ABSTRACT

Title of Document: HIGH-THROUGHPUT TIME-SERIES METABOLOMIC ANALYSIS OF A SYSTEMATICALLY PERTURBED PLANT SYSTEM.

Harin Haridas Kanani
Doctor of Philosophy, 2007

Directed By: Dr. Maria I. Klapa
Assistant Professor
Department of Chemical and Biomolecular Engineering

In the post-genomic era, availability of high-throughput profiling techniques enabled the measurement of entire cellular molecular fingerprints. Major characteristics of the high-throughput revolution were that (a) studying biological problems did not have to rely on prior hypotheses, while (b) parallel occurring phenomena, previously assumed disconnected, could now be simultaneously observed. Metabolomics is the newest of the "omics" techniques. It enables the quantification of hundreds of free metabolite pools, providing a metabolic fingerprint. Considering the importance of cellular metabolism, which is the net effect of changes at the genomic, transcriptomic and proteomic levels and of the cell with its environment, the metabolomic profile, is a fundamental determinant of cellular physiology.

Obtaining accurate and complete metabolomic profiles is thus of great importance. However, being recent technology, metabolomics is currently at its

standardization phase. As part of my PhD thesis research, I focused on addressing several current challenges in metabolomics technology development. Specifically a novel data correction, validation and normalization strategy for gas chromatography-mass spectrometry (GC-MS) metabolomic profiling analysis was developed, which dramatically increased the accuracy and reliability of GC-MS metabolomic profiles. The optimized metabolomics protocol was applied to study the short-term dynamic response of systematically perturbed *Arabidopsis thaliana* liquid culture system to study regulation of its primary metabolism. The biological system was studied under conditions of elevated CO_2 stress, salt (NaCl) stress, sugar (trehalose) signal, and hormone (ethylene) signal, applied individually; the latter three stresses also applied in combination with the CO_2 stress. Analysis of the obtained results required the appropriate application of multivariate statistical analysis techniques, which are developed mainly in transcriptomic analysis, into metabolomics analysis for the first time.

The acquired results identified important new regulatory information about the biological systems resulting in new targets for metabolic engineering of plants. The large number of dynamic perturbation allowed re-construction of metabolic networks to identify possible novel metabolic pathways based on correlations between metabolic profiles. In addition, it demonstrates the advantages of dynamic, multiple-stress “omic” analysis for the elucidation of plant systems function. In this sense, it contributes in further advancing the computational and experimental metabolic engineering and systems biology toolbox.

HIGH-THROUGHPUT TIME-SERIES METABOLOMIC ANALYSIS OF A SYSTEMATICALLY PERTURBED PLANT SYSTEM

By

Harin Haridas Kanani

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Maria I. Klapa, Chair
Professor Evanghelos Zafiriou
Professor Nam Sun Wang
Professor Srinivasa Raghavan
Professor Zhongchi Liu

ACKNOWLEGMENTS

I think one of the most interesting part of my PhD research was that it allowed me to meet and interact with so many interesting people all of whom helped me in some way or other to learn various aspects of being a Metabolic Engineer and Systems Biologist. I started my research with the motivation of learning Metabolic Engineering; however during the course of my research I also got to be an Analytical Chemist, Plant Physiologist, Bioinformatician, a Programmer and finally back again as a Metabolic Engineer to put all this information together and identify engineering targets for the plant system. All this would not have been possible without the help and guidance of my many teachers and colleagues which helped me along during the journey towards this PhD Dissertation.

I would first and foremost like to thank my advisor Prof. Maria Klapa for the initial experiment design and providing the initial conceptual motivation for the current project. For my foray as an Analytical Chemist, I would like to thank Dr. Brian Bagley and Prof. Judd Nelson from University of Maryland – Mass Spec Facility, for allowing us to use their instrument for the initial trials till we got our own instrument and for their helpful suggestions and discussions during trouble-shooting. I would also like to thank Prof. Gerry Dietzer and Mr. John Korns of the University of Maryland Green House Facility for allowing us to use and make necessary modifications to the growth chambers in the Green House. I would also like to thank Ms. Linda Moy and Dr. Fenglong Liu for their help and initial training with the extraction protocols at The Institute of Genomic

Research (TIGR) in Prof. John Quackenbush's Lab. I would like to thank Mr. Ben Woodard of University of Maryland-Bio Scale-Up facility for allowing us to have access to their lab, house our GC-MS during the intermediate period of time and use some of the equipments required for our research. I would also like to thank my advisor Dr. Maria Klapa and National Science Foundation for the financial support provided for my research, without which this project would not have been possible.

I would also like to thank Prof. Nam Wang, Prof. Evanghelos Zafiriou, Prof. Srinivas Raghava, Prof. Zhongchi Liu and Prof. John Quackenbush for being on my committee and for their comments, feedback and wonderful questions. Even though due to university rules and schedule conflict Prof. Quackenbush could not be part of the final official committee I am really thankful to him for all his suggestions during the course of the PhD and allowing us to work in his lab for doing the extractions. I would also specially like to thank Prof. Wang, Prof. Raghavan and Prof. Zafiriou for helping me throughout my PhD, for being there for me when I wanted to talk to some one about my research and giving me very valuable feedbacks about improving my presentation skills, experimental designs and data analysis techniques. Especially from Dr. Wang's courses on Engineering Methods and Biological Detection Techniques were really very helpful in de-mystifying some of the seemingly complicated mathematical and biological techniques used frequently in the modern biological research. I also learnt a lot and had a lot of fun while working as a TA in Prof. Zafiriou's design course and I would like to thank him for giving me the wonderful opportunity. One of the regrets from my five years at UMD would be that due to some reason or other I was not able to take Prof. Zafiriou's optimization and control class. I have heard a lot about it from my seniors and

from Dr. Klapa and was really looking forward to the same. I hope some day I will get an opportunity to do this.

One of the important challenges in front of me during the project was to understand molecular biology, physiology and development of plants. I can not thank Prof. Zhongchi Liu and Prof. Hevan Sze for their help, wonderful courses and their encouragement. They really introduced me to the fascinating world of plant physiology, answered my numerous questions patiently, and provided me various opportunities to interact with graduate students and researchers in the area. Even though I was from Chemical Engineering Department, they made sure to make me feel included in the community and I will always be grateful for their help and generosity.

I would also like to thank this opportunity to thank all my graduate student colleagues, especially my friends Rinku, Aruna, Ye, Angela, Brandy, Stefanie, Rohan, Bani, Vivek, Diana and Patricia. However I would like to thank the most my colleague and lab mate Bhaskar Dutta, for being there at frustrating times, for helping me whenever I need help outside of the research, for the long nights in the green house and labs working on the posters and for all the fun time and discussions we have had during the research and conferences.

Finally, my individual transition from a Process Chemical Engineer to a Metabolic Engineer would not have been possible without constant guidance from my advisor Prof. Maria Klapa. It is indeed an honor to be among her first batch of PhD students. Apart from the long scientific discussions which were really critical when I was taking the first steps in this biology, she has also provided constant motivation to work hard and achieve

the desired results. I am especially grateful to her for taking care of us even when she was away in Greece, discussing with us till as late as 3 am for her after a hard day's work teaching and guiding students in her lab in Greece. Even though it seemed difficult at first she made sure we had all the opportunities (and even more) and required guidance to be successful in our careers. She provided a lot of advise and micro-managed in the beginning when I really needed the guidance and later when she was confident I can do it on my own - also gave me a free hand to peruse projects and collaborations which I found interesting. I really could not have defined or asked for a better advisor and I will always be grateful to her for accepting me as her graduate student and for being a great advisor.

On a personal level, I would like to thank my wife Jesal for her love, her support and her never failing faith in me, for providing me with a strong motivating force and encouraging me to perform at my best level. I could not have done and achieved all this without her help, understanding and motivation. I would also like to thank my pretty loving daughter Mihika, who has given me tremendous amount of joy. I can also never thank my parents Harish and Beena enough, for installing the spirit in me to always aim high in my life, to work hard to achieve what I believe in, for the high moral standards and for instilling the curiosity in me to wonder 'How Life Works'.

to… all my teachers,
my grandparents,
my loving parents,
my dearest jesal and mihika,
and most respected Gandhiji.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The analysis of biological systems had to traditionally rely on the monitoring of macroscopic physiological properties and few variables at the molecular scale, due to limitations in the available analytical techniques. Under these conditions, the selection of the few microscopic markers to be monitored required the use of initial hypothesis regarding the particular biological problem. Therefore, any conclusions or models derived from such analysis depended heavily on the validity of this initial hypothesis, i.e. whether the selected measurements were indeed the most sensitive markers of the process under investigation. In addition, being few and specific, the molecular measurements were not usually adequate for the researcher to observe and correlate simultaneously occurring phenomena that had not been included in the initial hypothesis.

Advances in the computational and robotic techniques, along with better understanding of biological processes, allowed for the development of the high-throughput ('OMICS') techniques. The latter enabled researchers to obtain very detailed and comprehensive information about the state of a biological system at the molecular level. In contrast to the conventional analysis, the high-throughput analysis of biological systems does not require the use of initial hypothesis. Moreover, allowing for the holistic analysis of cellular fingerprints, simultaneously occurring phenomena can now be observed and correlated leading to more detailed and accurate models of cellular function. Hence, high-throughput techniques can significantly upgrade the quality of information that is

obtained about a biological system.

Transcriptional profiling using cDNA micro arrays (Schena et al., 1995) or Affymetrix 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 the complex biological systems requires the quantitative integration of all cellular fingerprints, i.e. genome sequence, maps of gene and protein expression, metabolic output, and *in vivo* enzymatic activity (Idekar et al., 2001). In a systematically perturbed cellular system such integration can provide insight about the function of unknown genes, metabolic regulation and even the reconstruction of the gene regulation network (Klapa and Quackenbush, 2003).

Before such integrated analysis can be carried out, the challenges of quantitative high-throughput analysis at each level of cellular function need to be resolved. These challenges range from limitations in the available experimental protocols to lack of data analysis techniques for upgrading the information content of the acquired data. With this motivation, the proposed research will address major challenges in the quantitative high throughput analysis of the metabolic state of a biological system.

1.1 METABOLOMICS

Metabolic profiling (or Metabolomics) refers to the high-throughput analysis of the metabolic state of a biological system by the simultaneous measurement of the relative concentration of at least few hundred small molecules in the cellular biomass (Sumner and Dixon, 2003). Even though it's the changes in metabolic fluxes that directly reflects

reshuffling of the in-vivo enzymatic activity due to an applied perturbation and could be directly integrated with gene expression data, metabolic concentrations are also expected to change as a result of metabolic flux re-distribution. Therefore a metabolic profile of a biological system can still provide a significant fingerprint of the metabolic state of the cells.

