proteins, and metabolites – and elucidating all the causal relationships

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 169–182.

among them will require a vast effort. While most biological research to date has been the piecemeal elucidation of components and causal relationships of a very small and circumscribed subset of cellular pathways, several recent approaches have been based on "systems biology" in which a very large number of components are catalogued and statistical methods are used to try to infer correlations, which in turn suggest further types of investigation (Fiehn and Weckwerth, 2003; and see this volume). This approach has been most prominent in the sequencing of whole genomes, including two plant genomes (Arabidopsis and rice) followed by computer analysis of the coding information of these genomes, and the analysis of the expression of the entire set of genes by means of DNA microarrays.

DNA, RNA, and proteins each constitute a class of compounds with some structural properties in common, thus allowing for the development of analytical methods that apply to basically all members of the class. As pointed out by Trethewey (2004), the metabolites found in the cell have no shared chemical features on which general, combined isolation-separation-identification methods can be based – at least, no such features have been recognized so far. Analysis of metabolites typically starts with some method of extraction from the tissue, and different methods have to be used to extract different classes of compounds. In the next steps, compounds have to be separated and identified, and these processes too may involve different methodologies for different groups of compounds. Current metabolic profiling techniques are primitive and allow for the extraction and separation of only a small fraction of plant metabolites and only a fraction of those have been identified. Thus, it is not surprising that at present we have probably not yet identified the majority of compounds of plant primary metabolism, and our knowledge of specialized (secondary) metabolism in any species is either severely limited or non-existent.

2 PLANT VOLATILES: CHEMISTRY AND FUNCTION

Plant volatiles constitute a small segment of the total plant metabolite output, and they do share chemical properties – mainly their volatility – that allows us to apply common analytical techniques. Plant volatiles are organic molecules (typically less than 300 Da) that often contain oxygen functionalities and sometimes nitrogen or sulfur (Figure 12-1). These compounds are lypophylic in nature and although their boiling point is typically above ambient temperature, they have high vapour pressure and therefore easily vaporize.

Plant volatiles serve many functions (Pichersky and Gershenzon, 2002; Dudareva et al., 2004). Many floral scents are emitted to attract pollinators,

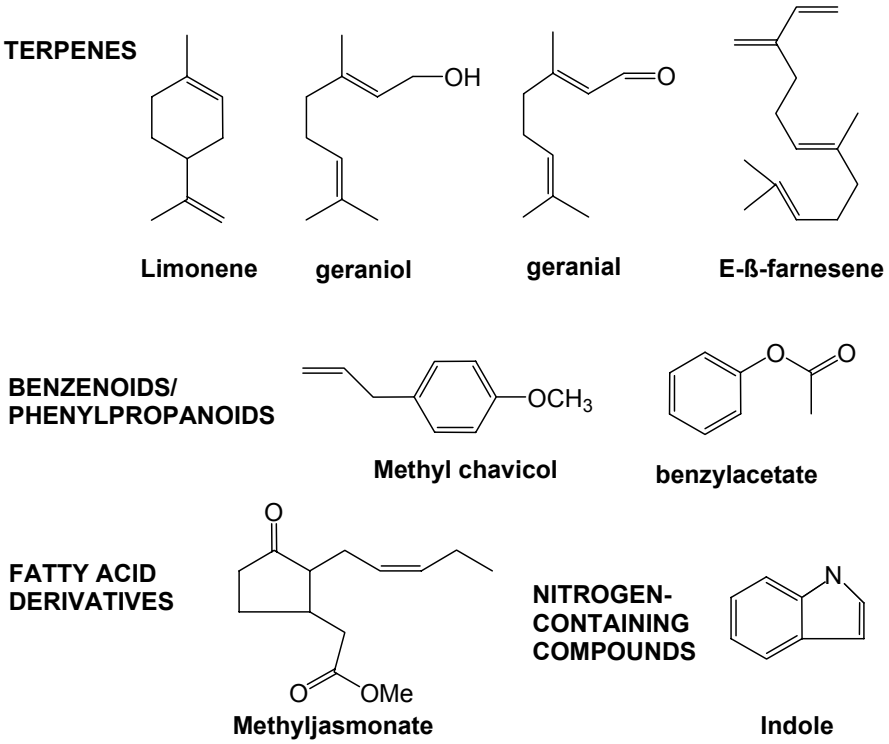


Figure 12-1. Examples of plant volatiles.

mostly but not only insects, although some odours may be used to deter unwanted visitors. Volatiles in fruits may directly attract animals, insects, or birds that eat the fruit and thereby disperse the seeds, or they may constitute a reward by contributing to the flavour. In vegetative tissues, volatiles are emitted following injuries inflicted during insect herbivory, and the emitted volatiles can attract predators of the herbivores. Some data even suggest that neighbouring plants are capable of detecting such "distress signals" and turn their own defense systems on. Finally, some volatiles are simply toxic compounds (and may be stored in specialized vegetative tissues or cells, such as glands), and exert their effect on herbivores after they are ingested when the herbivores feed on the plant. Some compounds can serve as both attractants and repellants/toxins, depending on which insect/animal is involved and even whether the same insect/animal is interacting with them through the olfactory or the digestive systems.

3 PLANT VOLATILES ARE DERIVED FROM A FEW BASIC PATHWAYS BY A LIMITED NUMBER OF MODIFICATION REACTIONS

Little is known about the pathways leading to the synthesis of the majority of plant volatiles. Our laboratories have focused on identifying and characterizing the enzymes that make volatiles and the genes encoding these enzymes. Our long-term goal has been to understand how the myriad species of the plant kingdom have evolved the ability to make so many different volatile compounds, estimated to be in the thousands. Our results, and the results from several other laboratories, have shown that while volatiles are diverse chemicals, most are derived from just a few modified biochemical pathways.

One such pathway is the terpene pathway in which a carbon skeleton is built up first into isoprene diphosphate (C5) units that are condensed into C10 or C15 diphosphate intermediates, which are finally converted into monoterpene (C10) and sesquiterpene (C15) volatiles, respectively, by enzymes encoded by a large family of genes termed terpene synthases (Figure 12-2) (larger terpenes are also produced in other branches of the pathway, but they are not generally volatiles). In contrast to the biosynthetic terpene pathway, most other volatile compounds are derived from two other classes of compounds, phenylpropanoids and fatty acids, through the shortening of a carbon skeleton, often followed by further modification, or simply by modification of the existing carbon skeleton. Compounds that are already somewhat volatile may also be modified, resulting in enhanced volatility or changed olfactory properties. The majority of these modifications involve the reduction or removal of carboxyl groups, the addition of hydroxyl groups, and the formation of esters and ethers (Figure 12-2). Each type of modification is catalysed by a group (or several groups) of related enzymes constituting protein families. Typically, these protein families contain enzymes involved in the synthesis of both volatile and non-volatile compounds. Some of our investigations into the identification of the enzymatic functions of specific members of such families are described below.

4 IDENTIFICATION OF THE ENZYMATIC FUNCTION OF MEMBERS OF THE TERPENE SYNTHASE FAMILY IN *ARABIDOPSIS THALIANA*

Until recently it was believed that *Arabidopsis thaliana*, a small weedy plant that is used as a model plant organism, produces only a few secondary

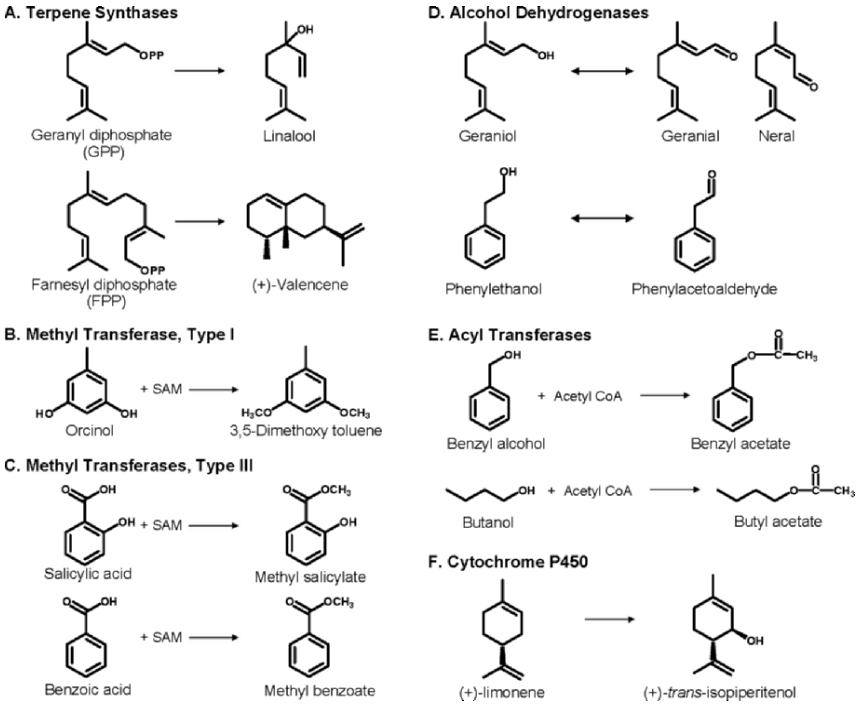


Figure12- 2. Some reactions catalysed by representatives of enzyme families.

metabolites. In particular, because it is believed to be mostly self-pollinating, it was thought to produce no floral volatiles, and casual sniffing of the flowers by the human nose does not indeed detect a noticeable bouquet. However, the recent determination of the complete genome sequence of Arabidopsis revealed about 30 genes with sequence similarity to known terpene synthases (TPSs) from other species (Aubourg et al., 2002). This observation prompted us to search more carefully for possible emission of floral volatiles. Using highly sensitive collection and detection methods, we were able to show that Arabidopsis flowers emit several monoterpenes (e.g., linalool, myrcene and limonene) and as many as 20 different sesquiterpenes (Figure 12-3A and B) (Chen et al., 2003a). The major floral volatile is β-caryophyllene, a sesquiterpene. The total emission of terpene volatiles is in the range of a few nanogram per hour per gram fresh weight of flowers, which is 2–3 orders of magnitude lower than the emission rate of some highly scented flowers (Raguso and Pichersky, 1995). Moreover, the human nose is not particularly sensitive to sequiterpenes. These two observations explain why humans cannot easily detect the scent of Arabidopsis flowers. Nonetheless, some insects might be able to detect these flowers by olfactory cues, and in fact Arabidopsis

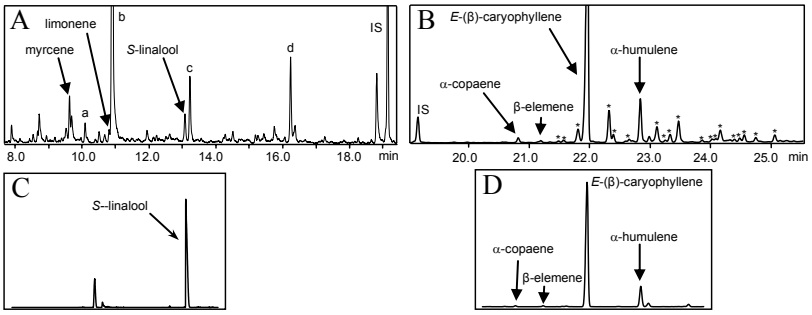


Figure 12-3. Identification of the products of two Arabidopsis terpene synthases. **A.** Gas chromatographic separation of the monoterpenes emitted by Arabidopsis flowers. **B.** Gas chromatographic separation of the sesquiterpenes emitted by Arabidopsis flowers. The amount of the internal standard (IS) is the same in both A and B chromatographs, showing the amount of the monoterpenes is much lower than the amount of sesquiterpenes. The peak labelled "a" in A is octanal, "b" is 2-ethyl-hexanol, "c" is nonanal and "d" is decanal. The peaks labeled with astericks in B are all sesquiterpenes. **C.** Gas chromatographic analysis of the product of the Arabidopsis TPS enzyme encoded by gene At1g61680, indicating that the enzyme catalyses the formation of the monoterpene S-linalool. **D.** Gas chromatographic analysis of the product of the Arabidopsis TPS enzyme encoded by gene At5g23960, indicating that the enzyme catalyses the formation of the four sesquiterpenes α-copaene, α-elemene, β-caryophyllene, and α-humulene. Unlabelled peaks in C and D are not terpenes, and are present in the control reactions as well.

flowers growing in the wild are visited by many types of insects (Hoffman et al., 2003).

Having established that Arabidopsis flowers do synthesize and emit terpenes, we next examined which of the members of the TPS gene family are involved. A set of RT-PCR experiments were carried out on all TPS genes in Arabidopsis and several genes were found to be expressed almost exclusively in the flowers. Complete cDNAs of these genes were isolated, spliced into a bacterial expression vector, and expressed in *E. coli*. The resulting proteins were assayed for activity with the substrate geranyl diphosphate (GPP), the universal precursor of monoterpenes, and farnesyl diphosphate (FPP), the universal precursor of sesquiterpenes (Figure 12-2A). These biochemical experiments identified three monoterpene synthases and two sesquiterpene synthases that are responsible for almost all of the Arabidopsis floral terpene volatiles. That only a small number of Arabidopsis TPS genes account for all the floral terpenes is explained by the observation that while some of these enzymes produced a single product (the linalool synthase is one such enzyme, Figure 12-3C), other enzymes can produce multiple products. For example, one Arabidopsis TPS gene turned out to encode a sesquiterpene synthase that catalyses the formation of four sesquiterpenes (Figure 12-3D), and another florally expressed Arabidopsis sesquiterpene synthase is responsible for the synthesis of as many as 15

compounds (unpublished). That some TPSs produce multiple products had been previously observed (Chen et al., 2003a).

In this investigation, we started with the observation that the coding capacity of the Arabidopsis genome contained potential TPS genes, but no prior information was available about terpene synthesis in Arabidopsis flowers or in any other organs (with the exception of the synthesis of gibberellins, which are diterpenes). The metabolic profiling results indicated that some of the TPS enzymes encoded in the Arabidopsis genome were likely to be active in floral scent biosynthesis. However, a simple comparison of these TPS sequences with functionally defined TPS proteins from other species was not sufficient to identify which Arabidopsis protein corresponds to each volatile (or volatiles), since an extensive body of research has demonstrated that rapid convergent evolution occurs in the terpene synthase gene family in separate plant lineages, so that, for example, the various linalool synthases known from distally related species are not very similar to each other and instead are more similar to other monoterpene synthases from the same lineage. Thus, the linalool synthase from Arabidopsis, for example, could not be identified based simply on sequence similarity to linalool synthases from other species, and the biochemical experiments were therefore crucial in identifying enzymatic functions.

5 IDENTIFICATION OF A BENZOIC ACID/SALICYLIC ACID METHYLTRANSFERASE IN *A. THALIANA*

We investigated a similar situation with regards to the biosynthesis of benzenoid methylesters in Arabidopsis. Methylsalicylate (MeSA) and methylbenzoate (MeBA) are two benzenoid methylesters that are commonly found in floral volatiles of diverse taxa, although not in flowers of Arabidopsis. The genes encoding salicylic acid methyltransferase (SAMT) and benzoic acid methyltransferase (BAMT) were first identified in flowers of *Clarkia breweri* and snapdragons, respectively (Ross et al., 1999; Murfitt et al., 2000), and were recognized to constitute a new type of methyltransferase family, designated the SABATH methyltransferase family (D'Auria et al., 2002). In Arabidopsis, there are 24 related genes (D'Auria et al., 2002; Chen et al., 2003b).

MeSA had been reported to be emitted from vegetative tissues of many plant species, including Arabidopsis, during herbivory (Van Poecke et al., 2001) or viral infection (Shulaev et al., 1997). Given that the Arabidopsis genome has 24 genes with homology to *C. breweri* SAMT, it was likely that at least one of them encodes a SAMT. However, protein sequence comparisons did not identify a single Arabidopsis gene among these 24

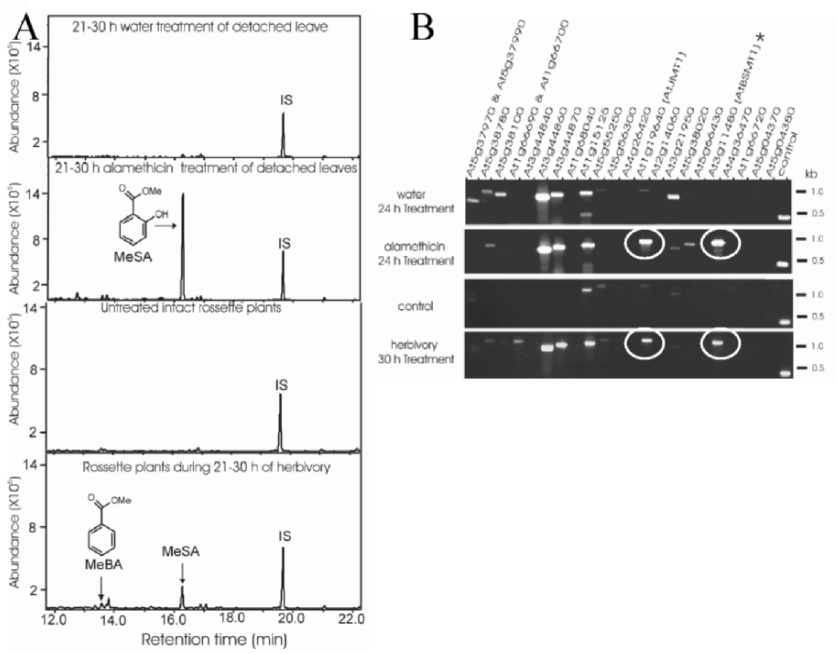


Figure12-4. Metabolic profiling and gene expression profiling to identify the AtSABATH gene involved in benzenoid methyl ester formation. **A**. Gas chromatograph analyses of plant samples under different conditions. **B**. RT-PCR gene expression profiling of all AtSABTH genes under same conditions. Circles show increased transcript levels in conditions eliciting emission of benzenoid methylesters compared with control conditions where no such emission was observed. The gene denoted with an asterisk is the one encoding the enzyme benzoic acid/salicylic acid methyltransferase (BSMT), as proven by subsequent *in vitro* enzyme assays with the purified protein.

SABATH genes which was more similar to *C. breweri* SAMT than to any other Arabidopsis gene in this family, and no Arabidopsis SABATH gene exhibited >50% identity to *C. breweri* SAMT.

Therefore, to identify the Arabidopsis SABATH gene(s) responsible for the synthesis of MeSA, we chose a combined approach of gene expression profiling with metabolic profiling. We first searched for conditions under which MeSA is emitted from Arabidopsis leaves. We examined herbivory in detail, and found out that when Arabidopsis leaves are attacked by the specialized herbivore *Plutella xylostella*, not only is MeSA emitted but also some MeBA is also released (Figure 12-4A). In addition, we established that MeSA (but not MeBA) is emitted from detached Arabidopsis leaves treated

with alamethicin, a fungal elicitor, but not from detached leaves treated with water alone (Figure 12-4A).

We next examined by RT-PCR the expression of the 24 Arabidopsis SABATH genes under these conditions, using specific oligonucleotide primer pairs for each gene. While several SABATH genes were induced during herbivory, and a few other SABATH genes were induced during alamethicin treatment, only two SABATH genes were induced under both of these treatments, and not induced in the controls (Figure 12-4B). One of these genes had previously been identified as jasmonic acid methyltransferase (JMT) (Seo et al., 2001). A full-length cDNA of the other gene was obtained, expressed in *E. coli*, and the protein shown to have the ability to catalyse the methylation of both SA and BA.

6 ENZYMES INVOLVED IN THE BIOSYNTHESIS OF PHENYLPROPENES AND TERPENES IN BASIL GLANDS

The two examples above deal with *A. thaliana*, where the task of identifying candidate genes is made easier due to the availability of the full genome sequence. However, the majority of investigations into the biosynthesis of plant volatiles, and plant secondary metabolites in general, are carried out in species for which very little genetic and genomic information is available. In such systems, metabolic profiling is still very important. In fact, metabolic profiling is usually done first as a surveying tool to identify plants of interests – those that may have interesting floral bouquets or desirable herbal spices (which are volatiles that impart distinct flavours to our foods). Once such volatiles are detected, several sets of tools are developed to make the identification, isolation, and characterization of genes possible. An example is illustrated below with our investigation into the biosynthesis of volatile flavour compounds in basil.

Basil plants have been used since antiquity to spice up food. Many cultivars of basil with distinct aromas have been bred. Basil (*Ocimum basilicum*) is in the Lamiaceae family, and is known for containing both terpenes and phenylpropenes (Figure 12-5). We have investigated in depth three such varieties, known as EMX1, SW, and SD. Metabolic profiling, done initially on material extracted from whole leaves, showed that EMX1 is particularly rich in methylchavicol (and has some methyleugenol as well), SW is rich in eugenol and linalool, and in SD the predominant volatile is citral, a mixture of geranial and neral.

Lamiaceae species, including basil, have numerous glandular trichomes, or glands, on the surface of their leaves. A very common type of gland in basil, called a peltate gland, consists of four cells connected to the epidermal

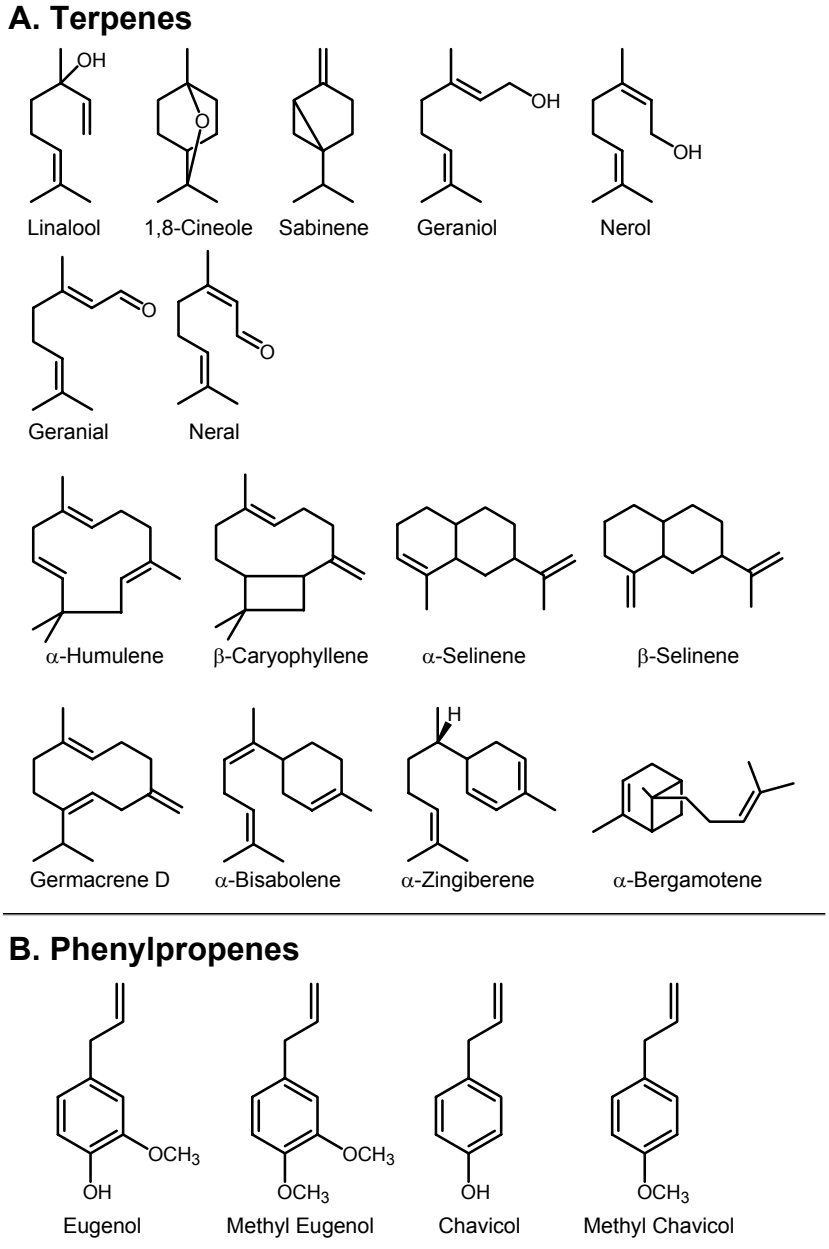


Figure 12-5. Some terpenes (**A**) and phenylpropenes (**B**) found in basil glands.

cell layer by a short-stalk cell. The four cells of the gland are covered by a thick cuticle that can expand into a "sac" to contain material secreted from the gland cells (Figure 12-6). Since previous reports showed that the volatile terpenes found in the mint plant, also in the Lamiaceae family, are synthesized and stored in the leaf peltate glands (Gershenzon et al., 2000), we examined the volatile contents of basil peltate glands by directly extracting material from individual sacs with a micropipette and analysing the material by gas chromatography-mass spectrometry (GC-MS) (Gang et al., 2001). This analysis showed that the volatiles stored in the glands were the same as those detected from whole leaf, whereas leaves devoid of peltate glands did not contain these volatiles, indicating that basil peltate glands, like mint glands, were the site of storage, and possibly synthesis, of these volatiles.

As outlined above, our basic goal is to identify the enzymes and genes responsible for the volatile biosynthesis in plants. But since no gene sequence information was available from basil plant, such information had to be obtained first. However, a genome sequencing approach is currently not feasible for every plant. To circumvent this problem, we chose to obtain sequence information from basil in a way that maximized the information on the desired genes and minimized the investment in obtaining sequence information that was not relevant to our particular interest. This was done by obtaining sequence information on genes that are specifically expressed in

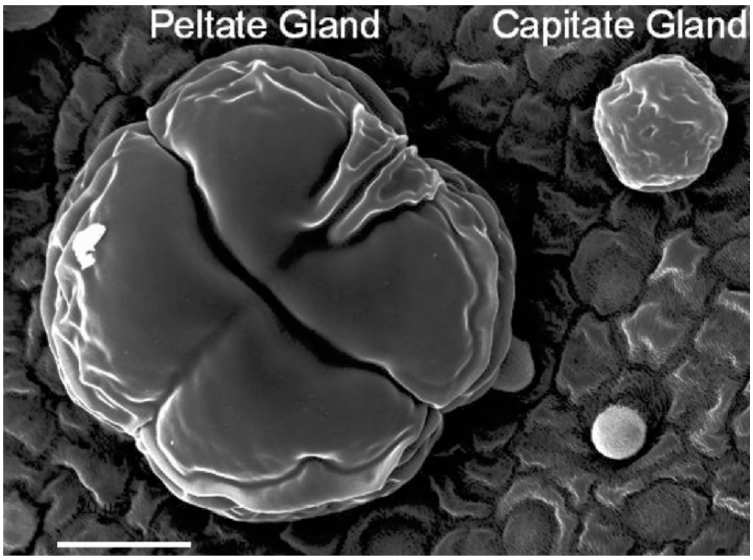


Figure 12-6. Scanning electron micrograph of a 4-celled peltate gland and a 2-celled capitate gland from the surface of a basil leaf.

the glands. To do so, we first adopted a procedure originally developed for isolating mint peltate glands to isolate basil peltate glands. Our procedure yielded intact 4-cell basil peltate glands that were completely separated from the rest of the leaf and from other types of glands, as well as devoid of gland stalks. RNA was extracted from peltate gland cells of each of the three basil cultivars, and cDNA libraries were constructed. Then, the DNA sequences of several thousand cDNAs from each library were determined and analysed for protein coding information, creating expressed sequence tag (EST) databases (Gang et al., 2001; Iijima et al., 2004).

Basil cultivar SD is rich in citral, which is the product of the oxidation of geraniol (Figure 12-2) (Iijima et al., 2004). Geraniol itself is a monoterpene alcohol, an isomer of linalool, and like linalool was believed to be synthesized from GPP, although no geraniol synthase had been identified. Examination of the EST databases of the three cultivars identified many potential terpene synthases, including several sequences that were uniquely found in the SD database. Each of the unique SD TPS sequences was expressed in *E. coli* and tested for enzymatic activity. One of them was found to be geraniol synthase, catalysing the exclusive formation of geraniol from GPP (Iijima et al., 2004).

To examine the synthesis of methylchavicol and methyleugenol in EMX1, the EST database of EMX1 was examined (Gang et al., 2002) for sequences with homology to known methyltransferases, including the enzyme that can methylate eugenol and isoeugenol in *C. breweri* flowers (Wang et al., 1997). Several sequences were identified based on this criterion, and in addition these sequences were not prevalent in the other two EST databases. Enzymatic assays of *E. coli*-produced proteins showed one of them to be eugenol methyltransferase, and another to be chavicol methyltransferase. Both proteins were highly similar to each other (>90% identical), but while they had some sequence similarity to *C. breweri* isoeugenol/eugenol methyltransferase (and no similarity at all to SAMT), they were more similar to other methyltransferases such as isoflavone methyltransferase. This observation indicates that the enzymes that methylate phenylpropenes in basil and *C. breweri* must have evolved their substrate specificity independently.

7 CONCLUSIONS

Metabolic profiling of whole plants or selected tissues and even cell types, combined with detailed information on the sequence of specific genes whose expression is correlated with the production of specific metabolites, is a powerful approach to identifying candidate genes involved in the biosynthesis of such metabolites. However, final identification of enzymatic

function must be achieved by biochemical experiments in which the candidate proteins are shown to possess the postulated catalytic activities.

REFERENCES

Aubourg, S., Lecharny, A., and Bohlmann, J., 2002, Genomic analysis of the terpenoid synthase (AtTPS) gene family of Arabidopsis thaliana, *Mol. Genet. Genomics* **267**:730–745.

Chen, F., Tholl, D., D'Auria, J.C., Farouk, A., Pichersky, E., and Gershenzon, J., 2003a, Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers, *Plant Cell* **15**:481–494.

Chen, F., D'Auria, J.C., Tholl, D., Ross, J.R., Gershenzon, J., Noel, J.P., and Pichersky, E., 2003b, An Arabidopsis thaliana gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense, *Plant J.* **36**:577–588.

D'Auria, J.C., Chen, F., and Pichersky, E., 2002, The SABATH family of methyltransferases in Arabidopsis thaliana and other plant species, in: *Recent Advances in Phytochemistry*, Elsevier Science Ltd, Oxford, vol. 37. pp. 253–283.

Dudareva, N., Pichersky, E., and Gershenzon, J., 2004, Biochemistry of plant volatiles, *Plant Physiol* **135**:1893–1902.

Fiehn, O. and Weckwerth, W., 2003, Deciphering metabolic networks, *Eur. J. Biochem.* **270**:579–588.

Hoffmann, M.H., Bremer, M., Schneider, K., Burger, F., Stolle, E., and Moritz, G., 2003, Flower visitors in a natural population of Arabidopsis thaliana, *Plant Biology* **5**:491–494.

Gang, D.R., Wang, J.H., Dudareva, N., Nam, K.H., Simon, J.E., Lewinsohn, E., and Pichersky, E., 2001, An investigation of the storage and biosynthesis of phenylpropenes in sweet basil, *Plant Physiol.* **125**:539–555.

Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P., and Pichersky, E., 2002, Characterization of phenylpropene O-methyltransferases from sweet basil: Facile change of substrate specificity and convergent evolution within a plant OMT family, *Plant Cell* **14**:505–519.

Gershenzon, J., McConkey, M.E., and Croteau, R.B., 2000, Regulation of monoterpene accumulation in leaves of peppermint, *Plant Physiol.* **122**: 205–213.

Iijima, Y., Gang, D.R., Fridman, R., Lewinsohn, E., and Pichersky, E., 2004, Characterization of geraniol synthase from the peltate glands of sweet basil (Ocimum basilicum), *Plant Physiol.* **134**:370–379.

Murfitt, L.M., Kolosova, N., Mann, C.J., and Dudareva, N., 2000, Purification and characterization of S-adenosyl-L-methionine: Benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methyl benzoate in flowers of Antirrhinum majus, *Arch. Biochem. Biophys.* **382**:145–151.

Pichersky, E. and Gershenzon, J., 2002, The formation and function of plant volatiles: perfumes for pollinator attraction and defense, *Curr. Op. Plant Biol.* **5**:237–243.

Raguso, R.A. and Pichersky, E., 1995, Floral volatiles from Clarkia breweri and C. concinna (Onagraceae): recent evolution of floral scent and moth pollination, *Plant Sys. Evol.* **194**:55–67.

Ross, J.R., Nam, K.H., D'Auria, J.C., and Pichersky, E., 1999, S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases, *Arch. Biochem. Biophys.* **367**:9–16.

Seo, H.S., Song, J.T., Cheong, J.J., Lee, Y.H., Lee, Y.W., Hwang, I., Lee, J.S., and Choi, Y.D., 2001, Jasmonic acid carboxyl methyltransferase: A key enzyme for jasmonate-regulated plant responses, *Proc. Natl. Acad. Sci. USA* **98**:4788–4793.

Shulaev, V., Silverman, P., and Raskin, I., 1997, Airborne signalling by methyl salicylate in plant pathogen resistance, *Nature* **386**:738–738.

Trethewey, R.N., 2004, Metabolite profiling as an aid to metabolic engineering in plants, *Curr. Op. Plant Biol.* **7**:196–201.

Van Poecke, R.M.P., Posthumus, M.A., and Dicke, M., 2001, Herbivore-induced volatile production by Arabidopsis thaliana leads to attraction of the parasitoid Cotesia rubecula: Chemical, behavioral, and gene-expression analysis, *J. Chem. Ecol.* **27**:1911–1928.

Wang, J., Dudareva, N., Bhakta, S., Raguso, R.A., and Pichersky, E., 1997, Floral scent Production in Clarkia breweri (Onagraceae). II. Localization and developmental modulation of the enzyme SAM:(Iso)Eugenol-O-methyltransferase and phenylpropanoid emission, *Plant Physiol.* **114**:213–221.

Chapter 13

PROFILING DIURNAL CHANGES IN METABOLITE AND TRANSCRIPT LEVELS IN POTATO LEAVES

Ewa Urbanczyk-Wochniak[1], Charles Baxter[2], Lee J. Sweetlove[2], and Alisdair R. Fernie[1]

[1]Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany; [2]Department of Plant Sciences, South Parks Rd, University of Oxford, Oxford OX1 3RB, UK

Abstract: The availability of sequence data is contributing immensely to the development of gene-expression resources, in parallel to these advances, several methods have been established for systematic analysis of metabolite composition. In this chapter, we illustrate the utility of parallel transcript and metabolic profiling analysis to study metabolic regulation during the day/night cycle. Recently we presented a gas chromatography-mass spectrometry-based metabolic profiling protocol, alongside spectrophotometric techniques, to follow changes in a broad range of potato leaf metabolites throughout the day/night cycle (Urbanczyk-Wochniak et al., 2005). In tandem, we profiled transcript levels using both a custom array containing approximately 2,500 cDNA clones representing predominantly transcripts involved in plant metabolism, and commercially available arrays containing approximately 12,000 cDNA clones that gave coverage of transcript levels over a broader functional range. The levels of many metabolites and transcripts varied during the day/night cycle. Whilst a large number of the differences might be expected based on earlier data, several novel changes were seen. Here we present novel description of changes in metabolites and genes associated with secondary metabolism. Profiling of diurnal patterns of metabolite and transcript abundance in potato leaves suggests that specific sets of metabolic pathway are strongly transcriptionally regulated, but revealed that the majority of the metabolic network is primarily under post-transcriptional control.

1 INTRODUCTION

Diurnal regulation of plant metabolism has been studied for many years, especially as regards metabolites such as carbohydrates like sucrose and starch (Geiger and Servaites, 1994; Kruger, 1997; Fernie and Willmitzer,

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 183–192.

2004) or key compounds in carbon–nitrogen interactions like 2-oxoglutarate, glutamate, and asparagine (Ferrario-Mery et al., 2001; Stitt and Fernie, 2003; Masclaux-Daubresse et al., 2002). Recently, transcript profiling has been used to analyse diurnal regulation at a more global level (Harmer et al., 2000; Thain et al., 2002). Whilst many light cycle-associated differences have been observed in these studies (e.g., phenylpropanoid metabolism, starch degradation, and cell wall elongation), the main focus has been on identifying genes associated with circadian rhythm maintenance. Furthermore, whilst several studies have examined the levels of specific leaf metabolites throughout the day/night cycle (Scheible et al., 2000; Stitt and Fernie, 2003), the development of broad-range approaches have not yet been widely documented for this purpose.

In a recent study, we performed a comprehensive characterization of changes in potato leaf metabolism throughout the diurnal period (Urbanczyk-Wochniak et al., 2005). To achieve this we used a gas chromatography-mass spectrometry (GC-MS)-based method to profile key primary metabolites and some secondary metabolites in leaf samples harvested at time points through a 24 h period. In addition, we used a custom-made *Solanaceous* macroarray, in combination with commercially available cDNA microarrays to profile changes in transcript levels in a subset of the samples used for metabolic profiling. The levels of many metabolites and transcripts varied during the diurnal period with 56 significant differences observed in metabolite contents and 832 significant differences in transcript levels. When analysed together, these results suggest that whilst some minor metabolites appear to correlate closely to changes in transcript levels of associated genes, the majority of tight metabolite regulation is exerted at the post-transcriptional level.

2 RESULTS AND DISCUSSION

For the analysis of metabolites and transcripts, samples of wild type, greenhouse-grown potato plants (*Solanum tuberosum* cv. Desiree) were harvested over a 24 h period (Urbanczyk-Wochniak et al., 2005). To classify a broad range of metabolites we utilized an established gas chromatography, mass spectrometry (GC-MS) protocol (Roessner-Tunali et al., 2003), with which we were able to quantify the content of over 70 metabolites. Diurnal changes in major and minor carbohydrates and amino and organic acids were found to be largely similar to those previously documented for tobacco or Arabidopsis plants (Masclaux-Daubresse et al., 2002; Carrari et al., 2005). Furthermore, we were able to analyse a few secondary metabolites, such as caffeate and chlorogenate.

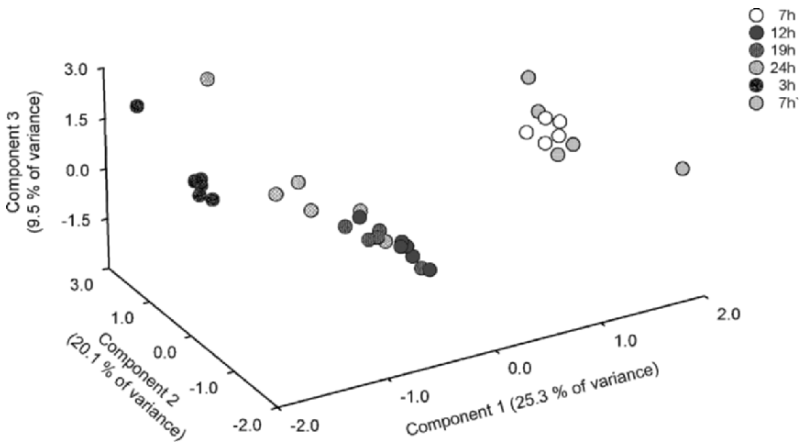


Figure 13-1. Principal component analysis (PCA) of metabolite profiles of samples harvested from wild-type potato leaves during a diurnal period. The distances between these populations were calculated as described in Roessner *et al*. (2001). The percentage of variance explained by each component is shown in parenthesis. Samples representing leaves collected at 07:00 (white and medium grey circle), 12:00 (dip dark grey circle), 19:00 (medium striped grey circle), 00:00 (light dotted grey circle), and 03:00 (black circle). Each data point represents an independent sample. With kind permission of the publisher Springer-Verlag GmbH (permission applied for).

In addition to point-by-point analysis of metabolites, the statistical tool, principal component analysis (PCA) was applied to the complete data set (Figure 13-1). As could be expected, two sets of samples harvested at the same time on consecutive days clustered together. Also, samples harvested at different time points, especially those collected during the light period (12:00 and 19:00), are easily resolved from samples harvested during the dark period (24:00 and 03:00) and indeed periodicity can be observed. This result suggests a tight coordination of metabolism throughout the diurnal period.

Having established the pattern of diurnal changes in metabolite levels in potato leaf, attention was turned to assessing transcript levels by performance of two sets of microarray experiments on identical plant material from a subset of the samples described above (Urbanczyk-Wochniak et al., 2005). In the combined data set, 455 clones were upregulated at 03:00 (146 of which were exclusively expressed at this time point); 18 of these clones are associated with amino acid metabolism, 36 with carbohydrate metabolism, 17 with cell wall metabolism, 42 with photosynthesis, and 14 with secondary metabolism. Conversely, 377 clones were downregulated (79 being exclusively expressed in the light period),

including 6 associated with amino acid metabolism, 10 with carbohydrate metabolism, 14 with cell wall metabolism, 53 with photosynthesis, and 5 with secondary metabolism).