The main advantage of metabolic profiling analysis over metabolic flux analysis is that the former is high-throughput while the later is not. The later requires extensive knowledge about the structure of the biochemical reaction network of a biological system and has been primarily applied to steady or pseudo-steady state conditions. These limitations render metabolic flux analysis cumbersome and challenging for the study of complex eukaryotic systems. In most cases for these systems, steady or pseudo steady state conditions are a risky assumption to make, while – especially in the case of plants – insufficient and/or incomplete biochemical information is available.

Since the initial metabolomic publications in plants in 1999 (Roessner et. al., 1999), the interest in metabolomic analysis has grown tremendously. Applications of metabolic profiling range from bacteria (Burja et al., 2003) to yeast (Castrillo et al., 2003), plants (Roessner et. al., 2000; Fiehn et. al., 2000b; Kanani, 2004; Hirai et. al., 2004; Dutta et. al., 2007b), rats (Gopaul et al., 2000) and humans (Gemesi et. al., 2001). Some of the recent papers also refer to integrated studies of metabolic profiling in combination with other high-throughput analysis such as proteomics (Mayr et al., 2004) and transcriptomics (Hirai et. al., 2004; Dutta et. ak., 2007). The increasing diverse studies involving metabolomic analysis of different biological systems is also targeted towards a

variety of commercial applications in Industries such as Agricultural, Agri-Biotech, Industrial (or White) Biotech, Pharmaceutical, Diagnostics and Fundamental Research of biological systems in Universities and research Institutes. Figure 1-1 summarizes various currently proposed applications of metabolomic analysis in these industries.

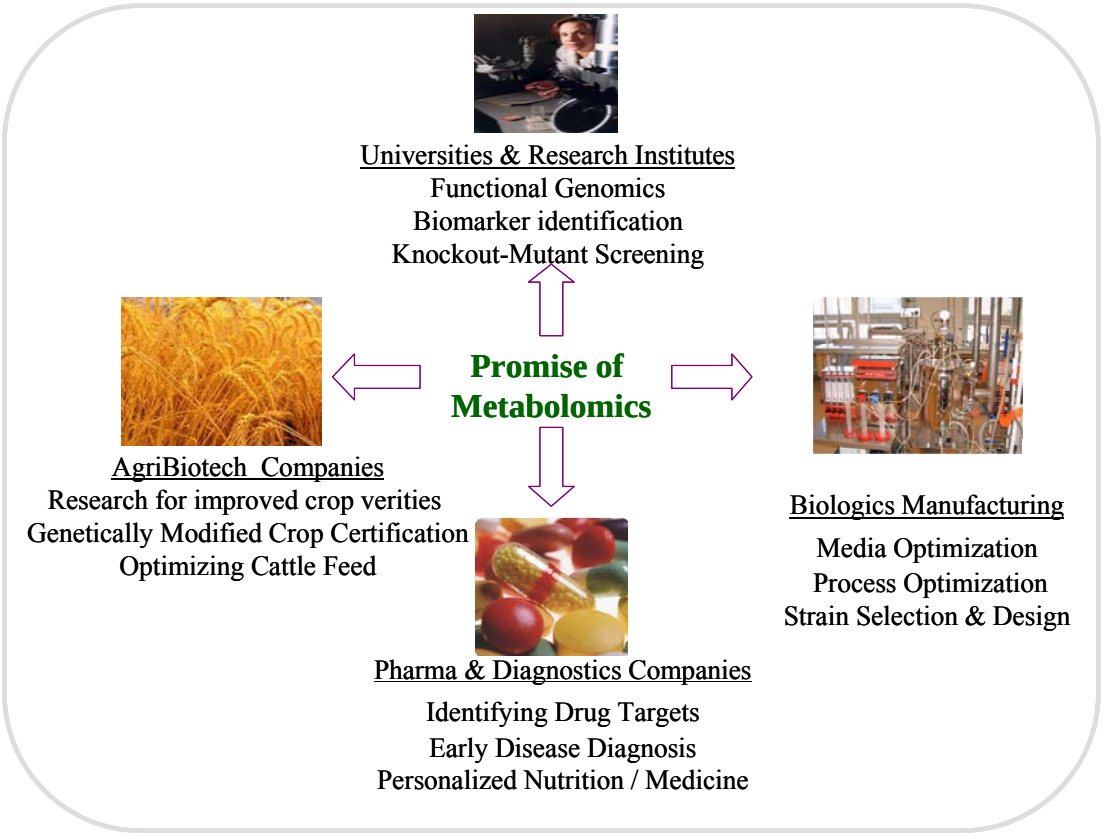


Figure 1-1: Emerging Applications of Metabolomics in Industry and Academia

The measurement of the metabolite concentrations is carried out primarily by gas (Roessner et. al., 2000, Fiehn et al., 2000b, Kanani, 2004, Kanani and Klapa, 2007) or liquid (Katz et. al., 2004) chromatography- mass spectrometry or nuclear magnetic resonance (Ratcliffe and Hill, 2004) spectroscopy (NMR). Even though each one of the techniques has advantages and disadvantages gas chromatography coupled with mass spectrometry (GC-MS) offers the maximum number of advantages and there by has

been the most commonly used for metabolomic profiling analysis (Fiehn, 2001; Kopka et al., 2004). More information about the differences between these techniques has been provided in Chapter 2 of this report. While the techniques used for the determination of the concentration of small molecules in a biological sample are not new, it is their use in a high-throughput manner along with sophisticated experimental design that introduces the main novelty to this type of analysis.

Most of the reported metabolic profiling studies have focused on the identification of environmental and genetic phenotype. Data analyses in these studies involve statistical analysis primarily using Principal component analysis (PCA), Hierarchical Clustering (HCL) and t-test. Other potential applications of metabolic profiling which have been discussed, but not yet fully addressed in the literature are:

- Metabolic network reconstruction from metabolic profiling data correlation analysis (Kose et. al., 2001; Steuer et. al.,2003).

- Identification of the function of unknown genes and reconstruction of metabolic gene regulation network through integration of metabolic profiling to gene expression data (Klapa and Quackenbush, 2003).

The main reasons that these issues have not yet been fully addressed are:

- Identification of unknown biochemical pathways and reconstruction of the metabolic regulation network requires time series metabolic profiling analysis (Steuer et al., 2003). However apart from the previous work in our lab (Kanani,

2004), very few time-series metabolomic experiments have been performed.

- Data analysis techniques for measuring and interpreting correlations between time series profiles of metabolite or gene expression data for reconstruction of a biochemical network are not available.

- For integrated analysis applications, metabolic flux information is required. However experimental design and data analysis methodology for obtaining metabolic flux information from a complex eukaryotic system at transient metabolic conditions in a high throughput manner is not available.

- The protocol for metabolic profiling using GC-MS is still not completely optimized (Kanani, 2004).

Moreover all recently published high-throughput metabolomic studies measure the response of a system to one perturbation at a time. However when two or more perturbations are applied in combination, the response of the system is not expected to be linearly related to the responses of the individual perturbations. To date, no highthroughput experimental design or data analysis strategies that would lead to derivation of conclusions from multiple perturbations of a biological system using high throughput techniques are currently available.

1.2 PLANT PRIMARY METABOLISM

Plants along with photosynthetic bacteria are the primary fixers of solar energy, inorganic Carbon (CO_2) and inorganic nitrogen (NO_3^- , N_2). This unique ability of plants plays an

important role in sustaining life on earth. Plants are highly sophisticated multi-cellular organisms which can produces thousands of chemicals through highly complex, compartmentalized chemical reaction networks. Plants are able to produce these numerous highly complex chemicals primarily using sunlight, CO_2, O_2, N_2 from the atmosphere; macro-nutrients such as $NO_3^-, SO_4^{-2}, PO_3^{-3}, Na^+, K^+, Ca^{+2}$ and several micronutrients (see Taiz, 2002 for complete list) normally available in ground. The role that the plants are foreseen to play in society, science and technology has changed considerably since the completion of the first plant genome sequence in 2000 (Arabidopsis Genome Initiative, 2000) and the initiation of the *Arabidopsis* 2010 functional genomics initiative (Somerville and Dangle, 2002). Beyond the traditional role of agriculture in providing food, neutraceuticals and natural polymers, for commercial, environmental and (bio)ethical reasons, plants are now taking central stage in bio-fuel (Ragauskas et. al., 2006), engineered bio-polymer (Slater et al., 1999) and chemical (Oksman-Caldentey, 2004) industry. For the traditional uses of the plants, the potential targets for genetic modifications so far have been mainly in the secondary metabolism, which is responsible for crop protection (Dixon et al., 1996) and neutraceutical production (Ajjawi et al., 2004). *Optimization of the upcoming agri-industrial applications will require plants that could efficiently fixate CO_2 and Nitrogen while withstanding stressful growth conditions with minimal nutrient requirements.* To engineer plants with these traits, extensive knowledge about the regulatory mechanisms governing plant primary metabolism will be of great importance.

A vast amount of studies in the last century, have significantly contributed to the research

aiming at unraveling the stoichiometry and regulation of plant primary metabolism (see summary in (Buchanan, 2000)). Even after such extensive efforts, detailed information about regulation, stoichiometry and compartmentalization of primary pathways are still putative to a greater extent as compared to the bacterial and in some cases even the mammalian systems. This currently limited information about plant primary metabolism usually hinders the application of the *in vivo* and *in Silico* flux balance models (Stephanopoulos, 1998) favored by metabolic engineering in other systems to elucidate the underlying regulatory mechanisms. In this case, the post-genomic high-throughput "omics" techniques and the multivariate statistics and systems engineering/biology analytical toolbox could be used to increase our understanding of plant primary metabolism.

1.3 RESEARCH OBJECTIVES

In this context, the main objective of the current project is the development of a methodology for the high-throughput, quantitative, dynamic metabolomic analysis of a systematically perturbed *A. thaliana* liquid culture system. In order to achieve the main objective, the following specific aims were pursued:

Specific Aim 1

To improve the available experimental protocol for the metabolic profiling analysis of plant biomass using Gas Chromatography-Mass Spectrometry. Improvement is expected to increase of the number and accuracy of acquired metabolic measurements as well as

guarantee the reproducibility.

Specific Aim 2

To design and perform experiments that enable the monitoring of the dynamic response of *A. thaliana* liquid cultures to various environmental stresses applied independently or in combination.

Specific Aim 3

To develop a data analysis strategy to upgrade the information content of the acquired data. Specifically the following issues will be addressed:

- Identification of metabolic fingerprints of a specific perturbation.
- Identification of metabolic fingerprints of specific perturbations that are carried over when two or more perturbations are applied simultaneously.
- Identification of changes observed between the metabolic profile of the combined perturbations versus those applied individually.
- Reconstruction of the metabolic regulation network from metabolic profiling data from multiple metabolic conditions.

Specific Aim 4

To interpret and validate the results obtained from data analysis in the context of the known *A. thaliana* physiology to understand the response of the plant system to the

applied environmental stresses.

1.4 DESCRIPTION OF THE THESIS

The thesis is organized in 7 Chapters.

Chapter 1: Describes the motivation for the thesis and importance of high-throughput analysis in biological systems, metabolomic analysis and regulation of plant primary metabolism is described. Specific objectives pursued in the current project are also stated.

Chapter 2: This chapter provides an overview of the history and current state of metabolomic analysis. It provides information about various platforms for metabolomic analysis, various stages of a metabolomic analysis and finally techniques available for this analysis.

Chapter 3: The chapter describes the methodology followed for optimizing the GC-MS metabolomic profiling analysis of plants. It describes the experiments carried out for optimizations and the resulting improvements achieved by optimization of each stage of metabolomic analysis.

Chapter 4: Description of the design of the biological experiments which were performed as part of the current project is provided in this chapter. Details about the choice of model system, choice of environmental perturbations, experimental conditions and time-points are also provided.

Chapter 5: Results obtained from the time-series metabolomic analysis of *Arabidopsis*

thaliana liquid culture response to four environmental perturbations (1) Elevated CO_2 (2) Salt (NaCl) Stress (3) Sugar (Trehalose) Signal (4) Hormone (Ethylene) Signal are presented. The results obtained from multivariate statistical analysis of the profiles are presented and their significance is discussed in the biological context. The results demonstrate the advantage of using high-throughput metabolomic analysis and time-series experimental design.

Chapter 6: Results obtained by comparing the individual stress response to the combined stress response are presented. Comparison of the results from individual perturbations is also presented to gain a system wide understanding of regulation of primary metabolism. Finally results obtained from re-construction of the metabolic network using metabolomic profiles from multiple dynamic perturbations.

Chapter 7: Conclusions from the current analysis along with recommendations for direction of future experimental design and data analysis techniques to significantly improve the interpretation of metabolomic data is provided.

2 GC-MS METABOLOMIC PROFILING

The metabolomic profile of a biological system refers to the concentration profile of all its free small metabolite pools ((Fiehn et al., 2000a; Roessner et al., 2000; Kanani and Klapa, 2007). Metabolites are defined as the small molecules that participate in the metabolic reactions as substrates or products; debate still exists regarding the maximum size of the "small" metabolites, which will also determine the size of the entire metabolome. Taking into consideration that the concentrations of the metabolites affect and are affected by the rates of the metabolic reactions (or metabolic fluxes), it becomes apparent that the metabolomic profile of a biological system provides a fingerprint of its metabolic state. As such, it is a phenotypic correspondent of the transcriptomic and proteomic profiles, which provide, respectively, the cellular fingerprint at the transcriptional (mRNA) and translational (protein) levels (Fiehn et al., 2000a). To-date, metabolomic profiling in plants has been mainly used to differentiate between metabolic states and/or identify an environmental or genetic phenotype (Broeckling et al., 2005; Cook et al., 2004; Dutta et al., 2007; Fiehn et al., 2000a; Kanani, 2004; Noguchi et al., 2003; Roessner et al., 2000; Sakai et al., 2004; Taylor et al., 2002; Weckwerth et al., 2004). The interest of metabolomic analysis in the current biological research could also be monitored by the dramatic increase in the number of publications in the recent years. Specifically Pub-med citations for articles containing any of the keywords "metabolic profiling", "metabolite profiling", "metabolomics", "metabolomic profiling" or "metabonomics", has increased from 14 in 2000, to 330 citations in 2006 (Figure 1-1).

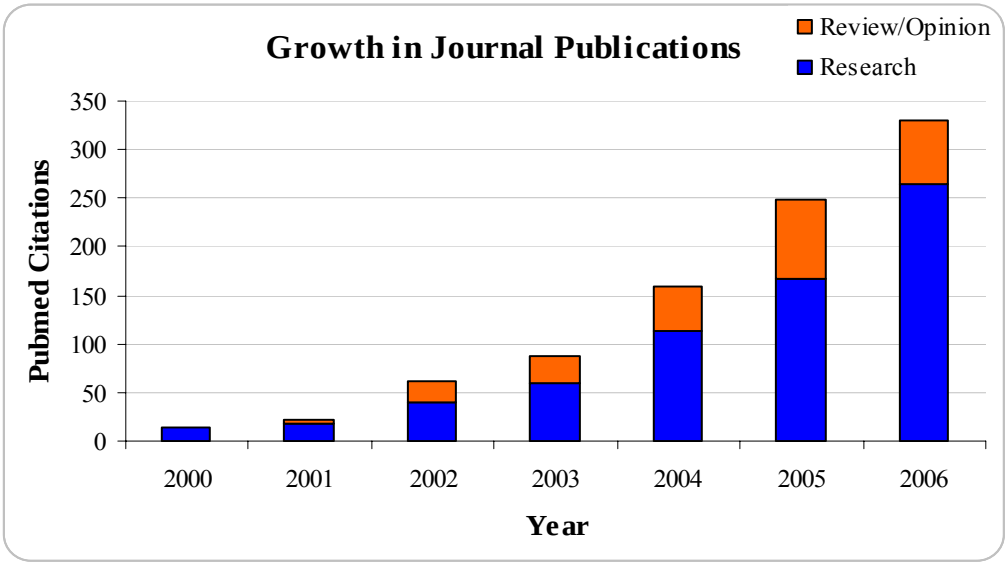


Figure 2-1: Growth of Publications matching keyword “Metabolomics” OR "Metabolomic Profiling" OR "Metabonomics" OR "Metabolic Profiling" OR “Metabolite Profiling” in Pubmed Database.

In the current research project metabolomic analysis is used to measure response of biological systems to external perturbations. In this chapter a brief review of the steps and tools available for performing metabolomic analysis are presented. (see Kanani, 2004 for a more detailed review of various steps involved).

2.1 PLATFORMS AND COMPARISON

High-throughput metabolomic analysis can be performed using several platforms. The most commonly platform are: being Gas or Liquid Chromatography- Mass Spectrometry (GC/LC-MS) and Nuclear Magnetic Resonance (NMR) Spectra. These techniques have recently compared and their conclusion has been summarized in Table 1 (Kopka et al., 2004, Kanani, 2004). GC-MS has been chosen as the technique to use in the current study due to high number of advantages that have ranked it the most commonly used technique

Table 2-1 Overview of analytical techniques available for metabolic profiling.

Analytical Technique	Compound Class measured	# of Peaks Detected	Advantages	Disadvantages
Gas chromatography – Mass Spectrometry	Mono, disaccharides, amino acids, organic acids, alcohols, monophosphates, volatiles (esters), lipids, sterols	400-700	Low cost, Large library, Better Separation, Differentiated mass-spectra	Derivatization Required, Very High molecular weight compounds not measured.
Liquid Chromatography-Mass spectrometry	All of above, less volatiles + pigments, diphosphates, alkaloids	Variable	No derivatization required	High Cost, Limited library, difficult separation of isomers
Nuclear Magnetic Resonance	Compounds that contain atoms with magnetic activity	150	Structure Identification, In-vivo studies possible	Low sensitivity, Magnetically active groups required
Others	For specific class of compounds OR tasks	50 – 100	Low cost, Specific for a given category	Absence of protocols and library

(Broeckling et al., 2005; Cook et al., 2004; Dutta et al., 2007; Fiehn et al., 2000a; Kanani, 2004; Noguchi et al., 2003; Roessner et al., 2000; Sakai et al., 2004; Taylor et al., 2002; Weckwerth et al., 2004) in metabolomic analysis. A detailed description of the equipment and its use for metabolomic analysis is available in (Kanani, 2004).

2.2 GC-MS METABOLOMIC ANALYSIS

Obtaining the metabolomic profile of the biological system starting from the initial biomass involves several steps which are summarized in Figure 2-2. In case of GC-MS metabolomic analysis, an additional step of derivatization is required in order to make metabolites volatile so that they can be separated in the gas phase by GC. The requirement of derivatization also adds an additional data correction step as shown in Figure 2-2 which restores the one-to-one proportionality between the original metabolic profile in the biological sample and the acquired metabolomic profile. The four steps involved in a typical metabolomic analysis using GC-MS are discussed below.

2.2.1 Metabolite Extraction

The small molecules targeted by metabolomic analysis can be obtained from the cellular biomass by three different methods of extraction:

(a) Vapor phase extraction

(b) Free metabolite extraction

(c) Total metabolite extraction.

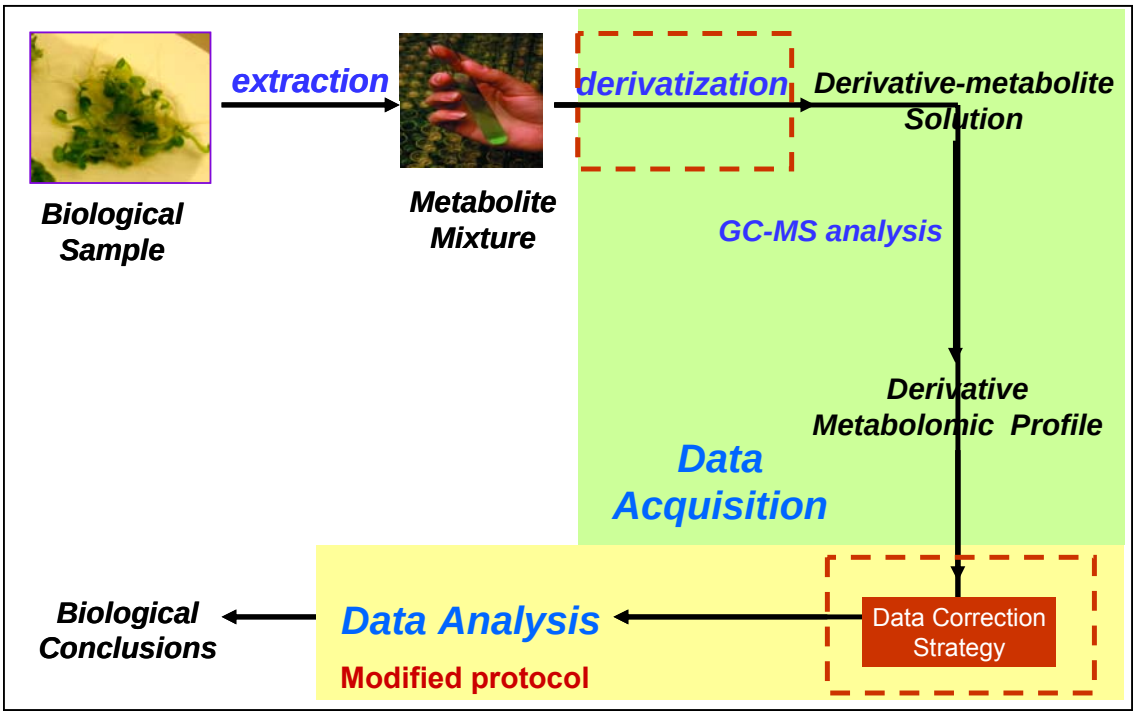


Figure 2-2: Steps involved in GC-MS Metabolomic Analysis. The addition step (as compared to LC-MS/NMR) of derivatization introduces bias into the analysis and hence requires an additional Data Correction step prior to Data Analysis.