As mentioned above, diurnal changes in amino acid contents observed here in potato were largely consistent with those previously found in tobacco (Matt et al. 1998; Ferrario-Mery et al., 2002), tomato (Carrari et al., 2003) and Arabidopsis (Carrari et al., 2005), with the majority of amino acids increasing during the day and decreasing during the dark period. As one might expect, the patterns of increase or decrease were largely conserved within those metabolites sharing common precursors. The majority of amino acid metabolism associated genes that displayed altered transcript levels were upregulated during the night. However, in general, changes in gene expression did not result in detectable increases in the abundance of associated amino acids.

The observed diurnal variations in carbohydrate content were typical for growth under a long-day regime in a range of species (see for example Matt et al., 1998; Lytovchenko et al., 2002; Ferrario-Mery et al., 2002; Chia et al., 2004). Variations in minor carbohydrate levels largely mirrored the pattern of the major sugars with significant increases in arabinose, mannose, fucose, ribose, rhamnose, and xylose during the light period. Furthermore, although the pattern of change in pool size of the minor carbohydrates was highly varied, all carbohydrates declined during the dark period.

Organic acids also showed similar trends to those reported previously (Scheible et al., 2000; Müller et al., 2001), with a tendency toward increasing organic acid contents upon the induction of the trycarboxylic acid cycle (TCA cycle) in the dark period. This is not true, however, for all organic acids; malate levels, for example, showed a completely different diurnal pattern.

In addition to the data discussed above, some secondary metabolites also display diurnal variation (Figure 13-2). Application of our GC-MS profiling method allowed the evaluation of changes in the levels of a handful of soluble secondary metabolites to be tracked throughout the course of the experiment. Despite quite some interest in these compounds (see for example Howles et al., 1996; Guo et al., 2001; Chen et al., 2003), such data has not been previously reported. This analysis revealed a minor, but statistically significant increase in the levels of chlorogenate, the major soluble phenylpropanoid in Solanaceous species (Matt et al., 2002), and of caffeate during the dark period. Conversely, levels of quinate and glycerate went down during the dark period (however, caution must be taken in integrating this data since glycerate levels determined by our method could include a large proportion of dephosphorylated 3-phosphosphoglycerate).

We next evaluated changes in the genes associated with secondary metabolism (Figures 13-3 and 13-4) and found that far more of them were

upregulated rather than downregulated in the dark period (Urbanczyk-Wochniak et al., 2005). Transcripts encoding the enzymes cinnamic acid

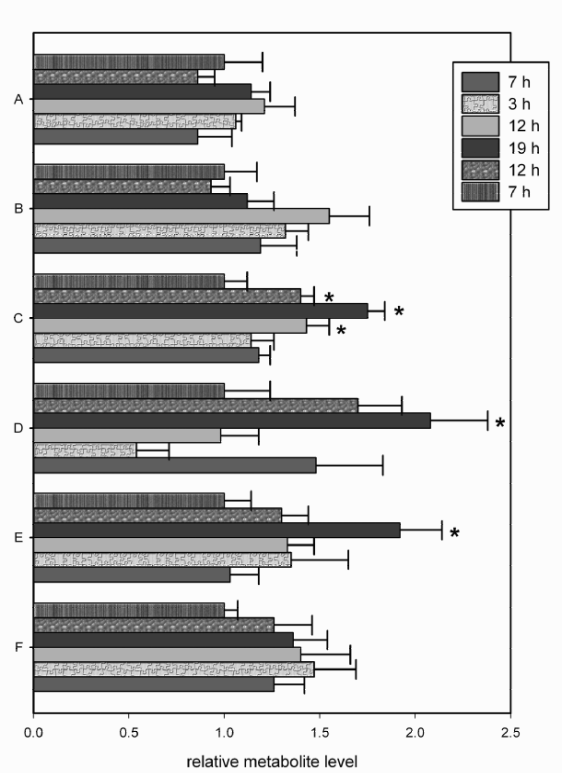


Figure 13-2. Diurnal changes in metabolites: chlorogenate (A), caffeate (B), quinate (C), glycerate (D), phenylalanine (E), tryptophan (F). Metabolites were determined as described in Urbanczyk-Wochniak et al. (2005). At each time point, samples were taken from mature source leaves and the data represent the mean ± SE of measurements of six plants. *Represents values that are significantly different from the first sampling point.

4-hydroxylase and phenylalanine ammonia-lyase which are involved in phenylpropanoid metabolism were elevated in the dark, as were those encoding the essential enzymes in the carotenoid biosynthetic pathway – namely, phytoene synthase and isopentenyl diphosphate isomerase.

Whilst we could not ascribe a direct mechanism for the changes in metabolite pool sizes through the course of the experiment, analysis of the differences in metabolite and transcript levels between the day and night allow us to construct several hypotheses.

First, these modulations may merely reflect the relative availability of precursor metabolites or cofactors during the course of the experiment. In keeping with this interpretation, levels of phenylalanine and tryptophan peak

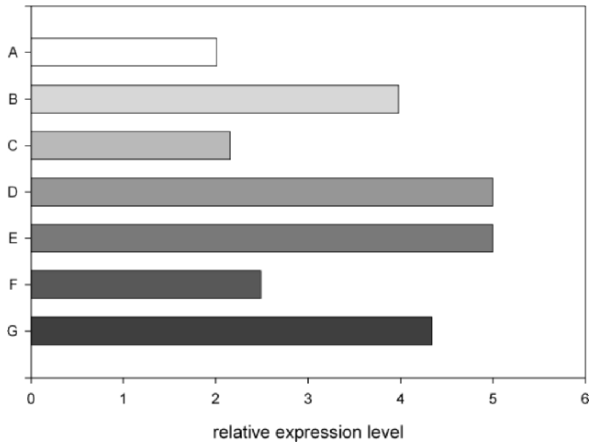


Figure 13-3. Ratio of transcription of selected genes between 3 A.M. determined by microrarray analysis. A) relative expression level of acyltransferase B) relative expression level of anthocyanin 3'-glucosyltransferase); C) relative expression level of caffeic acid O-methyltransferase; D) relative expression level of glucosyl transferase (signal was detected exclusively only at that particular time point); E) relative expression level of orcinol O-methyltransferase (signal was detected exclusively only at that particular time point) F) relative expression level of phenylalanine ammonia-lyase; G) relative expression level of phytoene synthase.

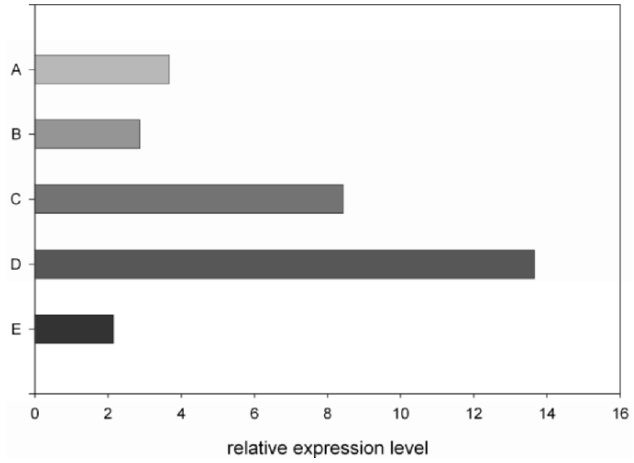


Figure 13-4. Ratio of transcription of selected genes between 12 P.M. determined by microrarray analysis. A) relative expression level of 4-hydroxyphenylpyruvate dioxygenase; B) relative expression level of chalcone synthase; C) relative expression level of gamma hydroxybutyrate dehydrogenase; D) relative expression level of neoxanthin synthase; E) relative expression level of uroporphyrinogen decarboxylase, chloroplast precursor.

at the end of the light period directly preceding the minor increases in caffeate and chlorogenate. Second, these changes could be direct effects of up- or downregulation of anabolic or catabolic pathways of phenylpropanoid metabolism. Such changes have been reported previously for several genes associated with circadian clock regulation (Harmer et al., 2000). Similarly, in this study, the observed changes in secondary metabolism-associated transcript levels included genes from the phenylpropanoid pathway (clones encoding phenylalanine ammonia-lyase and cinnamic acid 4-hydrolase), as well as two different clones encoding phytoene synthase. A third possibility is that the changes in these metabolites are an indirect effect of the modulation of transcript levels from a closely associated pathway during the night. One possible candidate would be transketolase, which is upregulated in the night, since it has been demonstrated previously that the activity of this enzyme is positively correlated to the chlorogenate content in the illuminated leaves of transgenic plants (Henkes et al., 2001). Given that transgenic experiments in tobacco have indicated that phenylalanine ammonia-lyase exhibits a large degree of control in the synthesis of chlorogenate (Howles et al., 1996), we tend to favour the second explanation. However, it is clear that further investigations are required to elucidate the exact mechanisms underlying the temporal changes in these metabolite levels.

3 CONCLUSION

There are still only a few applications of metabolomics and transcriptomic studies to plant metabolism of using both approaches in parallel (Urbanczyk-Wochniak et al., 2003; Hirai et al., 2004; Oksman-Caldentey et al., 2004). The data presented previously (Urbanczyk-Wochniak et al., 2005) and above provide a relatively comprehensive, although by no means complete, analysis of changes in metabolite and transcript levels over a diurnal period. Although a full genome chip has recently been described for rice (Zhu et al., 2003), no such tool is available yet for *Solanaceous* species. Given the obvious limitations of an incomplete array here we decided to characterize metabolite levels in parallel in order to give us greater confidence in the data obtained from the transcript profiling experiments. Qualitative comparison of the combined data sets obtained from the parallel analysis of transcripts and metabolites suggests that relatively few changes in transcript levels correlate strongly with changes in metabolite levels during the day/night cycle. The changes that occur are associated primarily with central metabolism. In contrast, principal component analysis of metabolite profiles revealed that the levels of many metabolites change progressively throughout the day/night cycle. These

results suggest that although leaf metabolism is regulated tightly throughout the cycle, this regulation is exerted primarily at the post-transcriptional level. Intriguingly, this appears to be a common motif in nature with similar patterns of regulation being observed in *Escherichia coli* (Almaas et al., 2004) and *Saccharomyces cerevisiae* (Daran-Lapujade et al., 2004).

ACKNOWLEDGMENTS

We thank Dr. Joachim Kopka for the discussion of various aspects of bioinformatics analysis. LJS acknowledges the financial support of the BBSRC, UK. EU-W and ARF acknowledge the financial support of the Max Planck Society. We are grateful to Megan McKenzie for her critical revision of this chapter.

REFERENCES

Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z.N., and Barabasi, A-L., 2004, Global organization of metabolic fluxes in the bacterium Escherichia coli, *Nature* **427**:839–843.

Carrari, F., Coll-Garcia, D., Schauer, N., Lytovchenko, A., Palacios-Rojas, N., Balbo, I., Rossi, M. and Fernie, A. R., 2005, Deficiency of a plastidial adenylate kinase in Arabidopsis results in elevated photosynthetic amino acid biosynthesis and enhanced growth, *Plant Physiology* **137**:70–82.

Carrari, F., Nunes-Nesi, A., Gibon, Y., Lytovchenko, A., Ehlers-Lourario, M., and Fernie, A. R., 2003, Reduced expression of aconitase results in an enhanced rate of photosynthesis and marked shifts in carbon partitioning in illuminated leaves of Lycopersicon pennellii, *Plant Physiology* **133**:1322–1335.

Chen, F., Duran, A.L., Blount, J.W., Sumner, L.W. and Dixon, R.A., 2003, Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis, *Pytochemistry* **64**:1013–1021.

Chia, T., Thorneycroft, D., Chapple, A., Messerli, G., Chen, J., Zeeman, S.C., Steven, M., Smith, S.M., and Smith, A.M., 2004, A cytosolic glucosyltransferase is required for conversion of starch to sucrose in Arabidopsis leaves at night, *The Plant Journal* **37**:853–863.

Daran-Lapujade, P., Jansen, M.L.A., Daran, J-M., van Gulik, W., de Winde, J.H., and Pronk, J.T., 2004, Role of Transcriptional Regulation in Controlling Fluxes in Central Carbon Metabolism of Saccharomyces cerevisiae: a chemostat culture study, *J Biol. Chem.* **279**:9125–9138.

Fernie, A.R. and Willmitzer, L., 2004, Carbohydrate metabolism, in: *The Handbook of Plant Biotechnology*, Christou, P. and Klee, H.K., eds., Wiley, Chichester, UK., in press.

Ferrario-Mery, S., Masclaux, C., Suzuki, A., Valadier, M.H., Hirel, B., and Foyer, C.H., 2001, Glutamine and alpha-ketoglutarate are metabolite signals involved in nitrate reductase gene transcription in untransformed and transformed tobacco plants deficient in ferredoxin-glutamine-alpha-ketoglutarate aminotransferase, Planta **213**:265–271.

Ferrario-Mery, S., Hodges, M., Hirel, B., and Foyer, C. H., 2002, Photorespiration-dependent increases in phosphoenolpyruvate carboxylase, isocitrate dehydrogenase and glutamate dehydrogenase in transformed tobacco plants deficient in ferredoxin-dependent glutamine-alpha-ketoglutarate aminotransferase, *Planta* **214**:877–886.

Geiger, D. R., and Servaites, J. C., 1994, Diurnal regulation of photosynthetic carbon metabolism in C-3 plants, *Annual Review of Plant Physiology and Plant Molecular Biology* **45**:235–256.

Guo, D.J., Chen. F., Inoue, K., Blount. J.W., and Dixon, R.A., 2001, Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: Impacts on lignin structure and implications for the biosynthesis of G and S lignin, *Plant Cell* **13**:73–88.

Harmer, S. L., Hogenesch, L. B., Straume, M., Chang, H. S., Han, B., Zhu, T., Wang, X., Kreps, J. A., and Kay, S. A., 2000, Orchestrated transcription of key pathways in Arabidopsis by the circadian clock, *Science* **290**:2110–2113.

Henkes, S., Sonnewald, U., Badur, R., Flachmann, R. and Stitt, M. A., 2001, Small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism, *Plant Cell* **13**:535–551.

Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., and Saito, K., 2004, Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana, *PNAS* **101**:10205–10210.

Howles, P. A., Sewalt, V. J. H., Paiva, N. L., Elkind, Y., Bate, N. J., Lamb, C. and Dixon, R. A., 1996, Overexpression of L-phenylalanine ammonia-lyase in transgenic tobacco plants reveals control points for flux into phenylpropanoid biosynthesis, *Plant Physiology* **112**:1617–1624.

Kruger, N.J., 1997, Carbohydrate synthesis and degradation, in: *Plant Metabolism*, Dennis, D.T., Turpin, D. H., Lefebvre, D. D., and Layzell, D. B., eds., Harlow, UK, Longman, pp. 83–104.

Lytovchenko, A., Sweetlove, L.J., Pauly, M., and Fernie, A.R., 2002, The influence of cytosolic phosphoglucomutase on photosynthetic carbohydrate metabolism, *Planta* **215**:1013–1021.

Masclaux-Daubresse, C., Valadier, M.H., Carrayol, E., Reisdorf-Cren, M., and Hirel, B., 2002, Diurnal changes in the expression of glutamate dehydrogenase and nitrate reductase are involved in the C/N balance of tobacco source leaves, *Plant Cell and Environment* **25**:1451–1462.

Matt, P., Krapp, A., Haake, V., Mock, H. P., and Stitt, M., 2002, Decreased Rubisco activity leads to dramatic changes of nitrate metabolism, amino acid metabolism and the levels of phenylpropanoids and nicotine in tobacco antisense RBCS transformants, *Plant Journal* **30**:663–677.

Matt, P., Schurr, U., Klein, D., Krapp, A., and Stitt, M., 1998, Growth of tobacco in short-day conditions leads to high starch, low sugars, altered diurnal changes in the Nia transcript and low nitrate reductase activity, and inhibition of amino acid synthesis, *Planta* **207**: 27–41.

Müller, C., Scheible, W. R., Stitt, M., and Krapp, A., 2001, Influence of malate and 2-oxoglutarate on the NIA transcript level and nitrate reductase activity in tobacco leaves, *Plant Cell and Environment* **24**:191–203.

Oksman-Caldentey, K.-M., Inze, D., and Oresic, M., 2004, Connecting genes to metabolites by a systems biology approach, *PNAS* **101**:9949–9950.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., Fernie, A., 2001, Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems, *Plant Cell* **13**:11–29.

Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D., Fernie, A. R., 2003, Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development, *Plant Physiology* **133**:84–99.

Scheible, W.-R., Krapp, A., and Stitt, M., 2000, Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate

synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves, *Plant Cell and Environ.ment* **23**:1155–1168.

Stitt, M. and Fernie, A.R., 2003, From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions, *Current Opinion in Biotechnology* **14**:136–144.

Thain, S.C., Murtas, G., Lynn, J.R., McGrath, R. B., and Millar, A.J., 2002, The circadian clock that controls gene expression is tissue specific, *Plant Physiol.ogy* **130**:102–110.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., and Fernie, A.R., 2003, Parallel analysis of transcript and metabolic profiles: a new approach in systems biology, *EMBO Rep.* **4**:989–93.

Urbanczyk-Wochniak, E., Baxte, C.J., Kolbe, A., Kopka, J, Sweetlove, L.J., and Fernie, A.R., 2005, Profiling of diurnal patterns of metabolite and transcript abundance in potato (Solanum tuberosum) leaves, *Planta* **221**:891–903.

Zhu, T., 2003, Global analysis of gene expression using GeneChip microarrays, *Curr. Opin. Plant Biol.* **6**:418–425.

Chapter 14

GENE EXPRESSION AND METABOLIC ANALYSIS REVEAL THAT THE PHYTOTOXIN CORONATINE IMPACTS MULTIPLE PHYTOHORMONE PATHWAYS IN TOMATO

Srinivasa Rao Uppalapati and Carol L. Bender
Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

Abstract: Coronatine (COR) is a phytotoxin produced by several pathovars of *Pseudomonas syringae* and consists of coronafacic acid (CFA), an analogue of methyl jasmonic acid (MeJA), and coronamic acid (CMA), which resembles 1-aminocyclopropane-1-carboxylic acid (ACC), a precursor to ethylene. An understanding of how COR functions, is perceived by different plant tissues, and the extent to which it mimics MeJA remain unclear. In this study, COR and related compounds were examined with respect to structure and function. cDNA microarrays were utilized to understand the molecular processes that are regulated by MeJA, COR, CFA, and CMA in tomato leaves. A comparison of COR- and MeJA-regulated transcriptomes revealed that COR regulated 35% of the MeJA-induced genes. There was significant overlap in the number of COR and CFA-regulated genes with CFA impacting the expression of 39.4% of the COR-regulated genes. Collectively, our results demonstrate that: (1) the intact COR molecule impacts signaling in tomato *via* the jasmonic acid, ethylene, and auxin pathways; (2) CMA does not function as a structural analogue of ACC; (3) COR has a broader range of functions than either CFA or CMA; and (4) COR and MeJA share similar, but not identical activities and impact multiple phytohormone pathways in tomato.

1 INTRODUCTION

Coronatine (COR) is a non-host-specific phytotoxin produced by several pathovars of *Pseudomonas syringae* (Bender et al., 1999; Mitchell, 1982). The toxin acts as a virulence factor in *P. syringae* pv. *tomato*, allowing the pathogen to obtain higher population densities and develop larger lesions than COR-defective strains (Bender et al., 1987; Brooks et al., 2004; Mittal and Davis, 1995). In addition to chlorosis, COR induces a wide array of

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 193–208.

effects in plants including anthocyanin production, alkaloid accumulation, ethylene emission, accumulation of proteinase inhibitors, tendril coiling, inhibition of root elongation, and hypertrophy (Bender et al., 1999; Feys et al., 1994; Lauchli and Boland, 2003; Weiler et al., 1994).

COR consists of the polyketide coronafacic acid (CFA) (Parry et al., 1994), and coronamic acid (CMA), a cyclized derivative of isoleucine (Mitchell, 1985). CFA and CMA function as biosynthetic intermediates and are joined together by an amide linkage to form the parent compound, COR (Ichihara et al., 1977) (Figure 14-1a). CMA is a structural analogue of 1-aminocyclopropane-1-carboxylic acid (ACC) (Figure 14-1b), an intermediate in the pathway to ethylene in higher plants (Ecker, 1995). It has also been noted that COR is a structural and functional analogue of jasmonic acid (JA) and related signalling compounds such as methyl jasmonate (MeJA) and 12-oxo-phytodienoic acid (12-OPDA), the C_{18} precursor of JA/MeJA (Feys et al., 1994; Weiler et al., 1994). OPDA, JA, MeJA (Figure 14-1b), and other octadecanoids impact the regulation of diverse plant responses including biotic stress (Farmer et al., 2003), wounding (Howe and Schilmiller., 2002), abscission (Burns et al., 2003), and volatile production (Weber, 2002). The identification of the Arabidopsis *coi1* (*co*ronatine *i*nsensitive) and a JA insensitive mutant (*jai1*) of tomato mutant further supports the hypothesis that COR is a functional analogue of MeJA (Feys et al., 1994; Li et al., 2004).

Although COR is involved in various physiological responses, we do not understand how COR is perceived in different tissues, precisely how it

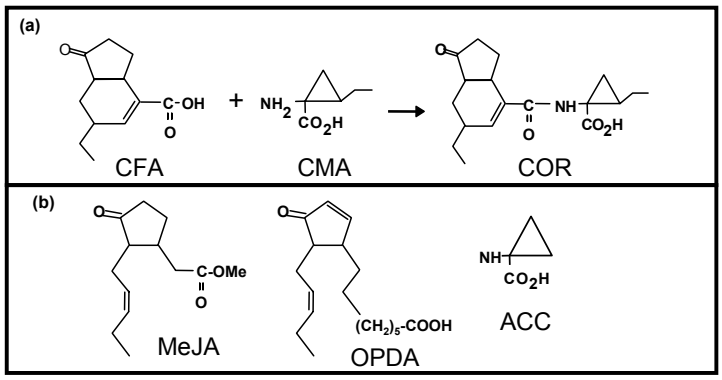


Figure 14-1. Structures of COR derivatives and analogues discussed in the text. (a) CFA, coronafacic acid; CMA, coronamic acid; and COR, coronatine. (b) MeJA, methyl jasmonate; OPDA, 12-oxo-phytodienoic acid; and ACC, 1-aminocyclopropane-1-carboxylic acid.

functions, and to what extent it mimics MeJA. Previous reports have documented the production of ethylene in COR-treated tissue (Ferguson and Mitchell, 1985; Kenyon and Turner, 1992), a response that may be attributed to the structural similarities between CMA and ACC. In the present study, we investigate whether each component of COR has biological function *in planta*. We set out to answer these questions using cDNA microarrays to identify the molecular responses associated with COR, CFA and CMA.

2 RESULTS

2.1 Visual observations of treated tomato leaves

COR-treated (20 nmol) leaves exhibited moderate to severe yellowing with chlorosis spreading 5–10 mm from the application site five days after treatment. Plants treated with CMA or ACC showed slight burning at the inoculation site, but were not chlorotic. No chlorosis or burning was observed on plants treated with H_2O, CFA, or MeJA (Table 14-1). Previously, it was reported that COR inhibits root growth and induces anthocyanin accumulation in Arabidopsis (Feys et al., 1994). We wondered whether this response was unique to Arabidopsis, or whether it also occurs in tomato. COR, CFA, and MeJA each inhibited root growth and induced anthocyanin accumulation in seedlings (Table 14-1; Uppalapati et al., 2005). ACC induced root inhibition and a typical "triple response" in tomato seedlings (reduced elongation, thickened hypocotyl, and thickened apical hook) at 0.2 and 200 μmol (Table 14-1; Uppalapati et al., 2005). Unlike ACC, CMA did not inhibit root growth or induce a triple response (Table 14-1).

In summary, COR was much more effective in inhibiting root growth and inducing anthocyanin accumulation than CFA and MeJA. In contrast, CMA and ACC were relatively ineffective in stimulating anthocyanin accumulation (Table 14-1; Uppalapati et al., 2005). ACC (but not CMA) inhibited root growth but only at levels 10,000-fold higher than COR (Table 14-1; Uppalapati et al., 2005), suggesting that CMA does not behave as a functional analogue of ACC in these assays.

2.2 cDNA microarray analysis of COR, CFA CMA, and MeJA treated tomato leaf tissues

The identification of COR, CFA, CMA, and MeJA-responsive genes offers an opportunity for studying the potential functions of these compounds; therefore, we conducted gene expression profiling in tomato tissue treated with these compounds.

Table 14-1. Effects of COR derivatives and analogues on tomato leaf tissue and seedlings

Compound	*Chlorosis*[a]	*Root inhibition*[b]	*Anthocyanin accumulation*[c]
COR	+++	+++	+++
CFA	ND	++	++
MeJA	ND	++	++
CMA	ND	none	–
ACC	ND	+++	–

[a] +++, chlorotic zone was 5–10 mm in diameter; and ND, no detectable chlorosis.
[b] (+++) ≥ 70% and (++) ≥ 40% root growth inhibition in comparison to untreated controls.
[c] (+++) ≥ 7% and (++) ≥ 5% anthocyanin accumulation/mg fresh weight in comparison to untreated controls.

The 12 h time point was chosen because maximal differential expression was observed at this time point when compared to H_2O-inoculated control tissue. Differential regulation of gene expression in COR, CFA, and CMA-treated tissue was compared using Venn diagrams (Figure 14-2). The complete list of genes and expression ratios are presented elsewhere (Uppalapati et al., 2005). When a twofold induction ($P < 0.05$) relative to the control was used, 256, 231, and 143 genes were identified as induced by COR, CFA, and CMA, respectively (Figure 14-2). The largest set of upregulated genes were those induced by COR (256 genes; Figure 14-2).

The identification of genes downregulated by COR, CFA, and CMA is equally important in understanding the response to these compounds. When a twofold decrease ($P < 0.05$) in expression was applied as a cut-off, 274, 292, and 176 genes were identified as downregulated by COR, CFA, and CMA, respectively (Figure 14-2).

These results indicate the greatest degree of overlap exists between COR- and CFA-regulated genes. However, COR was more active in regulating the genes investigated in this study, and the COR holo-toxin regulates a greater array of genes in tomato than CFA or CMA.

Our observations using cell biology demonstrated that COR induces chlorosis and alters the structure of the chloroplast (Table 14-1, Uppalapati et al., 2005). Consistent with our observations, COR downregulated a large number of genes belonging to chloroplast metabolism (e.g., genes encoding chlorophyll a/b binding proteins, nicotinamide adenine dinucleotide phosphate-oxidase (NADPH):protochlorophyllide oxidoreductase, thylakoid luminal proteins) (Figure 14-3, group I). Interestingly, MeJA did not induce any visible chlorosis (Table 14-1) and was less active than COR and CFA in repressing genes involved in chloroplast metabolism (Figure 14-3, group I).

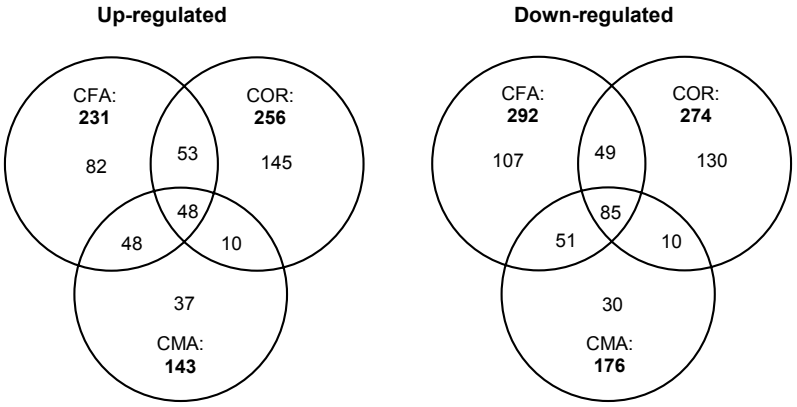


Figure 14-2. Venn diagrams showing the number of COR, CFA, and CMA (12 h post-treatment) regulated genes. Numbers in bold font denote the total number of genes regulated by each compound.

JA biosynthesis is known to be regulated by a JA-mediated positive feedback loop (Sasaki et al., 2001; Stenzel et al., 2003). Consistent with observations in Arabidopsis, MeJA positively stimulated genes involved in JA biosynthesis and JA responsiveness in tomato (Figure 14-3, groups II and III). COR was more active than MeJA, CFA, or CMA in upregulating JA biosynthesis genes, including lipoxygenase (*LOXD*), allene oxide cyclase (*AOC*), and oxophytodienoate reductase (*OPR3*) (Figure 14-3, group II).

Several JA/wound responsive genes were induced by MeJA, COR, CFA, and CMA (e.g., wound-inducible serine proteinase inhibitors I and II) (Figure 14-3, group III). However, CMA was generally less active in inducing this group of genes, especially polyphenol oxidase and multicystatin.

COR stimulates ethylene production in both bean and tobacco, leading to speculation that the CMA portion might stimulate ethylene production (Ferguson and Mitchell, 1985; Kenyon and Turner, 1992). COR strongly induced genes involved in ethylene biosynthesis and/or ethylene-responsiveness (Figure 14-3, group IV), suggesting that the holotoxin can directly stimulate ethylene production. In contrast, these genes were not modulated by CMA. These results indicate that CMA does not stimulate ethylene production or functionally mimic ACC. Furthermore, our cDNA microarray experiments identified genes that were not previously known to be regulated by COR. For example, COR induced the expression of a set of auxin-related genes, including IAA-conjugate hydrolases (e.g., *IAR3*) and an auxin-regulated protein (Figure 14-3).

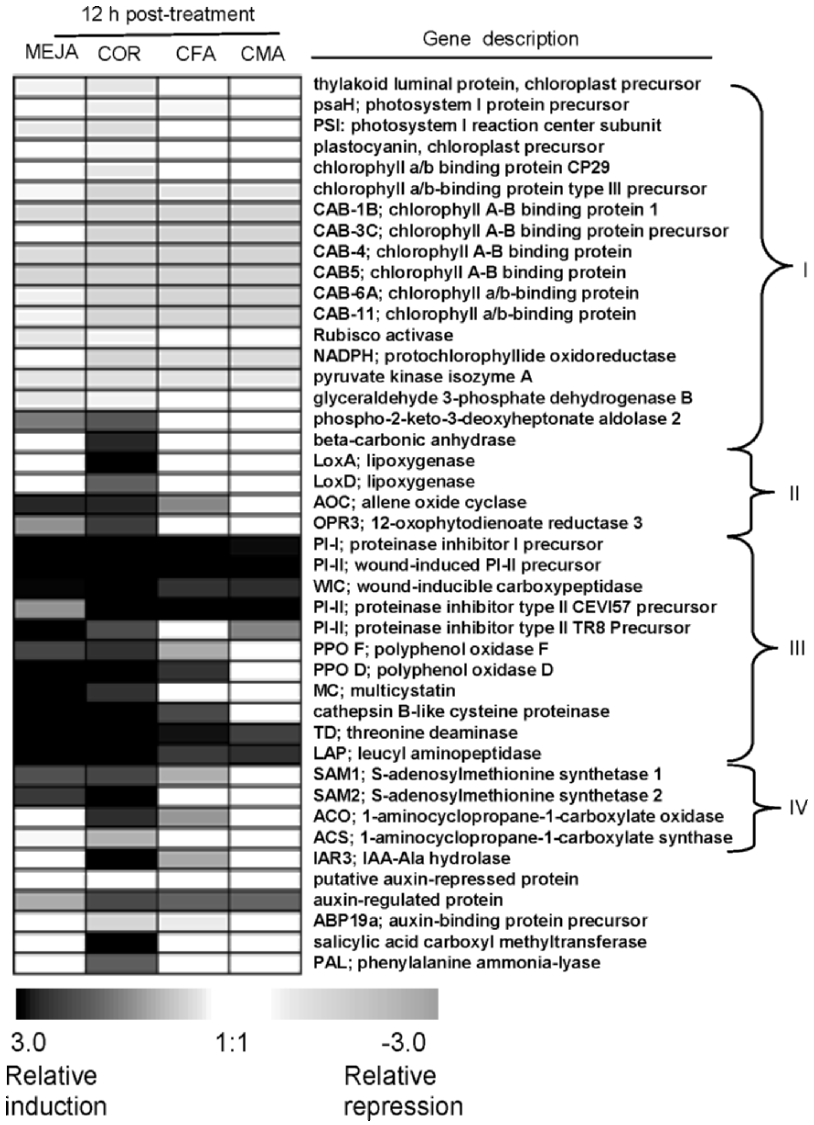


Figure14-3. Expression diagram of selected genes that are differentially regulated by MeJA, COR, CFA, or CMA. Functional groups discussed in text include: group I, genes belonging to chloroplast metabolism; group II, genes involved in oxylipin biosynthesis; group III, JA and/or wound-responsive genes; group IV, genes involved in ethylene biosynthesis.

2.3 Metabolic analysis of COR-treated tomato leaf tissues

Our transcriptional profiling studies have shown that COR imparts multiple phytohormones in tomato leaf tissues. To further confirm these results we have utilized metabolic analysis. In this study, we observed that treatment with COR stimulates the accumulation of endogenous levels of JA in tomato leaves (Figure 14-4). Thus both COR and endogenous MeJA may modulate genes involved in JA biosynthesis. Simultaneously, increased accumulation of COR and MeJA were previously reported during infection of Arabidopsis by *P. syringae pv. tomato* (Schmelz et al., 2003).

Furthermore, we show that the treatment of tomato leaves with COR stimulated the accumulation of endogenous levels of IAA in tomato leaves (Figure 14-4). These results suggest that COR may increase the free IAA levels within the plant to increase virulence. Taken together, transcriptional and metabolic analysis showed that COR impacts multiple phytohormone pathways in tomato.

2.4 Transcriptional profiling of COR and MeJA in tomato leaf tissues

To further investigate to what extent COR modulates gene expression in a manner similar to MeJA, transcriptome analysis of MeJA-treated tissues was performed and compared with COR-treated tomato tissue. A total of 256 and 320 genes were identified as induced by COR and MeJA, respectively (Figure 14-5a). Approximately 40% (128) of the MeJA-induced genes were

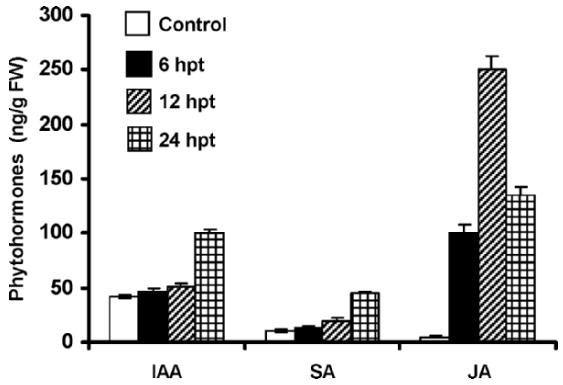


Figure 14-4. Changes in levels of SA, JA, and IAA in tomato leaf tissue treated with COR (mean ± SD, n = 3).

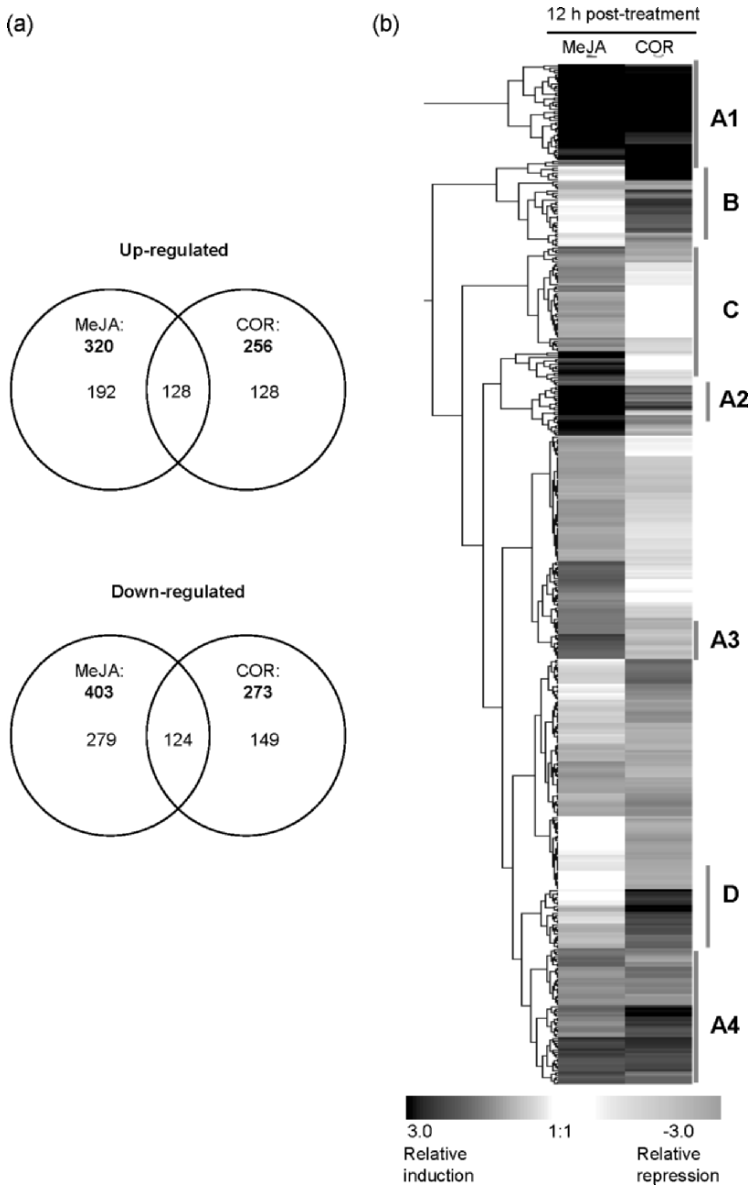


Figure 14-5. Transcriptional profiles of COR- and MeJA-treated tomato leaf tissues. (a) Venn diagram showing the total number of genes regulated by COR and MeJA at 12 h. Numbers represent genes with ≥twofold induction or repression in response to COR or MeJA. Bold font indicates the total number of genes regulated by each compound. (b) Hierarchical clustering of COR or MeJA inducible gene expression patterns in treated leaf tissues at 12 h post-inoculation.

also induced by COR, and 30.7% (124) of the MeJA-repressed genes were downregulated by COR (Figure 14-5a).

In addition to identifying many unknown genes as MeJA/COR-responsive (Figure 14-5b; Uppalapati et al., 2005), our results correlate well with previously reported JA or wound-regulated transcriptomes (Schenk et al., 2000; Sasaki et al., 2001; Cheong et al., 2002; Goossens et al., 2003). Based on average linkage hierarchical clustering, we selected four groups for further comparison of COR- and MeJA-induced genes (Figure 14-5b, A–D). Cluster A (A1–A4) consists of genes induced by both COR and MeJA. This group consisted of genes involved in JA responsiveness and JA biosynthesis (Figure 14-5b) and includes genes encoding leucine aminopeptidase, wound-inducible serine proteinase inhibitors I and II, multicystatin, polyphenol oxidase, threonine deaminase, AOC, and OPR3. Interestingly, many genes implicated in wound and/or cellular signalling also clustered with the JA-responsive genes.

Both COR and MeJA induced a gene encoding cysteine protease (Figure 14-5b, A4), an enzyme implicated in pathogen-induced cell death (Navarre and Wolpert., 1999). This may be relevant in the context of nutrient pools in the apoplast; for example, the release of nutrients from dying cells may facilitate pathogen multiplication in the necrogenic stage of its life cycle. Furthermore, COR and MeJA impacted the expression of genes involved in polyamine biosynthesis (Figure 14-5b, clusters A3 and A4). Similarly, MeJA is known to alter polyamine metabolism in barley (Walters et al., 2002). Cluster B consisted of genes induced by COR that were either suppressed or not differentially expressed in response to MeJA, including genes encoding lipoxygenase, auxin-related protein, NAC domain protein (NAC2), JA transcription factor 2, receptor-like protein kinase, homeobox protein 1, and abnormal inflorescence meristem 1. Cluster C consisted of genes induced by MeJA, but repressed in COR-treated tissues; this cluster included *AIR12*, which is involved in lateral root development, and a tuberization-related gene. Cluster D consisted of genes that were induced by COR and were not differentially regulated by MeJA. This cluster contained genes potentially involved in the ubquitin-proteasome pathway, including ubiquitin-related protein (RUB1) and ubiquitin-conjugating enzyme. Genes involved in ethylene (*ACO1*) and auxin metabolism (*IAR3*) were also represented in this cluster.

To summary, the transcriptional profiles of MeJA and COR-treated tomato leaf tissues showed substantial overlap with respect to genes involved in JA biosynthesis and JA signaling, ethylene biosynthesis, and auxin metabolism. Functional analysis and determination of the biological relevance of the novel genes identified in this study will help us understand how COR and MeJA function in tomato.

3 DISCUSSION

3.1 Comparison of CFA, CMA, COR, and MeJA: tomato seedling assays and transcript profiling

In biological assays with tomato leaf tissue and seedlings, CFA and MeJA induced proteinase inhibitors (Uppalapati et al., 2005), stimulated anthocyanin production, and inhibited root growth in tomato (Table 14-1). In transcript profiling experiments, we observed that MeJA and CFA regulated most of the JA-responsive genes (Figure 14-3), but were generally less active than COR in inducing the above-mentioned activities (Table 14-1; Figure 14-3).