Typically vapor phase extraction has been used for:

- The identification of components e.g. esters. As quality control markers in plants distained for food use (Deng et al., 2004)
- The study of the signaling between plants through secondary metabolites (Deng et al., 2004)

However for most general studies, water/methanol (for polar metabolites) (Roessner et al., 2000) and chloroform (for non polar metabolites) extraction has been used (Fiehn et al., 2000a). Using this method of extraction only the free small metabolite pools are obtained. Total Metabolite Extraction is used if in addition to free metabolite pool it is

important to measure the amount of the small precursors which are a part of cellular macromolecules However even though protocols for extracting the total (attached and free) metabolite pools for specific metabolite categories exist (Kitson et al., 1996), no protocol that enables the total extraction of multiple categories of metabolites has yet been developed. In addition extracting the total pool of a metabolite might require special type of analysis in the case of dynamic models of biological systems in order to relate the rate of incorporation of a metabolite into a macromolecular pool to the rate of production of the free pool. For all these reasons free metabolite extraction has been used in this study.

2.2.2 Metabolite Derivatization

Derivatization is imperative for the conversion of the small metabolites to volatile, non-polar 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) & Methoxime (MEOX) derivative(s) (Roessner et al., 2000). The chemical reactions involved can be seen from Figure 2-3. To ensure accuracy of the metabolomic analysis, the derivatization time should be optimized (see next Chapter for the optimization strategy).

2.2.3 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. These derivatized metabolite mixture when injected into GC-MS, is separated by gas chromatography using an inert

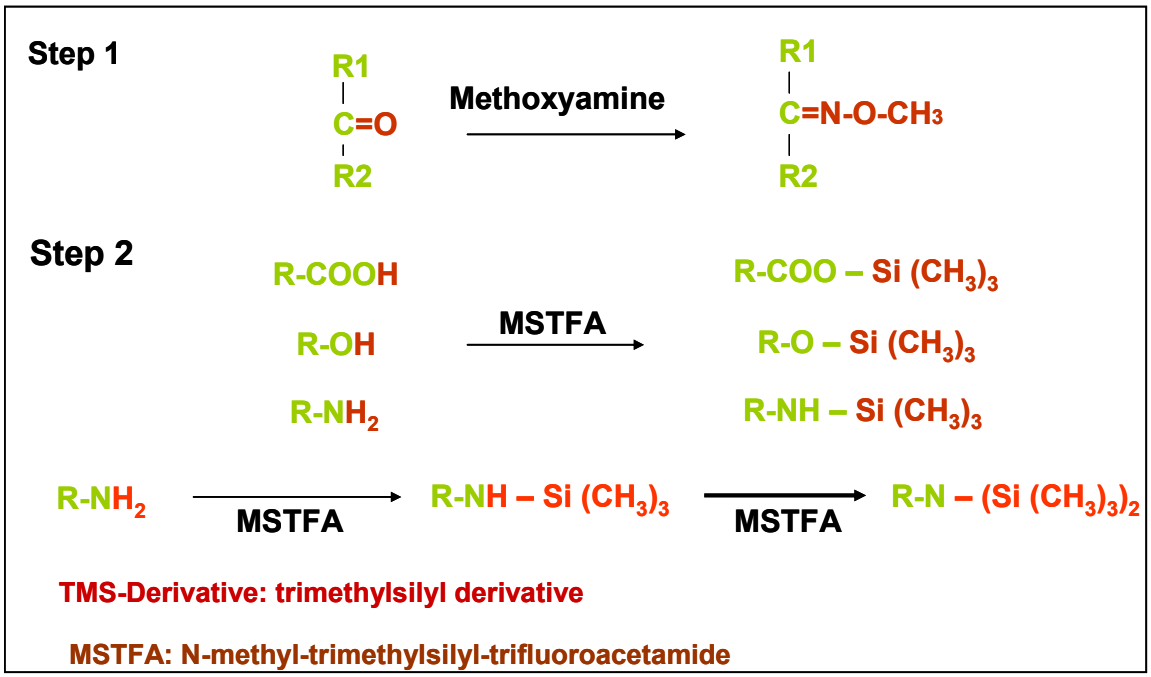


Figure 2-3: Most commonly used two step-derivatization process for producing Methoxime, trimethylsilyl derivatives of the metabolites in GC-MS Metabolomic analysis.

carrier gas and temperature programming. The separated metabolites enter the mass spectrometer, where they are ionized by electron bombardment. The mass to charge (m/z) and intensity of the fragmented ions are than recorded by mass spectrometer generating a mass spectrum of a given scan. Details concerning the actual GC-MS operating conditions are provided in (Kanani, 2004).

2.2.4 Metabolite Identification & Quantification

Identification of metabolites: The mass spectrum recorded in each scan by the mass spectrometer is than integrated to generate a chromatogram containing peaks representing one of more metabolite. Using the retention time and reference spectra of known derivatized metabolite available in the library, the peaks in the chromatogram are

identified. Rest of the peaks whose spectra/retention time can not be conclusively matched in the available library are treated as unknown metabolites (see Kanani, 2004 for more details).

Quantification of metabolites: Metabolic profiling analysis only obtains relative concentration of extracted metabolites. The peak area of the unique m/z fragment for a given metabolite at a give time is used for quantification. Prior to attempting any quantification, the linear range of operation of the instrument is determined to ensure optimum sample quantity and split ratio. It is important to ensure all measurements are within the linear range (typically spans 3-4 orders of magnitude). In metabolomic analysis however it is always the relative concentration of the metabolite which is of interest. In order to obtain relative quantification, fixed amount of internal standard (typically a compound not produced by plant) is added to the plant sample. The relative concentration of each metabolite is than determined by taking the ratio of peak area of a marker ion of the metabolite, to that of marker ion of the internal standard (see Kanani, 2004 for more details).

2.3 DATA NORMALIZATION

Metabolomic analysis consists of number of steps involving different chemical processes and equipments. Thus during sample preparation, significant variability is likely to occur. In addition in case of GC-MS analysis, 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 4 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 3 categories, for each of which a specific normalization strategy is suggested.

2.3.1 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 distorts 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., 2000a). 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.

2.3.2 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. 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, also Chapter 3 of this report). In order to account for these biases, a novel data correction, validation and normalization algorithm was developed which is presented in the Chapter 3 of this report.

2.3.3 Process/Setup 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.

2.3.4 Normalization with Reference Biological States

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, 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 time points 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 on 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 4-step data normalization and validation procedure could now be used in further data analysis.

2.4 MULTIVARIATE STATISTICAL TECHNIQUES

2.4.1 Clustering Techniques

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 snap-shots (Fiehn et al., 2000a; Roessner et al., 2001; Taylor et al., 2002), where presence of clusters represents different states. In case of time-series metabolomic profiling also, HCL and PCA can identify if the given perturbation alters the metabolism of the biological system significantly or not.

Principal component analysis is a technique which projects a large set of data onto a smaller set of variables (called principal components) which are a linear combination of all the initial variables. The principal components are chosen in a way so that the first component accounts for the largest variation in the samples. For data sets with high degree or correlation between variables, the first three components together can account for more than 50% of the total variability of the system. Under such conditions by plotting the metabolic profiles in the first three principal component plane, a large part of the variations (or the difference) between the plant samples can be visually identified. The Principal components required for the analysis are obtained by identifying the eigen vectors of the metabolomic data sets. For a given set of metabolomic data, the eigen vectors for the data are obtained and the scores of each biological sample along the three

eigen vectors is plotted. In most metabolomic time-series analysis, the first three eigen vectors together account for about 60-85% of the total variation. These is thus a useful tool in time-series analysis to identify and compare overall effects on the metabolism of the biological system in response to a given perturbations. In the present analysis, principal component analysis algorithm incorporated in MEV-TM4 (Saeed et al., 2003) was used. A more detailed description of the use of PCA and HCL analysis can also be obtained from Kanani, 2004.

2.4.2 Identifying significant metabolites

PCA analysis or any other clustering algorithm allows us to distinguish between biological states and to determine if in response to a perturbation in the system if the metabolomic profile was altered significantly. 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 snap-shot "-omic" studies (Fiehn et al., 2000a; Roessner et al., 2001; Taylor et al., 2002), this has been achieved through t-test or fold change (FD) analysis. These methods, however, do not include any distinct threshold characterizing significance and also do not allow estimation of false detection rate accurately.

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 (Tusher et al., 2001; Kanani, 2004; Xiang et al., 2004; Hirai 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.

2.5 TIME SERIES METABOLOMIC DATA

Paired-SAM thus is a useful technique to understand the important biological changes which are taking place across all time points. However in case of time-series analysis, the objective is not only the identification of significant metabolites from overall analysis, but also at individual time-point. However this can not be achieved using SAM when only two replicates are available at each timepoint. Hence a new method was recently developed (Dutta et al., 2007) which was modified to identify significant metabolites at individual timepoint which is described in more detail in the next section.

2.5.1 MiTimeS for Metabolomic Data – General Definition

MiTimeS Analysis is a modified paired-SAM algorithm (Dutta et. al., 2007), which identifies number of significant metabolites at each individual time point. Specifically when using paired-SAM analysis, the observed score for each gene is estimated as follows:

$$d(i) = \frac{(\overline{X}_1(i) - \overline{X}_2(i))}{S(i) + S_0} \qquad (2\text{-}1)$$

where $\overline{X}_1(i)$: the mean expression of gene i in group 1 (Perturbed set); $\overline{X}_2(i)$: the mean expression of gene i in group 2 (Control Set); S(i) : the standard deviation of i^{th} gene's expression; S_0 : fudge factor. It is used to eliminate numerical biases at low values of S(i).(Dutta et al., 2007). For MiTimeS analysis, the observed score is redefined as:

$$d_t(i) = \frac{(X_{1t}(i) - X_{2t}(i))}{S(i) + S_0} \qquad (2\text{-}2)$$

Where subscript t indicates the observed score, and expression at a given time point t. The d_t distribution is then compared with the expected score d_e which is generated from the overall time point just as in paired-SAM analysis. A cutoff value of the observed score is obtained for a fixed delta value at each and every time point using the following equations:

$$\text{lower-cut-off}(\Delta)_t = \min\{di, \text{ where } de(i) - d_t(i) > \Delta\} \qquad \textbf{(2-3)}$$

$$\text{upper-cut-off}(\Delta)_t = \min\{di, \text{ where } d_t(i) - de(i) > \Delta\} \qquad \textbf{(2-4)}$$

where de(i) is the expected score distribution obtained by calculating observed scores for a population of comparisons generated by randomly switching time points between the control and perturbed groups. For 8 time points – 256 such random observed score distribution – are generated which are first sorted and subsequently averaged to obtain the expected score distribution. Equations 2-3 and 2-4 thus allow us to obtain the lower and upper cutoff at each time point. Any metabolite with an: observed score lower than the

lower cutoff is considered negatively significant metabolite at that time point; and having observed score at a time point higher than the upper-cut-off at that time point is considered positively significant metabolite at that time point.

In addition to calculating the positively and negatively significant metabolite, the MiTimeS analysis also allows calculation of Significance Variability (SV) Score (Dutta et al., 2007). For this calculation each positively significant metabolite is assigned +1 and negatively significant metabolite is assigned a value -1. The SV score is calculated by taking the mean of the absolute difference between consecutive time points. Hence if a metabolite remains at the same state at all the time points, the SV score value for the same is 0, where as if the metabolite changes state from positively significant to negatively significant and vice versa, at each an every time point the metabolite has an SV score value of 2. Thus SC score of the metabolite indicates variability in its significance level across time point and gives a sense of dynamics of the metabolite significance. Finally in order to identify time points which show a correlation in their response i.e. which have a high number of intercept between their positively or negatively significant metabolites, MiTimeS algorithm (Dutta et al., 2007) allows calculation of significance correlation matrix for both positively and negatively significant metabolites. In this report MiTimeS analysis was incorporated into a Mathcad algorithm to perform these calculations for metabolites and the same was used to identify significant metabolites at different time points, SV scores and SCM networks.

2.5.2 Pattern Matching

Methods to construct a known biochemical network by measuring the correlation between various metabolites at the same metabolic state have been proposed in past (Kose et al., 2001, Steuer et al., 2003). The assumption supporting such methodology is that metabolites which are "close" in terms of the position or function in a metabolic network should show correlated variations. However in the absence of time-series metabolic data, the proposed techniques were unable to re-construct biochemical pathways using metabolic profiling data. Demonstrating the ability to reconstruct known biochemical networks from metabolic profiling data would lead eventually to the discovery of unknown biochemical pathways, or the identification of metabolites that might be related due to the structure of the regulatory and not the stoichiometric network per se.

In the time series metabolic profiling analysis previously conducted in our lab (Kanani, 2004), it was attempted to measure and interpret correlations between different metabolites. Even though this analysis indicated metabolites belonging to the tricarboxylic acid cycle (TCA cycle) and metabolites belonging to sugar production pathways clustering together those clusters also contained a large number of apparently unrelated metabolites. It is expected that combination of data from multiple perturbations of the same metabolic network could provide a higher resolution picture of the correlated metabolites. Combining all the information would decrease the uncertainty about the metabolic network reconstruction in the context of quantitative techniques like Relevance Network (Butte et al., 2000) and Pavlidis Template Matching (PTM) (Pavlidis and Noble,

2001). From these Pavlidis template matching (PTM) analysis was the chosen technique for this project. PTM, uses one of the variables which can be chosen by the user, as template and identifies all other variables in the data set which show a correlation to the template for a given p value. Successively increasing stringency by lower p-values allows creating of network around the template metabolite showing which the metabolites which are most closely related to the queried metabolite. Results from using these techniques are shown in Chapter 6 of the report.

3 DATA CORRECTION STRATAGY FOR GC-MS METABOLOMIC ANALYSIS

The metabolomic profile of a biological system – referring to the concentration profile of its free small metabolite pools (Fiehn et al., 2000a; Roessner et al., 2000) - provides a phenotypic correspondent of the high-throughput transcriptional and proteomic profiles (Fiehn et al., 2000a). Gas Chromatography–Mass Spectrometry (GC-MS) has to-date been the technique of choice for quantitative metabolomics (Broeckling et al., 2005; Cook et al., 2004; Fiehn et al., 2000a; Kanani, 2004; Noguchi et al., 2003; Roessner et al., 2000; Sakai et al., 2004; Taylor et al., 2002; Weckwerth et al., 2004), of polar metabolites in particular, due to its current advantages over other available techniques (e.g. Liquid Chromatography-Mass Spectrometry (LC-MS), Nuclear Magnetic Resonance Spectroscopy (NMR) or Capillary Electrophoresis – Mass Spectrometry (CE-MS)) (Kanani, 2004; Kopka et al., 2004).

3.1 MOTIVATION FOR DATA-CORRECTION

To be detected, however, through GC-MS, small metabolites have to be converted to a volatile, non-polar and stable derivative form (Roessner et al., 2000). To-date, the most commonly used derivatization method in GC-MS metabolomics involves the original metabolites' conversion into their trimethylsilyl (TMS) and methoxime (MEOX) derivative(s) (Broeckling et al., 2005; Cook et al., 2004; Fiehn et al., 2000a; Kanani,

2004; Noguchi et al., 2003; Roessner et al., 2000; Sakai et al., 2004; Taylor et al., 2002; Weckwerth et al., 2004). Hence, when GC-MS is used, the actually measured metabolomic profile is the peak area profile of the metabolite derivatives. In this case, quantitative metabolomics has to take into account biases that distort the one-to-one directly proportional relationship between the original metabolite concentration profile and the measured peak area profile of the metabolite derivatives.

There are two types of biases in GC-MS Metabolomic analysis. The first type of potential systematic biases in GC-MS metabolomics is common among all analytical techniques. They are responsible for the proportionality ratio between a metabolite's original concentration and the peak area of its derivative's marker ion to potentially vary among samples to identical fold-extent for all metabolites. Hence, in the presence of such biases only and in the case that there were a one-to-one relationship between the original metabolite concentration and the measured peak area profiles, the composition of the two profiles would be the same. However, quantitative comparison among samples would not be possible in the absence of an internal standard for appropriate normalization. The selection criteria and type of the internal standard(s) have already been the subject of previous published work (Fiehn et al., 2000a; Roessner et al., 2000; Gullberg et al., 2004; Jiye et al., 2005). The second type of systematic biases in GC-MS metabolomics are due to the required derivatization of the original metabolite sample. These biases concern restrictively any separation-molecular identification/quantification technique used in metabolomics or any other chemical compound analysis that requires the derivatization of the original sample. Specifically, these biases might:

- Distort the one-to-one relationship between the original and the derivative metabolite profiles.

- Change the proportionality ratio between a metabolite's original concentration and the peak area of its derivative's marker ion to a different fold-extent for the various metabolites in the sample.

Following limitation of the derivatization reactions give rise to these biases:

Formation of multiple Derivatives: Some metabolites form more than one derivatives (Gehrke et al., 1969; Gehrke and Leimer, 1971; Poole, 1978), despite efforts to ensure a single derivative per metabolite (Gehrke et al., 1969; Gehrke and Leimer, 1971),

Incomplete Derivatization: Unless the derivatization reaction has completed for all metabolites in the sample, the acquired derivative metabolite profile depends both on the composition of the original sample (Fogler, 2002) and the derivatization time at which it has been acquired (Gehrke and Leimer, 1971).

If the derivative peak area profiles are not corrected from the "derivatization" biases, identified profile differences that might be due only to chemical kinetics and/or experimental setup could be attributed biological significance leading thus to erroneous conclusions. Even though this type of errors in the GC-MS spectra of certain classes of molecules have been known since the late 60s (Gehrke et al., 1969; Gehrke and Leimer, 1971), the discussion about them in the metabolomics community has been quite limited (Gullberg et al., 2004; Halket et al., 2005). The MS community resolved the issue experimentally for certain classes of compounds by using a different derivatizing agent

per class, each producing only one derivative per compound of the specific class (Poole, 1978; Evershed, 1993). In the high-throughput metabolomic context, however, this would mean that multiple runs with different derivatizing agents and derivatization/data acquisition protocols would have been needed with all the obvious quantification problems that such differentiation could bring along.

Hence a streamlined data correction, normalization and validation strategy that does not jeopardize the high-throughput nature of GC-MS metabolomics was developed and presented in the subsequent sections.

3.2 DERIVATIZATION BIASES

3.2.1 Response Factor Definition

In GC-MS metabolomic (but also any other chemical compound) analysis that requires derivatization of the original sample, there are three quantities of interest, the interrelationship of which needs to be seriously taken into consideration for an accurate data acquisition, quantification and analysis process:

- The concentration of a metabolite in the original sample.
- The concentration of the metabolite derivative(s) in the derivatized sample
- The measured peak area(s) of the metabolite's derivative(s).

The last two are directly proportional, the proportionality ratio

(i.e. $\frac{\text{marker ion's peak area}}{\text{derivative's concentration}}$) defined as the response factor, RF, of this derivative under specific GC-MS equipment's conditions (Colón and Baird, 2004). Therefore, if the latter remain constant, a derivative's RF is the same among all measured samples. Because of their proportionality, quantities (b) and (c) tend to be used interchangeably. However, they are not the same, and as it will be indicated below, considered as such could lead to erroneous data acquisition and analysis protocols.

3.2.2 Metabolite Derivative Formation

Conversion of the original metabolites into their trimethylsilyl (TMS) and methoxime (MEOX) derivative(s) involves two reactions.

Methoxime Derivative: In the first reaction involves the reaction between the original metabolite mixture and the solution of methoxyamine hydrochloride in pyridine. Any metabolite containing ketone (-C=O) group is transformed to the more stable and non-polar methoxime (-C=N-O-CH_3) group (Poole, 1978; Laine and Sweeley, 1971; Evershed, 1993).

Trimethylsilyl Derivative: On completion of the first reaction, the oxime derivatives of the ketone-group containing original metabolites and the metabolites that did not react in lack of ketone groups react with a silylating agent, e.g. the N-methyl-trimethylsilyl-trifluoroacetamide (MSTFA), to produce their more volatile, non-polar and stable TMS-derivatives. TMS derivative formation involves the replacement of the active hydrogen atoms (-H) in the hydroxyl, carboxylic and amine (-OH,-COOH, NH_2) functional groups

by the TMS (-Si$(CH_3)_3$) group (Poole,1978; Evershed, 1993).