In this study, COR and MeJA, but not CMA, induced genes involved in ethylene biosynthesis and ethylene responsiveness (Figure 14-3). Ethylene plays an important role in the symptoms associated with bacterial speck of tomato and soybean (Lund et al., 1998; Weingart et al., 2001). In our transcript profiling experiments, COR induced genes associated with ethylene biosynthesis and responsiveness (Figure 14-3), suggesting that COR may modulate ethylene as a virulence strategy.

COR also induced the expression of a set of auxin-related genes (Figure 14-3), implying that auxin levels also play an important role in pathogenesis (Kunkel et al., 2004). In a study using potato tubers and mung bean hypocotyls, Sakai et al. (1979) concluded that auxin and COR have different primary sites of action but ultimately target the same physiological activities. Similarly, our results suggest that the COR-induced JA pathway may positively regulate auxin responses in tomato. This is consistent with the hypothesis that JA and auxin may function *via* a common signaling intermediate that modulates response to multiple plant hormones (Devoto and Turner, 2003).

Our results show that COR modulates genes involved in the pathways to JA, ethylene, and auxin. This raises an interesting question: should COR be considered a phytotoxin or a phytohormone mimic? It is not surprising that COR targets these particular phytohormone pathways, as both ethylene and JA are known to positively regulate susceptible interactions between tomato/Arabidopsis and *P. syringae* (Kunkel and Brooks, 2002). A popular hypothesis is that COR may act as a suppressor of defence response(s), possibly by suppressing salicylic acid-dependent defences in Arabidopsis and tomato (Kloek et al., 2001; Zhao et al., 2003) Mutual antagonism between JA- and SA-mediated defense pathways is well documented (Kunkel and Brooks, 2002); consequently, COR may stimulate the JA pathway at the expense of SA-dependent defense responses.

3.2 Comparison of COR- and MeJA-regulated transcriptional changes

One outcome of the present study was the identification of gene sets that respond differentially to COR and MeJA, supporting the contention that they have different activities based on the comparison of the expression profiles at a single time-point (Figures 14-3 and 14-5). Interpretation of these changes is complicated due to the differences in the kinetics of induction and by the fact that both primary and secondary transcriptional changes occur following exposure to MeJA or COR. For example, two recent reports document the existence of JA-modifying enzymes, including a MeJA esterase and a JA amino acid synthetase (Staswick and Tiryaki, 2004; Stuhlfelder et al., 2004). Presumably, a MeJA esterase could cleave exogenous MeJA to form JA, which could be further metabolized by a JA amino acid synthetase to form JA amino acid conjugates. These metabolized products of MeJA, along with the other MeJA-induced phytohormones (e.g., ethylene, IAA) could contribute to the secondary transcriptional changes in MeJA-treated leaves. This process may enable plant cells to "fine tune" the chemical signals that regulate plant growth and help maintain jasmonate homeostasis (Staswick and Tiryaki., 2004). However, it remains unclear whether COR is metabolized and forms conjugates with amino acids *in planta*. There are very striking differences in the structure of COR and MeJA (Figure 14-1a and d), and these changes might enable COR to "evade" MeJA-modifying enzymes. If COR is not further metabolized, this could lead to perturbations in JA homeostasis and result in phytotoxicity. Clearly, there are many unresolved questions regarding the activity of COR in modulating phytohormone pathways. Experiments are underway to further analyze COR/MeJA-responsive genes and the metabolites modulated by these compounds, which will help elucidate the mechanism of action for both COR and MeJA.

4 EXPERIMENTAL PROCEDURES

4.1 Plant material

Lycopersicon esculentum Mill. cv. Glamour was used in all experiments. Plants were grown from seed in a peat:soil mix in 10 cm diameter plastic pots and maintained in growth chambers (25^{o}C, 40–70% RH, 12 h photoperiod, photon flux density 150–200 μmol m^{-2} sec^{-1}). Plants were approximately four weeks old at the time of treatment.

4.2 Isolation and synthesis of coronatine-related compounds

COR, CFA, and CMA were prepared as described previously (Jones et al., 1997). Methyl jasmonate was obtained from Bedoukian Research Inc. (Danbury, CT, USA), and ACC was obtained from Sigma (St. Louis, MO, USA).

4.3 Plant treatments and RNA extraction for microarray analysis

COR, CFA, and CMA (0.2 nmol per inoculation site) and MeJA (100 μM in 0.001% ethanol) were suspended in H_2O, and 8 droplets were applied in 2 μL aliquots onto tomato leaves. Sterile distilled H_2O was applied to tomato leaves as a mock treatment. Two leaves per plant were harvested 12 h post-treatment (hpt). Total RNA was purified with TRIzol™ reagent (Invitrogen, Carlsbad, CA, USA) according to the instructions of the manufacturer. Approximately 50 μg of RNA from each biological replicate was reverse-transcribed to synthesize cDNA using Superscript II™ reverse transcriptase (Invitrogen).

Tomato cDNA microarrays (Tom1 arrays) were obtained from the Center for Gene Expression Profiling, Cornell University (Ithaca, NY, USA). A brief description of the Tom1 array architecture, EST source and the complete list of the spotted genes (gene ID file) are provided (Alba et al., 2004; Uppalapati et al., 2005). cDNA was hybridized to individual slides using a modified "2-step protocol" using the 3DNA™ Submicro Kit (Genisphere). Pre-processing of data was accomplished using GenePix Autoprocessor (GPAP) (P. Ayoubi, unpublished results). This analysis included: (1) removal of data points where signal was less than the background plus two standard deviations in both channels; (2) removal of poor quality spots; (3) removal of spots where the ESTs failed quality control; (4) log transformation of the background subtracted Cy3/Cy5 median ratios; and (5) averaging the technical replicates within and across the replicates. Following pre-processing, the expression results were normalized using global LOWESS normalization (Yang et al., 2002). Normalized ratio values for each probe were averaged across valid signals obtained from three or more replicates. For each probe, the fold-change, moderated t-statistics (Smyth, 2004) and *P* values were determined. A candidate list of differentially expressed genes was then generated using a 5% false discovery rate (FDR) and greater than twofold change between treatments. A total of six experiments were conducted. Expression images and average hierarchical gene clustering were generated using Genesis software, Release 1.4.0 (Sturn et al., 2002).

4.4 Phytohormone quantification

Tomato leaves (~350 mg) were extracted and analyzed for SA, JA, and IAA using methods described by Schmelz et al., (2004). This method uses a quadropole MS system (5890 GC, Agilent, Palo Alto, CA, USA) connected to a 5989B Mass Selective Detector (Agilent) with electron spray ionization and selective-ion monitoring (selected ion ± 0.5 mass unit). The analytes were separated on a DB-5 column (30 m × 0.25 mm × 0.25 mm, Agilent) using the conditions described by Schmelz et al., (2004). The retention times and mass units of the methyl esters analyzed were: SA-ME, 8.35 min, 152; JA, *trans* 12.30 min*/cis* 12.54 min, 224; and IAA, 13.63 min, 189. Internal standards used were: [$^2H_6$]SA-ME (8.33 min, 156), dhJA-ME (*trans* 12.31 min, *cis* 12.53 min, 226), and [$^2H_5$]IAA-ME (13.62 min, 191). The [$^2H_5$]IAA-ME was converted to [$^2H_2$]IAA-ME to produce a parent ion with a mass unit of 191. Isotopically labeled internal standards were purchased from CDN Isotopes (Pointe-Claire, Quebec, Canada), while dhJA was prepared from methyl dihydrojasmonate (Bedoukian Research Inc.) by alkaline hydrolysis.

ACKNOWLEDGEMENTS

We thank David Jones, Eric Schmelz, and Jack Dillwith for their help with GC-MS analysis and Dr. Barbara Kunkel for her helpful comments. C. L. Bender acknowledges support from the National Science Foundation (IOB-0620469), the Oklahoma Center for Advancement of Science (AR031-005), and the Oklahoma Agricultural Experiment Station. The OSU Microarray Core Facility is supported by grants from NSF (EOS-0132534) and NIH (1P20RR16478-02 and 5P20RR15564-03).

REFERENCES

Alba, R., Fei, Z., Payton, P., Liu, Y., Moore, S.L., Debbie, P., Gordon, J.S., Rose, J.K.C., Martin, G., Tanksley, S.D., Bouzayen, M., Jahn, M.M., and Giovannoni, J., 2004, ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development, *Plant Journal* **39**:697-714.

Bender, C.L., Alarcón-Chaidez, F., and Gross, D.C., 1999, *Pseudomonas syringae* phytotoxins: mode of action, regulation and biosynthesis by peptide and polyketide synthetases, *Microbiology and Molecular Biology Reviews* **63**:266-292.

Bender, C.L., Stone, H.E., Sims, J.J., and Cooksey, D.A., 1987, Reduced pathogen fitness of *Pseudomonas syringae* pv. *tomato Tn5* insertions defective in coronatine production, *Physiological and Molecular Plant Pathology* **30**:273-283.

Brooks, D.M., Guzman, G.H., Kloek, A.P., Alarcón-Chaidez, F., Sreedharan, A., Rangaswamy, V., Peñaloza-Vázquez, A., Bender, C.L., and Kunkel, B.N., 2004,

Identification and characterization of a well-defined series of coronatine biosynthetic mutants of *Pseudomonas syringae* pathovar *tomato* DC3000, *Molecular Plant-Microbe Interactions* **17**:162-174.

Burns, J.K., Pozo, L.V., Arias, C.R., Hockema, B., Rangaswamy, V. and Bender, C.L., 2003, Coronatine and abscission in citrus, *J. Am. Soc. Hort. Sci.* **128**:309-315.

Cheong, Y. H., Chang, H. S., Gupta, R., Wang, X., Zhu, T., and Luan, S., 2002, Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis, *Plant Physiology* **129**:661-677.

Devoto, A., and Turner, J. G., 2003, Regulation of jasmonate-mediated plant responses in Arabidopsis, *Annals of Botany* **92**:329-337.

Ecker, J. R., 1995, The ethylene signal transduction in plants, *Science* **268**:667-675.

Farmer, E. E., Almeras, E., and Krishnamurthy, V., 2003, Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory*, Current Opinion in Plant Biology* **6**:372-378.

Ferguson, I. B., and Mitchell, R. E., 1985, Stimulation of ethylene production in bean leaf discs by the pseudomonad phytotoxin coronatine, *Plant Physiology* **77**:969-973.

Feys, B. J. F., Benedetti, C. E., Penfold, C. N., and Turner, J. G., 1994, Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen, *Plant Cell* **6**:751-759.

Goossens, A., Hakkinen, S. T., Laakso, I., Seppanen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A. M., Soderlund, H., Zabeau, M., Inze. D., and Oksman-Caldentey, K. M., 2003, A functional genomics approach toward the understanding of secondary metabolism in plant cells, *Proceedings of National Academy of Science (USA)* **100**:8595-8600.

Howe, G.A., and Schilmiller, A. L., 2002, Oxylipin metabolism in response to stress, *Current Opinion in Plant Biology* **5**:230-236.

Ichihara, A. K., Shiraishi, K., Sato, H., Sakamura, S., Nishiyama, K., Sasaki, R., Furusaki, A., and Matsumoto, T., 1977, The structure of coronatine, *Journal of American Chemical Society* **99**:636-637.

Jones, W. T., Harvey, D., Mitchell, R. E., Ryan, G. B., Bender, C. L., and Reynolds, P. H. S., 1997, Competitive ELISA employing monoclonal antibodies specific for coronafacoyl amino acid conjugates, *Food and Agricultural Immunology* **9**:67-76.

Kenyon, J. S., and Turner, J. G., 1992, The stimulation of ethylene synthesis in *Nicotiana tabacum* leaves by the phytotoxin coronatine, *Plant Physiology* **100**:219-224.

Kloek, A. P., Verbsky, M. L., Sharma, S. B., Schoelz, J. E., Vogel, J., Klessig, D. F., and Kunkel, B. N., 2001, Resistance to *Pseudomonas syringae* conferred by an *Arabidopsis thaliana* coronatine-insensitive (*coi1*) mutation occurs through two distinct mechanisms, *Plant Journal* **26**:509-522.

Kunkel, B. N., and Brooks, D. M., 2002, Cross talk between signaling pathways in pathogen defense, *Current Opinion in Plant Biology* **5**:325-331.

Kunkel, B.N., Agnew, A., Collins, J., Cohen, J., and Chen, Z., 2004, Molecular genetic analysis of AvrRpt2 activity in promoting virulence of *Pseudomonas syringae*, in: *Genomic and Genetic Analysis of Plant Parasitism and Defense*. Leach, J., Tsuyumu, S., Wolpert, T., and Shirashi, T., eds., APS Press, St. Paul, Minnesota, in press.

Lauchli, R. and Boland, W., 2003, Indanoyl amino acid conjugates: tunable elicitors of plant secondary metabolism, *Chemical Records* **3**:12-21.

Li, L., Zhao, Y., McCaig, B.C.,Wingerd, B.A., Wang, J., Whalon, M.E., Pichersky, E. and Howe, G.A., 2004, The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development, *Plant Cell* **16**:126-143.

Lund, S. T., Stall, R. E., and Klee, H. J., 1998, Ethylene regulates the susceptible response to pathogen infection in tomato, *Plant Cell* **10**:371-382.

Mitchell, R. E., 1982, Coronatine production by some phytopathogenic pseudomonads, *Physiological Plant Pathology* **20**:83-89.

Mitchell, R. E., 1985, Coronatine biosynthesis: incorporation of L-[U-^{14}C]isoleucine and L-[U-^{14}C]threonine into the 1-amido-1-carboxy-2-ethylcyclopropyl moiety, *Phytochemistry* **24**:247-249.

Mittal, S. M., and Davis, K. R., 1995, Role of the phytotoxin coronatine in the infection of *Arabidopsis thaliana* by *Pseudomonas syringae* pv. *tomato*, *Molecular Plant-Microbe Interactions* **8**:165-171.

Navarre, D. A., and Wolpert, T. J., 1999, Victorin induction of an apoptotic/senescence-like response in oats, *Plant Cell* **11**:237-249.

Parry, R. J., Mhaskar, S. V., Lin, M. T., Walker, A. E., and Mafoti, R., 1994, Investigations of the biosynthesis of the phytotoxin coronatine, *Canadian Journal of Chemistry* **72**:86-99.

Sakai, R., Nishiyama, K., Ichihara, A., Shiraishi, K., and Sakamura, S., 1979, The relation between bacterial toxic action and plant growth regulation, in: *Recognition and Specificity in Plant Host-parasite Interactions*, J.M. Daly, and I. Uritani, eds., University Park Press, Baltimore, MD., pp. 165-179.

Sasaki, Y., Asamizu, E., Shibata, D., Nakamura, Y., Kaneko, T., Awai, K., Amagai, M., Kuwata, C., Tsugane, T., Masuda, T., Shimada, H., Takamiya, K., Ohta, H., and Tabata, S., 2001, Monitoring of methyl jasmonate-responsive genes in Arabidopsis by cDNA macroarray: self-activation of jasmonic acid biosynthesis and crosstalk with other phytohormone signaling pathways, *DNA Research* **8**:153-161.

Schenk, P. M., Kazan, K., Wilson, I., Anderson, J. P., Richmond, T., Somerville, S. C., and Manners, J. M., 2000, Coordinated plant defense responses in Arabidopsis revealed by microarray analysis, *Proceedings of National Academy of Science (USA)* **97**:11655-11660.

Schmelz, E. A., Engelberth, J., Alborn, H. T., O'Donnell, P., Sammons, M., Toshima, H., and Tumlinson, J. H., 2003, Simultaneous analysis of phytohormones, phytotoxins, and volatile organic compounds in plants, *Proceedings of National Academy of Science (USA)* **100**:10552-10557.

Schmelz, E. A., Engelberth, J., Tumlinson, J. H., Block, A., and Alborn, H. T., 2004, The use of vapor phase extraction in metabolic profiling of phytohormones and other metabolites, *Plant Journal* **39**:790-808.

Smyth, G. K., 2004, Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, *Statistical Applications in Genetics and Molecular Biology* **3**:Article 3.

Staswick, P. E., and Tiryaki, I., 2004, The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in Arabidopsis, *Plant Cell* **16**:2117-2127.

Stenzel, I., Hause, B., Maucher, H., Pitzschke, A., Miersch, O., Ziegler, J., Ryan, C.A. and Wasternack, C., 2003, Allene oxide cyclase dependence of the wound response and vascular bundle-specific generation of jasmonates in tomato – amplification in wound signaling, *Plant Journal* **33**:577-589.

Stuhlfelder, C., Mueller M. J., and Warzecha, H., 2004, Cloning and expression of a tomato cDNA encoding a methyl jasmonate cleaving esterase, *European Journal of Biochemistry* **271**:2976-2983.

Sturn, A., Quackenbush, J., and Trajanoski, Z., 2002, Genesis: Cluster analysis of microarray data, *Bioinformatics* **18**:207-208.

Uppalapati, S. R., Patricia A., Weng, H. P., Palmer, D. A., Mitchell, R. E., Jones, W., and Bender, C. L., 2005, The phytotoxin coronatine and methyl jasmonate impacts multiple phytohormone pathways in tomato, *The Plant Journal* **42**(2):201-217.

Walters, D., Cowley, T., and Mitchell, A., 2002, Methyl jasmonate alters polyamine metabolism and induces systemic protection against powdery mildew infection in barley seedlings, *Journal of Experimental Botany* **53**:747-756.

Weber, H., 2002, Fatty acid-derived signals in plants, *Trends Plant Science* **7**:217–224.

Weiler, E. W., Kutchan, T. M., Gorba, T., Brodschelm, W., Neisel, U., and Bublitz, F., 1994, The *Pseudomonas* phytotoxin coronatine mimics octadecanoid signaling molecules of higher plants, *FEBS Letters* **345**:9-13.

Weingart, H., Ullrich, H., Geider, K., and Völksch, B., 2001, The role of ethylene production in virulence of *Pseudomonas syringae* pvs. *glycinea* and *phaseolicola*, *Phytopathology* **91**:511-518.

Yang, Y.H, Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P., 2002, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Research* **30**:e15.

Zhao, Y., Thilmony, R., Bender, C. L., Schaller, A., He, S.Y., and Howe, G.A., 2003, Virulence systems of *Pseudomonas syringae* pv. *tomato* promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway, *Plant Journal* **36**:485-499.

Chapter 15

PROFILING OF METABOLITES AND VOLATILE FLAVOUR COMPOUNDS FROM SOLANUM SPECIES USING GAS CHROMATOGRAPHY-MASS SPECTROMETRY

Tom Shepherd, Gary Dobson, Rhoda Marshall, Susan R. Verrall, Sean Conner, D. Wynne Griffiths, Derek Stewart, and Howard V. Davies
Scottish Crop Research Institute, Invergowrie Dundee DD2 5DA, Scotland UK

Abstract: Methods are described for analysis of metabolites in potato tubers using GC-time-of-flight (TOF) MS and for analysis of flavour volatiles released from raw and cooked potato tubers by automated thermal desorption (ATD)-GC-MS.

Key Words: flavour; gas chromatography; mass spectrometry; metabolites; potato tuber; thermal desorption; time-of-flight; volatiles.

1 INTRODUCTION

We are using high throughput profiling techniques linked to automated data processing to study substantial equivalence and unintended effects of genetic modification in Solanum species. In addition, metabolite variation within Solanum germplasm collections is being measured with the objective of exploring phytochemical diversity. An example of this is the investigation of the role of tuber metabolites and volatile compounds in relation to the organoleptic properties of potato as characterised by specialist taste panels.

Flavour molecules formed when potatoes and other foods are cooked arise from several sources (Maarse, 1991). Some compounds including aldehydes, alcohols and alkyl furans originate from lipids via enzymatic and non-enzymatic processes (Figure 15-1). Aldehydes are also derived from amino acids and dicarbonyl compounds via the Strecker reaction.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 209–219.

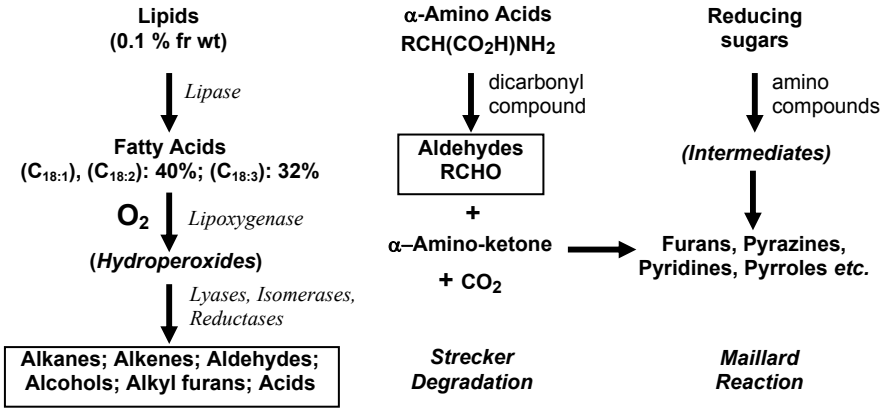


Figure 15-1. Origin of flavour volatiles released from cooked potato. Precursor metabolites are shown at the top, the main products formed on boiling are shown in solid boxes.

Heterocyclic compounds are formed from carbohydrates and amino compounds via the Maillard reaction under more forcing conditions such as baking or roasting.

Here we outline the methods we developed to study the relationship between potato tuber metabolites and flavour compounds. The method for analysis of polar and non-polar metabolites using gas chromatography-time-of-flight mass spectrometry (GC-(TOF)-MS) was adapted from Roessner et al. (2000). Volatiles were collected using an entrainment system based on that of Robertson et al. (1993), and were analysed using automated thermal desorption (ATD) coupled with GC-MS.

2 MATERIAL AND METHODS

2.1 Plant material

Potato tubers (6 or 7, average size) selected at random from storage were washed, blotted dry with tissue paper, weighed, and cut into eighths. Two opposite eighths from each tuber were taken for freeze-drying (Griffiths et al., 2001), immediately immersed in liquid N_2 and bulked by replicate. The sample was freeze-dried (FD), ground in a laboratory mill fitted with a 1 mm screen and stored in the dark at −20°C until used for metabolite analysis. The remaining segments were used in the cooking experiments.

2.2 Preparation and derivatization of tuber metabolites

Powdered FD tuber (100 mg) plus the internal standards (IS) for polar (100 μL aqueous ribitol, 2 mg mL^{-1}) and non-polar (100 μL methanolic methyl nonadecanoate, 0.2 mg mL^{-1}) components were shaken at 30°C for 30 min each with methanol (3 mL), water (0.75 mL) and chloroform (6 mL) in a glass culture tube. Water (1.5 mL) was added; the mixture was shaken by hand and separated on a centrifuge into upper (polar) and lower (non-polar) fractions.

An aliquot (250 μL) of the polar fraction was evaporated to dryness and oximated with methoxylamine hydrochloride in pyridine (80 μL, 20 mg mL^{-1}) at 50°C for 4 h and then silylated at 37°C for 30 min with 80 μL of MSTFA (*N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide). A subsample (40 μL) was taken and added to an autosampler vial containing an alkane retention index (R_t) mixture (compositional details are shown in Figure 15-4). After dilution (1:1) with pyridine the sample was analysed by GC-(TOF)-MS.

The non-polar fraction was evaporated to dryness and transesterified at 50°C overnight with methanolic sulfuric acid (2 mL, 1%). After addition of sodium chloride (5 mL, 5%) and chloroform (3 mL) the mixture was shaken and left to separate into two layers. The lower chloroform layer was shaken with potassium bicarbonate (3 mL, 2%) and then passed through a short column of anhydrous sodium sulfate. The column was washed with more chloroform and the combined chloroform fractions were evaporated to dryness and silylated with MSTFA (80 μL), chloroform (50 μL) and pyridine (10 μL) at 37°C for 30 min. A subsample (40 μL) was prepared for analysis by GC-(TOF)-MS as described for the polar fraction.

2.3 Analysis of tuber metabolites by GC-TOF-MS

Polar and non-polar samples were analysed similarly using a Thermo Finnigan Tempus GC-(TOF)-MS system. Samples (1 μL) were injected into a programmable temperature vaporizing (PTV) injector with a split of 167:1. The PTV conditions were injection temperature 132ºC for 1 min, transfer rate 14.5ºC s^{-1}, transfer temperature 320ºC for 1 min, clean rate 14.5ºC $s^{-1,}$ and clean temperature 400ºC for 2 min. Chromatography was effected on a DB5-MS column (15 m x 0.25 mm x 0.25 μm) using helium (He) at 1.5 mL min^{-1} (constant flow). The GC temperatures were 100ºC for 2.1 min, 25ºC min^{-1} to 320ºC then isothermal for 3.5 min. The GC-MS interface temperature was 250ºC. MS acquisition conditions were electron impact (EI) ionisation at 70 eV, solvent delay 1.3 min, source 200ºC, mass range 35–900 a.m.u. at 4 spectra s^{-1}. Data were acquired using the Xcalibur software package.

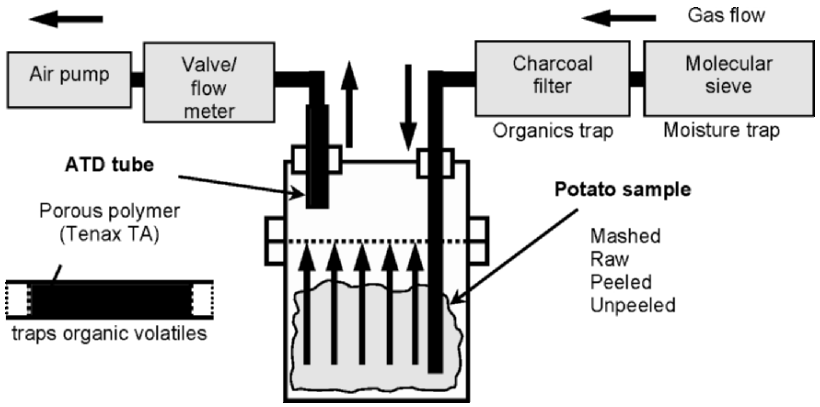


Figure15-2. Apparatus for entrainment of flavour volatiles from boiled potato tubers. Filtered air was drawn through the glass collection vessel and sample then through the stainless steel ATD tube via PTFE tubing. The vessel was sealed with PTFE tape at all connections.

2.4 Preparation and collection of flavour volatiles

Tuber segments (approximately 0.5 – 1.0 kg) and distilled water (1L) in a pyrex saucepan were heated on a 400 W hotplate for about 30 min at the maximum setting until the water started to boil. The tubers were simmered at reduced heat and periodically checked to assess the degree of cooking. *Solanum tuberosum* required a further 10 min, whereas *S. phureja* was ready for sampling after the initial 30 min. Water was drained off and six one eighth segments were taken for freeze drying. The remaining material was mashed and transferred to the entrainment vessel (Figure 15-2) which was sealed and allowed to cool for 1 h. A stainless steel ATD tube containing the porous polymer Tenax TA was connected to the system and filtered air was passed through at 100 mL min^{-1} for up to 24 h. The ATD tube was removed and back flushed with dry nitrogen at 25 mL min^{-1} for 30 min and then loaded onto the ATD autosampler for analysis by GC-MS.

2.5 Analysis of flavour volatiles by ATD-GC-MS

Full details of instrumentation and analytical conditions are given in Robertson et al. (1993). The ATD tube was heated at 200°C for 15 min in a flow of He and volatiles released from the tube were cryofocused at –25°C onto a cold trap containing Tenax-TA. Volatiles were then transferred to the GC by very rapid heating of the cold trap to 240°C. Chromatography was

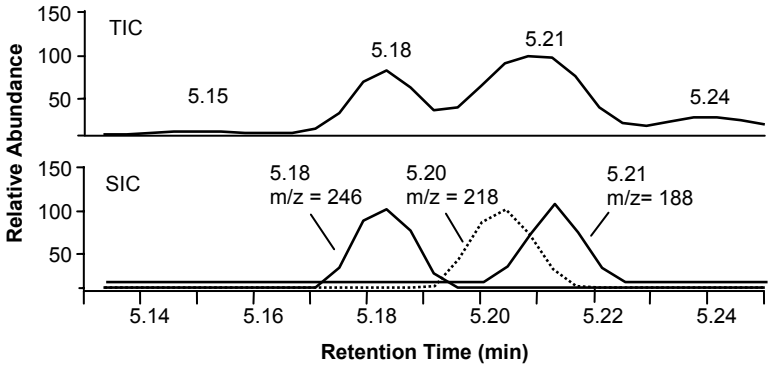


Figure 15-3. Total ion chromatogram (TIC) and selected ion chromatogram (SIC) traces for three co-eluting amino acid derivatives extracted from tubers of *Solanum tuberosum*. Glutamine TMS_3, m/z = 246; phenylalanine TMS_2, m/z = 218; asparagine TMS_3, m/z = 188.

performed on a J & W DB 1701 column (60 m × 0.25 mm × 1.0 μm) with a GC oven temperature programme from 40°C to 240°C at 5° min^{-1}, followed by a 20 min isothermal period. The GC-MS interface temperature was 250ºC, the source temperature was 200ºC. Data were acquired over the mass range 20–400 a.m.u at 1 scan s^{-1} using the MassLab software package.

2.6 Data analysis

Representative examples of Xcalibur raw data files for polar and non-polar metabolites were used with the AMDIS software package to verify the presence of individual analytes, to deconvolute co-eluting peaks and to help identify ions(s) characteristic of each. Having selected suitable ion(s) for compound detection, data processing methods were created using Xcalibur. Time windows were defined for each component, including an appropriate IS, relative to an adjacent R_t standard. For each metabolite a selected ion chromatogram (SIC) was generated within the appropriate time window. The use of this method to deconvolute three co-eluting amino acids in the polar metabolite fraction is shown in Figure 15-3. Response ratios were calculated for each analyte relative to the IS using the calculated SIC areas for both components. These values were used directly during subsequent data analysis.

After initial acquisition using Masslab, data for flavour volatiles was analysed using HP MS ChemStation after conversion from Masslab using the MassTransit file conversion software. Compositional analysis was based on integration of individual total ion chromatogram (TIC) peaks, and data for each component was expressed as a percentage of the total. Identification of compounds was based on the analysis of standards, comparison with MS libraries and literature data, and by extrapolation from known compounds.

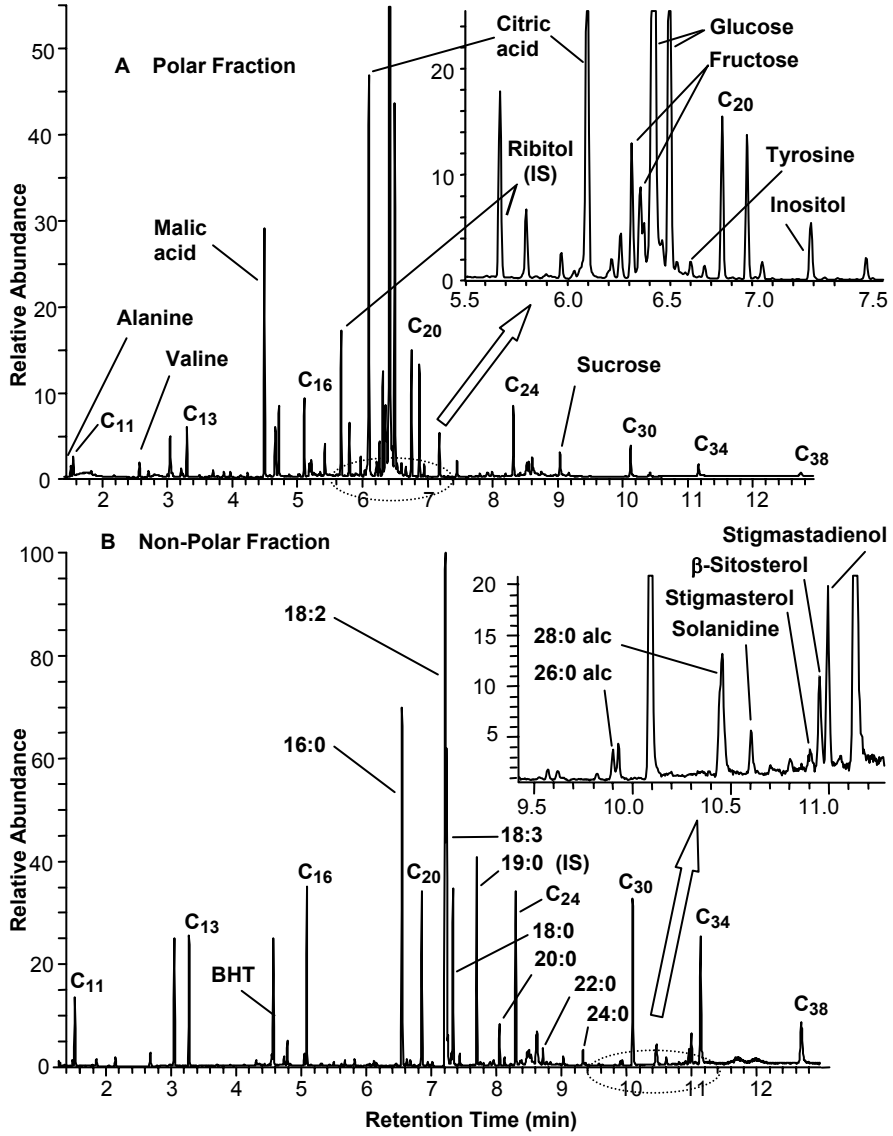


Figure 15-4. Representative TIC of the polar (A) and non-polar (B) fractions of potato tuber extract including alkane R_t standards (C_{11}–C_{38}) and an IS, with selected metabolites named. Polar metabolites are silylated and additionally some sugars are also oximated. Fatty acids are present as methyl esters. Long-chain alcohols (alc), phytosterols, etc. are silylated.

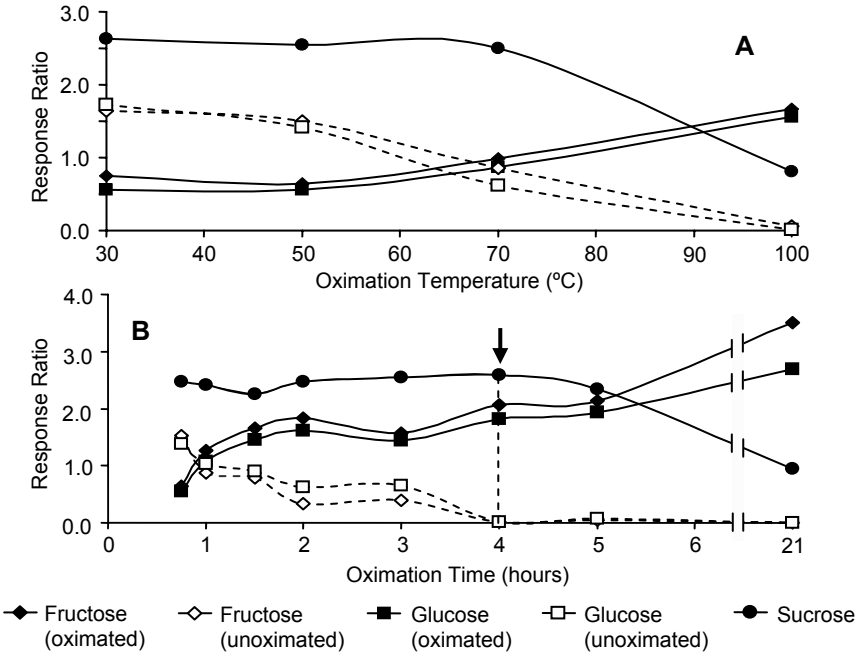


Figure 15-5. Effects of: (A) temperature on oximation of potato tuber sugars for 45 minutes and (B) time on oximation at 50ºC, followed by silylation with MSTFA. Derivatives shown are fructose O-methyloxime $(TMS)_5$; fructose $(TMS)_5$; glucose O-methyloxime $(TMS)_5$; glucose $(TMS)_5$; sucrose $(TMS)_8$. The arrow indicates optimum oximation conditions.

Processed data was subject to appropriate statistical treatment such as principal components analysis (PCA) or analysis of variance (ANOVA).

3 RESULTS AND DISCUSSION

3.1 Tuber metabolites

In its final developed form, the profiling method was used to characterize a total of about 200 major and minor polar and non-polar metabolites in potato tubers (Figure 15-4), of which 50% could be identified. During derivatization of non-polar components, methyl oximes of aldose and ketose sugars such as fructose and glucose are formed from the condensation reaction between the carbonyl group of the acyclic form of the sugar and methoxylamine hydrochloride. This locks the sugar in the acyclic form to give two positional isomers per sugar and in theory avoids the formation of anomeric furanose and pyranose forms. In practice, the extent of oximation is temperature and time dependent. The presence of appreciable amounts of unoximated derivatives complicates both the chromatography and data

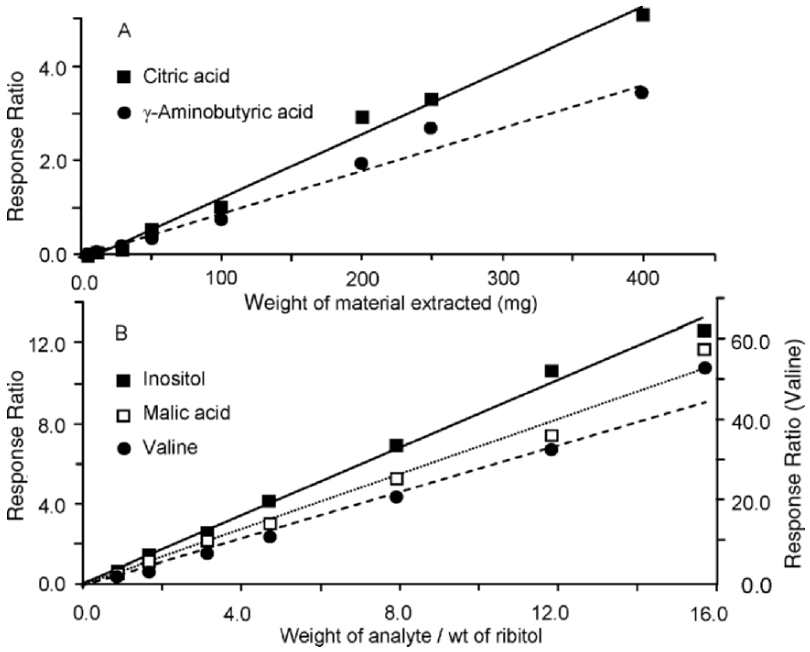


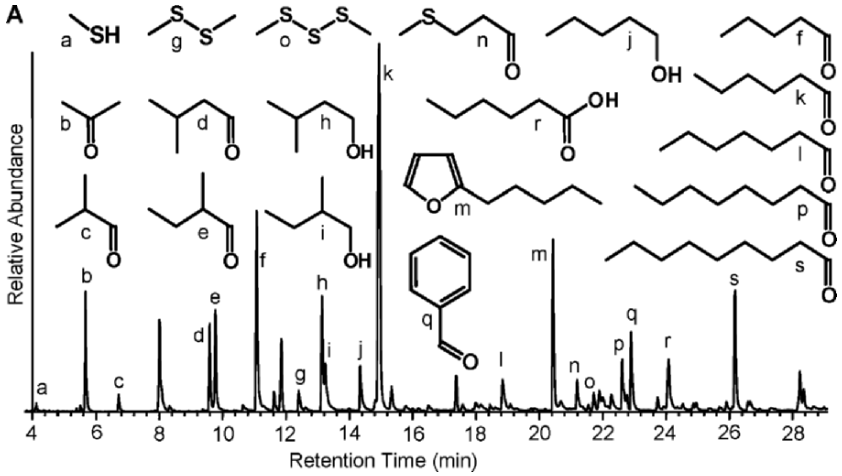
Figure 15-6. Linearity of response ratio for polar metabolites in (A) potato extracts and in (B) standards relative to the internal standard, ribitol. All metabolites as TMS derivatives.

analysis. An additional consideration is the stability of other metabolites, in particular sucrose, which degrades to glucose and fructose under more extreme conditions. Therefore we investigated the effects of temperature (Figure 15-5A) and time (Figure 15-5B) on oximation to find optimal conditions. At lower temperatures and shorter times the reaction is incomeplete. At higher temperatures and longer times oxime formation is essentially complete but sucrose degrades. Optimum conditions were found to be 50^{o}C for 4 h.

The linearity of the profiling method was shown in two ways. Individual compounds were analysed at a range of concentrations, and the analyses were conducted of extracts made from different amounts of FD potato (Figure 15-6). Measurements were made relative to a fixed quantity of IS.

3.2 Flavour volatiles

Of the 83 compounds identified in the potato volatile profile, the major components included straight (*n*-) and branched-chain (*br*)- aldehydes and alcohols, acetone, hexanoic acid, sulphur compounds, benzaldehyde, and



B	Compound	P	UP	R	Sniff
f	Pentanal (L)	6.39	12.02	0.91	Green/Pungent
k	Hexanal (L)	47.51	30.93	9.94	Green/Hay
l	Heptanal (L)	2.44	2.80	7.60	Green/Burnt
p	Octanal (L)	2.74	2.08	17.77	Fatty
c	2-Methylpropanal (S)	0.05	0.83	nd	Wet fur
d	3-Methylbutanal (S)	0.46	4.08	7.32	Toasted/Sweet
e	2-Methylbutanal (S)	0.90	5.95	3.77	Toasted/Sweet
h	3-Methylbutanol (S)	2.67	6.89	0.81	Sweet
i	2-Methylbutanol (S)	1.15	2.07	0.61	Sweet
j	Pentanol (L)	2.80	3.35	0.46	Car exhaust
m	2-Pentylfuran (V)	13.28	9.82	4.22	Cooked
q	Benzaldehyde (V)	5.07	4.96	13.15	Almonds
n	3-Methylthiopropanal	2.62	1.72	0.34	Cooked potato
a	Methanethiol	0.17	0.32	nd	Onion
g	Dimethyldisulfide	2.87	1.47	0.42	Onion
o	Dimetyltrisulfide	5.94	1.22	2.43	Onion
r	Hexanoic Acid (V)	1.54	4.44	27.42	Raw potato
b	Acetone (V)	1.40	5.04	2.90	Sweet
	Relative Total	**34**	**16**	**1**	

Figure 15-7. (A) TIC trace of potato flavour volatiles entrained on Tenax TA showing the structures of the major components. (B) Composition of Volatiles from cooked unpeeled (UP), cooked peeled (P), and raw unpeeled (R) tubers of *Solanum tuberosum* cv. Montrose. Compounds originate from: lipid oxidation (L); amino acids via the Strecker reaction (S); various sources (V). Sniff test characterisations are base on published literature data (see main text). Relative totals are based on combined TIC peak areas.

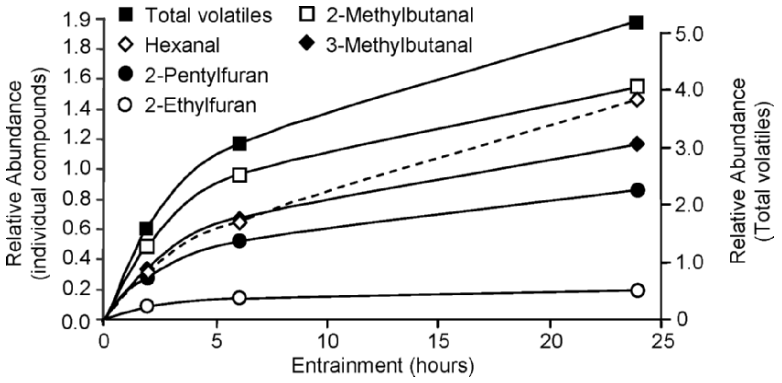


Figure 15-8. Accumulation of potato flavour volatiles from cv. Montrose with time.

alkyl-furans (Figure 15-7A). The distribution of such compounds varied between cooked (peeled and unpeeled) and raw (unpeeled) tubers (Figure 15-7B). Volatiles were entrained directly from raw tuber segments by omitting the cooking step. Cooked material generated relatively more pentylfuran, 3-methylthiopropanal and more of the shorter C_5 and C_6 *n*-aldehydes, whereas hexanoic acid and the longer *n*-C_7 and *n*-C_8 aldehydes were characteristic of raw tubers. It is of note that in the sniff test evaluation of potato flavour by specialist panels, 3-methylthiopropanal and hexanoic acid were characterised as smelling of cooked and raw potato respectively (Petersen et al., 1998; Ulrich et al., 2000). Considerably more volatile compounds were released from cooked and mashed tubers than from raw tubers, and the quantities generated from cooked material was greater if they were peeled prior to cooking. The relative abundance of *br*-aldehydes and alcohols was significantly greater from unpeeled tubers, suggesting that the precursor metabolites may be particularly associated with regions close to the outer epidermis.

In a series of test experiments, similar quantities of volatiles accumulated during entrainment over each of the periods 0–2 h, 2–6 h, and 6–24 h (Figure 15-8) indicative of reduced rates of volatile production with time. Although the overall composition remained similar, the abundance of some components, such as hexanal, increased over the 6–24 hour period, indicating that some processes generating volatiles were still active. Consequently, entrainment for 24 h was selected for definitive experiments.

In Chapter 19 we report how the methodology described here was used to study phytochemical diversity between two solanum species in terms of their tuber metabolites and flavour volatiles (Dobson et al., 2007).

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Scottish Executive Environment and Rural Affairs department.

REFERENCES

Dobson, G., Shepherd, T., Marshall, R., Verrall, S.R., Conner, S., Griffiths, D.W., McNicol, J.W., Stewart, D., and Davies, H.V. 2007, Application of metabolite and flavour volatile profiling to studies of biodiversity in solanum species. Following article.

Griffiths, D.W. and Dale, M.F.B., 2001, Effect of light exposure on the glycoalkaloid content of Solanum phureja tubers, *J. Agri. Food Chem.* **49**:5223–5227.

Maarse, H., 1991, Volatile Compounds in Foods and Beverages, Marcel Dekker, New York, USA, pp. 764.

Petersen, M. A., Poll, L., and Larsen, L. M., 1998, Comparison of volatiles in raw and boiled potatoes using a mild extraction technique combined with GC odour profiling and GC-MS, *Food Chem.* **61**:461–466.

Robertson G.W., Griffiths, D.W., MacFarlane Smith, W., and Butcher, R.D., 1993, The application of thermal desorption-gas chromatography-mass spectrometry to the analyses of flower volatiles from five varieties of oilseed rape (*Brassica napu*s spp. *oleifera*), *Phytochem. Analysis* **4**:607–612.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000, Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.* **23**:131–142.

Ulrich, D., Hober, E., Heugebaur, W., Tiemann, H., and Darsow, U. 2000, Investigation of the boiled potato flavour by human sensory and instrumental methods, *Amer. J. of Potato Res*. **77**:111–117.

Chapter 16

METABOLOMIC ANALYSIS OF LOW PHYTIC ACID MAIZE KERNELS

Jan Hazebroek, Teresa Harp, Jinrui Shi, and Hongyu Wang
Pioneer Hi-Bred International, Inc., a DuPont company, P.O. Box 1004, Johnston, IA 50131-1004 USA

Abstract: Phytic acid, or hexaphosphorylated *myo*-inositol, is the major storage form of phosphorous (P) in maize kernels. Phytic acid in foods or animal feeds can complex with proteins and mineral cations resulting in reduced bioavailablility of important nutrients. Classic mutation breeding has been used to develop maize plants that produce kernels with significantly less phytic acid. An extensive survey of the low phytate phenotype in different maize genetic backgrounds grown in five field locations revealed that an increase in inorganic P correlated with a decrease in phytic acid P, but the increased amount of inorganic P did not consistently account for the P reduction noted in the low phytate lines. There were no quantitative phosphorous differences in phospholipids or starch. In follow-up experiments using a metabolomics approach, both mutant and wild type kernels were obtained from a single segregating ear, minimizing variability. Individual mature kernels were lyophilized and ground. Kernel phenotype was determined by using a simple colorimetric test for inorganic P content. Kernels of similar phenotype were pooled and extracted in aqueous methanol and partitioned into polar and nonpolar fractions. Metabolites were derivatized and subjected to GC/TOF/MS, and raw data was processed using the Leco ChromaTof peak deconvolution software. Compounds were identified *via* coelution and/or mass spectrum matching with authentic standards. Each of these metabolites was semi-quantified by calculating the ratio of the peak area of a characteristic extracted ion against that of the internal standard and correcting for sample weight. P-containing metabolites were recognized easily by a prominent m/z 299. Several P-containing metabolites were more abundant in low phytic acid kernels, although it is unlikely that they are responsible for the reduced yield associated with this phenotype.

1 INTRODUCTION

Phytic acid (*myo*-inositol 1,2,3,4,5,6-hexakisphosphate, Figure 16-1) is a very abundant molecule in the seeds of many cereals and legumes (Shi et al.,

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 221–238.

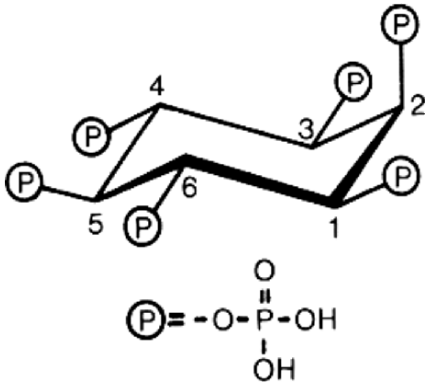


Figure 16-1. Phytic acid, or hexaphosphorylated *myo*-inositol.

2003). The negatively charged molecule is associated *in planta* with cations such as K^+, Mg^{2+}, and Ca^{2+}. As such, phytic acid is a major storage form of *myo*-inositol, phosphate as well as, several mineral cations, all needed to sustain seedling development. Phytic acid is also believed to be central to the control of inorganic phosphate levels in both developing seeds and growing seedlings (Strother, 1980).

The unique chemical properties of phytic acid have significant consequences for human and animal nutrition. Phytic acid absorption by the digestive track depends largely on microbial phytase activity, which is essentially lacking in nonruminant animals, including humans (Holm, 2002). Furthermore, the intact phytic acid molecule will interfere with absorption of nutritionally important minerals such as iron, zinc, and magnesium, resulting in suboptimal animal weight gain, or affect adversely human nutrition in communities dependent on a high grain diet (Zhou and Erdman, 1995). For this reason, phytic acid can be categorized as an anti-nutritional component in many maize-based foods and feeds. Of great concern to the livestock industry is the potential for substantial amounts of phosphorous (P) in the form of undigested phytate to be excreted in animal manure, contributing to environmentally damaging levels of P in runoff from high density livestock operations (Cromwell and Coffey, 1991). Clearly, reducing the amount of seed phytic acid in cereal grains commonly fed to animals and/or people while maintaining the amount required for normal seed and seedling development could be advantageous.

To address these needs, mutant and transgenic low phytic acid maize, barley, rice, wheat, and soybean have been developed (Larson et al., 1998; Wilcox et al., 2000; Hitz et al., 2002; Raboy, 2002; Guttieri et al., 2004). Low phytic acid (*lpa1*) maize is a chemically (EMS) induced mutant with a

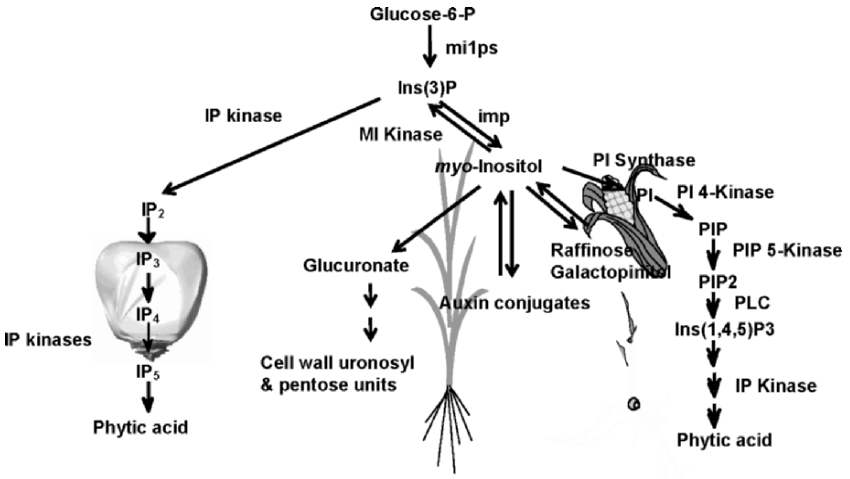


Figure16-2. Putative phytic acid biosynthesis in maize. mi1ps, 1D-*myo*-inositol-1-phosphate synthase; Ins(3)P, *myo*-inositol-3-phosphate; Ins(1,4,5)P3, *myo*-inositol-1,4,5-triphosphate; IP, *myo*-inositol-phosphate; IP_2, diphosphorylated *myo*-inositol-phosphate; IP_3, triphosphorylated *myo*-inositol-phosphate; IP_4, tetraphosphorylated *myo*-inositol-phosphate; IP_5, pentaphosphorylated *myo*-inositol-phosphate; imp, *myo*-inositol monophosphatase; MI, *myo*-inositol; PI, phosphatidylinositol; PIP, phosphatidylinositol-phosphate; PIP2, phosphatedylinositol-diphosphate.

65% reduction in seed phytic acid content and about a tenfold increase in free inorganic phosphorous (Pi). The *myo*-inositol phosphates with fewer than six esterified phosphates ("lower *myo*-inositol phosphates") do not accumulate. We have determined that the *lpa1* mutation essentially does not change the amounts of total P, oil, protein, starch, K^+, Mg^{2+}, Ca^{2+}, Mn^{2+}, Zn^{2+}, and Fe^{3+}. Although *lpa1* seed development, seed desiccation, seed germination, seedling development, and seedling vigor are all normal, there is typically up to 15% unexplainable loss of kernel dry weight (yield).

The *LPA1* gene has been mapped to Chromosome 1 (Raboy et al., 2000) but its function is not known. Other low phytic acid maize traditional mutants have been isolated subsequently. Kernels of the *lpa2* mutant accumulate significant amounts of *myo*-inositol-P_3, *myo*-inositol-P_4, and *myo*-inositol-P_5 (Shi et al., 2003). A third mutant (*lpa3*), like *lpa1*, does not accumulate the lower myo-inositol phosphates in their seeds (unpublished observation).

The genetics and biochemistry of phytic acid synthesis in maize is incompletely understood despite intensive analyses of several *lpa* mutants (Raboy et al., 2000; Shi et al., 2003). It is clear that the first step in committed phytic acid biosynthesis is the conversion of glucose-6-P to *myo*-inositol-3-P catalyzed by *myo*-inositol-P synthase (Milps, Figure 16-2). In

developing kernels, phytic acid is synthesized by sequential kinase-catalysed phosphorylations. *Myo*-inositol-P can also be dephosphorylated and the liberated sugar alcohol can be incorporated into phosphatidylinositol followed by an alternative phosphorylation pathway to phytic acid. *Myo*-inositol is also the precursor to various raffinosaccharides, cell wall components, and auxin conjugates. However, the metabolic flow through these pathways in wild type and mutant kernels is unknown.

We hypothesize that phosphorous-containing metabolite(s) that accumulate in low phytic acid maize kernels are associated with reduced kernel dry weight. A targeted analysis of phosphorylated compounds was done initially to identify those that might be tied to reduced yield. This effort was followed up with a more comprehensive metabolomics approach. We also anticipate that comparison of results from wild type and low phytic acid type might define better our incomplete understanding of phytic acid biosynthesis in these mutant kernels.

2 METHODS

2.1 Analysis of phytic acid, inorganic, and total P in field-grown seeds

Twenty inbred and hybrid lines with wild-type phenotypes were planted alongside their *lpa1* conversions at five locations within the US Midwest Corn Belt. Plots at each location were harvested at the same time, and the seeds were cleaned, dried, and analyzed for phytic acid, inorganic P, and total P. Phytic acid was measured by anion exchange HPLC. Seeds were ground using a Kleco ball mill (Visalia, CA). Samples were weighed (500 mg) into 20 mL scintillation vials. A 5 mL of 0.4M HCl was added and the samples were shaken on a gyratory shaker at room temperature for 2 h. Extracts were filtered through a 0.45 μm PVDF syringe filter attached to a 5 mL plastic disposable syringe barrel. A 450 μL aliquot was filtered through a 0.2 μm microcentrifuge spin filter unit and then transferred to a 2 mL glass autosampler vial fitted with a 400 μL glass insert. A Dionex DX 500 HPLC equipped with a Thermo Separation Products AS3500 autosampler was used. Extracts were injected in 25 μL amounts onto a Dionex OmniPac™ PAX-100 analytical column (4 × 250 mm) in line with an OmniPac™ PAX-100 guard column (4 × 50 mm) and an ATC-1 anion trap column. A Dionex conductivity detector module II was used with an anion self-regenerating suppressor (ASRS-Ultra II) set up in the external water regeneration mode and operated with a current of 300 mA. Phytic acid was eluted at 1 mL min^{-1} with the following mobile phase program: H_2O/200 mM NaOH/50% aqueous *iso*-propanol (68/30/2) for 4.0 min, followed by a step gradient to H_2O/200 mM

NaOH/50% aqueous *iso*-propanol (39/59/2) at 14.1 min, then a step gradient return to initial conditions at 15.1 min, followed by equilibration for 15 min. The separation was performed at room temperature. The concentration of phytic acid P was calculated from the concentration of phytic acid by dividing the former by the molecular weight of the sodium phytate standard, multiplying by 6 (P per phytic acid molecule), and multiplying by 31 (molecular weight of P). Inorganic P was measured spectroscopically using modifications of the method of Chen et al. (1956) (Shi et al., 2003). Total P was determined by a contract laboratory using inductively coupled plasma spectroscopy.

2.2 Targeted analysis of P-containing kernel constituents

Bulk samples of mature kernels from Pioneer Hybrid 3730 (wild type) and its *lpa1* hybrid conversion were used for targeted analysis of P-containing constituents. These two seed sources were grown at the same field location. Phospholipids were extracted twice from 1.0 g ground seeds in two 3 mL aliquots of ice cold methanol:chloroform:formic acid (10:10:1) with centrifugation for 5 min at 2500 revolutions per minute (rpm) after each extraction. The pellet was re-extracted twice with two 3 mL aliquots of methanol:chloroform:H_2O (5:5:1), again with centrifugation for 5 min at 2500 rpm after each extraction. The supernatants from all four extractions were combined, and 3.6 mL of a solution containing 1.16 mL 85% H_3PO_4 and 7.455 g KCl in a total volume of 100 mL were added. The solution was vortexed and centrifuged for 5 min at 2500 rpm. Major phospholipids in the lower layer were determined by normal phase HPLC with evaporative light scattering detection adapting the method of Picchioni et al. (1996).

Phosphorous was measured in starch extracted from isolated endosperm, the kernel tissue where the majority of starch is found (Perry, 1988). The extraction and purification of starch was according to the method of Bechtel and Wilson (2000) with modifications. The endosperm was ground to a fine powder in a Kleco ball mill. One and one-half grams ground endosperm were incubated for 60 min at 37°C with 25 mL H_2O and 10 mL 0.8% Pepsin A in 0.04N HCl. Five milliliters 0.08% hemicellulase (1500 units/g activity) in 0.1M sodium acetate were added, and the reaction mix was incubated an additional 3 h at 45°C. Five milliliters of detergent mix (5% Triton X-100, 5% Tween 40, 5% SDS, and 5% Triton X-15) were added, and the reactions were vortexed and centrifuged for 5 min at 2500 × g at 20°C, and the supernatant was discarded. The pellet was resuspended in 25 mL H_2O, vortexed, centrifuged, and decanted as before. The water washing, vortexing, centrifuging, and decanting steps were repeated three times. The pellet was resuspended in 2 mL H_2O and applied on a 53 μm or 75 μm-mesh screen and washed with approximately six volumes of water. The filtrate was

centrifuged for 5 min at 2500 × g at 20°C, and the supernatant was discarded. The resulting pellet was dissolved in 3 mL 70% ethanol, vortexed, centrifuged, and the supernatant was discarded. The final purified starch pellet was lyophilized for a minimum of 48 h. The entire procedure was performed on 20 wild type and 20 *lpa1* 1.5 g samples, and the purified starch from each phenotype was pooled to accumulate sufficient material for subsequent total P determination.

2.3 Metabolomic analysis of individual kernels

Wild type and *lpa1*or *lpa3* plants were crossed to obtain kernels on F1 ears segregating 1:1 for the mutant genotype. At physiological maturity, individual kernels were removed from the cob and frozen immediately in liquid nitrogen. Only kernels from the central portion of the ear were harvested; the butt and tip kernels were discarded. The kernels were lyophilized and ground to a fine powder in a Genogrinder 2000 ball mill (SPEX CertiPrep, Metuchen, NJ). The low phytic acid phenotype of individual kernels was determined indirectly by measuring the amount of Pi spectroscopically using modifications of the method of Chen et al. (1956) (Shi et al., 2003).

Metabolites were extracted from 30 mL ground material from each of 10 wild type and 10 low phytic acid kernels. Extraction and chemical derivatization were performed according to the method of Fiehn et al. (2000) (Figure 16-4). Both *lpa1* and *lpa3* mutants and their wild-type controls were analyzed. The unique experimental design minimizes greatly the environmental influence on metabolite expression, since every sample developed within the same ear. The nonpolar fractions, after methylation and trimethylsilylation, and the polar fractions, after methoxyamination and trimethylsilylation, were subjected to GC/TOF/MS. The trimethylsilyl derivatives were separated by gas chromatography on a Supelco 30 M × 0.25 mm I.D. × 0.25 mm film thickness SPB-50 column. One-microliter injections were made with a 1:10 split ratio using an Agilent 7683 autosampler. The polar extracts were rerun in the splitless mode in order to improve sensitivity for some phosphorous-containing metabolites.

An Agilent 6890N gas chromatograph was programmed for an initial temperature of 70ºC for 5 min, increased to 310°C at 5º min^{-1} where it was held for 1 min. The injector and transfer line temperatures were 230ºC and 250ºC, respectively, and the source temperature was 200ºC. He was used as the carrier gas with a constant flow rate of 1 mL min^{-1} maintained by electronic pressure control. Mass spectra were obtained online with a Leco Pegasus III time-of-flight (TOF) mass spectrometer. An electron beam of –70eV was

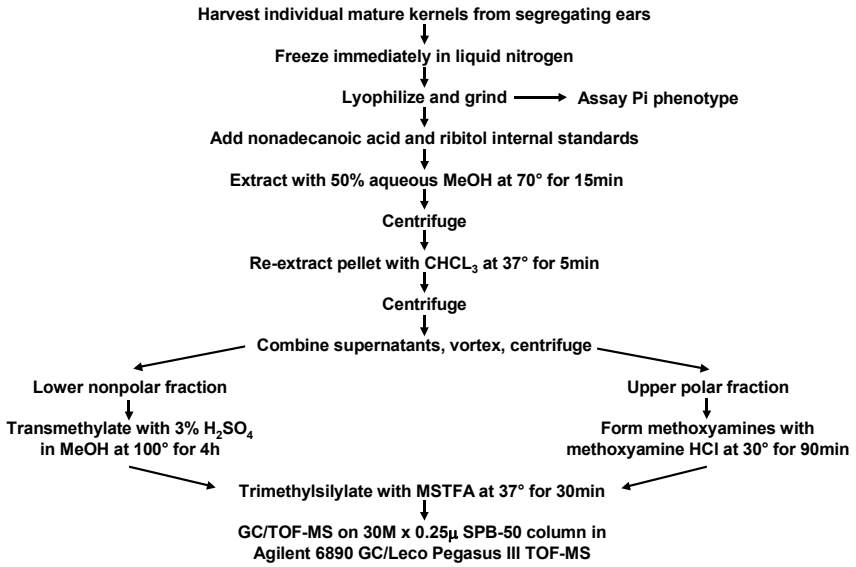


Figure 16-3. Extraction and derivatization scheme for metabolomics analysis of maize kernels. MeOH, methanol; $CHCL_3$, chloroform; MSTFA, N-methyl-N-trimethylsilyl-trifluoro-acetamide.

used to generate spectra with a mass range of m/z 41–999 at a sampling rate of 5 spectra s^{-1}. Metabolites were identified based on a match to both the mass spectrum and retention of appropriately derivatized authentic standards. The relative amount of each metabolite was based on the hand-curated area of the extracted ion chromatogram of a characteristic quantifying m/z value. All quantifications were normalized to the peak area of quantifying m/z value of the internal standard and the initial sample dry weight. Student's T-tests were performed to evaluate the statistical significance of the mean relative amounts of each metabolite in wild type and low phytic acid kernels.

3 RESULTS AND DISCUSSION

3.1 Phosphorous balance in field-grown seeds

The reduced phytic acid content of the *lpa1* mutant was evident in all combinations of genetic background and planting location (Figure 16-4). There were no significant effects of either genetics or location on this relationship. There was about twice the amount of phytic acid in *lpa1*

kernels of sample 20 compared with those of the other genetic backgrounds, but still significantly less than in sample 20 wild-type kernels. This increase is undoubtedly due to the high oil character of sample 20. A proportionally larger embryo characterizes these high oil kernels. Thus, more phytic acid is to be expected on a whole-kernel weight basis since it accumulates preferentially in the embryo (O'Dell et al., 1972). We did not measure phytic acid in isolated embryos. As expected, Pi contents in these kernels exhibited an inverse relationship to that of phytic acid. Total P was relatively constant in kernels of samples 1–19, but was slightly higher in those of the high oil type (sample 20) due to their proportionally larger embryos (Figure 16-4).

The lack of variability seen in measured phytate and Pi of the low phytic acid phenotype observed in this field trial suggests that this trait could be used potentially in a breeding program. However, significant reductions in seed yield are observed consistently with *lpa1* plants compared with wild-type plants. Although the physiological basis for this yield reduction is unknown, an obvious suggested cause would be a disrupted P balance. Yield reduction could be attributed to the reduction in phytic acid and/or elevation in Pi, although it is possible that other P-containing metabolites are involved. Our data show about 32% of the total P in *lpa1* kernels is unaccounted for vs 9% in the wild-type kernels (Table 16-1). There was a very consistent and significant increase in the amount of organic P not associated with phytic acid in low phytic acid kernels in all genetic backgrounds and planting locations. To better understand where the unaccounted P has accumulated, we attempted to quantify P in the major P-containing biomolecules in *lpa1* and wild-type kernels using a targeted analysis approach. For practical reasons, we used a single seed source for this more in-depth analysis.

3.2 Targeted analysis of P-containing kernel constituents

As expected, all four of the major membrane-associated phospholipids (phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, and phosphatidylserine) were found in the nonpolar kernel extracts. However, there were no significant differences in the amounts of any of these phospholipids between wild type and *lpa1* kernels, suggesting that altered membrane function is not associated with reduced *lpa1* kernel weight. We also did not find a significant difference in the amount of P associated with endosperm starch between wild type and *lpa1* kernels. This is perhaps not surprising since a mutation affecting P incorporation into phytic acid in the embryo would influence P content of starch in the endosperm. Although scenarios can be suggested to account for this, it is more likely that the effect of the *lpa1* mutation would be restricted to P metabolism in the embryo. To

investigate this possibility, we plan to measure the P content of embryo-associated proteins. Regardless, the lack of an obvious candidate for the unaccounted P in *lpa1* kernels led us to conduct a more comprehensive metabolomics approach.

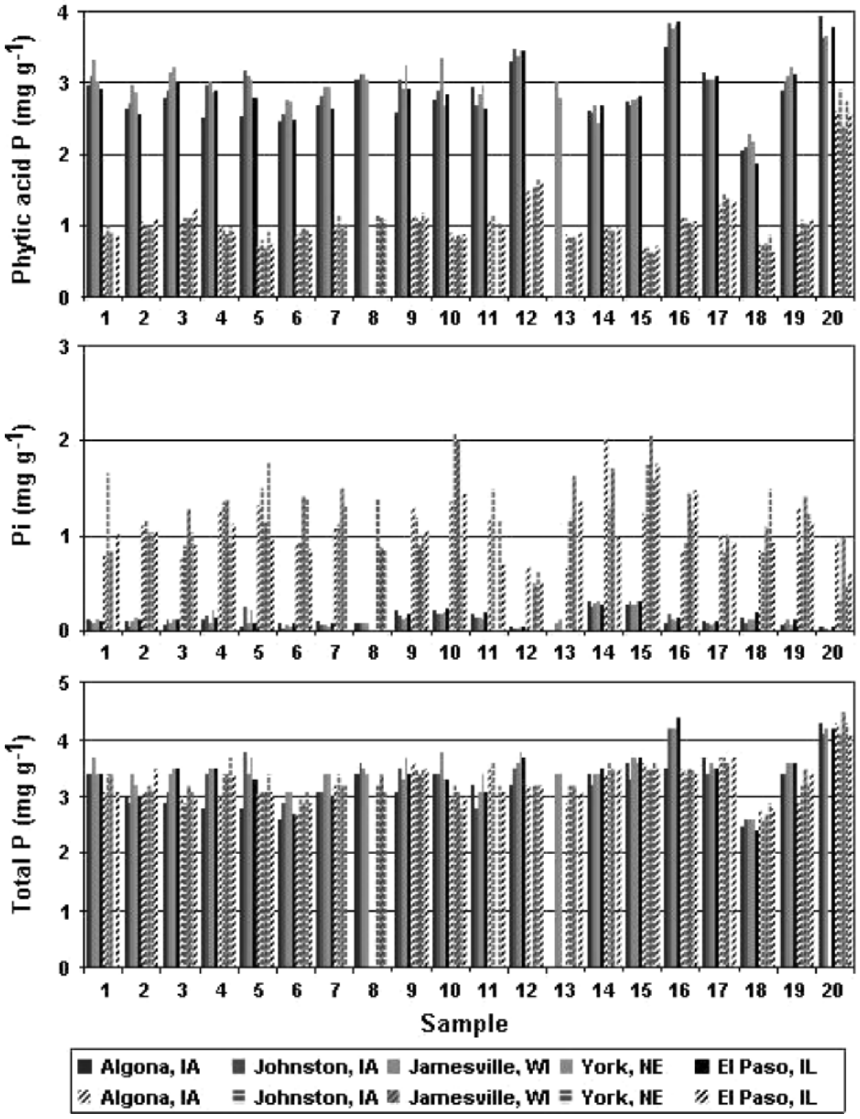


Figure 16-4. Phytic acid phosphorus, Pi, and total P in whole kernels from wild-type plants (solid bars) and their *lpa1* conversions (hashed bars) of different genetic backgrounds grown at five different locations within the US Midwest.

Table 16-1. Phosphorus accounting in wild type and *lpa1* maize kernels

Component	Wt	*lpa1*
	mg g^{-1}	
Total P	3.38	3.25
Phytic acid P	2.93	1.11
Inorganic P	0.14	1.11
Remainder	0.31	1.03

3.3 Metabolomic analysis of individual kernels

Total ion chromatograms from the polar extraction of wild type and *lpa1* kernels were fairly similar (Figure 16-5). Approximately 24 clearly defined peaks were apparent in both samples, with relatively few quantitative differences between the two. However, the high data collection rate of the TOF analyzer coupled with the uniformity of mass spectra across a peak affords automated peak deconvolution, resulting in reliable identification and reproducible quantitation of even very closely eluting metabolites. Total ion chromatograms from the nonpolar extraction of wild type and *lpa1* kernels were also fairly similar to each other (data not shown).

Since we are interested particularly in P-containing metabolites, we took advantage of the fact that m/z 299 in our electron impact mass spectra is diagnostic of such compounds due to instability of the ester-linked trimethylsilylphosphate moiety. This allowed us to identify likely sinks for the unaccounted P in lpa1 kernels, even if these molecules exist at very low relative abundances. Not surprisingly, most of the peaks characterized by an m/z 299 fragment were significantly more abundant in lpa1 kernels compared with wild-type kernels (Figure 16-6). These peaks include the TMS derivatives of phosphoric acid, glycerol-3-phosphate, phosphatidylinositol, and two unknown phosphorylated metabolites (Table 16-2). In addition, an unknown metabolite of phosphatidylinositol, defined as such since it was identified when an authentic phosphatidylinositol standard was subjected to the sample preparation and derivation process, was significantly more abundant in wild-type kernels. However, all of the organic P-containing metabolites were present at low levels in *lpa1* kernels, too low to account for

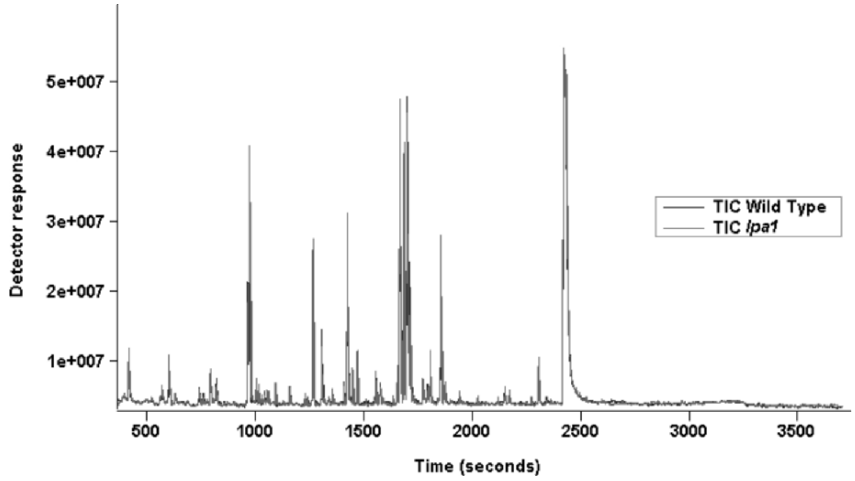


Figure 16-5. Total ion chromatograms (TIC) of polar extractions of wild type and *lpa1* kernels.

a significant proportion of the unaccounted P. Interestingly, our findings are not consistent with the presumed biosynthetic pathway of phytic acid in *lpa1* mutant kernels. This may not be surprising since we worked with fully mature kernels that (1) should exhibit much less active phytic acid synthesis than developing kernels and (2) contain a significant amount of endosperm that does not accumulate phytic acid, thus effectively diluting metabolites involved directly in phytic acid synthesis. For these reasons, we plan to repeat this study with developing embryos.

The relative expression of all the identified polar and nonpolar metabolites in *lpa1* compared with wild-type kernels is presented in Tables 16-2 and 16-3, respectively. There were only two metabolites, glycerol and phosphoric acid, that were found in both polar and nonpolar extracts. Both metabolites were far more abundant in the polar fraction, as expected for molecules with multiple hydroxy groups. Their presence in the nonpolar fraction could be due to hydrolysis of hydrophobic glycerolipids and phospholipids during sample preparation, probably during the high temperature methylation step. Intact glycerolipids or phospholipids were not found in the nonpolar metabolite profiles since they are not detectable with the GC conditions employed. Regardless, it is important to recognize such supposed artifacts of the analytical process when evaluating the biological significance of metabolite profiles. As expected, metabolites represented by two different derivatives (i.e., different numbers of methyl or trimethylsilyl groups) showed similar expression between the two phenotypes, indicating consistent extraction and detection. Aside from the aforementioned P-containing metabolites, the *lpa1* phenotype was associated with little change

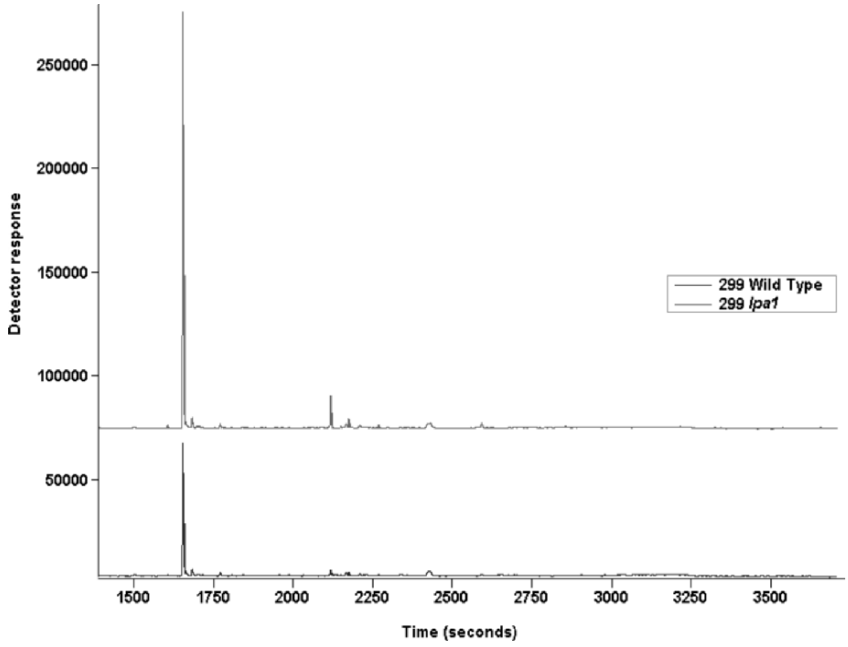


Figure 16-6. Partial extracted ion chromatograms for m/z 299 of polar extractions of wild type and *lpa1* kernels.

in the amounts of the primary metabolites measured. This result suggests either a limited perturbation of primary metabolism by the *lpa1* mutation, or alternatively, the analytical precision and number of replicate samples were insufficient to uncover more subtle changes. A somewhat different picture emerges from the relative expression of identified polar and nonpolar metabolites in *lpa3* kernels compared to their wild-type controls (Tables 16-4 and 16-5, respectively). As in *lpa1* kernels, the relative amounts of several P-containing metabolites were correlated with the *lpa3* genotype. These metabolites were phosphoric acid, glycerol-3-phosphate, sucrose-6-phosphate, phosphatidylinositol, and all three unknown phosphorylated metabolites observed in *lpa1* kernels. As in *lpa1* kernels, all of the organic P-containing metabolites in *lpa3* kernels were present at levels too low to account for a significant portion of the unaccounted P. As in *lpa1* kernels, our results are not consistent with the presumed biosynthetic pathway of phytic acid in *lpa3* kernels. We also plan to repeat this study with developing embryos.

The metabolic profile of *lpa3* kernels exhibits far more differences in non-P-containing metabolites compared to wild type than that of *lpa1*

kernels. Several amino acids, organic acids, sugars, fatty acids, fatty alcohols, alkanes, phenolic acids, and phytosterols were affected. However, the physiological significance of the altered amounts of these diverse primary metabolites is unknown. Several phenolic acids were found in both polar and nonpolar extracts, often as different derivatives. For example, the TMS ester, TMS ether derivatives of caffeic and ferulic acids were found in the polar fraction, while the methyl ester, TMS ether derivatives appeared in the nonpolar fraction that underwent transmethylation. Although there were

Table 16-2. Differential polar metabolite expression in *lpa1* and wt maize kernels

Metabolite[a]	*lpa1*:wt	Metabolite [a]	*lpa1*:wt
Alanine,N,O TMS	0.9	Malic acid TMS	1.1
β-Alanine,N,N,O TMS	0.8	Succinic acid TMS	1.1
4-Aminobutyric acid TMS	0.5**	Glycerol-3-phosphate TMS	1.5**
Asparagine,N,N,O TMS	1.0	*myo*-Inositol-1/3-phosphate TMS	1.3
Asparagine,N,N,N,O TMS	0.5	Phosphatidylinositol TMS	1.7**
Asparatic acid,N,O,O TMS	0.7*	Phosphatidylinositol metabolite TMS	0.8**
Glutamic acid,N,O,O TMS	0.9	Phosphoric acid,O,O,O TMS	1.8***
Glutamine,N,N,O TMS	1.0	Unknown phosphorylated metabolite TMS 1	4.3***
Glycine,N,N,O TMS	0.8	Unknown phosphorylated metabolite TMS 2	1.8**
Homoproline,O TMS	0.9	Arabinitol TMS	1.2
Homoproline,N,O TMS	0.9	Erythritol TMS	0.8
2-Hydroxyglutaric acid TMS	1.2	myo-Inositol TMS	1.7**
Isoleucine,N,O TMS	1.0	Sorbitol TMS	0.6
Leucine,N,O TMS	0.9	Caffeic acid TMS	b
Methionine,N,O TMS	0.8	Ferulate acid TMS	0.2
Phenylalanine,N,O TMS	1.0	Adenine TMS	1.5
Proline,N,O TMS	0.9	Arabinose MeOX2 TMS	0.9
Pyroglutamic acid,N,O TMS	0.9	Fructose MeOX1 TMS	0.8*
Serine,O,O TMS	0.0	Fructose MeOX2 TMS	0.8*
Serine,N,O,O TMS	0.9	Galactose MeOX1 TMS	0.5
Threonine,N,O,O TMS	0.7	Galactose MeOX2 TMS	2.0
Tyrosine,N,O TMS	0.8	Glucose MeOX1 TMS	0.8
Tyrosine, N,O,O TMS	1.0	Glucose MeOX2 TMS	0.8
Valine,N,O TMS	1.0	Glucuronic acid MeOX1 TMS	0.9
Glyceric acid TMS	1.0	Glucuronic acid MeOX2 TMS	b
Glycerol TMS	0.9	Raffinose TMS	0.8
Citric acid TMS	0.9	Sucrose TMS	1.0
Fumaric acid TMS	0.7		

[a]TMS, trimethylsilyl ester; MeOX, methoxyamine.
[b]Not found in wild type.
*Means are significantly different at P<0.1.
**Means are significantly different at P<0.05.
***Means are significantly different at P<0.01.