In the rest of the text, MSTFA will represent the silylating agent in the second reaction only for demonstration purposes, without affecting the validity of the description for any other silylating agent selected to act in a TMS-derivatization process. Due to the mechanism and kinetics of the methoximation and silylation reactions, a metabolite might form multiple (oxime-)TMS derivatives (Gehrke et al., 1969; Gehrke and Leimer, 1971; Poole, 1978; Laine and Sweeley, 1971; Evershed, 1993; Halket, 1993).

3.3 METABOLITE CLASSIFICATION

Based on the number and type of their TMS-derivatives, the metabolites can be classified into three categories, as listed below. This grouping assists in the identification of the derivatization biases' sources, but also in the determination of ways to either avoid or correct for these biases in the acquired metabolomic profiles, as will be subsequently explained in this chapter.

3.3.1 Metabolites forming only one derivative (MD)

Metabolites in this category comprise only hydroxyl (–OH) and/or carboxylic (-COOH) functional groups. At the silylation step, the active hydrogen atoms of these groups react simultaneously leading to the formation of a single TMS-derivative. Schematically, the derivatization reaction of a category-1 metabolite M into its derivative MD could be depicted as follows:

$$M + MSTFA \xrightarrow{k} MD$$

where k represents the derivatization rate constant. Independent of the order of the derivatization reaction kinetics, the MD concentration in the derivatized sample becomes equal to the M concentration in the original sample after derivatization time t_1^*. For category-1 metabolites, t_1^* coincides with the time required for the completion of the original metabolite M's conversion, t_M. Figure 3-1(A) illustrates the time profile of the M and MD concentrations in the case of first-order derivatization reaction kinetics.

According to the principles of chemical reaction kinetics (Fogler, 2002), only the size of the plateau of the time profile of the MD concentration is independent of the derivatization reaction kinetics, depending only on the concentration of metabolite M in the original sample. Specifically, if all other derivatization parameters remain constant, replicates of the same sample might correspond to a different derivative concentration, and thus peak area, for metabolite j, if measured at different derivatization times (at least one) shorter than $t_{1,j}^*$. Moreover, samples of different original metabolite composition, measured at the same derivatization time shorter than the $t_{1,j}^*$ in (at least one of) the samples, might correspond to a different derivatization stage for metabolite j. The latter holds true, because metabolite's j derivatization kinetics, $t_{1,j}^*$ included, depends also on the original sample's composition (see Fogler, 2002 for kinetics of parallel-occurring compound reactions with the same reactant). Therefore, to avoid non-correctable biases affecting the comparison of a category-1 metabolite's j concentration among various samples, barring changes in the GC-MS operating conditions, i.e. for constant RF of

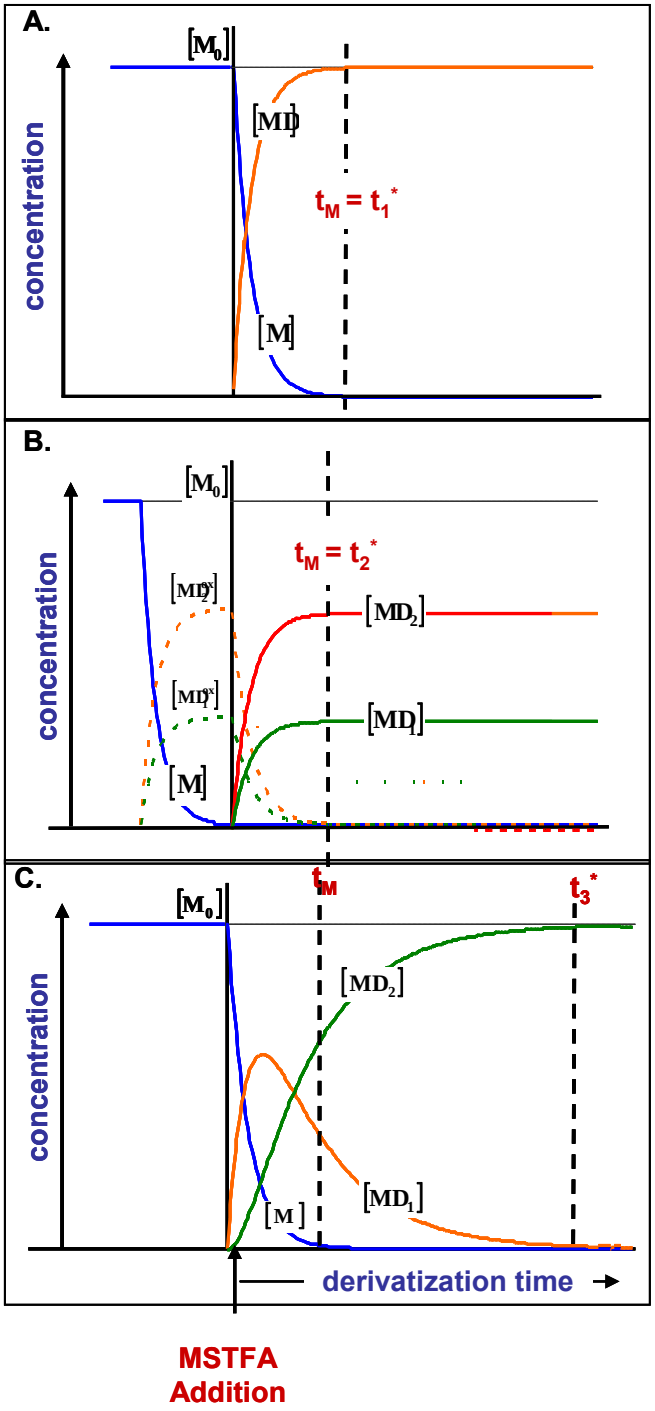
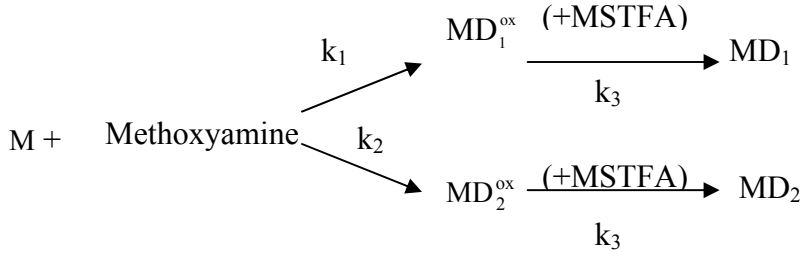


Figure 3-1: Variation in the original metabolite and TMS-derivative concentration with derivatization time, assuming first-order derivatization kinetics. (a) Category-1 (b) Category-2 and (c) Category-3 Metabolites. The final steady-state in each category is independent of the derivatization kinetics.

metabolite's j derivative, the TMS-derivative metabolomic profile of each sample should be acquired at derivatization time equal to or longer than the $t^*_{1,j}$ in this sample.

3.3.2 Metabolites forming two isomeric oxime-TMS derivatives

Metabolites in this category contain also ketone apart from the hydroxyl and carboxylic groups. At the methoximation step, these metabolites react through parallel reactions to produce two, *syn* and *anti,* geometric isomers (Laine and Sweeley, 1971; Poole, 1978; Halket, 1993). After silylation, these are converted through a catagory-1 type of reaction into two isomeric oxime-TMS derivatives that give rise to separate peaks in the metabolomic profile (Poole, 1978; Halket, 1993). Glucose, fructose, Mannose are examples of category-2 metabolites. Schematically, the methoximation and silylation reactions of a category-2 metabolite M into its two derivatives MD_1 and MD_2 could be depicted as follows:



where, k_1, k_2 depict the oxime formation and k_3 the silylation rate constants, respectively. Independent of the order of the oxime formation and silylation reaction kinetics, the concentrations of MD_1 and MD_2 are of constant ratio $k_o = \frac{k_1}{k_2}$ throughout the silylation reaction and reach final values, summing up to the initial M concentration, at derivatization (silylation) time t^*_2. In this case, t^*_2 coincides with the time required for

the complete conversion of the intermediate methoxime derivatives ($MD_{1,2}^{ox}$), t_M. Figure 3-1(B) depicts the time profile of the M, $MD_{1,2}^{ox}$ and $MD_{1,2}$ concentrations in the case of first-order oxime formation and silylation reaction kinetics. Under the same argument stated for category-1 metabolites, only the ratio of and the size of the plateau in the time profiles of the $MD_{1,2}$ concentrations is independent of the derivatization kinetics, the former being equal to k_o, a characteristic of metabolite M, and the latter depending only on the concentration of metabolite M in the original sample and k_o. In this context, three conclusions can be derived regarding the metabolomic profile of the category-2 metabolites:

(a) to avoid non-correctable biases affecting the comparison of a category-2 metabolite's j concentration among various samples, barring changes in the GC-MS operating conditions, i.e. for constant RF's of metabolite's j $MD_{1,2}$ derivatives, the TMS-derivative peak area profile of each sample should be acquired at derivatization time equal to or longer than the $t_{2,j}^{*}$ in this sample.

(b) Both MD_1 and MD_2 peak areas ($PA_{MD1,2}$) are in directly proportional relationship with the concentration of metabolite M in the original sample. Specifically:

$$k_o = \frac{k_1}{k_2} = \frac{[MD_1]}{[MD_2]} = \frac{1/RF_{MD_1}}{1/RF_{MD_2}} \quad \frac{PA_{MD_1}}{PA_{MD_2}} = \frac{RF_{MD_2}}{RF_{MD_1}} \quad \frac{PA_{MD_1}}{PA_{MD_2}} = \frac{RF_{MD_2}}{RF_{MD_1}} \quad \alpha \qquad (3\text{-}1)$$

$$[M] = [MD_1] + [MD_2] = (1 + \frac{1}{k_o}) \cdot \frac{1}{RF_{MD_1}} \cdot PA_{MD_1} \qquad (3\text{-}2A)$$

$$= (1 + k_o) \cdot \frac{1}{RF_{MD_2}} \cdot PA_{MD_2} \qquad (3\text{-}2B)$$

where $RF_{MD_{1,2}}$ represent, respectively, the response factors of $MD_{1,2}$.

Hence, the two derivative peak areas are not independent. Therefore, considering both in the metabolomic profile of a particular biological sample is expected to add double weight to the change in the concentration of category-2 metabolites compared to the other metabolite categories in any multivariate statistical analysis. In this case, only one of the two peak areas should be considered, the largest and less susceptible to noise being the better choice.

c) Ratio α of the $MD_{1,2}$ peak areas (see Eq. (3-1)) could be used as a criterion to verify whether the GC-MS operating conditions remained constant throughout the data acquisition of a sample batch. k_o being a characteristic of metabolite M, ratio α would be different for certain samples only in the case that the GC-MS operating conditions, and thereby the $MD_{1,2}$ RF's, changed before the acquisition of these samples. No other such data validation criterion exists currently in the metabolomics literature.