Table 16-3. Differential nonpolar metabolite expression in *lpa1* and wild type maize kernels

Metabolite[a]	*lpa1*:wt
14:0 Me	1.2
16:0 Me	1.0
16:0 TMS	1.2
16: 1cisΔ7 Me	1.1
17:0 Me	1.1
18:0 Me	1.1
18:1cisΔ9 Me	1.0
18:1cisΔ9 TMS	0.8
18:2cisΔ9,12 Me	1.0
18:2cisΔ9,12 TMS	1.6
18:3cisΔ9,12,15 Me	0.9
20:0 Me	1.5
20:1cisΔ11 Me	2.0
22:0 Me	0.4
23:0 Me	0.8
24:0 Me	1.1
25:0 Me	3.7
26:0 Me	2.6
2HO-20:0 MeTMS	1.3
2HO-22:0 MeTMS	2.2
2HO-24:0 MeTMS	0.5
ρ-Coumaric acid MeTMS	0.7
3,5-Di-tert-butyl-4-hydroxybenzoic acid Ee	1.2**
4-Methoxy, 3-hydroxycinnamic acid Me	1.1
Campesterol TMS	0.6
β-Sitosterol Me	1.1
β-Sitosterol TMS	1.2
Stigmasterol TMS	0.2
Glycerol TMS	b
Phosphoric acid,O,O,O TMS	0.3

[a]Me, methyl ester; TMS, trimethylsilyl ester; MeTMS, methyltrimethylsilyl ester; Ee, ethyl ester.
[b]Not found in *lpa1*.
**Means are significantly different at $P<0.05$.

some free phenolic acids in the polar fraction, as a class they were far more abundant in the nonpolar fraction. This finding suggests that the phenolic acids were liberated by hydrolysis of hydrophobic conjugates, presumably phenolic acids esterified to phytosterols and/or various acyl groups, during sample preparation, most likely during transmethylation. As with glycerol and phosphoric acid in *lpa1* kernels, the biological significance of phenolic

Table 16-4. Differential polar metabolite expression in *lpa3* and wild type maize kernels

Metabolite[a]	*lpa3*:wt	Metabolite [a]	*lpa3*:wt
β-Alanine,N,N,O TMS	0.9	Indoleacetic acid TMS	0.7
ρ-Coumaric acid MeTMS	1.0	Homoproline,N,O TMS	0.2**
2-Hydroxyglutaric acid TMS	1.7**	Isoleucine,N,O TMS	0.7**
4-Aminobutyric acid TMS	1.2	Leucine,N,O TMS	0.6***
4-Hydroxybenzoic acid TMS	0.9	Lysine,N,N,N',O TMS	1.2
5-Hydroxyindoleacetic acid TMS	1.0	Malic acid TMS	0.9
5-Hydroxynorvaline,N,O,O TMS	1.1	Methionine,N,O TMS	0.6**
Adenine TMS	1.0	*myo*-Inositol TMS	1.4
Alanine,N,O TMS	0.9	*myo*-Inositol-1/3-phosphate TMS	1.2
Arabinitol TMS	0.6	Ornithine,N,N,N',O TMS	1.0
Arabinose MeOX2 TMS	1.2	Ornithine,N,N,O TMS	1.2
Asparagine,N,N,N,O TMS	0.8	Phenylalanine,N,O TMS	0.8
Asparagine,N,N,O TMS	1.0	Phenylalanine,O TMS	0.9
Asparatic acid,N,O,O TMS	0.7*	Phosphatidylinositol metabolite TMS	1.5**
Benzoic acid TMS	1.1	Phosphatidylinositol TMS	16.4***
Caffeic acid TMS	0.8	Phosphoric acid,O,O,O TMS	3.5***
Citric acid TMS	3.4***	Proline,N,O TMS	0.4
Erythritol TMS	0.9	Pyroglutamic acid,N,O TMS	1.0
Ferulate acid TMS	0.9	Raffinose TMS	1.1
Fructose MeOX1 TMS	1.0	Ribose MeOX2 TMS	4.6
Fructose MeOX2 TMS	1.0	Serine,N,O,O TMS	0.2**
Fumaric acid TMS	1.1	Serine,O,O TMS	0.9
Galactose MeOX1 TMS	0.9	Sorbitol TMS	1.0
Galactose MeOX2 TMS	1.0	Succinic acid TMS	1.1
Gluconic acid TMS	1.0	Sucrose TMS	1.2**
Glucose MeOX1 TMS	1.0	Sucrose-6-phosphate TMS	2.8***
Glucose MeOX2 TMS	0.9	Threonine,N,O TMS	1.1
Glucuronic acid MeOX1 TMS	0.7	Threonine,N,O,O TMS	0.4**
Glutamic acid,N,O,O TMS	0.7	Tyrosine, N,O,O TMS	1.2
Glutamine,N,N,O TMS	0.7	Tyrosine,N,O TMS	1.0
Glyceric acid TMS	1.2	Unknown phosphorylated metabolite TMS	42.4***
Glycerol TMS	1.4**	Unknown phosphorylated metabolite TMS	7.6***
Glycerol-3-phosphate TMS	2.3***	Uracil TMS	0.8
Glycine,N,N,O TMS	0.8**	Urea,N,N TMS	1.0
Histidine, N,O TMS	1.4	Valine,N,O TMS	0.8
Homoproline,O TMS	0.6*	Valine,O TMS	1.3

[a]TMS, trimethylsilyl ester; MeTMS, methyltrimethylsilyl ester; Ee, ethyl ester; MeOX, methoxyamine.
*Means are significantly different at P<0.1.
**Means are significantly different at P<0.05.
***Means are significantly different at P<0.01.

Table 16-5. Differential nonpolar metabolite expression in *lpa3* and wild type maize kernels

Metabolite[a]	*lpa3*:wt	Metabolite[a]	*lpa3*:wt
14:0 Me	1.3**	1HO-22:0 TMS	3.3
15:0 Me	1.5*	2HO-18:0 MeTMS	1.2
16:0 Me	1.2*	2HO-20:0 MeTMS	2.6**
16:0 TMS	1.3**	2HO-22:0 MeTMS	7.2**
16: 1cisΔ7 Me	1.3*	2HO-24:0 MeTMS	3.5**
17:0 Me	1.1	2HO-25:0 MeTMS	2.3**
17: 1cisΔ10 Me	1.3*	25:0	1.1
18:0 Me	1.1	27:0	1.3***
18:0 TMS	1.1	28:0	0.7
18:1cisΔ9 Me	1.1	29:0	1.4*
18:1cisΔ9 TMS	1.3*	30:0	1.0
18:2cisΔ9,12 Me	1.1	31:0	1.4**
18:2cisΔ9,12 TMS	1.5*	33:0	1.2**
18:3cisΔ9,12,15 Me	1.1	ρ-Coumaric acid MeTMS	1.0
20:0 Me	1.4***	β-Sitosterol Me	1.2*
20:1cisΔ11 Me	1.3**	β-Sitosterol TMS	1.7**
20:2cisΔ11,14 Me	1.4	3-Methoxy, 4-hydroxybenzaldehyde TMS	b
21:0 Me	1.1	4-Hydroxybenzene acetic acid MeTMS	1.5
22:0 Me	1.2	4-Hydroxybenzoic acid MeTMS	6.9**
23:0 Me	1.2*	4-Methoxy, 3-hydroxybenzoic acid MeTMS	1.3
24:0 Me	1.4***	4-Methoxy, 3-hydroxycinnamic acid MeTMS	1.0
25:0 Me	1.1	Caffeic acid MeTMS	1.1
26:0 Me	1.4**	Campesterol TMS	1.8**
1HO-12:0 TMS	1.5**	Ferulic acid MeTMS	1.0
1HO-14:0 TMS	1.2	Stigmasterol Me	2.1**
1HO-18:0 TMS	1.2	Stigmasterol TMS	2.1**

[a]Me, methyl ester; TMS, trimethylsilyl ester; MeTMS, methyltrimethylsilyl ester; Ee, ethyl ester.
[b]Not found in wild type.
*Means are significantly different at P<0.1.
**Means are significantly different at P<0.05.
***Means are significantly different at P<0.01.

metabolites in the polar and nonpolar fractions should be interpreted with the presumed effect of the analytical process in mind. We will understand this phenomenon better when we determine the relative amounts of free and bound phenolic acids in maize kernels by LC/MS.

Our targeted and metabolomic analyses revealed several P-containing metabolites that were much more abundant in both *lpa1* and *lpa3* kernels compared to their wild-type controls that developed within the same ear. Several other metabolites were also affected differentially; these are potential targets for possible metabolic rescue of suboptimal yield. The

metabolomic data are not consistent with the presumed phytic acid biosynthetic pathway in either mutant. This is not surprising for less metabolically active mature kernels, thus the need to extend the analyses to embryos isolated from developing kernels. We also plan to analyze *lpa2* kernels in order to investigate the effects of all three low phytic acid mutations on additional P-containing (and other) metabolites measured by GC/MS and LC/MS. LC/MS will allow us to semi-quantify additional metabolites, and determining their empirical formula with high-resolution mass spectrometry will facilitate their identification.

ACKNOWLEDGMENTS

The authors acknowledge the Organizing Committee of the 3rd International Conference on Plant Metabolomics for the inclusion of this paper in an oral symposia and these Proceedings Oliver Fiehn and the Max Planck Institute of Molecular Plant Biology for sharing of their analytical methods and GC/MS libraries, and Jonathan Lightner, Bruce Orman, and Pioneer/DuPont management for their useful discussions and support.

REFERENCES

Bechtel, D.B. and Wilson, J., 2000, Variability in a starch isolation method and automated digital image analysis system used for the study of starch size distributions in wheat flour, *Cereal Chemistry* **77**:401-405.

Chen P.S., Toribara, T.Y., and Warner, H., 1956, Microdetermination of phosphorus, *Analytical Chemistry* **28**:1756-1758.

Cromwell, G.L. and Coffey, R.D., 1991, Phosphorus: a key essential nutrient, yet a possible major pollutant. Its central role in animal nutrition, in: *Biotechnology in the Feed Industry*, TP Lyons, ed., Alltech Tech Publishers, Nicholasville, KY, pp. 133-145.

Fiehn, O., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry, *Analytical Chem.* **72**:3573-3580.

Guttieria, M., Bowena, D., Dorsch, J.A., Raboy, V., and Souza, E., 2004, Identification and characterization of a low phytic acid wheat, *Crop Science* **44**:418-424.

Hitz, W.D., Carlson, T.J., Kerr, P.S., and Sebastian, S.A., 2002, Biochemical and molecular characterization of a mutant that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds, *Plant Physiology* **128**:650-660.

Holm, P.B., Kristiansen, K.N., and Pedersen, H.B., 2002, Transgenic approaches in commonly consumed cereals to improve iron and zinc content and bioavailability, *Journal of Nutrition* **132**:514S-516S.

Larson, S.R., Young, K.A., Cook, A., Blake, T.K., and Raboy, V., 1998, Linkage mapping of two mutations that reduce phytic acid content of barley grains, *Theoretical and Applied Genetics* **99**:27-36.

O'Dell B.L., de Boland, A.R., and Koirtyohann S.R., 1972, Distribution of phytate and nutritionally important elements among the morphological components of cereal grains, *Journal of Agricultural and Food Chemisty* **20**:718-721.

Picchioni, G.A., Watada, A.E., and Whitaker, B.D., 1996, Quantitative high-performance liquid chromatography analysis of plant phospholipids and glycolipids using light-scattering detection, *Journal of the American Oil Chemists Society* **31**:217-221.

Raboy, V., Gerbasi, P.F., Young, K.A.,, Stoneberg, S.D., Pickett, S.G., Bauman, A.T., Murphy, P.P., Sheridan, W.F., and Ertl, D.S., 2000, Origin and seed phenotype of maize *low phytic acid 1-1* and *low phytic acid 2-1*, *Plant Physiology* **124**:355-368.

Raboy, V., 2002, Progress in breeding low phytate crops, *Journal of Nutrition* **132**:503S-505S.

Shi, J., Wang, H., Wu, Y., Hazebroek, J., Meeley, R.B., and Ertl, D.S., 2003, The maize low-phytic acid mutant lpa2 is caused by mutation in an inositol phosphate kinase gene, *Plant Physiology* **131**:507-515.

Perry, T.W., 1988, Corn as a livestock feed, in: *Corn and Corn Improvement*, Sprague, G.F. and Dudley, J.W., eds., 3rd ed., American Society of Agronomy, Madison, WI, pp.941-963.

Strother S., 1980, Homeostasis in germinating seeds, *Annals of Botany* **45**:217-218.

Wilcox, J., Premachandra, G., Young, K., and Raboy, V., 2000, Isolation of high seed inorganic P, *low phytic acid* soybean mutants, *Crop Science* **40**:1601-1605.

Zhou J.R. and Erdman J.W. Jr., 1995, Phytic acid in health and disease, *Critical Reviews in Food Science and Nutrition* **35**:495-508.

Chapter 17

THE LOW TEMPERATURE METABOLOME OF *ARABIDOPSIS*

Gordon R. Gray[1] and Doug Heath[2]
[1]*Department of Plant Sciences, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5A8, Canada;* [2]*Phenomenome Discoveries Inc., 204-407 Downey Road, Saskatoon, Saskatchewan S7N 4L8, Canada*

1 INTRODUCTION

Low temperature represents an environmental variable which significantly affects plant performance, causing losses in productivity and limiting geographical distribution of many species (Boyer, 1982). Low temperature exposure has consequences for most biological processes, and freezing temperatures often lead to severe damage due to the cellular dehydration which occurs upon ice formation (Thomashow, 1999; Xin and Browse, 2000). However, the exposure of certain plant species, including *Arabidopsis*, to low temperatures (5–10°C), initiates a series of events which, over a varying period of time, results in these plants acclimating to the lower growth temperature and becoming more freezing tolerant (Browse and Xin, 2001; Stitt and Hurry, 2002). This is referred to as cold acclimation.

Cold acclimation is complex and involves numerous molecular, physiological and biochemical changes. Due to its agricultural importance, considerable effort has been directed at understanding the phenomenon of cold acclimation at the molecular genetic level (Thomashow, 2001; Fowler and Thomashow, 2002). Equally important are the biochemical changes which occur at the level of the metabolome (Cook et al., 2004; Kaplan et al., 2004). Examination at the metabolic level offers a direct link between a gene and function, as well as the elucidation of relationships that occur through complex biochemical regulation (Fiehn, 2002).

Our goal was to examine the effects of cold acclimation on metabolome from a global perspective; incorporating changes from all metabolic pathways using an unbiased, non-targeted approach afforded us by Fourier transform ion

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 239–246.

cyclotron mass spectrometry (FTMS) technology (Aharoni et al., 2002; Brown et al., 2005).

2 MATERIALS AND METHODS

2.1 Plant material, growth conditions and experimental design

Seeds of *Arabidopsis thaliana* (L.) Heynh., ecotype Columbia were germinated from seed under controlled environment conditions at 23°C with an 8 h photoperiod and growth irradiance of 90 μmol quanta $m^{-2}s^{-1}$ as described previously (Gray et al., 2003). Plants were allowed to grow under these non-acclimating conditions for 27 days and then shifted to cold acclimating conditions at 4°C with the same photoperiod and irradiance as the non-acclimated control plants. Leaves were sampled in triplicate biological replicates, flash frozen in liquid nitrogen and ground to powder.

2.2 Non-targeted analyses of metabolites using FTMS

2.2.1 Sample extraction and preparation

Fifty mg of ground leaf material was triple extracted using 1 mL of 1% (v/v) formic acid and 3 × 3 mL of ethyl acetate. The aqueous fractions were centrifuged for 10 min, the supernatant removed and stored at –80°C until analysis. The combined ethyl acetate fractions were evaporated to dryness under nitrogen, reconstituted in 1 mL of 100% (v/v) methanol and also stored at –80°C prior to FTMS analysis. Samples were diluted 1:19 prior to electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) analyses. Dilution for all negative and positive ion ionization analyses occurred using methanol: 0.1% (v/v) ammonium hydroxide (50:50, v/v) and methanol: 0.1% (v/v) formic acid (50:50, v/v) as mobile phases, respectively.

2.2.2 Instrument operating conditions

All analyses were performed on a Bruker Daltonics APEX III Fourier transform ion cyclotron resonance mass spectrometer equipped with a 7.0-Tesla actively shielded superconducting magnet (Bruker Daltonics, Billerica, MA, USA). Samples were introduced separately by direct injection into ESI or APCI sources. Flow rates were 600 μL h^{-1} for both ESI and APCI. Ionization (ESI and APCI) and ion transfer/detection parameters were optimized using a standard mix of serine, tetra-alanine, reserpine, Hewlett-

Packard tuning mix, and the adrenocorticotrophic hormone fragment 4–10. In addition, the instrument conditions were tuned to optimize ion intensity and broadband accumulation over the mass range of 100 to 1000 a.m.u. according to the instrument manufacturer recommendations. A mixture of the above-mentioned standards was used to internally calibrate each sample spectrum for mass accuracy over the acquisition range of 100 to 1000 a.m.u. Using a linear least squares regression line, mass axis values were calibrated so that each internal standard mass peak had a mass error of <1 ppm compared to its theoretical mass.

2.2.3 Raw spectra processing and data alignment

Using XMASS software (v 6.0.3) from Bruker Daltonics Inc., data file sizes of 1 megaword were acquired and zero-filled to 2 megawords. A simm data transformation was performed prior to Fourier transform and magnitude calculations. The mass spectra from each analysis were integrated, creating a peak list that contained the accurate mass and absolute intensity of each peak. In order to compare and summarize data across different ionization modes and polarities, all detected mass peaks were converted to their corresponding neutral masses assuming hydrogen adduct formation.

A self-generated 2-dimensional (mass versus sample intensity) array was then created using *DISCOVA*rray software (Phenomenome Discoveries Inc., Saskatoon, SK, Canada). The data from multiple files were integrated and this combined file was then processed to determine all of the unique masses. The average of each unique mass was determined, representing the y-axis. A column was created for each file that was originally selected to be analysed, representing the x-axis. The intensity for each mass found in each of the files selected was then filled into its representative x, y coordinate. Coordinates that did not contain an intensity value were left blank. Once in the array, the data was further processed, visualized, and interpreted, as well as a putative chemical identity assigned.

2.2.4 Statistical analyses

The array was imported as a text file into GeneLinker Gold v. 3.0 (Predictive Patterns Software Inc., Kingston, ON, Canada) for statistical analyses. An *F*-Test was used to create a list of masses that had significant intensity changes ($p \leq 0.05$) between any two sample means generated from biological triplicates. These masses were designated as separate component names, and the corresponding sample peak intensities were used for subsequent hierarchical cluster analysis (HCA) by components.

3 RESULTS

Typically, studies examining cold acclimation grow plants under non-acclimating (23°C) conditions to a certain developmental age and then shift them to the cold acclimating temperature (4°C). In the present study, we employed a similar experimental design, examining shifted leaves for an extensive time course (up to 49 days under cold acclimating conditions).

In total, 1187 compounds were found in the *Arabidopsis thaliana* leaf extracts from all sampling points. These data were filtered such that only those compounds that were observed to significantly change during the experiment were retained (*F*-test, $p \leq 0.05$). These components (593 compounds) were further subjected to pair wise analysis (data not shown).

We employed HCA to further examine the differences and similarities between the leaf putative metabolite profiles from our filtered data sets. The results of the HCA demonstrate that the cold acclimated leaves present a constantly changing metabolic phenotype and this became more distinct

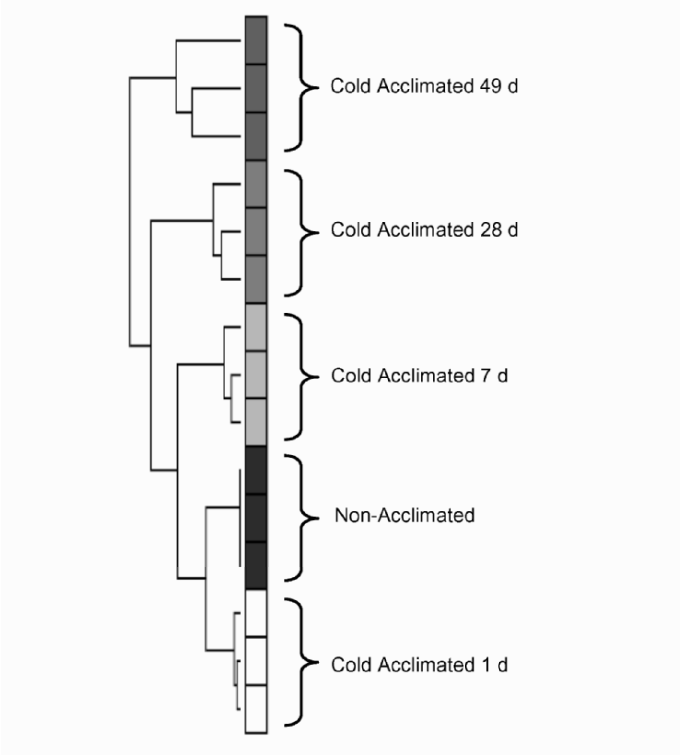


Figure 17-1. The effect of cold acclimation on the metabolic profile of shifted leaves in comparison to the non-acclimated control as determined by HCA.

from the non-acclimated control the longer the shifted leaves remained at low temperature (Figure 17-1). This is indicative of a complete reprogramming of the metablome in response to low temperature. This reorganization of the metabolome is further supported by the pair-wise comparisons of the changing compounds (data not shown).

To confirm and validate the results of our global analysis with compounds known to change during cold acclimation, we examined the metabolites associated with photosynthetic carbon metabolism. In *Arabidopsis* and numerous other cold-tolerant plant species, a reprogramming of photosynthetic carbon metabolism is frequently observed which results in the preferential accumulation of soluble sugars (Stitt and Hurry, 2002). These are thought to be an essential element for acclimation to low growth temperatures and the attainment of maximal freezing tolerance for winter survival (Stitt and Hurry, 2002; Strand et al., 2003).

The responses observed for the total hexose (Figure 17-2a), di-hexose (Figure 17-2c) and hexose-phosphate (Figure 17-2e) pools are consistent with those obtained from previous studies examining the individual compounds which would comprise these pools (Strand et al., 1997, 1999; Hurry et al., 2000). Representative spectra for these pools are presented in Figures 17-2b, d, and f and correspond to the detect ion masses obtained from the negative ESI mode of highly polar fraction in each case. These were 179.0562, 341.1083 and 259.0222 a.m.u respectively.

4 DISCUSSION

FTMS allows for the separation of metabolites in a sample solely by mass resolution (Brown et al., 2005). Based on accurate mass determination, the elemental composition is determined which can then lead to the putative identification of the metabolite. Relative quantification is achieved by comparing absolute intensities of each mass (Aharoni et al., 2002). This technology does not allow us (in most cases) to unequivocally identify specific metabolites. However, it does allow us to detect a comprehensive list of masses (based on *m/z* values) which are reflective of individual components (or putative metabolites).

The shift from non-acclimating growth conditions to cold acclimating temperatures is characterized by transient, physiological, biochemical and molecular perturbations. These transient stress responses lead to stable, long-term adjustments that reflect developmental responses to the new growth temperature (Huner et al., 1993). Leaves shifted to low temperature present putative metabolite profiles which are constantly changing in an attempt to reach a cold acclimated metabolic state. Thus, metabolome analysis indicates that the metabolic alterations which occur in *Arabidopsis* leaves

subjected to low temperature are representative of a global reprogramming of metabolism.

Our results are consistent with the notion that photosynthetic carbon metabolism is reprogrammed in response to low temperature (Stitt and Hurry, 2002). Whereas previous conclusions were the result of studies which examined each metabolite in isolation, our data sets are reflective of the entire metabolome. By measuring an entire spectrum of compounds versus an individual or group of metabolite(s), a global unbiased assessment of metabolic processes relative to cold acclimation was determined. Clearly, the

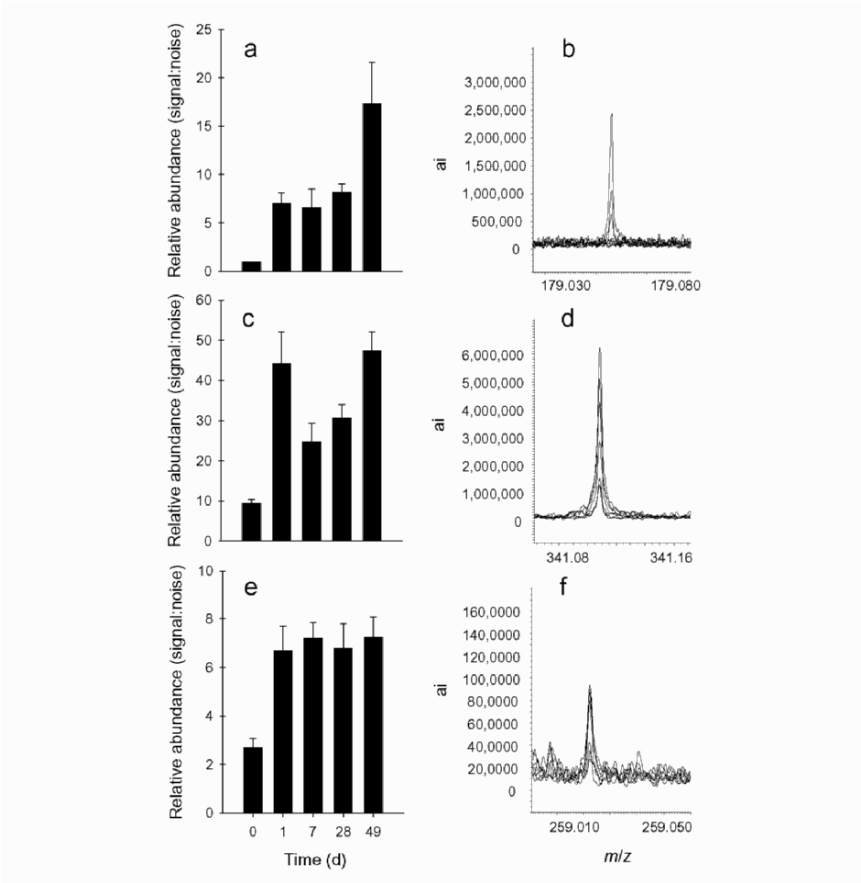


Figure 17-2. Abundance of the total hexose pool (*a*), total di-hexose pool (*c*), and total hexose-phosphate pool (*e*) in leaves shifted to cold acclimating conditions. Values represent means ± SD (n = 23). Representative spectra from (*a*), (*c*), and (*e*) are shown in (*b*), (*d*), and (*f*) respectively.

regulation or reprogramming of metabolism within the leaf during cold acclimation extends beyond that of photosynthetic carbon metabolism.

Techniques allowing a full description of the metabolome status of an organism can strongly complement existing functional genomic approaches (Sumner et al., 2003). Several studies relate stress conditions to changes in gene expression patterns at the mRNA (or protein) level (Fowler and Thomashow, 2002). However, care must be taken in the interpretation of these studies as our data demonstrate that there are fundamental differences at the level of the metabolome, which are dependent on the duration of low temperature exposure. Our results highlight the importance of proper experimental design and the significance of leaf prehistory (Krol et al., 1984; Huner et al., 1993; Gray et al., 2003) when studying complex environmental stress responses.

ACKNOWLEDGEMENTS

We thank Ms. Carmen L. Whitehead for technical assistance and Drs. N.P.A. Huner and V.M. Hurry for their valuable discussions during the course of these studies. This work was supported by a Research Grant to G.R.G. from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

Aharoni, A., De Vos, C.H.R., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., and Goodenowe, D.B., 2002, Non-targeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry, *OMICS*. **6**:217-234.

Boyer, J.S., 1982, Plant productivity and environment, *Science.* **218**:443-448.

Brown, S.C., Kruppa, G., and Dasseux, J-L., 2005, Metabolomics applications of FT-ICR mass spectrometry, *Mass Spec. Rev.* **24**:223-231.

Browse, J., and Xin, Z.G., 2001, Temperature sensing and cold acclimation, *Curr. Opin. Plant Biol*. **4**:241-246.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F., 2004, A prominent role for the CBF cold responsive pathway in configuring the low-temperature metabolome of *Arabidopsis*, *Proc. Natl. Acad. Sci. USA*. **101**:15243-15248.

Fiehn, O., 2002, Metabolomics: The link between genotypes and phenotypes, *Plant Mol. Biol*. **48**:155-171.

Fowler, S., and Thomashow, M.F., 2002, *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway, *Plant Cell.* **14**:1675-1690.

Gray, G.R., Hope, B.J., Qin, X., Taylor, B.G., and Whitehead, C.L., 2003, The characterization of photoinhibition and recovery during cold acclimation in *Arabidopsis* thaliana using chlorophyll fluorescence imaging, *Physiol Plant*. **119**:365-375.

Huner, N.P.A., Öquist, G., Hurry, V.M., Krol, M., Falk, S., and Griffith, M., 1993, Photosynthesis, photoinhibition and low temperature acclimation in cold tolerant plants, *Photosynth Res*. **37**:19-39.

Hurry, V., Strand, Å., Furbank, R., and Stit, M., 2000, The role of inorganic phosphate in the development of freezing tolerance and the acclimatization of photosynthesis to low temperature is revealed by the *pho* mutants of *Arabidopsis thaliana*, *Plant J*. **24**:383-396.

Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y., and Guy, C.L., 2004, Exploring the temperature-stress metabolome of Arabidopsis, *Plant Physiol*. **136**:4159-4168.

Krol, M., Griffith, M., and Huner, N.P.A., 1984, An appropriate physiological control for environmental temperature studies: comparative growth kinetics of winter rye, *Can J Bot*. **62**:1062-1068.

Stitt, M., and Hurry, V., 2002, A plant for all seasons: alterations in photosynthetic carbon metabolism during cold acclimation in *Arabidopsis*, *Curr. Opin. Plant Biol*. **5**:199-206.

Strand, Å., Foyer, C.H., Gustafsson, P., Gardeström, P., and Hurry, V., 2003, Altering flux through the sucrose biosynthesis pathway in transgenic *Arabidopsis thaliana* modifies photosynthetic acclimation at low temperature and the development of freezing tolerance, *Plant Cell Environ*. **26**:523-535.

Strand, Å., Hurry, V., Gustafsson, P., and Gardeström, P., 1997, Development of *Arabidopsis thaliana* leaves at low temperature releases the suppression of photosynthesis and photosynthetic gene expression despite the accumulation of soluble carbohydrates, *Plant J*. **12**:605-614.

Strand, Å., Hurry, V., Henkes, S., Huner, N., Gustafsson, P., Gardeström, P., and Stitt, M., 1999, Acclimation of Arabidopsis leaves developing at low temperatures. Increasing cytoplasmic volume accompanies increased activities of enzymes in the Calvin cycle and in the sucrose-biosynthesis pathway, *Plant Physiol*. **119**:1387-1397.

Sumner, L.W., Mendes, P., and Dixon, R.A., 2003, Plant metabolomics: large scale phytochemistry in the functional genomics era, *Phytochem*. **62**:817-836.

Thomashow, M.F., 1999, Plant cold acclimation: freezing tolerance genes and regulatory mechanisms, *Annu. Rev. Plant Physiol Plant Mol. Biol*. **50**:571-599.

Thomashow, M.F., 2001, So what's new in the field of plant cold acclimation? Lots! *Plant Physiol*. **125**:89-93.

Xin, Z., and Browse, J., 2000, Cold comfort farm: the acclimation of plants to freezing temperatures, *Plant Cell Environ*. **23**:893-902.

Chapter 18

CLONING, EXPRESSION AND CHARACTERIZATION OF A PUTATIVE FLAVONOID GLUCOSYLTRANSFERASE FROM GRAPEFRUIT (CITRUS PARADISI) LEAVES

Tapasree Roy Sarkar, Christy L. Strong, Mebrahtu B. Sibhatu, Lee M. Pike, and Cecilia A. McIntosh
Department of Biological Sciences, Box 70703, East Tennessee State University, Johnson City, Tennessee 37614, USA

Abstract: As part of an ongoing effort to understand the regulatory role of glucosylation in grapefruit bitter compound production and overall flavonoid secondary metabolism, PSPG box gene-specific primers were designed and used to "fish" out potential secondary product glucosyltransferase (GT) clones. This is a report on the isolation of the first full-length putative GT clone, its expression, and evaluation of its activity using common flavonoids or aglycones of flavonoids commonly found in grapefruit tissue. While sequence analysis strongly supports this clone being a secondary product GT, it did not transfer labeled glucose from UDP-14C-glucose to any of the flavonoid substrates tested.

Key Words: glucosyltransferase; expression; secondary product; grapefruit; flavonoid.

1 INTRODUCTION

A unique biochemical characteristic of higher plants is the production of a wide variety of natural products or "secondary metabolites." Many types of compounds fall into this category; the major groups are phenolics, alkaloids, and terpenoids. The roles of these substances in plant biochemistry and function are extremely diverse. Furthermore, many of these compounds have found widespread utilization in the livelihood of human society. The roles and uses for secondary metabolites include (but are not limited to) flower colorations and UV patterning, antibiotic, and other medicinal uses,

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 247–257.

dyes, pesticides, gums, detergents, and flavoring agents. While virtually all higher plants produce "secondary metabolites" and some of these compounds are fairly ubiquitous, in many cases specific compounds or classes of compounds are made and/or accumulated during the growth and development of specific plant groups. As a result, many plant families possess characteristic natural product profiles.

The research focus of our laboratory has been to study the regulation of biosynthesis of specific flavonoids (a major group of phenolics) and to elucidate factors that control flavonoid synthesis and accumulation during plant development and growth. Our specific focus is elucidation of the regulation of glucosylation of different subclasses of flavonoids resulting in production of the derivatives (e.g., glycosides) actually found in plant tissues. In order to gain a basic knowledge of the regulatory system, it is critical to understand factors involved in regulating biosynthesis and accumulating secondary metabolites, as well as roles of specific compounds made by a plant in its normal physiology and development. This information is important for understanding potential repercussions of altering these factors during production of transgenic plants.

With the exception of the "flavonoid" 3-O-glucosyltransferases involved in flavonol and/or anthocyanin synthesis, relatively little is known of regulation of the myriad of enzymes involved in flavonoid ring glucosylation and subsequent removal of flavonoid aglycones from the "ring converting" metabolic pool. *Citrus paradisi* (grapefruit) is well-known for the presence of high levels (up to 40–70% dry weight) of a bitter flavanone diglycoside (naringin) in very young leaves and fruits and accumulation in specific tissues of mature fruit (Jourdan et al., 1985; McIntosh and Mansell, 1990 and ref. therein; McIntosh and Mansell, 1997). Grapefruit and other citrus are known for their accumulation of flavanone and flavone glycosides in addition to the more common flavonol glycosides. This makes them an excellent source for the study of flavonoid metabolic regulation and the role glucosylation plays in that regulation.

Previous efforts to elucidate potential levels of control of flavanone glycosylation in grapefruit have indicated that regulation and control of ring-substituting branch points have a significant role in this process (McIntosh and Mansell, 1990; McIntosh et al., 1990; Durren and McIntosh, 1999; Pelt et al., 2003). Of special interest is a flavanone-specific 7-O-glucosyltransferase (EC 2.4.1.185) with some unique biochemical characteristics, and up to four additional flavonoid glucosyltransferases (GTs) from young grapefruit leaves have been at least partially characterized (McIntosh et al., 1990).

We recently initiated efforts to "fish" out putative secondary product GT clones from a young grapefruit leaf cDNA library for subsequent expression and characterization as a prelude to structure/function analysis of flavonoid GTs. We report here on the isolation of the first full-length putative GT

clone as well as the expression and analysis of its activity with flavonoid substrates.

2 METHODS

2.1 RNA isolation

Seeds of *Citrus paradisi* (v. Duncan) were obtained from the Citrus Budwood Registry, FDACS (Winter Haven, Florida, USA), soaked under running water overnight, and grown under greenhouse conditions. Total RNA was extracted from very young, metabolically active light green leaves from 2- to 3-month-old seedlings using the RNeasy Plant Mini Kit with shredder column (Qiagen).

2.2 Preparation of cDNA and amplification of putative plant secondary product glucosyltransferases

A SMART RACE cDNA amplification kit (Clonetech) was used according to manufacturer's instructions to construct both 3′ and 5′ RACE-ready cDNA from total cellular RNA in order to increase likelihood of obtaining cDNA with intact 5′ and 3′ ends. Gene-specific primers (GSPs) were designed (OLIGO Primer Analysis, version 5) using a highly conserved portion of the plant secondary product glucosyltransferase (PSPG) box, a 44 amino-acid long consensus sequence and a component of the UDP-glucose binding domain. Since nothing is known about possible preferential codon usage in grapefruit, the GSPs (20-mers) were designed to include universal bases (*) or degeneracy in key positions. GSP1F: 5′ACG(TA)CAT(C)TGC(T)GG*TGGAAT(C)TC-3′ and GSPR: 5′GAA(G)TTCCA*CCG(A)CAA(G)TGC(AT)GT-3′. These primers were synthesized by Integrated DNA Technologies (IDT) and used to PCR amplify putative GT sequences from SMART RACE cDNA using a gradient PCR in order to determine optimal annealing conditions.

Several PCR products were obtained, gel-purified (GenElute Minus EtBr Spin columns; Sigma) and cloned using a TOPO-TA cloning kit (Invitrogen) and transformed into PCR-4-TOPO cells (Invitrogen) for further analysis. Colonies were blue/white selected for transformants. Cyclo-Prep Miniprep Plasmid DNA purification kit (Amresco) was used for DNA isolation. Several candidate partial clones representing 5′ ends were obtained. A clone specific primer (CSP2; 5′GTGGTCTTCCCCTGACGAGTA-3′) was designed to obtain clones corresponding to the remaining 3′ portion of putative GT1 (PGT1) with sufficient overlap to confirm sequence segments belonged to the same clone. Sequencing was performed by the University of Tennessee

Molecular Core Facility (Knoxville), and clones were analyzed for overlap quality using the BioEdit Sequence Alignment editor program.

2.3 Sequence analysis and identification

The 1106 bp PGT1 sequence (Figure 18-1) was compiled from the sequence of the 5′ clone obtained using GSP1F and the 3′ clone obtained using CSP2. PGT1 was examined first for presence of a continuous open reading frame that would contain a potential PSPG box with critical conserved residues. Subsequently, PGT1 nucleotide and inferred amino acid sequences were used in BLAST (Altschul et al., 1990) and FASTA (Pearson, 1999) searches to evaluate whether the clone corresponded to any previously reported. Results showed similarity to plant GTs in general, although no specific absolute matches were obtained.

2.4 Amplification and expression of full-length PGT1 clone

In order to obtain a full-length clone of PGT1 for further study, primers were designed from a 5′ region from bp 7 through bp 24 (Figure 18-1) before the start codon (GTSP5F; 5′GGGATGAAGTTGGCACTA-3″) and from a 3′ region from bp 1082 through bp 1062 (Figure 18-1) just after the stop codon (GTSP6R; 5′-TTAGAGTTTAAAGGCCTGTGG-3′) and used with the SMART RACE cDNA library to amplify a full-length clone. A single PCR product was obtained and TOPO-cloned as previously described. Clone identity was confirmed by sequencing.

Amplification of PGT1 from the TOPO clone was done using primers designed to incorporate an NcoI restriction site at the 5′ end that encompassed the start codon and a SalI restriction site at the 3′ end after the stop codon. A single PCR product was obtained, gel purified, cut with NcoI and SalI, directionally cloned into expression vector pCD1 and the sequence verified. The recombinant pCD1 vector with PGT1 insert was subsequently transformed into expression host *Escherichia coli* BL21(DE3)RIL. Protein induction was performed at 27C and 37C and expressed protein levels monitored after 4.5 hrs and 8 hrs. Optimal production of soluble PGT1 was found at 27C for 4.5 hrs. Expressed PGT1 was isolated in soluble form from induced cells using a lysozyme method adapted from Novagen (2002).

2.5 Test of PGT1 flavonoid GT activity

PGT1 was tested for flavonoid GT activity by using a method adapted from McIntosh et al. (1990) and an extract from young grapefruit leaves was used as a positive control. All enzyme sources were buffered at pH 7.5 and contained 14 mM βME. Representative flavanone, flavone, flavonol, and

chalcone aglycones were chosen for testing (naringenin, naringenin chalcone, hesperetin, apigenin, kaempferol, and quercetin) using information on presence of naturally occurring derivatives in grapefruit or compounds that were shown to be acceptable substrates for grapefruit flavonoid glucosyltransferases (McIntosh and Mansell, 1990; McIntosh et al., 1990). Reactions were as follows: 5 µL aglycone (50 nmol in ethylene glycol monomethylether), 10 µL UDP-^{14}C-glucose (100,000 dpm; 100 nmol), and enzyme sample (30 µL for the extracted cell pellets; 60 µL for culture supernatants and crude grapefruit leaf extract) in a total reaction volume of 75 µL. Extracts from uninduced cultures were used as negative controls. Reactions were incubated at 30C for the times specified. Reactions were stopped and incorporation of labeled glucose into flavonoid glycosides was determined as previously described (McIntosh et al., 1990).

3 RESULTS

3.1 Obtaining PGT1 full-length clone and inferred amino acid sequence alignment with other GTs

Use of GTSP5F and GTSP6R primers with young grapefruit leaf cDNA gave a single PCR product that was cloned and sequenced. Confirmation of having a full-length clone was confirmed by location of an unambiguous start codon and a contiguous open reading frame that was preceded by 2 in-frame stop codons at the 5′ end (Figure 18-1) and with the PSPG box inframe. The nucleotide sequence and an amino acid sequence deduced by analysis of the continuous ORF were subject to FASTA searches. The top 4 amino acid alignments for proteins of known function (Kita et al., 2000; Hirotani et al., 2000; Ford et al., 1998) were flavonoid 3-O-GTs from *Ipomoea purpurea* ($E = 8.5e^{-33}$) and *Vitis vinifera* ($E = 3.2e^{-31}$), a flavonoid-7-O-GT from *Scutellaria baicalensis* ($E = 6.2e^{-36}$), and a limonoid GT from *Citrus unshiu* ($E = 4e^{-39}$). Amino acid alignment of PGT1 with 3 of these GTs is shown in Figure 18-2. Within the PSPG box, PGT1 has 63% identity with the *Citrus* limonoid GT and the *Vitis* F3-O-GT, 65% identity with the *Impomoea* F3-O-GT, and 61% identity with the *Scutellaria* F7-O-GT. Outside of the PSPG box, the sequences are quite different from each other.

3.2 Production of PGT1 and test for flavonoid GT activity

PGT1 production was induced at 37C and 27C for 4.5 h and 8 h as previously described and results analyzed by SDS-PAGE using 10% gels followed by staining with Coomassie. Induction at 37C resulted in PGT1 being deposited in inclusion bodies (data not shown). Induction at 27C for

4.5 h gave optimal production of soluble PGT1 protein within the cells (Figure 18-3). The approximately 34 kDa pPGT1 protein is indicated by an arrow.

```
                ...
1     ACGCGGGGGA TGAAGTTGGC ACTACCTCTC CTCAGAGAAT ATTTTGACTC
                                     ...
51    ATCCTCTCAG CCATAAACTA ACAAGTTAAA ACTATTTTTG TGTTTCAGAT

101   GAAAAATTCC TTACGGATGA CACTCTTGAG AAACCTATTG ATTGGATCCC

151   CGGCATGAGC AATATTCGGC TCAGGGATTT ACCAAGCTTT ATCAGAACCA

201   CCGACCCTAA CGAAATTATG TTCGATTTCA TGGGCTCAGA AGCACAAAAT

251   TGCTTCAGAT CTTCTGCAAT CATATTTAAC ACATTCGATG AGTTTGAACA

301   TGAAGCTTTA GAGGTTATTG CTTCGAAATT TCCTAACATT TACACCGTAG

351   GTCCACTCCC GTTGCTCTGC AAGCAAGTGG ATGAAACCAA ATTTAGGTCA

401   TTTGGATCAA GCTTGTGGAA GGAAGACACT GACTGTCTCA AATGGCTCGA

451   CAAAAGAGAC GCCAATTCAG TTGTGTACGT TAATTATGGC AGCGTGACTG

501   TGATGTCAGA GTAACACTTG ACAGAATTTG CATGGGGTCT TGCAAATAGC

551   AAGCGTCCAT TTTTATGGAT TCTTAGGCCG GACGTTGTGA TGGGCGACTC

601   CGTGGTCTTG CCTGACGAGT ATTTTGAAGA GATCAAGGAT AGAGGATTGA

651   TAGTTAGCTG GTGCAACCAA GAGCAAGTGC TGTCGCACCC CTCAGTTGGA

701   GCTTTTCTGA CACATTGCGG ATGGAACTCT ACAATGGAGA GTATTTGCGG

751   TGGCGTGCCT GTAATTTGCT GGCCTTTCTT TGTTGAGCAA CAAACAAATT

801   GCAGATATGC ATGCACAACT TGGGGCATTG GCATGGAAGT CAATCATGAT

851   GTGAAGCGTG GTGACATTGA AGCTCTTGTT AAGGAAATGA TGGAAGGAGA

901   TGAAGGAAAG AAAATGAGGC AGAAGGCTTG GGAATGGAAA AAGAAAGCTG

951   AAGCAGCGAC TGCCGTTGGA GGTCAGTCTT ACAATAATTT TGACAGATTA

1001  GTTAAGATGG TTCTTCACCA AGGAAATTGG ACCGGAAACG AAACCCTTCA

1051  CTAGTCCGTC GCCACAGGCC TTTAAACTCT AATAAATATC TTCTTGGAGT

1101  TAAAAACAAA AAAAAAAAAA AAAAAAAAA
```

Figure 18-1. Compiled Full-Length Nucleotide Sequence of PGT1 Clone. Sequences from GSP generated 5′end clone 110-2A2 and 3′ clone 126-1C obtained using a CSP designed from 110-2A2. The start and stop codons are in bold typeface and overlined, the PSPG box region is in bold typeface, and stop codons preceding the start codon are marked by **...** above. The subsequent full-length clone sequence corresponded to bp 7 through bp 1082.

```
                10        20        30        40        50        60
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     ------------------------------------------------------------MRKAG
FGT     ---------MGQLHIVLVPMIAHGHMIPMLDMAKLFSSRGVKTTIIATPAFAEPIRKARE
3GT     ------------------------------------------------------------
PGT1    ------------------------------------------------------------
                70        80        90        100       110       120
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     NFTYEPTPVGDGFIRFEFFEDGWDEDDPRREDLDQYMAQLELIGKQVIPKIIKKSAEEYR
FGT     SGHDIGLTTTKFPPKGSSLP.NIRSL.QVTD..LPHFFRALELLQEPVEE.MEDLKPDCL
3GT     -----------------------------------MAMPGNYVKA.AEAEAETGTKFG
PGT1    ------------------------------------------------------------
                130       140       150       160       170       180
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     PVSCLINNPFIPWVSDVAESLGLPSAMLWVQSCACFAAYYHYFHGLVPFPSEKEPEIDVQ
FGT     VSDMFLPWTTDSAAKFGIPR.LFHGTS.FARCF.EQMSIQKPYKNVSSDSEPFVLRGLPH
3GT     CFLTDAFLW.GGDLAAERGGVPWIALWTAGACSISAHL.TDFVRS.AAATPTGNGNVLE.
PGT1    ------------------------------------------------------------M
                190       200       210       220       230       240
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     LPCMPLLKHDEMPSFLHPSTPYPFLRRAILGQYENLGKPFCILLDTFYELEKEIIDYMAK
FGT     EVSFVRTQIPDYELQEGGDDAFSKMAKQMRDADKKSTGDVINSFEELESEYADYNKNVFG
3GT     KLKVIPGMSEISIGEMPGEILAKD.QEPFP.MIY.MALKLPGANAVVINSFQNLEPTVTD
PGT1    SNIRLRDLPSFIRTTDPNEIMFD.MGSEAQNCFRSSAII.NTFDEFEH.ALEV.ASKFPN
                250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     ICPIKPVGPLFKNPKAPTLTVRDDCMKPDECIDWLDKKPPSSVVYISFGTVVYLKQEQVE
FGT     KKAWHIGPLKLF.NR.EQKSSQRGKESAIDDHEC.AWLNSKKPNSVVYMCFGSMATFTPA
3GT     DIRS.LQKVFNIG.MILRQAAAATPGP.ISDDHNCIPWVD.LPPASPPAVYLSFGSGLTP
PGT1    .YTVG.LPL.C.QVDETKFRSFGSSLWKEDTDCLKWLDKRDANSVVYVNYGSVTVMSEQH
                310       320       330       340       350       360
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     EIGYALLNSGISFLWVMKPPPEDSGVKIVDLPDGFLEKVGDKGKVVQ<b>WSPQEKVLAHPSV</b>
FGT     QLHETAVGLES.GQDFIWVVRNGGENEDWLPQGFEERIK.KGLMIRG.<b>A..VMI.D...T</b>
3GT     PPDEIVALAEALEAKRAPFLWSLKPHGVKH..E....RTKEF..I.P.<b>A..VQ..S..G.</b>
PGT1    LTEF.WGLANSKRPFLWILR.DVVMGDS.V...ETF.EIK.R.LI.S.<b>CN..Q..S....</b>
                370       380       390       400       410       420
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     <b>ACFVTHCGWNSTMESLASGVPVITFPQWGDQ</b>VTDAMYLCDVFKTGLRLCRGEAENRIISR
FGT     <b>GA.........L.GICA.L.MV.W.VFAE</b>.FYNEKLVTE.L...VSVGNKKWQRVGEGV
3GT     <b>GA.........L.AISF..CL.CR.FY</b>...QINSRFVES.WEI.VKVEG.KFTKDETLK
PGT1    <b>GA.L.........ICG.....CW.FFVE</b>.Q.NCR.A.TTWGI.MEVNHDVKRGDIEAL
                430       440       450       460       470       480
        ....|....|....|....|....|....|....|....|....|....|....|....|
LGT     DEVEKCLLEATAGPKAVALEENALKWKKEAEEAVADGGSSDRNIQAFVDEVRRTSVEIIT
FGT     GSEAVKEAVERVMVGDG.A.MRSRALYYKEMARK.VEEGGSSYNNLNALIEELSAYVPPM
3GT     AINVVLDSDRGKLL.ENVVKLKGEAMEAVKPHGSCTKEFQELVHLLNGY
PGT1    VKEMMEGDEGKKMRQKAWEWKKKAEAATAVGGQSYNNFDRLVKMVLHQGNWTGNETLH
                490       500       510
        ....|....|....|....|....|....|...
LGT     SSKSKSIHRVKELVEKTATATANDKVELVESRRTEVQY
FGT     KQGLN
```

Figure 18-2. Amino Acid Alignment of 3 Known GTs and PGT1. LGT=Citrus unshiu limonoid UDP-GT (BAA93039); FGT=Scutellaria baicalensis UDP-glucose flavonoid 7-O-GT (BAA83484); 3GT=Ipomoea purpurea UDP-glucose: Flavonoid 3-O-GT (AAB86473). The Plant Secondary Product Glucosyltransferase (PSPG) box of each sequence is shown in **bold**.

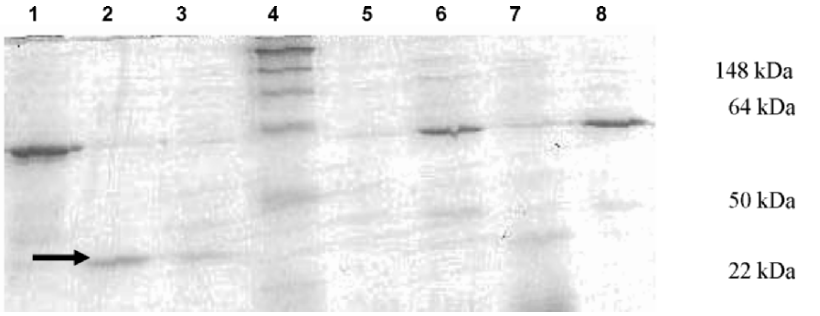


Figure 18-3. Expression of PGT1 Protein. Cells were grown at 27C, collected by centrifugation, and lysed as described. Lane 1=4.5 hr induction lysed cell pellet, Lanes 2-3 = 4.5 hr induction lysed cell soluble protein (2 different concentrations loaded), Lane 4=markers, Lane 5=uninduced culture lysed cell soluble protein; Lane 6=uninduced culture lysed cell pellet; Lane 7=8 hr induction lysed cell soluble protein; Lane 8=8 hr induction lysed cell pellet. Arrow indicates a band corresponding to the predicted size of PGT1.

Soluble PGT1 was tested for flavonoid GT activity using 6 different flavonoid aglycones (Figure 18-4). Previous analyses of flavonoid GTs isolated from young grapefruit leaf tissue (McIntosh and Mansell, 1990; McIntosh et al., 1990) indicated the presence of at least four flavonoid GTs: flavonol 3-O-GT (glucosylated kaempferol and quercetin), flavonol 7-O-GT (glucosylated kaempferol), highly specific flavanone 7-O-GT (glucosylated naringenin and hesperetin), and a GT that glucosylated naringenin chalcone and apigenin. All of these substrates were used to determine whether PGT1 had any GT activity toward flavonoid aglycones. As a positive control, a crude protein extract from young grapefruit leaves was tested using naringenin; the negative control was using proteins from transformed but uninduced, lysed bacterial cells. Aglycone substrate and UDP-glucose levels were in excess and the specific activity of the UDP-glucose was sufficient to be able to detect even low levels of activity.

Results from the crude grapefruit leaf extract showed the presence of naringenin GT activity that was linear for 60 minutes as previously shown (McIntosh et al., 1990). PGT1 did not show activity with any of the flavonoid substrates tested (Figure 18-5). PGT1 sample was mixed with grapefruit leaf extract to determine if any inhibitory compounds were present; this mixture did show GT activity using naringenin as substrate (data not shown).

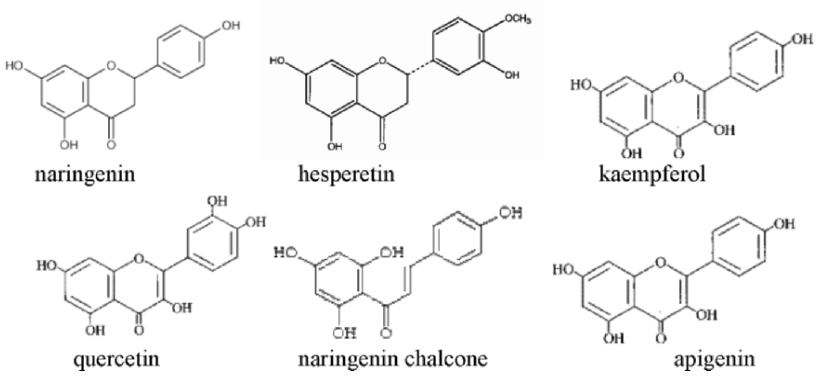


Figure 18-4. Structures of Flavonoids Tested as Potential Substrates for PGT1. Six different compounds were evaluated: naringenin and hesperitin (flavanones), kaempferol and quercetin (flavonols), naringenin chalcone and apigenin (flavone).

4 DISCUSSION

We used information from the amino acid sequence of conserved PSPG boxes of known flavonoid glucosyltransferases to design gene specific primers to cast a wide net to "fish" out potential grapefruit GT sequences by PCR. Using clone-specific primers designed from 5′ partial clones, we subsequently were able to use PCR to amplify and obtain clones corresponding to the 3′ end of the putative GT, PGT1.

A compiled PGT1 sequence was used for analysis of potential reading frame as well as a basis for BLAST and FASTA searches. These searches confirmed that PGT1 had high correlation with plant secondary product GTs already in the databases. Amino acid alignment of PGT1 with the highest matching GTs showed relatively high structural identity (61–65%) within the PSPG box itself, and low identity outside the PSPG box region. This supports the identification of PGT1 as a secondary product GT, although it does not give a good indication of specific function with regard to possible substrates. Currently, amino acid sequence alone is not a good predictor of specific function of plant secondary product GTs. Indications are that enzymes of the same function from different plants show low amino acid identity outside of the PSPG box area. For example, alignment of sequences of flavonoid 3-O-GTs from *Vitis vinifera* (AAB81682), *Ipomoea purpurea* (AAB86473), and *Perilla frutescens* (BAA19659), showed 66–75% amino acid identity within the PSPG box but only 25–31% identity overall (data not shown). Therefore, analysis of activity of heterologously expressed proteins should be performed to determine or verify specific enzyme function.

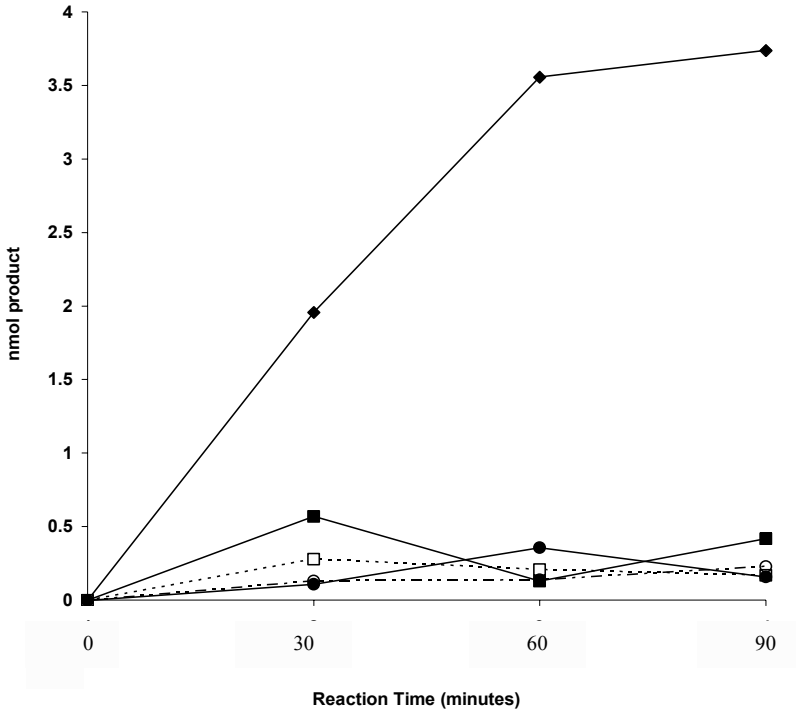


Figure 18-5. Reaction Kinetics with Flavonoid Aglycones and UDP-14C-Glucose. Reactions were run using either crude grapefruit leaf extract, PGT1 soluble protein from cultures induced 4.5 hr at 27C, or control sample from uninduced cultures. ♦ = crude leaf extract, naringenin (control); ■ = PGT1, naringenin; --□-- = uninduced sample, naringenin; ● = PGT1, kaempferol; --○-- = uninduced sample, kaempferol. PGT1 and uninduced samples gave nearly identical results with the other flavonoids (data not shown).

The predicted size of the PGT1 protein, approximately 34 kDa, was much smaller than the 49.5–55 kDa range of sizes for the native flavonoid GTs isolated and characterized from young grapefruit leaves (McIntosh et al., 1990). However, it was still evaluated for possible GT activity towards flavonoid aglycone substrates. As a first screen, we used flavonoids that are naturally produced in grapefruit plants and/or those that were shown to be glucosylated by the GT enzymes previously isolated from grapefruit leaf tissue. PGT1 did not demonstrate activity toward any of the flavanone, flavone, chalcone, or flavonol substrates tested. This supports the idea that PGT1 is not likely to be a flavonoid GT and its biochemical function is unknown at the present time.

We are in the process of completing analyses of clones for 2 additional candidate GTs, obtained in the manner described here, to confirm that we have complete sequence information. Once confirmed, we will use the

compiled sequences to design primers to obtain full-length clones that will be used for subsequent expression and analysis for flavonoid GT activity.

ACKNOWLEDGMENTS

This work was supported by the following sources: ETSU Research Development Council Grant awarded to CAM and LMP, NRI of USDA CSREES grant 2003-35318-13749 awarded to CAM and LMP, a Sigma-Xi Grant-in-Aid of Research awarded to TRS, ETSU W.H. Fraley and N.M. Fraley Memorial Award for Graduate Research awarded to MBS, ETSU Denise Pav Graduate Research Awards given to MBS and CLS. The authors acknowledge B. Winkel for the kind gift of pCD1 vector.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., 1990, Basic local alignment search tool, *J. Molec. Biol.* **215**:403-410.

Durren, R.L. and McIntosh, C.A., 1999, Flavanone-7-O-glucosyltransferase activity from *Petunia hybrida*, *Phytochemistry* **52**:793-798.

Ford, C.M., Boss, P.K. and Hoj, P.B., 1998, Cloning and characterization of *Vitis vinifera* UDP-glucose: flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by maize bronze-1 locus that may primarily serve to glucosylate anthocyanins *in vivo*, *J. Biol. Chem.* **273**:9224-9233.

Hirotani, M., Kuroda, R., Suzuki, H., and Yoshikawa, T., 2000, Cloning and expression of UDP-glucose: flavonoid 7-O-glucosyltransferase from hairy root cultures of *Scutellaria baicalensis*, *Planta* **210**:1006-1013.

Jourdan, P.S., McIntosh, C.A., and Mansell, R.L., 1985, Naringin levels in Citrus tissues: II. Quantitative distribution of naringin in *Citrus paradisi* Macfad, *Plant Physiology* **77**:903-908.

Kita, M., Hirata, T., Moriguchi, T., Endo-Inagaki, T., Matsumoto, R., Hasegawa, S., Suhayda, C.G., and Omura, M., 2000, Molecular cloning and characterization of a novel gene encoding limonoids UDP-glucosyltransferase in Citrus, *FEBS Lett.* **469**:173-178.

McIntosh, C.A. and Mansell, R.L., 1990, Biosynthesis of naringin in *Citrus paradisi*: UDP-glucosyltransferase activity in grapefruit seedlings, *Phytochemistry* **29**:1533-1538.

McIntosh, C.A., Latchinian, L., and Mansell, R.L., 1990, Flavanone-specific O-7-glucosyltransferase activity in *Citrus paradisi* seedlings: Purification and characterization, *Arch. Biochem. Biophys.* **282**:50-57.

McIntosh, C.A. and Mansell, R.L., 1997, Three-dimensional analysis of limonin, limonoate A-ring monolactone, and naringin in the fruit of three varieties of *Citrus paradise*, *J. Agric. Food Chem.* **45**:2876-2883.

Novagen, 2002, pET System Manual, 10th edition, www.novagen.com

Pearson, W.R., 1999, Flexible sequence similarity searching with the FASTA3 program package, in: *Bioinformatic methods and protocols*, Human Press, Totowa, NJ, pp. 185-219.

Pelt, J.L., Downes, W.A., Schoborg, R.V., and McIntosh, C.A., 2003, Flavanone 3-hydroxylase expression in *Citrus paradisi* and *Petunia hybrida* seedlings, *Phytochemistry* **64**:435-444.

Chapter 19

APPLICATION OF METABOLITE AND FLAVOUR VOLATILE PROFILING TO STUDIES OF BIODIVERSITY IN SOLANUM SPECIES

Gary Dobson[1], Tom Shepherd[1], Rhoda Marshall[1], Susan R. Verrall[1], Sean Conner[1], D. Wynne Griffiths[1], James W. McNicol[2], Derek Stewart[1], and Howard V. Davies[1]
[1]*Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, UK;*
[2]*Biomathematics & Statistics Scotland, Dundee Unit, Dundee DD2 5DA, Scotland, UK*

Abstract: Volatile flavour compounds produced when potato tubers are boiled have been related to polar and non-polar metabolites present in raw tubers.

Key Words: boiling; cooking; flavour; gas chromatography; mass spectrometry; metabolite profiling; potato tuber; *Solanum tuberosum*; *S. phureja*; volatiles.

1 INTRODUCTION

Potato is a globally important foodstuff and source of nutrition and has been developed for agronomic traits such as yield and disease resistance. To meet changes in consumer demands, much effort is being put into improving other characteristics such as nutritional value and organoleptic properties. We are using a wide range of potato germplasm to explore phytochemical diversity. The diploid species *Solanum phureja*, developed from the Andean cultivated potato, has a yellower flesh than *S. tuberosum* due to higher carotenoid levels, has distinctive mouth-feel characteristics (high in smooth, and low in grainy and floury traits) and has more intense favourable flavour attributes (creamy, sour, earthy) than *S. tuberosum* (De, Maine et al., 2000). *S. phureja* genotypes are the subject of studies concerning the relationships between tuber metabolites, volatile flavour compounds and taste. The simultaneous analysis of polar metabolites from potato tubers has been carried out by gas chromatography-mass spectrometry (GC-MS) (Roessner

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 259–268.

et al., 2000). In this study, GC-MS has been used to compare the polar and non-polar metabolites and volatile compounds from four *S. phureja* genotypes and four cultivars of *S. tuberosum.*

2 MATERIALS AND METHODS

2.1 Plant material

The plants used in this study were chosen on the basis of taste attributes determined by taste panels. The *S. tuberosum* cultivars Ailsa, Cara, Maris Piper and Pentland Dell represented "bland" cultivars, whilst the *S. phureja* genotypes (DB 257/28, DB 333/16, DB 337/37 and DB 358/23) were selected for their more "distinct" flavour (De, Maine et al., 2000).

Plants were field-grown using normal agronomic practices at a trial site (Mylnefield, Dundee, UK) in 2000. Tubers were harvested two weeks after foliage burn-down, kept at ambient temperature (ca. ~8–12°C) for 4 weeks to allow for skin set, then transferred to a 4°C store. At 4, 10 and 21 weeks post-harvest, *i.e.* 0, 6 and 17 weeks storage, two replicate samples of each genotype were taken from storage and used in the cooking experiments.

2.2 Isolation and analysis of tuber metabolites and volatile flavour compounds

For each replicate experiment, six average-sized tubers were chopped into eighths, two opposite eighths were taken for freeze-drying and the remainder were cooked by boiling. A further sub-sample was taken for freeze-drying after cooking, and the remaining tubers were sampled for volatile compounds. The freeze-dried samples were extracted and analysed for both polar and non-polar tuber metabolites by GC-MS. Volatile compounds produced during cooking were also analysed by GC-MS. Details of the procedures used for sample preparation, extraction, isolation and analysis are given in Chapter 15 (Shepherd et al., 2007).

3 RESULTS AND DISCUSSION

Comparisons were made between the four *S. phureja* genotypes and four cultivars of *S. tuberosum* on the basis of their polar and non-polar metabolite contents and compositions of volatile flavour compounds. The effect of low temperature storage was also studied.

3.1 Polar metabolites

The relative concentrations of 142 polar metabolites, including amino acids, sugars and organic acids were measured in each potato tuber sample. Data from replicate analyses of the four cultivars of *S. tuberosum* and four genotypes of *S. phureja* (raw and cooked) at all three storage dates were analysed by PCA and the two species were clearly separated (Figure 19-1A).

An examination of the specific metabolites responsible for the separation revealed that some amino acids (β-alanine, γ-amino butyric acid and proline) were elevated in *S. tuberosum,* whereas some aromatic (tyrosine and phenylalanine) and branched (*br*-) amino acids (leucine and isoleucine) were elevated in *S. phureja* (Figure 19-2), and the levels of valine, methionine and

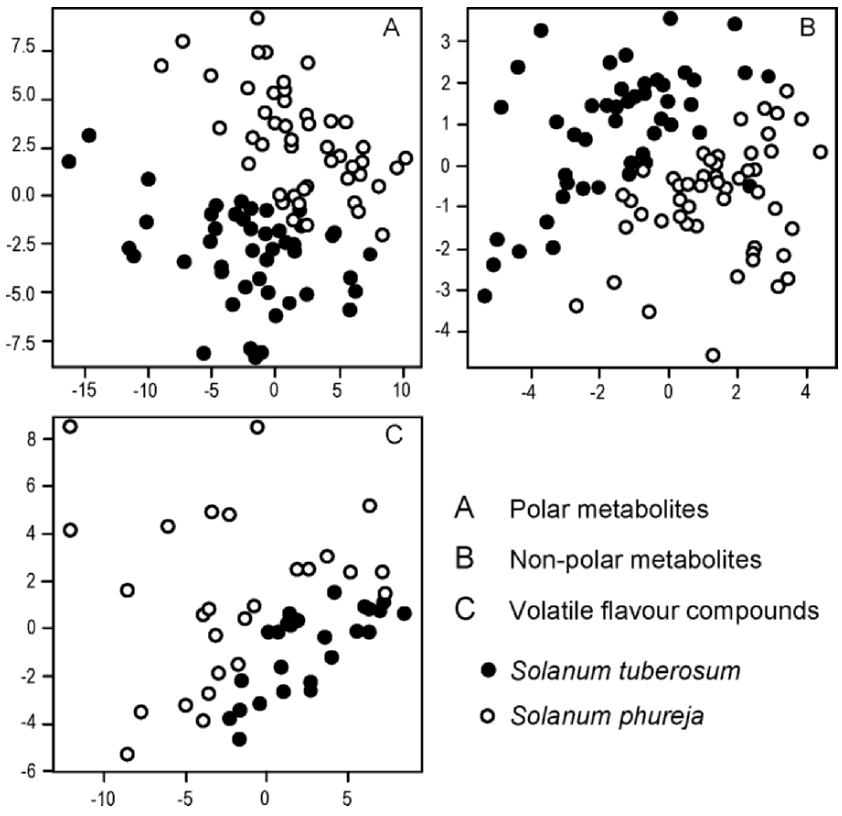


Figure 19-1. PCA plots showing separation of *Solanum tuberosum* and *S. phureja* in terms of (A) polar and (B) non-polar metabolites in cooked and non-cooked tubers and (C) volatiles from cooked tubers. Vertical axis: score (3); horizontal axis: score (1) for (A, C), score (2) for (B).

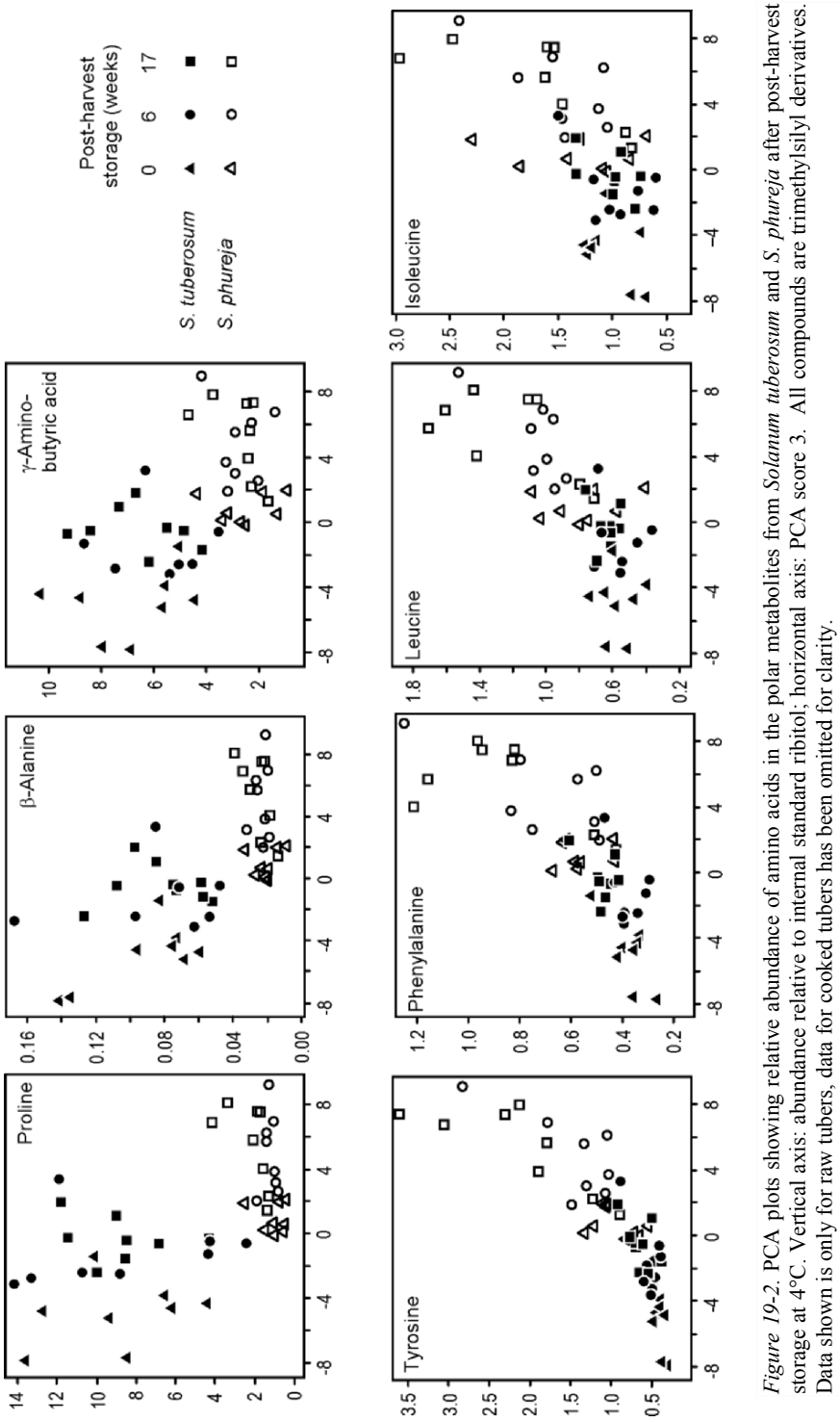


Figure 19-2. PCA plots showing relative abundance of amino acids in the polar metabolites from *Solanum tuberosum* and *S. phureja* after post-harvest storage at 4°C. Vertical axis: abundance relative to internal standard ribitol; horizontal axis: PCA score 3. All compounds are trimethylsilyl derivatives. Data shown is only for raw tubers, data for cooked tubers has been omitted for clarity.

lysine were higher in *S. phureja* line 333/16 only. With the exception of piperidine carboxylic acid (pipecolic acid) and 2,3,4-trihydroxybutyric acid (threonic acid), which were higher in *S. phureja* and *S. tuberosum* respectively, the levels of the other metabolites were similar in both species.

When data for raw tubers alone was considered over all storage periods, all four cultivars of *S. tuberosum* could be separated by PCA. The levels of some amino acids were different between the cultivars; M. Piper was high in γ-amino butyric acid and low in proline, P. Dell was high in β-alanine and Cara was low in lysine. Within *S. phureja*, over all storage periods, all genotypes except 257/28 could be separated from the others. 333/16 was higher in some amino acids (alanine and proline in addition to methionine, valine, and lysine), and citric acid was elevated in 333/16 and 337/37, whereas piperidine carboxylic acid was high in 358/23.

The most striking change with storage was an increase in fructose, glucose, and sucrose. This trend was evident for both species, and PCA of the data for raw tubers clearly separated those samples that were not stored at 4°C from those stored for 6 and 17 weeks, which in turn were not clearly separated. This is not surprising as low temperature sweetening is a well documented phenomenon in potato tubers (Brown et al., 1990). Other changes included increases in the levels of aspartic acid and serine, and a decrease in the level of fumaric acid, after storage.

3.2 Non-polar metabolites

Separation of *S. phureja* and *S. tuberosum* was achieved by PCA of data for the relative concentrations of 35 non-polar metabolites in all samples, both raw and cooked and at all 3 storage dates (Figure 19-1B). Levels of the fatty acids *n*-hexadecanoic acid, 15-methylhexadecanoic acid and *n*-heneicosanoic acid were all elevated in *S. phureja* relative to *S. tuberosum* (Figure 19-3).

On analysis of data for raw tubers, all four cultivars of *S. tuberosum* could be clearly separated by PCA over all storage periods. Among the fatty acids, 15-methylhexadecanoate was low in Cara, *n*-octadecanoate was high in Cara and low in Ailsa, and *n*-hexacosanaote was high in M. Piper. Among the straight chain alcohols, *n*-heptacosanol and *n*-nonacosanol were elevated in M. Piper whereas *n*-heneicosanol was high in Cara and low in Ailsa and M. Piper respectively. The separation between *S. phureja* genotypes was less distinct than with *S. tuberosum*. Over all dates, 337/37 and 358/23, but not 257/28 and 333/16, occurred as discrete groups. *n*-Hexadecanoic acid was particularly high in 358/23, and the order of abundance of 15-methyl hexadecanoic acid was 358/23>337/37>257/28>333/16.

For both *S. phureja* and *S. tuberosum,* separation according to storage date was evident at least for 0 and 17 weeks storage at 4°C. None of the

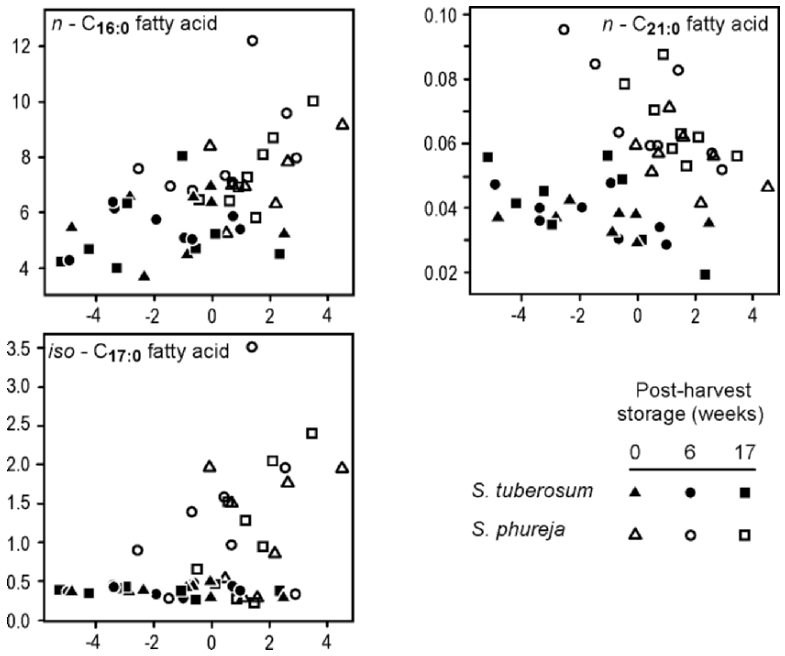


Figure 19-3. PCA plots showing fatty acid levels (as methyl esters) in the non-polar metabolites from *Solanum tuberosum* and *S. phureja* after post-harvest storage at 4°C. Vertical axis: abundance relative to internal standard methyl nonadecanoate; horizontal axis: PCA score 2.

metabolites showed any striking difference in relative abundance between the two storage dates. The levels of *n*-octacosanoic acid tended to decrease with storage in *S. tuberosum*, and in *S. phureja*, *n*-tricosanoic acid tended to increase. There was no evidence for a decrease in linoleic acid and an increase in α–linolenic acid with storage as observed in a previous targeted study on the fatty acids of the same materials (Dobson et al., 2004).

3.3 Volatiles

S. phureja and *S. tuberosum* were separated by PCA of the total ion count area percent compositions of 83 volatiles from cooked tubers (Figure 19-1C). The levels of some *br*-aldehydes (2-methylpropanal, 2-methyl-butanal, 3-methylbutanal), 3-methylthiopropanal, methyl esters of short-chain *br*-acids (2-methylpropanoic acid methyl ester and 2-methylbutanoic acid methyl ester), four sesquiterpenes and several other volatiles (2-methyl-furan, methanethiol, 2-butanone and 2-hydroxybenzoic acid) were higher in *S. phureja*. Hexanal and 2,3-pentadione were elevated in *S. tuberosum*. PCA plots for some of the compounds elevated in *S. phureja* are shown in Figure 19-4.

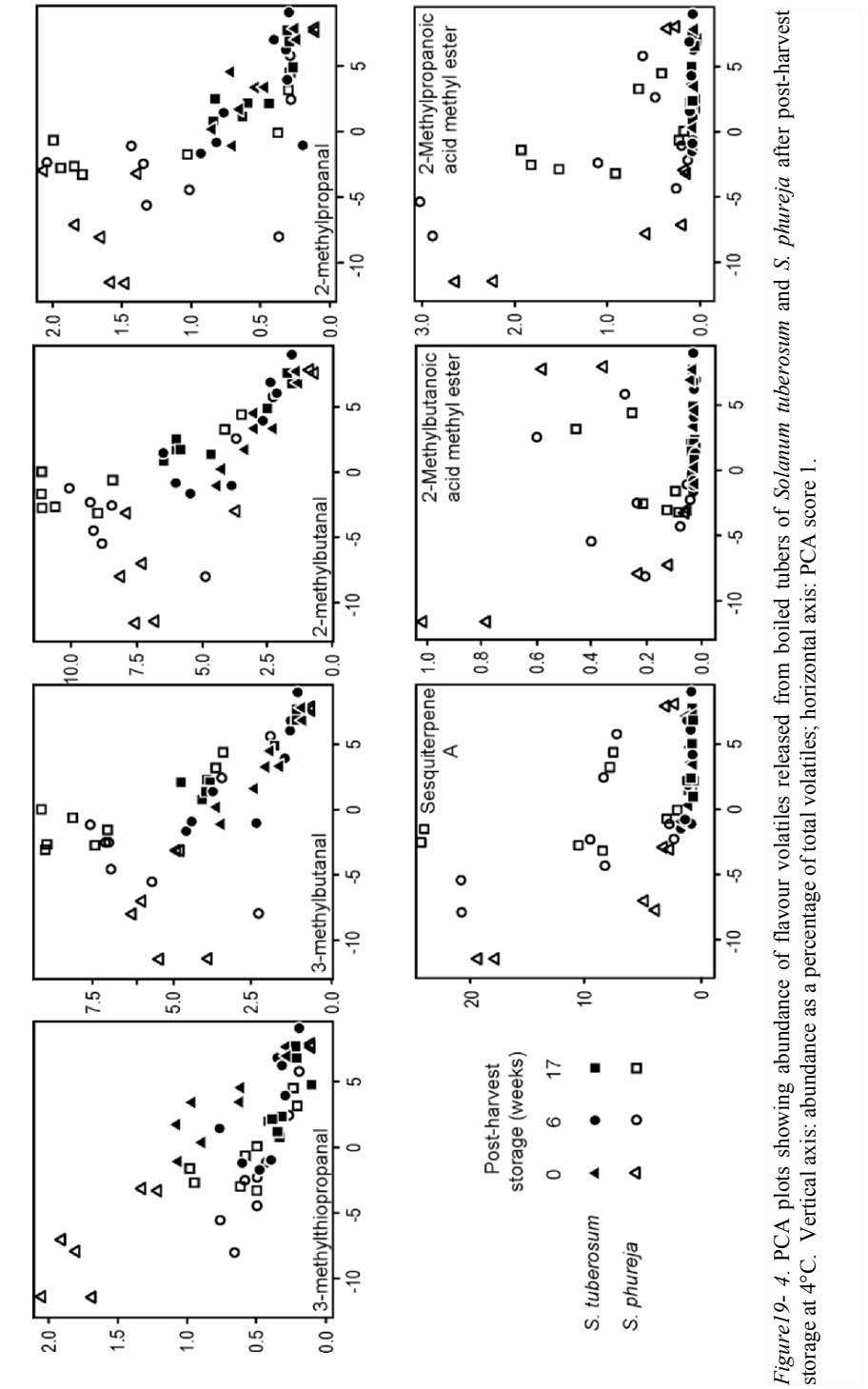


Figure19- 4. PCA plots showing abundance of flavour volatiles released from boiled tubers of *Solanum tuberosum* and *S. phureja* after post-harvest storage at 4°C. Vertical axis: abundance as a percentage of total volatiles; horizontal axis: PCA score 1.