3.3.3 Metabolites forming Multiple Derivatives

Category 3 comprises metabolites forming multiple derivatives differing in number of derivatives or Chemical structures. Category 3 comprises metabolites containing at least one amine ($-NH_2$) group. Protons in the ($-NH_2$) group react sequentially and slower than those in the carboxylic (-COOH) and hydroxyl (-OH) functional groups (Poole, 1978), leading thus to the formation of multiple derivative forms (Gehrke et al., 1969; Leimer,

1977; Poole, 1978). In addition, the silylation of a particular metabolite derivative might lead to the formation of a TMS-derivative of another chemical compound, the formation of TMS-pyroglutamate from glutamate-3 TMS being a characteristic example (Gehrke et al., 1969;Leimer et al., 1977). Schematically, the silylation reaction of a category-3 metabolite M into its multiple derivatives could be depicted as follows:

$$M + MSTFA \xrightarrow{k} M(TMS)_x \xrightarrow[k_1]{(+MSTFA)} M(TMS)_{x+1} \dashrightarrow \xrightarrow[k_n]{(+MSTFA)} M(TMS)_{x+n}$$

(reaction sequence A)

or

$$M + MSTFA \xrightarrow{k} M(TMS)_x \xrightarrow[k_1]{(+MSTFA)} M(TMS)_{x+1} \dashrightarrow \xrightarrow{k_n} M\_new(TMS)_?$$

(reaction sequence B)

where k's and x depict, respectively, derivatization rate constants and number of TMS-groups after all carboxylic (-COOH) and hydroxyl (-OH) groups of the original metabolite M have reacted.

In theory then, because of the sequential nature of the derivatization reactions for a category-3 metabolite M, only one derivative with concentration equal to the M's concentration in the original metabolite mixture will be present in the derivatized sample after derivatization's completion, at time t_3^*. At any derivatization time shorter than t_3^*,

more than one derivatives of M may be present in the derivatized mixture. In contrast to the other two metabolite categories, t_3^* does not coincide with, but is (much) longer than, the time required for the complete conversion of the original metabolite M, $t_{M.}$ Figure 3-1(C) depicts the time profile of M and $MD_{x+1,x+2}$ concentrations in the case of first order derivatization reaction kinetics and two derivatives of metabolite M. From this description, the following conclusions can be derived regarding the derivative peak area profile of a category-3 metabolite:

- The multiple derivative peak areas of a category-3 metabolite are not independent and should not be considered as such in any multivariate statistical analysis of the metabolomic profiles;

- In contrast to the category 2 metabolite's two derivatives, the concentration of the category-3 metabolite's multiple derivatives is not each directly proportional to the concentration of metabolite M in the original sample; at times equal to or longer than t_M only their sum is.

Hence, none of the derivative peak areas can be used as single representative of the metabolite's M concentration in the original metabolite mixture. This is a major issue to be addressed in quantitative metabolomics, considering that, in the largest to-date publicly available curated retention-time library of TMS-derivatives (it is Max-Planck based [http://www.mpimp-golm.mpg.de/mms-library/index-e.html] and in the rest of the text will be referred to as MPL), out of 167 polar metabolites for which at least one derivative has been identified, 47 contain at least one ($-NH_2$)-group. Among those are the

amino acids, a class of major significance, because they are often used as markers of biological change (e.g. Noguchi et al., 2003; Sakai et al., 2004), implying that high accuracy in their quantification is indeed necessary.

Taking into consideration that the response factors of a category-3 metabolite's derivatives are not equal, at times equal to or longer than t_M *it is the weighted sum of the derivative peak areas that is directly proportional to the concentration of the metabolite M in the original sample, the weights being the inverse of the response factors of M's derivatives*. No algorithm enabling the estimation of the response factors of a category-3 metabolite's derivatives, and thereby the estimation of the weighted sum of its derivative peak areas, had been reported to-date. The first such algorithm is presented in the next section.

To avoid non-correctable biases affecting the comparison of a category-3 metabolite's j concentration among various samples, barring changes in the GC-MS operating conditions, the TMS-derivative peak area profile of each sample should be acquired at derivatization time equal to or longer than either (a) the $t^*_{3,j}$ in this sample, when only one derivative is present, or (b) the $t_{M,j}$ in this sample, but in this case the comparison should be made between the weighted sums of the metabolite's derivative peak areas in each sample. It is underlined that in the latter case acquiring the metabolomic profiles at the same time does not alleviate the constraint of using the weighted sums. The kinetics of metabolite's j derivatization reactions depend on the original metabolite sample's composition (Fogler, 2002). Therefore, profiles acquired at the same derivatization time,

but shorter than $t^*_{3,j}$ in (at least one of) the samples, might not necessarily correspond to the same derivatization stage for metabolite j.

3.4 ALGORITHM FOR ESTIMATING CUMULATIVE PEAK AREA

At any time equal to or longer than the time required for the complete conversion of a category-3 metabolite M into its derivatives, $t_{M,}$ the sum of the concentrations of all M's derivatives that are at the given time present in the derivatized sample is equal to the M's concentration in the original metabolite mixture, $[M_o]$. Therefore, if N is the number of all category-3 metabolite M's derivatives that are observed throughout the practically relevant derivatization period, the following equation is valid at any derivatization time within this period that is equal to or longer than t_M:

$$[M_0] = [MD_1] + ... + [MD_N] \tag{3-3}$$

where MD_i is the i-th derivative of metabolite M.

If $[IS_o]$, and $[MD_{IS}]$ are, respectively, the concentration of the category-1 internal standard (IS) added in the original sample and the concentration of IS's derivative in the derivatized sample after the completion of IS's derivatization, then, according to the description provided in the previous section regarding the category-1 metabolites, these two concentrations are equal. Therefore, if RF_{MD_i} and $RF_{MD_{IS}}$ are, respectively, the response factors of the i-th derivative of metabolite M and of the single derivative of the

internal standard, Eq. (3-3) is equivalent to:

$$\frac{[M_0]}{[IS_o]} = \frac{[MD_1]}{[MD_{IS}]} + \ldots + \frac{[MD_N]}{[MD_{IS}]}$$

$$= \frac{RF_{MD_{IS}}}{RF_{MD_1}} \cdot \frac{PA_{MD_1}}{PA_{MD_{IS}}} + \ldots + \frac{RF_{MD_{IS}}}{RF_{MD_N}} \cdot \frac{PA_{MD_N}}{PA_{MD_{IS}}}$$

$$= w_1^M \cdot RPA^{MD_1} + \ldots + w_N^M \cdot RPA^{MD_N} \qquad (3\text{-}4)$$

where PA_{MD_i} and $PA_{MD_{IS}}$ are, respectively, the peak area of the i-th derivative of metabolite M and of the internal standard's derivative, measured at the particular derivatization time to which Eq. (3-3) refers; RPA^{MD_i} is the relative (with respect to the peak area of the internal standard) measured peak area corresponding to the i-th derivative of metabolite M at the same derivatization time; w_i^M is the inverse of the relative (with respect to the response factor of the internal standard's derivative) response factor of the i-th derivative of metabolite M.

It is obvious that w_i^M depends only on the selection of the marker ion(s) for the i-th M's derivative and the GC-MS operating conditions. Therefore, considering the same marker ion(s) for each category-3 metabolite M's derivative in all runs and barring change in the GC-MS operating conditions, if the same original metabolite sample is measured at V different derivatization times longer than t_M, the following system of equations holds true:

$$\begin{bmatrix} RPA_{t_1}^{MD_1} & . & . & . & RPA_{t_1}^{MD_N} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ RPA_{t_V}^{MD_1} & . & . & . & RPA_{t_V}^{MD_N} \end{bmatrix} \bullet \begin{bmatrix} w_1^M \\ . \\ . \\ . \\ w_N^M \end{bmatrix} = \begin{bmatrix} \frac{[M_o]}{[IS_o]} \\ . \\ . \\ . \\ \frac{[M_o]}{[IS_o]} \end{bmatrix} \quad (3\text{-}5)$$

where $RPA_{t_j}^{MD_i}$ is the relative (with respect to the peak area of the internal standard) peak area corresponding to the i-th derivative of metabolite M at derivatization time t_j.

Eqs. (3-5) can thereby be used to determine the w_i^M's of a category-3 metabolite's M derivatives at particular GC-MS operating conditions by simple regression analysis. Specifically:

- *only one* representative metabolite sample containing metabolite M should undergo the repetitive measurement process for the w_i^M's estimation. The same w_i^M's could be used afterwards to "correct" the derivative peak area profile of metabolite M in any other biological sample as long as the GC-MS operating conditions remain constant;
- the representative metabolite sample should be run at V different derivatization times longer than t_M throughout a practically relevant derivatization period. The selection of the V derivatization times should be based on the following criteria:
    - the acquired peak area profiles of M should be clearly different from each other to exclude any mathematical artifacts in regression analysis

- V should be by at least one greater than the number of the metabolite M's derivatives that are observed throughout the measured derivatization period to allow for data reconciliation in the regression analysis
- No derivative degradation should have yet occurred;

If M's concentration in the representative metabolite sample, $[M_o]$, is not known, any constant C could be, in theory, used in Eq. (3-5) instead; in metabolomic analysis it is the relative change between profiles due to a particular perturbation that matters. To avoid mathematical artifacts, C should be of the same order of magnitude as the largest relative peak area observed for any metabolite M's derivative in any of the samples measured at the particular GC-MS operating conditions. If a certain constant C is used in the regression analysis instead of the actual concentration $[M_o]$, the estimated weights, w_i^M, do not represent the exact inverse of the relative response factors of metabolite's M derivatives, but a certain proportionality ratio between the relative concentrations of metabolite's M derivatives and their measured relative peak areas.

The estimated w_i^M values can subsequently be used to determine the "cumulative" relative peak area of metabolite M in any other metabolomic profile, as long as the GC-MS operating conditions (and the selected marker ions of the metabolite's M derivatives) remain constant, based on the following equation:

$$RPA_{s_a}^{M} = \sum_{i=1}^{N} w_i^M \cdot RPA_{s_a}^{MD_i} \qquad (3\text{-}6)$$

where $RPA_{s_a}^{M}$ and $RPA_{s_a}^{MD_i}$ represent, respectively, the "cumulative" relative peak area of

metabolite M and the measured relative peak area of the i-th derivative of metabolite M in sample S_a.