All genotypes of *S. phureja* could be separated by PCA, even when all storage times were considered together. Levels of pentanal, 2-methyl-propanoic acid methyl ester and two of the sesquiterpenes were highest from 333/16. Several aldehydes from 337/37 (hexanal, 5-methylhexanal, and 2,4-heptadienal) were elevated relative to the other genotypes, whereas other aldehydes (2-methylpropanal, 2-methylbutanal, 3-methylbutanal, 2-propenal, 2-methyl-2-butenal, 3-methylthiopropanal, and benzaldehyde) and some other volatiles (2-methylfuran, methanethiol, carbon disulfide, and 2-butanone) were low. 2-Pentylfuran was high from 358/23 and 257/28, whereas two aldehydes (nonanal and decanal) and two alcohols (3-methyl-1-butanol and 2-methyl-1-butanol) were higher in 257/28 and 358/23, respectively.

The unstored cultivars of *S. tuberosum* were all separated by PCA and when all storage periods were considered together, Ailsa and Cara, but not P. Dell and M. Piper, were readily distinguished. The proportions of hexanal, 2-pentenal, and 2,3-pentadione decreased in the order Ailsa > Cara > M. Piper and P Dell, whereas the reverse order was evident for 3-methylbutanal, 2-methylbutanal, 2-pentylfuran, and methanethiol. 3-Methyl-1-butanol and 2-methyl-1-butanol were higher in P. Dell and M. Piper and lower in Ailsa and Cara. Decanal was particularly low in Ailsa whereas a sesqiterpenoid was high in P. Dell.

Considering both species together there was clear separation by PCA according to storage date. Samples stored for 0 and 17 weeks were well separated and those stored for 6 weeks were intermediate in position. The proportions of some alkanes (nonane, decane, and undecane) were higher at 17 weeks, whereas others, of longer chain length (tetradecane, pentadecane, hexadecane, heptadecane, and octadecane), together with some aldehydes (heptanal, undecanal, 2-heptenal, 2-octenal, 2,4-nonadienal, and 2,4-decadienal) were higher at 0-week storage.

4 CONCLUSIONS

It is of interest to catalogue the differences in relative abundances of metabolites between species and among cultivars or genotypes, and changes with duration of storage. Some of these differences, notably the increase in sugars with storage, are well documented, but the majority are not. The real challenge is to identify relationships between different metabolites in terms of changes in metabolic pathways (Figure 19-5). There is a relationship between the elevated abundance of certain *br*- amino acids in tuber of *S. phureja* relative to *S. tuberosum* and similarly elevated levels of short-chain *br*-aldehydes and methyl esters of short chain *br*- acids in the volatile profiles from *S. phureja*. The aldehydes are generated from the amino acids *via* the

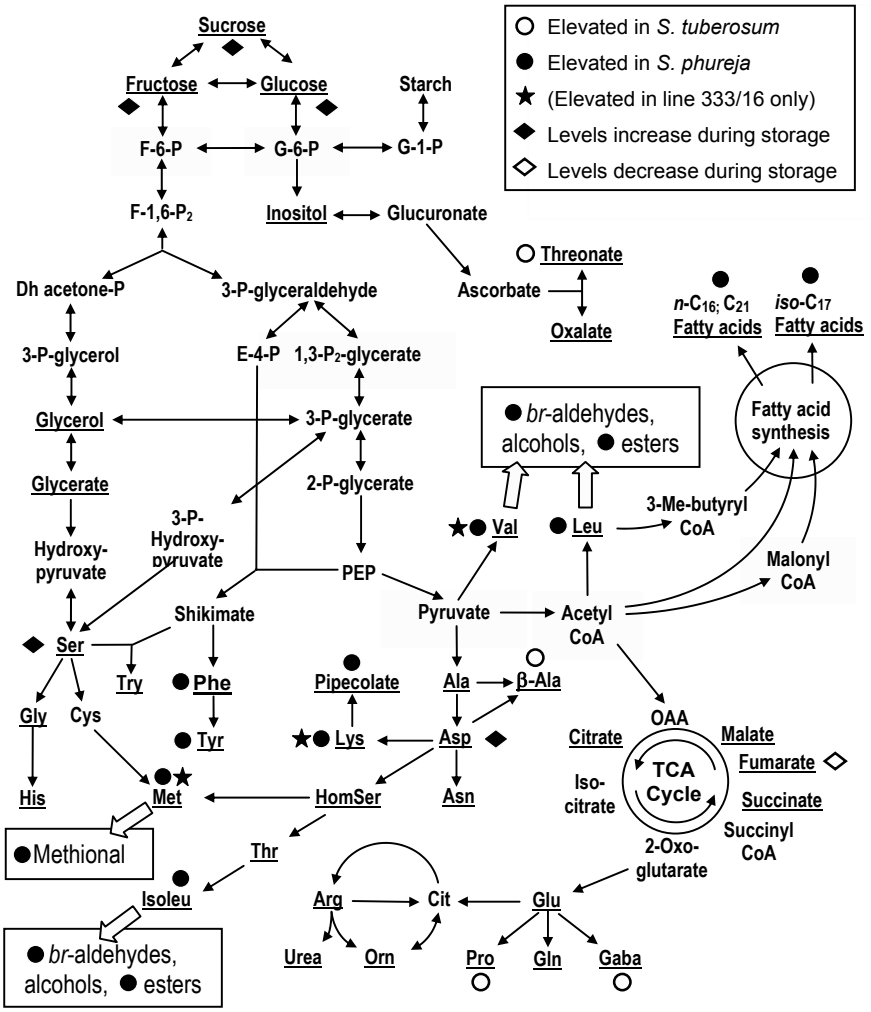


Figure 19-5. Simplified schematic representation of interrelationship between polar and non-polar metabolites isolated from tubers of *Solanum tuberosum* and *S. phureja*. Individual metabolites showing differences in abundance between *S. tuberosum* and *S. phureja* are indicated by open and closed circles. Metabolites which show changes in abundance during storage at low temperature are indicated by closed and open diamonds. Several of the metabolites shown are the source of a number of volatile flavour-related compounds generated when tubers are boiled. These volatiles, shown in the outlined boxes, and their precursor metabolites were relatively more abundant when sampled from tubers of *S. phureja.*

Strecker reaction (Shepherd et al., 2007) and the methyl esters are derived from the equivalent branched acylCoA thioester derivative which in turn is derived from an amino acid. 2-Methylbutanal and 2-methylbutanoic acid

methyl ester, and 2-methylpropanal and 2-methylpropanoic acid methyl ester are derived from isoleucine and valine, respectively, and all these compounds tend to be more abundant in samples from *S. phureja* (valine is elevated in genotype 333/16 only). 3-Methylbutanal and 3-methylbutanoic acid methyl ester are derived from leucine but although the amino acid and aldehyde were elevated from *S. phureja*, the methyl ester was not detected in either species. However, 15-methylhexadecanoic acid, which is derived from 3-methylbutyryl-CoA, the acyl starter unit used during formation of *iso*-branched odd chain fatty acids by the synthesis *de novo*, was more abundant in tubers from *S. phureja*. There were no differences between the species in the abundance of the *br*-alcohols 2-methyl- and 3-methylbutanol which also originate from isoleucine and leucine respectively. For individual genotypes different patterns are evident. For example, the levels of aldehydes in *S. phureja* 337/37 were lower than those of the other *S. phureja* genotypes and were comparable with the levels in *S. tuberosum*.

The increased abundance of *br*-aldehydes and *br*-short-chain esters in the volatile profiles from the four genotypes of *S. phureja* may be significant factors in their favourable flavour assessment by specialist taste panel.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Scottish Executive Environment and Rural Affairs Department.

REFERENCES

Brown, J., Mackay, G. R., Bain, H., Griffiths, D. W., and Allison, M. J., 1990, The processing potential of tubers of the cultivated potato, *Solanum tuberosum* L., after storage at low temperatures. 2. Sugar concentration, *Potato Res*. **33**:219-227.

De, Maine, M. J., Lees, A. K., Muir, D. D., Bradshaw J. E., and Mackay, G. R., 2000, Long-day adapted phureja as a resource for potato breeding and genetic research, In: *Potato, Global Research and Development-Volume1*, S.M.P. Khurana, G.S. Shekhawat, B.P. Singh, and S.K. Pandey, eds., Indian Potato Association, Shimla, India, pp. 134-137.

Dobson, G., Griffiths, D. W., Davies, H. V., and McNicol, J. W., 2004, Comparison of fatty acid and polar lipid contents of tubers from two potato species, *Solanum tuberosum* and *Solanum phureja*, *J. Agric. Food Chem.* **52**:6306-6314.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., and Willmitzer, L., 2000, Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.* **23**: 131-142.

Shepherd, T., Dobson, G., Marshall, R., Verrall, S., Conner, S., Griffiths, D. W., Stewart, D., and Davies, H. V., 2007, Profiling of metabolites and volatile flavour compounds from *Solanum* species using gas chromatography-mass spectrometry, *Proceedings of the Third International Congress on Plant Metabolomics*, Ames, Iowa, June 2004.

Chapter 20

METABOLIC PROFILING HORIZONTAL RESISTANCE IN POTATO LEAVES (CVS. CAESAR AND AC NOVACHIP) AGAINST *PHYTOPHTHORA INFESTANS*

Y. Abu-Nada[1], A.C. Kushalappa[1], W.D. Marshall[2], S.O. Prasher[3], and K. Al-Mughrabi[4]

Departments of [1]Plant Science, [2]Food Sci. and Agric. Chemistry, and [3]Bioresource Eng., McGill University, Ste. Anne de Bellevue, QC, Canada H9X 3V9; [4]New Brunswick Dept. Agric., Wicklow, NB, Canada E7L 3S4

Abstract: Metabolic profiles were developed for potato leaves, cvs. Caesar and AC Novachip, inoculated with water (CW and AW) or *Phytophthora infestans* (CP and AP), using gas chromatography/mass spectrometry (GC/MS). The level of horizontal resistance was higher for cv. Caesar than for AC Novachip with an area under the lesion expansion curve (AULEC) of 334 and 857, lesion area of 86 and 224 mm^2, and sporulation of 4.4×10^3 and 13.7×10^3 sporangia $lesion^{-1}$, respectively. A total of 51 metabolites were detected consistently in all the four replicates of at least one treatment. Of the 51 relatively consistent metabolites, 33 were unique to a treatment and 18 were common to two or more treatments. A total of 7 and 29 PR-metabolites were identified, respectively, in Caesar and AC Novachip cultivars. Most of the phenolic compounds were associated with the AP. The metabolite heptadecanoic acid, 16-methy was detected only in the CP. In response to the *P. infestans* attack, the two cultivars appear to follow two different pathways. The susceptible cv. AC Novachip appears to follow the Shikimic acid-phenylpropanoid pathway as we have detected many phenolic metabolites and benzoic acids, the latter is a precursor of the signal molecule salicylic acid (SA), known to trigger phenolic compounds. On the other hand, the resistant cv. Caesar appears to follow mevalonic acid-methylerythritol pathway as we have detected heptadecanoic acid, a probable derivative of linolenic acid that is a precursor of a signal molecule jasmonic acid (JA), known to trigger terpenes. The factor analysis using principal components discriminated all the four treatments and the factor loading indicated which compound loaded significantly to a treatment. The possible function of these compounds in plant defense against biotic stress is discussed.

Key Words: AC Novachip; Caesar; GC/MS; late blight; plant metabolomics; *Phytophthora infestans*; *Solanum tuberosum*; horizontal resistance.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 269–285.

1 INTRODUCTION

Potato late blight, *Phytophthora infestans* (Mont.) de Bary, is the most important pathogen that attacks potato (*Solanum tuberosum*) (Flier et al., 2003). *P. infestans* is a heterothallic fungal-like organism requiring the A1 and the A2 mating types for sexual reproduction (Stromberg et al., 2001; Peters et al., 1999; Daayf and Platt, 1999). The A2 mating type was found to be more aggressive than the A1 (Fry and Smart, 1999). In Canada, the clonal lineage US-8 is the most aggressive and dominant on cultivated potato cultivars (Medina et al., 1999; Peters et al., 2001; Daayf and Platt 2003).

In the past, breeding programs have mainly considered vertical resistance, generally controlled by single or major R-genes, because transferring genes from wild types to cultivated is easy. However, vertical resistance increases the selection pressure (Keller et al., 2000) and recently, many new races of *P. infestans* have been detected in North America. This has made the breeders to consider the horizontal resistance in the breeding programs, as it is considered to be more durable than the vertical resistance due to the polygenic nature of inheritance (Simmonds and Wastie, 1987; Peters et al., 1999). However, the progress made in transferring horizontal resistance to cultivated potatoes has been very limited because of the difficulty in breeding for polygenic traits (Evers et al., 2003).

In Plant-pathogen interactions, the complete pathway involves the binding of an elicitor, a suppressor or an inducer to a specific receptor, a messenger that carries the signal and an effecter that activate the phenotypic expression. Pathogens produce different enzymes that can hydrolyze plant cell walls and their products act as signal molecules that evoke plant defense responses by the accumulation of phytoalexins (Esquerre-Tugaye et al., 2000), pathogenicity related (PR) proteins (Palva et al., 1993), the enforcement of cell wall by lignification (Robertsen, 1987) and the accumulation of hydroxyproline-rich glycoproteins (HRGP) in the cell wall (Boudart et al., 1995; Huang, 2001). Biochemical defense compounds produced by plants have been grouped into preformed phytoanticipins and induced phytoalexins that are synthesized following infection (Osbourn, 1996). Some of the phytoanticipins commonly detected in plants include phenols, phenolic glycosides, sulfur compounds, saponins, cyanogenic glycosides and glucosinolates. Phytoanticipins are commonly found in the outer cell layers of the plant tissues and they are usually stored in the vacuoles or other organelles in the healthy plants. Following insect feeding damage or a necrotroph invasion, some of these compounds defend the plant against the attacking pathogen. Saponins are glycosylated compounds that belong to triterpenoid, steroid and steroidal glycoalkaloid groups. The steroidal glycoalkaloids are abundantly found in the Solanaceae family that

includes potato. Some fungi such as *Pythium* and *Phytophthora* withstand saponins because of having a low concentration of sterols in their cell membrane. Other phytoalexins isolated from potato include: rishitin, phytuberin, lubimin, solavetivone (Huang, 2001). Glucosinolates are sulfur-containing compounds and are well known as mustard oil glycosides.

Plant metabolites are numerous and they are estimated to be between 90,000 and 200,000 (Fiehn, 2001; Dixon et al., 2002). Primary metabolites are essential for the growth and development of the plant, while the secondary metabolites are not, and most of them are usually associated with the defense response in the plant (Taiz and Zeiger, 2002). Although the genomic, transcriptomic, and proteomic information of plant-pathogen interaction is very useful in developing cultivars with novel traits, they do not always give sufficient information on the end products of plant defense, the metabolites produced in the host-pathogen interactions (Fiehn, 2001; Roessner et al., 2000). Metabolic profiling has been used to identify genetically and environmentally modified traits (Roessner et al., 2001). Roessner et al. (2000) were able to identify more than 150 metabolites from potato tubers using GC/MS technique. Major differences were found in the concentration of the amino acids such as glutamine, proline, and arginine. In vitro microtubers were found to have higher concentrations of the amino acids compared with the soil-grown tubers. Fiehn et al., (2000) used metabolic fingerprinting for the comparison of two homozygous ecotypes and two single gene mutants of *Arabidopsis thaliana*. Distinct metabolic phenotypes were reported for each genotype.

Breeding for quantitative disease resistance is problematic due to lack of tools to evaluate quantitative resistance phenotypes. In segregating populations varying in low to high levels of quantitative resistance though the disease severity on very resistant and very susceptible plant-pathogen interactions were consistent among trials over years, those on the intermediate interactions were quite inconsistent (Haynes et al., 2002). Quantitative resistance in potato against late blight has been measured based on multiple epidemiological disease parameters such as infection efficiency, latent period, lesion size, amount of sporulation, etc. (Carlisle et al., 2002). However, these measurements are time-consuming and expensive for use in breeding programs. Plant breeders are looking for tools to measure phenoltypes varying in quantitative resistance. Metabolic phenotyping of cultivars varying in resistance to disease may be a potential alternative. Accordingly, the main objective of this study was to develop metabolic profiles for potato cultivars with different levels of horizontal resistance against leaf infection by *P. infestans* and relate them to levels of resistance/disease severity.

2 MATERIALS AND METHODS

2.1 Potato plant production

Elite seed potato tubers of cvs. AC Novachip and Caesar were obtained from Bon Accord Elite Potato Center, NB and Global Agri. Services Inc., NB, respectively. AC Novachip is very susceptible and Caesar is moderately resistant to foliar infection by *P. Infestans* (CFIA, 2003). These tubers were stored at 4˚C and 90% RH until use. Seed tubers were sown singly in 16 cm diameter pots containing soil mixture of 1:1 ratio of soil and PRO-Mix BX (Premier Horticulture Ltd, Riviere-du-Loup, QC). Plants were fertilized weekly with 200 mL/pot of a solution (1.5g L^{-1}) of Plant-Prod® 20:20:20 and trace elements (Plant Products Co. Ltd., Ontario, Canada). Three stems per tuber were maintained. Potato plants were grown in the green house (20–25˚C) for 30–40 days to obtain a good foliage growth.

2.2 Inoculum production

P. infestans culture (clonal lineage US-8, A2 mating type, isolate No. 1661, obtained from AAFC, Charlottetown, PEI) was maintained at 4˚C. The pathogen was sub-cultured on rye-agar seed extract media at 15˚C. The sporangia were harvested after 2–3 weeks by flooding with sterile water containing 0.02% Tween 80. The concentration of the sporangia in the suspension was adjusted to 5×10^4 mL^{-1}.

2.3 Inoculation and incubation

Plants grown in greenhouse were transferred to a growth chamber maintained at 20°C, 16 h photoperiod and 90% relative humidity. Three days later, 12 well-formed leaflets were inoculated at either sides of midrib on the undersurface with 5 μL of the sporangial suspension or 0.02% Tween 80 which served as a control. Plants were misted with sterile water, covered with plastic bags to maintain high relative humidity, and transferred back to the growth chamber. The bags were removed 24 h later.

2.4 Disease severity and sporulation assessment

The lesion diameter was measured on 1, 2, 4, 6, 8 days after inoculation (DAI) from which the lesion area and the area under the lesion expansion curve (AULEC) (Shaner and Finney, 1977) were calculated.

Twenty-four hours after inoculation, leaf discs were cut using 15 mm cork borer, placed upper surface down in Petri dish lined with a moist filter

paper and incubated at 20°C and 16 h photoperiod. The plates were completely randomized inside the incubator. Six days after inoculation, the leaf discs were transferred into a test tube containing 5 mL of 0.02% of Tween 80 in water, vortexed, and the number of sporangia was counted using a haemocytometer. Each sample was counted six times from which the average number of sporangia per lesion was derived.

2.5 Metabolite extraction and analysis

Leaf discs containing the inoculated sites were cut at 24 h after inoculation, using a 15 mm cork borer and crushed in liquid nitrogen using a mortar and pestle. The powdered samples were stored in Eppendorf tubes at –80°C until use. The polar metabolites were extracted following a method developed by Roessner et al. (2000) where 100 mg of the powdered plant tissue was extracted with 1.4 mL of 99.93% methanol, for 15 min at 70°C, vigorously mixed with one volume of water and centrifuged at 2,200 *g* for 10 min. 1 mL of the methanol/water supernatant was dried in Speed Vacuum, methoximated in 80 µL of 20 mg of methoxyamine hydrochloride in pyridine for 90 minutes at 30°C, and derivatized in 80 µL of MSTFA at 37°C for 30 minutes. Ribitol (50 µL solution of 2 mg ribitol mL^{-1} water) was added as an internal standard.

2.5.1 GC/MS analysis

100 µL samples of leaf extracts, in tubes with septa, were loaded to an auto sampler connected to GC (model 3400, Varian, Canada) which injected 1 µL sample in to the GC injection port. The GC was equipped with a capillary column (DB-5MS, 30 m and 0.25 mm diameter). The GC injector temperature was healed at 250°C. Helium was used as a carrier gas with a flow rate of 1 mL s^{-1}. The oven temperature was held at 50°C for 3 min and then increased at the rate of 3°C min^{-1} until 200°C, then at 12°C min^{-1} until 250°C and was held at this temperature for 2 min. The compounds were ionized and the mass spectra from 50 to 600 *m/z* were recorded using an ion trap analyzer. Data was analyzed using Saturn Lab Software, and NIST library was used to identify the metabolites. The data for the entire experiment were borrowed in to an EXCEL (Microsoft Corporation) spread sheet, and sorted based on retention time using Pivot Table procedure. The mass spectra of individual peaks/compounds across four blocks of a treatment were manually compared among themselves using the retention time (±0.01 min) as a reference, and with the top ten choices of NIST library search program. The output consisted of a list of compounds and their relative abundance of mass ions (ion trap detector output).

2.5.2 Experimental design

A completely randomized block design was used for the metabolic profiling study using GC/MS. The experiment consisted of two main factors of two cultivars of potato (Cesar and AC Novachip) and two sub-factors of inoculations (pathogen or water). The entire experiment was conducted four times over time. Each experimental unit for metabolic profiling consisted of 12 discs cut from 6 leaflets in 3 stems produced from one tuber. The data on metabolites and their abundance were used in statistical analysis. From the same single-tuber-plants, used for GC/MS analysis, the lesions on another 6 leaflets were used for the assessment of disease severity over an 8-day period. The data on lesion diameter was used to calculate the lesion area and area under the lesion expansion curve. A completely randomized design with two treatments of two cultivars and four replicates was used for the sporulation assessment. Each experimental unit consisted of single-tuber-plants from which 10 leaflet-discs containing inoculation sites were cut. Spore suspensions were prepared 6 DAI and the data on number of sporangia per inoculation site were used in statistical analysis.

2.6 Data processing and statistical analysis

2.6.1 Disease severity and sporulation

The data on average lesion area in mm^2, AULEC, and the number of sporangia per inoculation site, for each experimental unit, were subjected to ANOVA using completely randomized design procedure of SAS.

2.6.2 Metabolic profile

The metabolic profiles consisted of GC/MS output on retention time, names and abundances of compounds. The data for the entire experiment were borrowed in to an EXCEL (Microsoft Corporation) spreadsheet and the frequency of each metabolite occurrence among blocks was determined. The metabolites unique to a treatment (cultivars inoculated with water or pathogen) or metabolites common to two or more treatments were identified. The metabolites novel or increased in abundance following pathogen inoculation were designated here as pathogenesis-related (PR) metabolites.

The compounds that were common to two or more treatments and their average mass ion abundances were normalized by dividing each by the total for all the metabolites considered and designated as metabolic fingerprints. The metabolic fingerprints were subjected to factor analysis (SAS), using principle component method, to assign a compound or combinations of

compounds that significantly loaded to a treatment and to discriminate resistance/treatments based on factor scores.

3 RESULTS

3.1 Disease severity and amount of sporulation

The average lesion areas on the 8 DAI were 224 and 86 mm^2, and the average AULEC were 857 and 334, for AC Novachip and Cesar, respectively. The amount of sporulation was higher in susceptible AC Novachip (13.7×10^3 sporangia mL^{-1}) as compared to the resistant cultivar Caesar (4.4×10^3 sporangia mL^{-1}). An analysis of variance indicated that the treatments, of each of the three experiments, were highly significant at 1% level and the F-values for the blocks, for lesion size and AULEC, were not significant at 1% level.

3.2 Pathogenesis related (PR) metabolites and resistance level discrimination

A total of 875 metabolites were detected in the polar extracts of two potato cultivars, AC Novachip and Caesar, at 24 h after inoculation with *P. infestans* or water (control). Of these metabolites, 401 had mass ion abundances of ≥5000 (Ion Trap detector output), including 51 metabolites that were detected in all the four blocks, in at least one of the treatments (Table 20-1). Out of the 51 metabolites, 33 were unique to a treatment, 12 were common to two or more treatments, and the remaining 6 were common to all the treatments (Caesar pathogen (CP), Caesar water (CW), AC Novachip pathogen (AP) and AC Novachip water (AW) inoculated). Two metabolites 1,4-Dihydro-2-isopropyl-6 and 2-Hydroxy-3-methylanthraquin were detected only in the CW. Another two metabolites 1H,10H-Furo[3',4':4a,5]napht and heptadecanoic acid, 16-methy were specific to CP. Thirteen metabolites were unique to AW and sixteen were unique to AP. A metabolite, 2-Methoxy-4'-nitro-diphenyla was detected in all the four blocks of AW, but occurred only in one block of CW. Two metabolites, 2-Phenyl-3-methoxy-cycloprop and 4H-1-Benzopyran-4-one, 5-hyd were found in both the cultivars when pathogen was inoculated, but the frequency and the abundances of these metabolites were much higher in the AP. In addition, Myo-inositol, 1,2,3,4,5,6-he and Acetic acid, (trimethylsilyl) were found in all the treatments, though their abundances varied. A total of 7 and 29 PR-metabolites were detected in Caesar and AC Novachip, respectively, including 3 that were common to both the cultivars (Table 20-1). These

Table 20-1. Average abundance and frequencies of occurrence of metabolites (and PR-metabolites*) detected in leaves of two potato cultivars Caesar and AC Novachip inoculated with water or *P. infestans*

Metabolites (Chemical groups)[1]	Mass ion abundance (x10^3)			
	CW	**CP**	**AW**	**AP**
Alkaloids	**69.4 [3]**[2]	**5.8 [1]**	**328.2 [9]**	**249.5 [5]**
1,4-Dihydro-2-isopropyl-6	44.9 (4)[3]			
1H-Pyrimido[1,2-a]quinoline-			66.0 (4)	
2,3-Dihydro-1H-2-methylcyclo			36.7 (4)	
Conanin-3-amine, N,N-dimethy			24.3 (4)	
Morphinan-14-ol, 8-azido-6,7			73.4 (4)	
Pyrazolo[5,1-c]-as-triazine-			35.1 (4)	
1,3-Dicyano-2-phenyl-3-amino				25.4 (4)*
1,2-Dihydro-2,4-diphenyl-qui			54.0 (4)	8.9 (2)
1,4-Dihydro-2-isopropyl-6-ph			8.9 (2)	121.4 (4)*
Quinoline, 1,2-dihydro-1-(p-	6.7 (1)		22.4 (4)	62.0 (4)*
Isoxazolo[4,3-a]phenazine, 1	17.9 (2)	5.8 (1)	7.2 (2)	31.8 (4)*
Nitrogen	**12.4 [1]**		**99.1 [2]**	**48.5 [2]**
Methanamine, N-[4-(diphenylm			39.8 (4)	
2,3-Di-O-benzoyl-d,l-glycero				25.4 (4)*
Benzenesulfonamide, 4-(dimet				23.2 (4)*
2-Methoxy-4'-nitro-diphenyla	12.4 (1)		59.3 (4)	
Phenolics	**345.8 [3]**	**219.6 [3]**	**288.1 [4]**	**321.7 [9]**
2-Hydroxy-3-methylanthraquin	51.7 (4)			
Benzoic acid, 2,4-bis(trimethylsiloxy)-,			55.6 (4)	
Noradrenaline tetraTMS			51.3 (4)	
10-Dicyanomethylene-benz(a)a				23.3 (4)*
4H-1-Benzopyran-4-one, 6,7-d				30.7 (4)*
9,10-Anthracenedione, 1-phen				16.4 (4)*
Benzoic acid, 2,4-bis[(trimethylsilyl)oxy]-,				64.3 (4)*
Benzoic acid, 3,5-bis(1,1-di				22.4 (4)*
Luteoline (5,7,3',4'-tetrahy				22.1 (4)*
3-Hydroxy-4-(methylsulphonyl	5.8 (1)			33.2 (4)*
9,10-Anthracenedione, 1,3-di		23.8 (4)*	16.8 (3)	
4H-1-Benzopyran-4-one, 5-hyd		7.8 (1) *		26.8 (4)*
1,8-Dihydroxyanthraquinone d	288.3 (4)	188.0 (4)	164.5 (3)	82.5 (3)
Sulfur			**43.8 [1]**	**271.5 [2]**
Tetramethyl diphosphan-oxide			43.8 (4)	
2-Methyl-2-(4-methoxyphenyl)				212.4 (4)*

Table 20-1 (continued).

exo-2-Cyano-endo-2-methylthi				59.1 (4)*
Others	**2478.4 [5]**	**2144.8 [10]**	**1897.3 [11]**	**2424.6 [14]**
1H,10H-Furo[3',4':4a,5]napht		41.3 (4)*		
Heptadecanoic acid, 16-methy		56.2 (4) *		
2-Ethoxycarbonyl-5-isopropyl			41.0 (4)	
3-(t-Butylacetoxy)-3-methylb			9.3 (4)	
7,11-Dimethyloctadecane			11.3 (4)	
Cyclopenta[c]pyran-1,3-dione			29.5 (4)	
1,3,3-Trimethyl-6-hydroxy-2-				30.5 (4)*
2,3,4,6-Tetramethoxystyrene				30.7 (4)*
2-Propenoic acid, 3-(cyclohe				40.1 (4)*
6-Benzyloxy-2-ethoxy-2,3,4,5				50.0 (4)*
Octadecanoic acid, methyl ester				92.0 (4)*
2-Phenyl-3-methoxy-cycloprop		8.5 (2)*		113.8 (4)*
Cyclopenta[c]pyran-4-carboxy			26.9 (3)	106.8 (4)*
Hexadecanedioic acid diTMS	33.4 (2)	33.3 (1)		65.8 (4)*
3'-Methyl-2-benzylidene-coum		37.6 (3)*	63.9 (4)	61.4 (4)
Tricyclo[3.2.1.02,4]octane,		11.0 (2)*	20.2 (4)	58.5 (4)*
(3R,6R)-(+)-3-Isopropenyl-6-	159.6 (4)	99.7 (2)	97.5 (2)	102.5 (2)*
Acetic acid, [(trimethylsily	1246.7 (4)	1123.4 (4)	1166.4 (4)	1291.5 (4)*
Myo-Inositol, 1,2,3,4,5,6-he	907.3 (4)	633.5 (4)	148.2 (4)	172.7 (4)*
Propanoic acid, 3-(trimethyl	131.4 (4)	69.6 (2)	161.7 (4)	192.8 (4)*

1. Shortened names according to NIST library
2. Total number of metabolites per group for the corresponding treatment.
3. Frequency of metabolites among four blocks; * = Pathogenesis- Related (PR) metabolites (abundances of CP > CW; AP > AW)

PR-metabolites can be used to discriminate resistance between these two cultivars.

3.2.1 Factor analysis to discriminate resistance levels

The metabolic fingerprints based on normalized abundances of 18 metabolites that were common to two or more treatments (excluding 31 metabolites unique to a treatment, out of 51 relatively consistent metabolites, Table 20-1) were subjected to factor analysis to discriminate resistance based on factor loadings. The factor scores were used to discriminate resistance levels/treatments (Figure 20-1). Nine compounds loaded significantly to factor 1 that contributed mainly to AP (metabolites number 1-9, Table 20-2), where five belonged to phenolic and alkaloid groups.

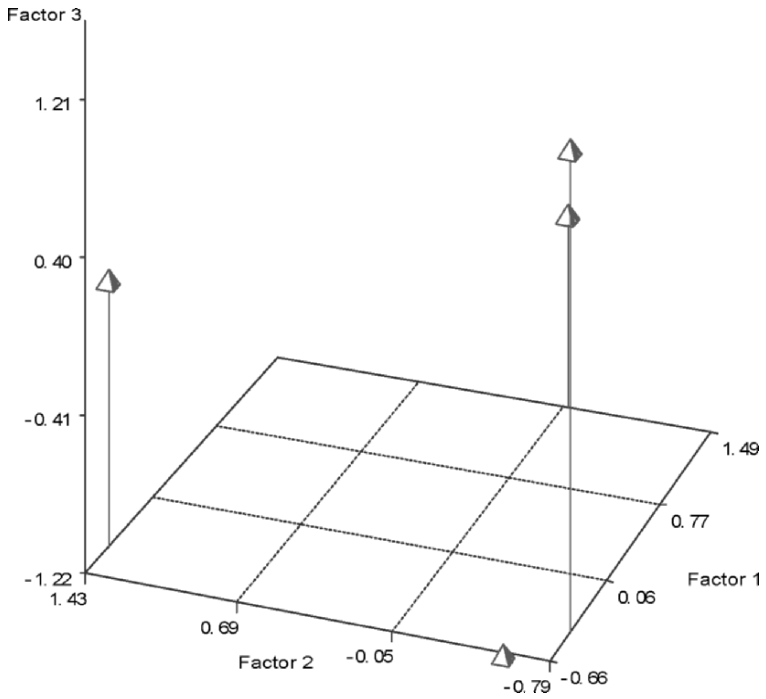


Figure 20-1. Scatter plot (using factor scores) of potato cultivars with contrasting levels of horizontal resistance, pathogen or water inoculated, based on factor analysis of normalized abundance of 18 compounds common to two or more treatments. AW, AP = water or pathogen inoculated AC Novachip and CW, CP = water or pathogen inoculated Caesar, respectively.

Five metabolites (12–16, Table 20-2) loaded significantly to factor 2 which contributed mainly to AW, and one metabolite (18, Table 20-2) loaded significantly to factor 3 which contributed to CP.

3.2.2 Metabolite groups to discriminate resistance levels

The metabolites detected were further classified into different chemical groups based on their function, such as, phenolics, alkaloids, sulfur-containing, nitrogen-containing metabolites, and others (Table 20-1). The most frequent and abundant groups were the phenolics followed by the alkaloids. The phenolics were found in all the treatments with some differences in their abundances. Three phenolic metabolites were detected in both CW and CP; however, the total abundance decreased following pathogen inoculation from 346×10^3 in CW to 220×10^3 in CP. On the

Table 20-2. Eigenvector loadings[a], based on factor analysis using principal component method, for normalized abundances of 18 metabolites that were common to two or more treatments (metabolites unique to treatments not included; details in Table 20-1)

No.	Metabolites	Factor 1	Factor 2	Factor 3
1	2-Phenyl-3-methoxy-cycloprop	0.994[a]	-0.094	-0.039
2	1,4-Dihydro-2-isopropyl-6-ph	0.994	0.041	-0.099
3	3-Hydroxy-4-(methylsulphonyl	0.962	-0.112	-0.245
4	Cyclopenta[c]pyran-4-carboxy	0.960	0.269	-0.075
5	4H-1-Benzopyran-4-one, 5-hyd	0.957	-0.233	0.167
6	Tricyclo[3.2.1.02,4]octane,	0.946	0.295	0.130
7	Quinoline, 1,2-dihydro-1-(p	0.910	0.385	-0.149
8	Isoxazolo[4,3-a]phenazine, 1	0.887	-0.106	-0.447
9	Hexadecanedioic acid diTMS	0.772	-0.635	-0.022
10	(3R,6R)-(+)-3-Isopropenyl-6-	-0.752	-0.059	-0.656
11	1,8-Dihydroxyanthraquinone d	-0.979	-0.116	-0.167
12	1,2-Dihydro-2,4-diphenyl-qui	-0.153	0.982	0.102
13	2-Methoxy-4'-nitro-diphenyla	-0.358	0.933	-0.009
14	Acetic acid, [(trimethylsily	0.116	0.861	0.493
15	Propanoic acid, 3-(trimethyl	0.512	0.819	-0.255
16	3'-Methyl-2-benzylidene-coum	0.394	0.747	0.535
17	Myo-Inositol, 1,2,3,4,5,6-he	-0.639	-0.758	-0.123
18	9,10-Anthracenedione, 1,3-di	-0.470	0.214	0.855

[a] Eigenvector (metabolite) loadings with high positive values indicate significant loading of compounds to each factor. Factor 1 contributed mainly to AC Novachip-pathogen-inoculated, Factor 2 = AC Novachip-water-inoculated; factor 3 = Caesar-pathogen-inoculated (see Figure. 20-1 for factor scores).

contrary, the total number of the phenolic metabolites in the AP increased from 4 in AW to 9 in AP. The total abundances of phenolic compounds increased from 288×10^3 in AW to 322×10^3 in AP. The total abundances of benzoic acid derivatives increased from 56×10^3 in AW to 87×10^3 in AP.

No benzoic acid derivatives were detected in either CP or CW. The total abundance of the phenolic metabolite 1,8-dihydroxyanthraquinone reduced in both the cvs. following pathogen inoculation, from 288×10^3 in CW to 188×10^3 in CP, and from 165×10^3 in AW to 83×10^3 in AP.

The number and abundances of alkaloids reduced following pathogen inoculation in both cultivars, however, both the number and abundances were higher for both AW and AP. Similarly, the abundances of nitrogen-containing compounds reduced in pathogen inoculated treatments. Two sulfur-containing metabolites with total abundances of 44×10^3 in AW increased following pathogen inoculation to 272×10^3. No sulfur-containing metabolite was detected in both treatments of the cv. Caesar.

4 DISCUSSION

In the present study, we were able to discriminate levels of resistance in two potato cultivars inoculated with *P. infestans* based on metabolic profiling of the polar portion of the plant extract using several criteria: (a) unique and combinations of PR-metabolites; (b) chemical groups of metabolites and their abundances; and (c) factor models based on metabolic fingerprints of normalized abundances of metabolites common to two or more treatments. The cv. AC Novachip produced more PR-metabolites, including phenolic compounds than cv. Caesar. The cv. Caesar produced high abundance of heptadecanoic acid, 16-methyl following pathogen inoculation. The chemical groups of metabolites, though not unique, the total abundances of phenolics were higher for AP than for CP. Also, the factor loadings and scores, based on common metabolites, discriminated the resistance levels.

The cv. Caesar was more resistant than cv. AC Novachip for leaf infection by *P. infestans* as confirmed by the lower disease severity (lesion area and AULEC) and lower amount of sporulation. This is in agreement with the assessment of resistance previously reported by CFIA (2003).

Many metabolites detected in our study were identified to be intermediate compounds of metabolic pathways of plant-pathogen interaction. Based on the compounds detected in this study, it appears that the two cultivars follow two different metabolic pathways to defend against *P. infestasns* attack. The cv. AC Novachip appears to follow shikimic acid pathway producing more of phenolics, while the cv. Caesar appears to follow mevalonic-acid pathway producing terpenoids, Heptadecanoic acid, 16-methy'. Majority of the compounds detected in this study belonged to the phenolics. This group contains many metabolites that are known for their antimicrobial activity (Dixon et al., 2002). Phenolic compounds are produced *via* Shikimic acid-phenylpropanoid pathway where the amino acid

phenylalanine is converted to cinnamic acid that in turn produces many metabolites belonging to different subgroups such as coumarins, flavones, flavanones, flavonols, isoflavans, isoflavones, anthocyanidins and many others (Dixon et al., 2002; Huang, 2001; Hopkins and Huner, 2004). The shikimic acid pathway also can increase the production of the aromatic amino acids phenylalanine, tyrosine and tryptophan (Dixon, 2001; Taiz and Zeiger, 2002; Lyon, 2003).

The activity of the phenolic compounds, total numbers and abundances, was higher in the cv. AC Novachip compared with the cv. Caesar. The total abundance of benzoic acid derivatives increased from 56×10^3 in the AW to 87×10^3 in the AP, while these were not detected in Caesar. Benzoic acid was reported to be a precursor of the signal molecule Salicylic acid (Dixon et al., 2002; Metraux, 2002; Nakane et al., 2003; Lyon, 2003). Although the SA was not detected in our study the presence of the benzoic acid derivatives that are precursors of SA may indicate its activity. Following pathogen attack plants produce SA, which in tern stimulated the production of PR-proteins (Durner et al., 1997) and phytoalexins (Nojiri et al., 1996) leading to local acquired resistance (LAR) and systemic acquired resistance (SAR) (Heil and Bostock, 2002). The abundance of the phenolic metabolite 1,8-dihydroxyanthraquinone d was reduced by about 35% in Caesar and 50% in AC Novachip following pathogen inoculation. This metabolite is an anthraquinone and could have been produced from the precursor isochorismic acid (Lyon, 2003). The reduction in anthraquinone production can stimulate the production of salicylic acid with the help of the enzyme pyruvate lyase. This pathway is proposed in *Arabidopsis*, but not yet proved in potato (Lyon, 2003). If this is true, then one could expect that the amount of salicylic acid produced *via* this pathway would be more in the AP than in the CP.

Although the activity of nitrogen-containing compounds and alkaloids were reduced in both cultivars when inoculated with the pathogen, the reduction was more in the cv. Caesar. The primary sources of these compounds are pyruvate, the tricarboxylic acid cycle (TCA) and the Shikimic acid pathway (Taiz and Zeiger, 2002; Nakane et al., 2003). Thus, we could hypothesize that either the activity of pyruvate, the TCA, or the activity of one branch of the shikimic acid pathway that produces aromatic amino acids was reduced.

Other metabolites which have phenol and nitrogen in their structures and might be synthesized *via* shikimic acid-phenylpropanoid pathway by AC Novachip were: (9,10-Anthracenedione, 1,3-di; 9,10-Anthracenedione, 1-phen; 2,3-Di-O-benzoyl-d,l-glycero; Benzenesulfonamide, 4-(dimet; Luteoline (5,7,3',4'-tetrahy; 10-Dicyanomethylene-benz(a)a; 4H-1-Benzopyran-4-one, 6,7-d). Factor analysis (Table 20-2) indicated significant loading of nine metabolites to factor 1 and these were significantly correlated with the

pathogen-inoculated AC Novachip. Five of these metabolites belong to the phenolics and alkaloids.

In pathogen inoculated Caesar Heptadecanoic acid, 16-methy was detected which is a fatty acid that could be a derivative of heptadecatrienoic acid (C17: 3) which is produced from linolenic acid (C18: 3) through the activity of α-dioxygenase, a pathogen inducible oxygenase (Lyon, 2003). In response to pathogen attack, it has been reported that in the mevalonic acid pathway lipoxygenase and other enzymes activate the conversion of linolenic acid to jasmonic acid (JA) and other derivatives (Liechti and Farmer, 2002). Although the plants can produce JA following insect attack or the infection by a pathogen, the end products of these interactions are different, and these affect the products of the local and systemic acquired resistance in the plant (Fidantsef et al., 1999). In Solanaceae, the jasmonates and oxylipins (linolenic acid derivative) elicit proteinase inhibitor accumulation and steroid glycoalkaloid synthesis responses, i.e., similar to mechanical wounds or chewing insect attack (Casey, 1995; Choi et al., 1994). On the contrary, fungal elicitors such as arachidonic acid (AA) and lipoxygenase metabolites stimulate the accumulation of PR-proteins and sesquiterpene phytoalexins, leading to programmed cell death (Bostock et al., 1986). Because high abundance of 'Heptadecanoic acid, 16-methy' was detected in cv. Caesar, it could be hypothesized that the mevalonic-acid pathway was active, which is signaled by JA. However, no JA and little lipophilic metabolites were detected in our study, but in reality, there may have been more as in the present investigation only polar phase of the plant extracts (methanol-water solvents), which excluded most lipophilic compounds including many terpenes, was analyzed.

P. infestans is a hemibiotroph and during the biotrophic phase, the fungus shows minimal secretary activates in order not to trigger the plant defense responses (Mendgen and Hahn, 2002). The pathogen elicits the hypersensitive response (HR) when infecting the nonhost plants, the partially resistant and the highly resistant plants. In the completely resistant *Solanum* species and nonhosts the HR was very fast, killing 1-3 infected plant cells. However, in the partially resistant plants, the HR was slow, gradually killing five or more plant cells. The major R-genes such as R1 were found to produce a strong HR response while the weak R-genes, i.e., R10 was found to produce a weak and late HR responses and the pathogen hyphae were able to grow beyond the HR lesion to start a new infection (Vleeshouwers et al., 2000). In our study, both cultivars are believed to have minor genes that are responsible for the horizontal resistance. They were found to have different levels of disease severity according to lesion size, sporulation rates and the AULEC, but there is a possibility that the cultivar Caesar could contain one of weak R-genes that mimic the horizontal resistance. Further studies are needed in this area.

In conclusion, the cultivars tested in this study, AC Novachip and Caesar seem to have evolved developing two separate mechanisms to suppress the invasion by *P. infestans*. The activity of phenolic metabolites increased in AC Novachip, in both numbers and abundances, which are known to be produced through the shikimic acid pathway. On the other hand, heptadecanoic acid, a potential precursor for the signal molecules jasmonic acid, was only detected in the inoculated Caesar. JA has been associated with the activation of many terpenes. An analysis of the lipophilic potion of the plant extract would shed more lights on the compounds that are involved in this pathosystem.

ACKNOWLEDGEMENTS

The research was financed by the CORPAQ, Quebec. We thank Dr. H. W. Platt, AAFC, PEI for providing a culture of *Phytophthora infestans* and Mr. H. Hamzehzarghani for assistance in statistical analysis.

REFERENCES

Bostock, R.M., Schaeffer, D.A., and Hammerschmidt, R., 1986, Comparison of elicitor activities of arachidonic acid, fatty acids and glucans from *Phytophthora infestans* in hypersensitive expression in potato tuber, *Physiological and Molecular Plant Pathology* **29**:349-360.

Boudart, G., Dechamp-Guillaume, G., Lafitte, C., Ricart, G., Barthe, J.P., Mazau, D., and Esquerre-Tugaye, M.T., 1995, Elicitors and suppressors of hydroxyproline-rich glycol-protein accumulation are solubilized from plant cell wall by endopolygalacturonase, *Eur. J. Biochem.* **232**:449-457.

Carlisle, D.J., Cooke, L.R., Wiatson, S., and Brown, A.E., 2002, Foliar aggressiveness of Northern Ireland isolates of *Phytophthora infestans* on detached leaflets of three potato cultivars, *Plant Pathology* **51**:424-434.

Casey, R., 1995, Sequence of cDNA clone encoding a potato (*Solanum tuberosum*) tuber lipoxygenase, *Plant Physiology* **107**: 265-266.

CFIA. 2003. Canadian Potato Varieties Descriptions. Available:http://www.inspection.gc.ca/english/plaveg/potpom/var/indexe.shtml 13/8/2003.

Choi, D., Bostock, R.M., Avdiushko, S., and Hildebrand, D.F., 1994, Lipid-derived signal that discriminate wound- and pathogen-responsive isoprenoids pathways in plants: methyl jasmonate and the fungal elicitor arachidonic acid induce different 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes and antimicrobial isoprenoids in Solanum tuberosum L, *Proceedings of the National Academy of Science USA* **91**:2329-2333.

Daayf, F., and Platt, H.W. (Bud), 2003, Differential pathogenicity on potato and tomato of *Phytophthora infestans* US-8 and US-11 strains isolated from potato and tomato, *Can. J. Pathol.* **25**:150-154.

Daayf, F., and Platt, H.W., 1999, Assessment of mating types and resistance to metalaxyl of Canadian populations of *Phytophthora infestans* in 1997, *American Journal of Potato Research* **76**:287-295.

Dixon, R.A., 2001, Natural products and plant resistance, *Nature* **411**: 843-847.

Dixon, R.A., Achnine, L., Kota, P., Liu, C.J., Reddy, M.S.S., and Wang, L., 2002, The phenylpropanoid pathway and plant defense-a genomic perspective, *Molecular Plant Pathology* **3**(5): 371-390.

Durner, J., Shah, J., and Klessing, D.F., 1997, Salicylic acid and disease resistance in plants, *Trends in Plant Science* **2**:266-274.

Esquerre-Tugaye, M.T., Boudart, G., and Dumas, B., 2000, Cell wall degrading enzymes, inhibitory enzymes, and oligosaccharides participate in the molecular dialogue between plants and pathogens, *Plant Physiol. Biochem.* **38**:157-163.

Evers, D., Ghislain, M., Hausman, J.F., and Dommes, J., 2003, Differential gene expression in two potato lines differing in their resistance to *Phytophthora infestans*, *J. Plant Physiol*. **160**: 709-712.

Fidantsef, A.L., Stout, M.J., Thaler, J.S., Duffy, S.S., and Bostock, R.M., 1999, Signals interaction in pathogens and insect attack: expression of lipoxygenase, proteinase inhibitor II, and pathogenesis-related protein P4 in the tomato *Lycopersicon esculentum*, *Physiological and Molecular Plant Pathology* **54**:97-114.

Fiehn, O., 2001, Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks, *Camp. Funct. Genom.* **2**: 155-168.

Fiehn, O., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry, *Anal. Chem.* **72**: 3573-3580.

Flier, W.G., van de Bosch, G.B.M., and Turkensteen, L.J., 2003, Stability of partial resistance in potato cultivars exposed to aggressive strains of Phytophthora infestans, *Plant Pathology* **52**:326-337.

Fry, W.E., and Smart, C.D., 1999, The return of *Phytophthora infestans*, a potato pathogen that just won't quit, *Potato Research* **42**: 279-282.

Haynes, K.G, Christ, B.J., Weingartner, D.P., Douches, D.S., Thill, J.C.A., Secor, G., Fry, W.E., and Lambert, D.H., 2002, Resistance to late blight in potato clones evaluated in national trials in 1997, *Amer. J. Potato Res.* **79**:451-457.

Heil, M., and Bostock, R.M., 2002, Induced systemic resistance (ISR) against pathogens in the context of induced plant defense, *Annals of Botany* **89**: 503-512.

Hopkins, W.G., and Huner, P.A., 2004, *Introduction to Plant Physiology*, John Wiley and Sons, Inc., 3rd ed., New Jersey, pp. 560.

Huang, J.S., 2001, *Plant Pathogenesis and Resistance, Biochemistry and Physiology of Plant-microbe Interactions*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.691.

Keller, B., Feuillet, C., and Messmer, M., 2000, Genetics of Disease Resistance, in: *Mechanisms of Resistance to Plant Diseases*. A.J. Slusarenko, R.S.S. Fraser, and L.C. van Loon, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 101-160.

Liechti, R., and Farmer, E.E., 2002, The jasmonate pathway, *Science* **296**: 1694-1650.

Lyon, G.D., 2003, Disease resistance-related proteins in potato, Scottish Crop Research Institute. Available: http://www.scri.sari.ac.uk/TiPP/PPS/protein.htm.

Medina, M.V., Platt, H.W., and Peters, R,D., 1999, Severity of late blight tuber infection caused by US-1 and US-8 genotypes of *Phytophthora infestans* in 12 potato cultivars, *Can. J. Plant Pathol.* **21**:388-390.

Mendgen, K.W., and Hahn, M., 2002, Plant infection and the establishment of fungal biotrophy, *Trends in Plant Science* **7**: 352-356.

Mertraux, J.P., 2002, Recent breakthroughs in the study of salicylic acid biosynthesis, *Trends in Plant Science* **7**: 332-334.

Nakane, E., Kawakita, K., and Doke, N., 2003, Elicitation of primary and secondary metabolism during defense in the potato, *J. Gen. Plant Pathology* **69**: 378-384.

Nojiri, H., Sugimori, M., Yamane, H., Nishimura, Y., Yamada, A., Shibuya, N., Kodama, O., Murofushi, N.. and Omori, T., 1996, Involvement of jasmonic acid in elicitor-induced phytoalexins production in suspension-cultured rice cells, *Plant Physiol*. **110**:387-392.

Osbourn, A.E., 1996, Preformed antimicrobial compounds and plant defense against fungal attack, *The Plant Cell* **8**:1821-1831.

Palva, T.K., Holmstrom, K.O., Heino, P., and Palva, E.T., 1993, Induction of plant defense response by exoenzymes of *Erwinia carotovora* subsp. *carotovora*. *Mol. Plant-Microbe Interact.* **6**:190-196.

Peters, R.D., Forster, H., Platt, H.W., Hall, R., and Coffey, M.D., 2001, Novel genotypes of *Phytophthora infestans* in Canada during 1994 and 1995, *Am. J. Potato Res*. **78**:39-45.

Peters, R.D., Platt, H.W., Hall, R., and Medina, M., 1999, Variation in aggressiveness of Canadian isolates of *Phytophthora infestans* as indicated by their relative abilities to cause potato tuber rot, *Plant Dis.* **83**:652-661.

Robertsen, B., 1987, Endo-polygalacturonase from *cladosporium cucumerinum* elicits lignification in cucumber hypocotyls, *Physiol. Mol. Plant Pathol.* **31**:361-374.

Roessner, U., Willmitzer, L., and Fernie, A.R., 2001, High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies, *Plant Physiology* **127**:749-764.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L., 2000, Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.* **23**: 131–142.

Shaner, G., and Finney, R.A., 1977, The effect of nitrogen fertilization on the expression of slow mildewing resistance in Knox wheat, *Phytopathology* **67**: 1051-1065.

Simmonds, N.W., and Wastie, R.L., 1987, Assessment of horizontal resistance to late blight of potatoes, *Annals of Applied Biology* **111**:213-221.

Stromberg, A., Bostrom, U., and Hallenberg, N., 2001, Oospore germination and formation by the late blight pathogen *Phytophthora infestans in vitro* and under field conditions, *J. Pathology* **149**:659-664.

Taiz, L., and Zeiger, E., 2002, *Plant Physiology*, Sinauer Association, Inc., 3rd ed., Sunderland, USA, pp. 691.

Vleeshouwers, V.G.A.A., van Dooijeweert, W., Govers, F., and Kamoun, S., 2000, The Hypersensitive Reaction Response is Associated with Host and Nonhost Resistance to *Phytophthora Infestans*, *Planta* **210**:853-864.

Chapter 21

IN VIVO ^{15}N-ENRICHMENT OF METABOLITES IN *ARABIDOPSIS* CULTURED CELL T87 AND ITS APPLICATION TO METABOLOMICS

Kazuo Harada, Ei-ichiro Fukusaki, Takeshi Bamba, and Akio Kobayashi
Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan

Abstract: A mass spectrometer is one of the best analytical tools for metabolomics. However, its quantitative accuracy can be compromised due to 'ion suppression' caused by insufficient separation during chromatography. Here we present a practical solution for quantitative analysis by means of stable isotope dilution, including *in vivo* ^{15}N-labeling. We employed *Arabidopsis thaliana* cultured cell T87 as a model plant cell. ^{15}N-enrichment was readily performed by cultivation with modified LS-media containing ^{15}N-labeled inorganic nitrogen sources, $K^{15}NO_3$ and $^{15}NH_4{}^{15}NO_3$. No significant morphological change in T87 cells was observed with ^{15}N-enrichment. A mixture of the extracts of ^{15}N-cultured cells and ^{14}N-cultured cells was subjected to capillary LC/MS analysis. Sufficient linearity was obtained in the relative quantification system. In addition, time-course sampling revealed an apparent turnover rate of metabolites including nitrogen atoms. The time course was started from the zero time at which culture media were changed from ^{14}N-media to ^{15}N-media. Interesting variations in nitrogen turnover rate among the metabolites was observed. This ^{15}N *in vivo* labeling system should become a powerful tool for both metabolomics and flux analysis.

1 INTRODUCTION

Metabolomic analytical procedures that have been developed to-date have the following problems: (i) Mass spectrometry is generally used for detection in metabolic profiling, e.g., GC/MS, LC/MS, CE/MS. However, mass spectrometric measurements are rarely quantitative due to fluctuations in the ionization efficiency of analytes (King et al., 2000; Mëller et al., 2002). This fluctuation is caused by the presence of material other than the target compound during ionization. Therefore, researchers have not yet obtained accurate quantitative data during metabolic profiling. (ii) Metabolic

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 287–297.

profiling procedures measure accumulated amounts of metabolites. This means information about metabolic flux is lacking in the profiling data (Matsuda et al., 2003). Therefore, it is difficult to discuss the dynamics of metabolism using profiling data.

It should be possible to utilize isotopes to overcome these problems. For (i), the stable isotope dilution method, which uses stable isotope labeled analogs (isotopomers) of analytes as internal standards, compensates for fluctuations in ionization efficiency (Dube et al., 2001). This method can also correct variations due to experimental procedures (i.e., extraction, preparation, injection into the analytical instrument) other than ionization in the mass spectrometer. Therefore, this allows the accurate quantification of metabolite levels. For (ii), the isotopic distribution of metabolites following incorporation of isotopes into samples over certain periods reveals apparent turnovers of metabolites. This helps us to interpret the dynamics of metabolism.

Moreover, plants can use inorganic ions, nitrate and ammonium, as their sole nitrogen source. This means that ^{15}N-labeling of these inorganic ions in culture media will cause isotopic labeling of every nitrogen-containing compound. Here, we present a novel analytical method for plant metabolomics using *in vivo* ^{15}N labeling in order to overcome above-mentioned problems.

2 ISOTOPE DILUTION METHOD USING *IN VIVO* ^{15}N-LABELING

2.1 Background

The stable isotope dilution method has been widely used to accomplish accurate quantification in mass spectrometry (Dube et al., 2001; Matuszewski et al., 2003). The method compensates for fluctuations caused by ion suppression or matrix effects, because the ionization efficiencies of isotopomers are identical under all conditions (i.e., including compounds other than analytes).

When applying the stable isotope dilution method to metabolic profiling, it is important to consider how isotopomers corresponding to the vast array of metabolites are to be prepared. The utility of commercial isotopomers is limited because isotopic compounds corresponding to most metabolites are expensive or unavailable, and organic or enzymatic syntheses are very laborious and time-consuming to prepare all the isotopic analogs of multitarget metabolites.

Thus two strategies are considered to be practical for stable isotope dilution. One is postharvest derivatization of sample mixtures using certain

reagents and their isotopomers. In the proteomics field, one of the most useful commercially available reagents for stable isotope dilution is the isotope coded affinity tag (ICAT) (Gygi et al., 1999). Although such methods have also been reported in the metabolomics field (Fukusaki et al., 2005), no practical methods have been developed because three problems remain: (i) procedures enabling exhaustive derivatization are limited; (ii) a loss of analytes caused by derivatization occurs; and (iii) the methods are not able to compensate for the difference in derivatization efficiencies using certain reagents and their isotopomers.

The other strategy is *in vivo* labeling method, in which samples uptake isotopes from culture media (Figure 21-1) (Ong et al., 2002; Steen et al., 2002). This method avoids the problems associated with postharvest derivatization. Thus *in vivo* labeling method should be appropriate for exhaustive metabolic profiling. Accordingly we verified the utility of *in vivo* labeling method by the following experiments.

2.2 Materials and experiments

As sample material, we selected *Arabidopsis thaliana* cultured cell T87 (Axelos et al., 1992). The ^{15}N-labeled cells were grown in modified Linsmaier and Skoog medium containing $K^{15}NO_3$ and $^{15}NH_4{}^{15}NO_3$ (^{15}N-LS medium) instead of $K^{14}NO_3$ and $^{14}NH_4{}^{14}NO_3$ under continuous light at 23ºC. Every 7 days, mother cell suspensions were transferred into new ^{15}N-LS medium.

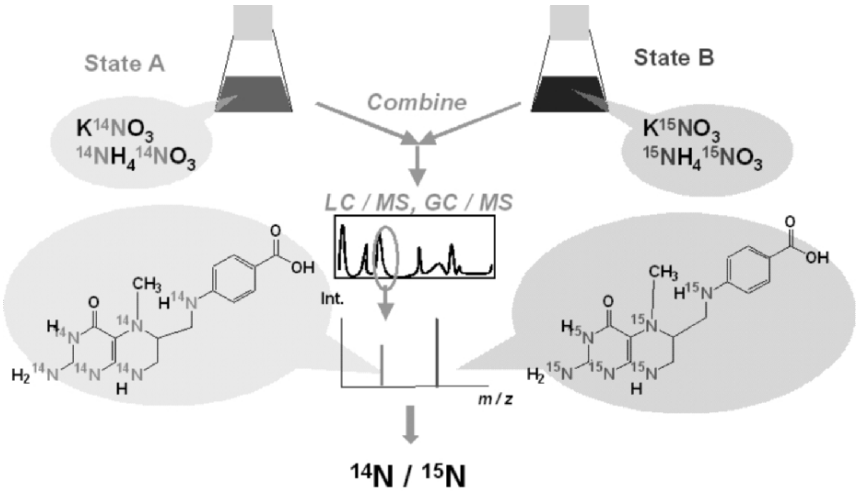


Figure 21-1. Concept of isotope dilution method using *in vivo* ^{15}N-labeling.

Folate determination was performed as follows. Unlabeled (^{14}N) and labeled (^{15}N) cells were harvested by filtration and immediately ground in liquid nitrogen. These were combined and immersed in an extraction buffer of 25 mM ammonium acetate containing sodium 2% ascorbate and 0.02 M 2-mercaptoethanol (pH 7.3). The sample suspensions were placed in a boiling water bath for 10 min. Subsequently, the extracts were rapidly cooled and centrifuged. Recovered extracts were purified by affinity chromatography (Konings, 1999). The eluents were incubated with carboxypeptidase for 4 h at 30ºC. The samples were freeze-dried, dissolved in 200 μL of CH_3COONH_4 (pH 7.3, 15 mM), filtered through a 0.45 μm filter and subjected to capillary LC/MS.

The instrument for capillary LC was a Famos-Switchos II- Ultimate (LC Packings, Amsterdam, The Netherlands), mass spectrometry was by an Esquire 3000 plus (Bruler Daltonics, Billerica, MA, USA) using electrospray ionization (ESI). An ODS capillary monolithic column (0.2 mmi.d.×750 mm, Kyoto Monotech Corp. Kyoto, Japan) was used as the analytical column. The solvent for LC/MS analysis was H_2O-acetonitrile-formic acid, at a flow rate of 3.2 μL/min.

2.3 Results and discussion

First, we confirmed whether or not all nitrogen in plant cells could be substituted by the stable isotope ^{15}N. Consequently, complete incorporation of ^{15}N in amino acids, folates, *S*-adenosylmethionine, and *S*-adenosylhomosysteine occurred after culturing for 21 days. Although the heavy ^{15}N-isotope should lead to a kinetic isotope effect, the weight, morphology and rate of growth of the ^{15}N-labeled cells were indistinguishable from those of the reference cells. This shows complete ^{15}N-enrichment of T87 cells is not detrimental to plant cells.

Next we study the accuracy of stable isotope dilution using *in vivo* labeling. Folates were chosen as targets, because they are difficult to analyze due to their low amounts and instability.

5-Methyltetrahydropteroate (5-CH_3-H_4pteroate) that is related to folate should be detected as [M+H, at *m/z* 331]$^+$ in the mass spectrometer using ESI. Because this compound contains 6 nitrogen atoms,^{15}N-labeled 5-CH_3-H_4 pteroate should be detected as [M+H, at *m/z* 337]$^+$. When the sample obtained by combining unlabeled and labeled cell equally was subjected to capillary LC/MS (Figure 21-2), both peaks in mass chromatogram of *m/z* 331 and 337 were detected with almost the same peak areas.

To validate the isotope dilution method using *in vivo* labeling as a quantitative method, a mixing experiment was performed using known

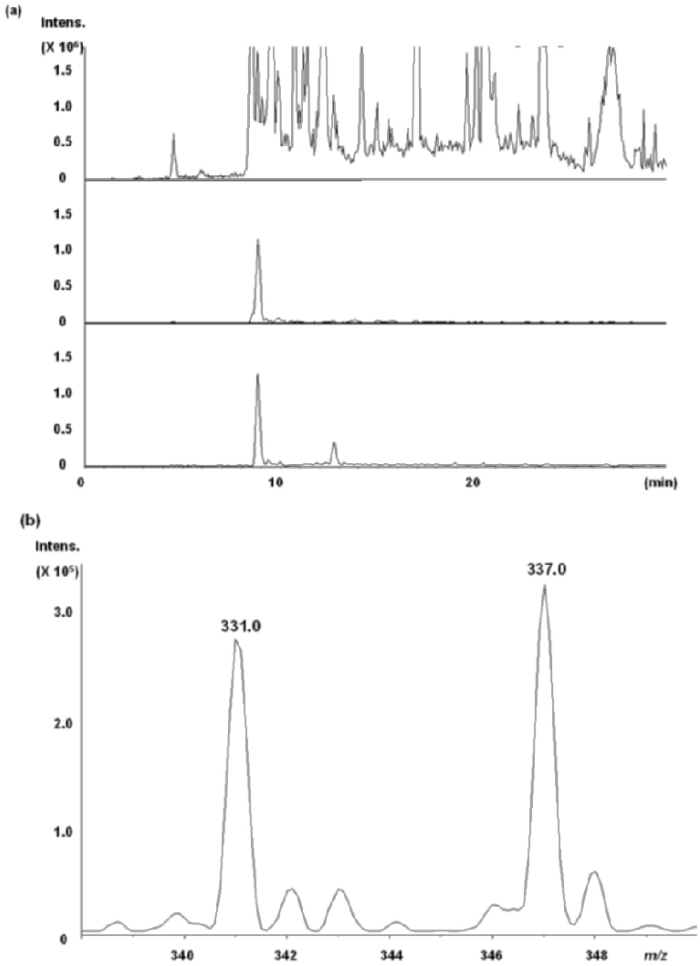


Figure 21-2. Capillary- LC /MS analysis of mixed extract from cultured cell in ^{14}N and ^{15}N medium **(a)** Base peak chromatogram (Upper), mass chromatogram of 331 *m/z* (Middle) and 337 *m/z* (Lower) **(b)** Mass spectrum of normal (331.0 *m/z*) and all nitrogen-^{15}N-labeled (337.0 *m/z*) 5-CH_3-H_4folate.

weights of T87 cells (Fig. 21-3). Unlabeled and labeled T87 cells were mixed in various ratios. The experimentally determined peak area ratios were found to be linear ($r^2 = 0.998$) over an abundance ratio from 1 to 3 (^{15}N cell / ^{14}N cell). This result should guarantee the linearity of the *in vivo* labeling method.

We applied the isotope dilution method to an analysis of metabolic responses against a metabolic inhibitor. Methotrexate is widely used as an anticancer, anti-inflammatory and immunosuppressive agent (Schweitzer

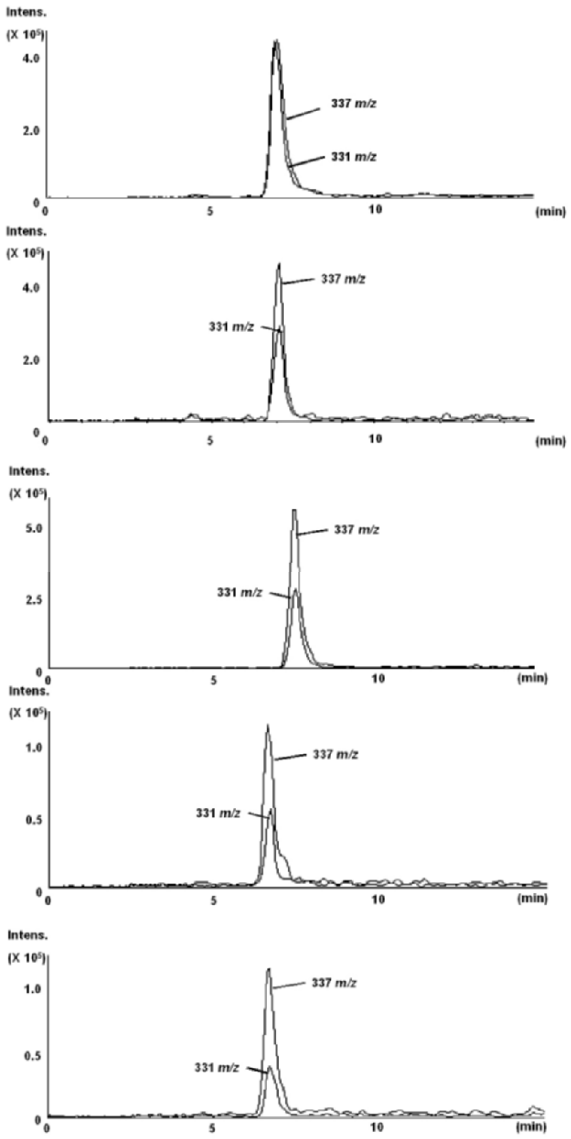


Figure 21-3. Mass chromatogram of 331 m/z and 337 *m/z* caused by capillary – LC /MS analysis. The mixing ratios (g-cell in ^{15}N medium / g- cell in ^{14}N medium) are 1, 1.5, 2, 2.5, 3 in order from top to bottom.

et al., 1990) and bonds dihydropteroate reductase and inhibits folate metabolism (Prabhu et al., 1998). T87 cells both treated with methotrexate and untreated were combined with ^{15}N-labeled cells. Folate compounds were then extracted and subjected to capillary LC/MS to compare the $^{14}N/^{15}N$

ratios between (treated cells/labeled cells) and (untreated cells/labeled cells). It was found that 5-CH_3-H_4pteroate was reduced to 5.8% and 5-formyltetrahydropteroate to 23% by methotrexate treatment. This indicates the stable isotope method allows us to determine metabolic changes quantitatively.

Although we present experiments using cultured cells in this report, the same experiment using plant seedlings should also be performed. We believe the isotope dilution method will be a useful and practical method for improving the accuracy of quantitative metabolic profiling data.

3 MEASUREMENT OF ^{15}N-LABELING RATIOS OF METABOLITES

3.1 Background

Information on metabolic flux is very important to analyze metabolic changes caused by genetic or environmental perturbation. However, metabolic profiling or metabolic fingerprinting, currently widely used, measures accumulated amounts of metabolites and does not reveal metabolic flux directly (Matsuda et al., 2003). The measurement of the isotopic distribution of metabolites following the incorporation of an isotope into a sample for a certain time should be possible for metabolic flux analysis. In microbiology, a methodology for metabolic flux analysis using ^{13}C-glucose has been developed (Stephanopoulos et al., 1998). Even in plant biology, analyses of metabolic fluxes using ^{13}C-glucose or $^{13}CO_2$ have been performed (Kruger et al., 2003); however, unidentified metabolic pathways or intracellular compartmentation of metabolism make such analyses difficult.

In contrast, a method using inorganic ^{15}N, which does not enable the direct analysis of metabolic flux either, allows measurement of apparent metabolite turnover. The apparent turnover helps us to estimate metabolic flux. Accordingly, we verified that *in vivo* ^{15}N-labeling could be applied to analyzing metabolic dynamics.

3.2 Materials and experiments

Materials and analytical instrument were the same as in 2.2 above. For amino acid analysis, labeled cells were harvested by filtration and were immediately ground in liquid nitrogen. The samples were immersed in a methanol/chloroform/water (2.5/1/1) extraction buffer at 37ºC for 30 min. The suspensions were centrifuged and upper phases (polar phase) were

recovered. These phases were added to water and centrifuged. Then upper phases were recovered and dried *in vacuo*. The residues were dissolved with lithium carbonate (80 mM, pH9.5) and derivatized with dansyl chloride (1.5 mg/ml in acetonitrile) for 1 hour (Tapuchi et al., 1981). Samples were filtered through a 0.45 μm filter, and then subjected to capillary LC/MS. The LC eluent was water/methanol containing 20 mM ammonium acetate.

3.3 Results and discussion

First, we confirmed whether or not a difference in isotopic distribution among metabolites was observed. The extracts derived from cells cultured in ^{15}N-labeled and unlabeled LS media under continuous light for 24 h were subjected to capillary LC/MS analysis. Figure 21-4 shows mass spectra of dansylated glutamine and serine extracted from the two types of cells. In the mass spectra of amino acids derived from unlabeled cell, peaks of monoisotopic mass (glutamine: 380 *m/z*, serine: 339 *m/z*) possessed the highest intensity, whereas those of isotopomers were very low. In contrast, mass distributions of amino acids from ^{15}N-labeled cell were shifted to higher *m/z*. This shows nitrogen in amino acids was substituted by ^{15}N stemming from inorganic nitrogen in the media. Moreover, differences in the degree of mass peak shift were observed. (The degree of peak shift for glutamine was higher than for serine as shown in Figure 21-4.) This indicates the incorporation ratios of ^{15}N are different among amino acids, which means mass distribution should be an indicator of metabolic flux.

Next, ^{15}N-labeling ratios of amino acids were compared between cells cultured under light and dark conditions. Incorporation of ^{15}N was performed

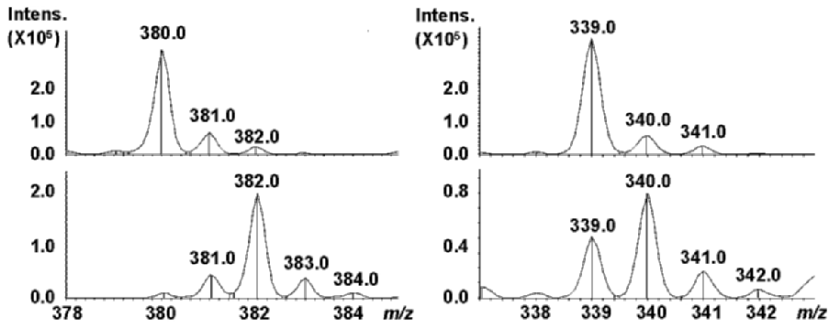


Figure 21-4. Mass spectra of dansylated glutamine (left) and serine (right) (Top) spectra obtained from extract of cultured cell in ^{14}N medium (bottom) spectra from extract of cell cultured in ^{15}N medium for 24 hours.

for 24 h. ^{15}N-labeling ratios of each amino acid were obtained by subtracting contributions of natural abundance ^{13}C, ^{15}N, ^{18}O, ^{34}S, ^{36}S from the observed mass distributions.

The results are shown in Figure 21-5. Under light conditions, the ^{15}N-labeling ratios of glutamine, glycine, and aliphatic amino acids were relatively high, whereas the ratios of asparagine, aspartate, and aromatic amino acids were low. This indicates the length of the pathway for nitrogen incorporation into each amino acid. In comparison to the light conditions, most ^{15}N-labeling ratios were decreased under dark conditions. Lysine, histidine, and aromatic amino acids were especially remarkable. This showed biosynthesis of these amino acids was inhibited. Previously it was shown using DNA microarray experiments that the transcription of genes related to histidine and tryptophan biosynthesis is induced by light (Ma et al., 2001), whereas the ^{15}N-labeling ratios of arginine and asparagine were increased under dark conditions. It is well known that biosynthesis of asparagine is activated under dark conditions (Ireland and Lea, 1999). Thus the results obtained by this experiment were consistent with previous findings, thereby establishing the validity of this method.

This study indicates that comparing ^{15}N-labeling ratios among extracts of cells cultured in different conditions should enable us to estimate the activity of metabolism. This method can be applied not only to amino acids but also other metabolites containing nitrogen, and proteins. Moreover, the method can also be applied to functional analysis of unknown genes using transformed cell lines or a transient RNAi system (An et al., 2003).

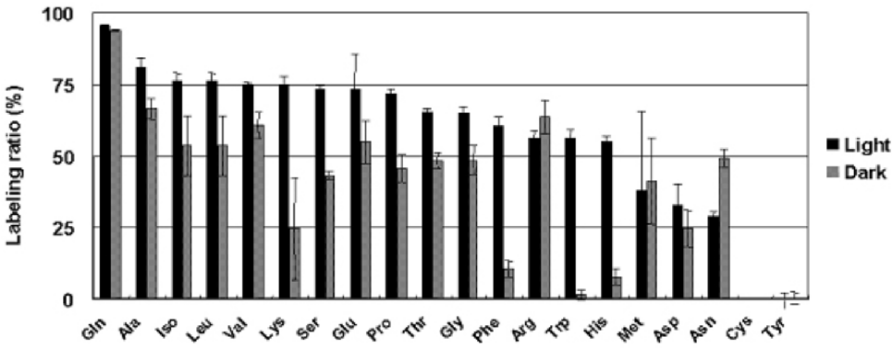


Figure 21-5. ^{15}N labeling ratio of amino acid pool derived from T87 cultured in light and dark condition (error bars represent standard deviations, n=6).

ACKNOWLEDGMENTS

We would like to thank Dr. Hiroyoshi Minakuchi (Kyoto Monotech Corp.) for his kind gift of an octadecylated capillary monolithic silica column. This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

An, C.I., Sawada, A., Fukusaki, E., and Kobayashi, A., 2003, A transient RNA interference assay system using Arabidopsis protoplasts, *Biosci Biotechnol Biochem*. **67**:2674-2677.

Axelos, M., Curie, C., Mazzolini, L., Bardet, C., and Lescure, B., 1992, A protocol for transient gene expression in Arabidopsis thaliana protoplasts isolated from cell suspension cultures, *Plant Physiol Biochem*. **30**:123-128.

Dube, G., Henrion, A., Ohlendorf, R., and Vidal, C., 2001, Application of the combination of isotope ratio monitoring with isotope dilution mass spectrometry to the determination of glucose in serum, *Rapid Commun Mass Spectrom*. **15**:1322-1326.

Fukusaki, E., Harada, K., Bamba, T., and Kobayashi, A., 2005, An isotope effect on the comparative quantification of flavonoids by means of methylation-based stable isotope dilution coupled with capillary liquid chromatograph/mass spectrometry, *J Biosci Bioeng*. **99**:75-77.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R., 1999, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat Biotech*. **17**:994-999.

Ireland, R.J., and Lea, P.J., 1999, The enzymes of glutamine, glutamate, asparagine, and aspartate metabolism, in: *Plant amino acids, Biochemistry and Biotechnology*, B.K. Singh, ed., Mercel Dekker Inc, New York, pp. 49-109.

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., and Olah, T., 2000, Mechanistic investigation of ionization suppression in electrospray ionization, *J Am Soc Mass Spectrom*. **11**:942-950.

Konings, E.J.M., 1999, A validated liquid chromatographic method for determining folates in vegetables, milk powder, liver, and flour, *J AOAC Int*. **82**:119-127.

Kruger, N.J., Ratcliffe, R.G., and Roscher, A., 2003, Quantitative approaches for analyzing fluxes through plant metabolic networks using NMR and stable isotope labeling, *Phytochemistry Reviews*. **2**:17-30.

Ma, L., Li, J., Qu, L., Hager, J., Chen, Z., Zhao, H., and Deng, X.W., 2001, Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways, *Plant Cell*. **13**:2589-2607.

Matsuda, F., Morino, K., Miyashita, M., and Miyagawa, H., 2003, Metabolic flux analysis of the phenylpropanoid pathway in wound-healing potato tuber tissue using stable isotope-labeled tracer and LC-MS spectroscopy, *Plant Cell Physiol*. **44**:510-517.

Matuszewski, B.K., Constanzer, M.L., and Chavez-Eng, C.M., 2003, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, *Anal Chem*. **75**:3019-3030.

Mëller, C., Schaefer, P., Störtzel, M., Vogt, S., and Weinmann, W., 2002, Ion suppression effects in liquid chromatography-electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries, *J Chromatogr. B*. **773**:47-52.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M., 2002, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol Cell Proteomics*. **1**:376-386.

Prabhu, V., Chatson, K.B. Lui, H., Abrams, G.D., and King, J., 1998, Effects of sulfanilamide and methotrexate on 13C fluxes through the glycine decarboxylase/serine hydroxymethyltransferase enzyme in Arabidopsis, *Plant Physiol*. **116**:137-144.

Schweitzer, B.I., Dicker, A.P., and Bertino, J.R., 1990, Dihydrofolate reductase as a therapeutic target, *FASEB J*. **4**:2441-2452.

Steen, H., and Pandey, A., 2002, Proteomics goes quantitative: measuring protein abundance, *Trends Biotechnol*. **20**:361-364.

Stephanopoulos, G.N., Aristidoum, A.A., and Nielsen, J., 1998, Metabolic *Engineering*: Principles and Methodologies, Academic Press, San Diego.

Tapuhi, Y., Schmidt, D.E., Lindner, W., and Karger, B.L., 1981, Dansylation of amino acids for high-performance liquid chromatography analysis, *Anal Biochem*. **115**:123-129.