The last two sections discussed the potential sources of derivatization biases in the metabolomic profile along with suggestions of how to account for them and appropriately "correct" the metabolomic profile in the context of a single metabolite or a specific category of metabolites at most. In the metabolomics context, all these suggestions should be combined into a systematic data correction/normalization strategy that does not jeopardize the high-throughput nature of the analysis.

3.5 DATA CORRECTION STRATAGY

Let us consider a biological sample comprising P, Q and R, respectively, metabolites in each of the categories 1, 2, and 3 as these were previously described. Then, according to the previous sections, the derivative peak area and the original metabolite concentration profiles would be in *one-to-one* directly proportional relationship with the proportionality ratio depending only on the GC-MS operating conditions, if (a) the metabolomic profile were acquired at derivatization time equal to or longer than T, the latter being defined as follows:

$$T = \max \left\{ T_1^*, T_2^*, T_3^* \right\} \tag{3-7}$$

where $T_1^* = \max_{i=1,2,..,P} \left\{ t_{1,i}^* \right\}$; $T_2^* = \max_{j=1,2,..,Q} \left\{ t_{2,j}^* \right\}$; $T_3^* = \max_{l=1,2,..,R} \left\{ t_{3,l}^* \right\}$ (3-8)

and $t_{1,i}^*$, $t_{2,j}^*$, $t_{3,l}^*$ depict, respectively, the time required for the completion of the

derivatization of the i-th metabolite in category 1, of the j-th metabolite in category 2 and of the l-th metabolite in category 3; and (b) only one of the two derivative peak areas of the category-2 metabolites were considered.

In theory then, T would have been the optimal derivatization time for high-throughput GC-MS metabolomics analysis. T_3^*, however, might be much longer than 30h. Besides the practical difficulties associated with an experimental protocol of this long duration, derivative degradation might also be taking place at such long derivatization times (unpublished data from our lab; Roessner et al., 2000). On the contrary, T_M, which is defined as the time at which all original metabolites have been completely transformed, is usually in the range of 2-6h.

According to the previous sections, to avoid derivatization biases due to incomplete conversion of the original metabolites, *T_M should be the lower bound for the derivatization time in an optimized metabolomics protocol*. The exact identification of T_M is relatively easy when few compounds of certain molecular categories are measured. This remains, however, a major challenge of quantitative high-throughput GC-MS analysis in general. In high-throughput metabolomics, T_M could be approximately estimated by observing the shape of the measured peak area profiles at various derivatization times for all metabolites. Identification, thereby, of the optimal derivatization time requires preliminary runs of the particular type of biological samples at multiple derivatization times. For 12-13 days old *A. thaliana* liquid cultures, for example, this time was identified to be 6 hours after the addition of MSTFA. Clearly, the identification of the optimal derivatization time should be performed only once at the

stage of the optimization of the experimental protocol for a particular biological system.

If all available derivatized samples are run at derivatization time equal to or longer than the optimal derivatization time T_M, remaining sources of derivatization biases are (a) potential change in the GC-MS operating conditions, and (b) the multiple derivative peak areas of category-3 metabolites. Neither could be avoided by acquiring all metabolomic profiles at the same derivatization time, as explained in the previous sections. Therefore, an automated derivatization system is not the solution. To identify and properly account for these biases, the following steps should be followed:

1. 4 and the measured peak areas of the metabolite's derivatives in the particular metabolomic profile.
2. Finally, the dataset representing the "annotated" part of each acquired metabolomic profile, which should be considered for further statistical analysis to extract biologically relevant conclusions, comprises: (a) the relative peak area of each known category-1 metabolite, (b) one of the two relative peak areas of each known category 2 metabolite; as less susceptible to noise, the largest of the two should be preferably selected, and (c) the "cumulative" peak area of each known category-3 metabolite, as estimated in Step 5.

With respect to the unknown part of the metabolomic profiles, the "molecular origin" of each peak should be determined, so it could be categorized in one of the 3 classes described in this paper. In this case, only the peak areas of category-1 metabolites could be "safely" used in subsequent statistical analysis. Regarding the category-2 metabolites, it is currently impossible to select one of the two peak areas to use in further analysis,

because no algorithm pairing unknown peaks corresponding to the same metabolite has yet been reported. Therefore, derivative peak areas of unknown category-2 metabolites may be included in the metabolomic dataset taking, however, into consideration that any change in these metabolites' concentration will be assigned double weight in any statistical analysis. Obviously, the bias introduced to the analysis by these peaks' inclusion increases monotonically with the number of unknown category-2 metabolites in the biological samples. Finally, regarding the derivative peaks of unknown category-3 metabolites, these could be identified from the change in their area profile with increasing derivatization time and/or from fragment peaks at characteristic m/z values (e.g. 156 or 174), indicating the presence of amine functional groups. Unless, however, an algorithm that can combine these peaks into groups representing the same metabolite is invented, enabling thus the estimation of their "cumulative" peak areas based on the presented normalization strategy, these peak areas should not be used in further statistical analysis. As it was previously discussed in the context of the category-3 metabolite derivatization kinetics and will be proven in the context of real samples, including derivative peak areas corresponding to category-3 metabolites as independent elements in the metabolomic profile vector might lead to highly erroneous statistical results. Assigning biological meaning to the latter could prove quite negative.

3.6 MATERIALS AND METHODS

In order to demonstrate the effectiveness of the algorithm and to estimate the w values for the known category-3 metabolites, a series of experiments were carried as described in

the sections below.

3.6.1 Sample Preparation

Category-3 Metabolite Standards: Vacuum-dried 200μL equal-volume mixture of 1 mg/mL of a particular category-3 metabolite solution in 1:1 (v/v) methanol and water and 1 mg/mL ribitol (as internal standard) solution in water. Table 3-1 includes a list of all category-3 metabolites whose standard sample was prepared and measured. In the case of cysteine, arginine, histidine and tryptophan, ~1mg pure compound was also derivatized directly, without prior treatment with methanol-water solution and subsequent drying.

Standard Metabolite Mix 1: Vacuum-dried 600 μL solution of 27 metabolites (16 amino acids, 4 organic acids, 7 sugar/sugar alcohols) and ribitol (as internal standard) in 1:1 (v/v) methanol and water (for exact composition, see Appendix I, Table A1-1).

Standard Metabolite Mix 2: A mixture of ~1mg from each of the 10 category-3 metabolites flagged with asterisk in Table 3-1.

Plant Samples: Vacuum-dried polar extracts (see extraction protocol in Roessner et al. (2000); Kanani (2004)) from ~125 mg of ground *Arabidopsis thaliana* liquid cultures. The cultures were grown in 200 mL of Gamborg media (Gamborg et al., 1976) with 20 g/L sucrose under constant light (80-100 μmole/m^2.s) and temperature (23°C) in the controlled environment of an EGC M-40 growth chamber. Two cultures were used in present analysis; plant sample 1 was 12 days and 9 hour old, while plant sample 2 was 13 days and 6 hours old. All reagents were procured from Sigma.

Table 3-1: All observed TMS-derivatives of 26 metabolites containing (-NH2) groups (including all 20 protein amino acids) in the metabolomic profiles of the real plant, pure metabolite and standard metabolite mix samples that were acquired at derivatization times spanning a period from 6h to 30h.

Amino Acid	Derivative 1	Derivative 2	Derivative 3
Alanine	Alanine N O	Alanine N N O	
Arginine**#	Ornithine N_2 N_5 N_5 O	Ornithine N_2 N_2 N_5 O [2]	Ornithine N_2 N_2 N_5 N_5 O
Asparagine	Asparagine N N O	Asparagine N N N O	Asparagine N N N N O[2,3] (putative)
Aspartate	Aspartate O O[2,3]	Aspartate N O O	
Cysteine*	Cysteine N O[2]	Cysteine N S O	Cysteine N N O
Glutamate	Glutamate N O O	Pyroglutamate N O[1]	
Glutamine	Glutamine N N O	Glutamine N N N O	Pyroglutamine N N O[1,2,3] (putative)
Glycine	Glycine N O	Glycine N N O	
Histidine*	Histidine O[2] (putative)	Histidine N O	Histidine N N O
iso-Leucine	iso-Leucine O	iso-Leucine N O	iso-Leucine N N O[2]
Lysine	Lysine N N O	Lysine N N N O	Lysine N N N N O[2] (putative)
Leucine	Leucine O	Leucine N O	Leucine N N O[2]
Methionine	Methionine N O	Methionine N N O[2]	
Proline	Proline N O		
Phenylalanine*	Phenylalanine O	Phenylalanine N O	
Serine	Serine O O	Serine N O O	Serine NNOO[2]
Threonine	Threonine O O	Threonine N O O	Threonine N N O O[2]
Tryptophan*	Tryptophan O[2] (putative)	Tryptophan N O	Tryptophan N N O
Tyrosine*	Tyrosine O[2] (putative)	Tyrosine O O	Tyrosine N O O
Valine	Valine O	Valine N O	Valine N N O [2,3]
Allantoin	Allantoin N N N	Allantoin N N N N	Allantoin N N N N N
β –Alanine*	B –Alanine O	β –Alanine N O	β -Alanine N N O
GABA*	GABA N O	GABA N N O	
Dopamine*	Dopamine N O O	Dopamine N N O O	
Homoserine*	Homoserine OO	Homoserine N OO	Homoserine N N O O
Ornithine*	Ornithine N_2 N_5 N_5 O	Ornithine N_2 N_2 N_5 O [2]	Ornithine N_2 N_2 N_5 N_5 O

[1] derivative forms produced by chemical transformation of one of the original metabolite's TMS derivatives

[2] derivative forms not yet reported in currently available major public MS libraries (i.e. MPL, CSB.DB, NIST)

[3] derivative forms matching reported peaks which have been currently assigned an unknown status in MPL:

- Asparagine Derivative 3 matched Potato Tuber 015 in MPL
- Valine Derivative 3 matched Potato Tuber 002 in MPL
- Glutamine Derivative 3 matched Tomato leaf 011 and Potato Tuber 007 in MPL
- Aspartate Derivative 1 matched Phloem *C. Max* 020 and Potato leaf 003 in MPL
- Threonine Derivative 3 matched Phloem *C. max* 028 in MPL

* Metabolites included in the Stdandard Metabolite Mix 2 # Arginine is converted to Ornithine

3.6.2 GC-MS Runs

Multiple replicates of the plant, standard metabolite mixes and pure metabolite samples were derivatized as described in (Roessner et al., 2000; Kanani, 2004) and run at various derivatization times over a period from 6 and 30 hours after addition of MSTFA, at 1:35 split ratio, using Varian 2100 GC-(ion-trap) MS fitted with 8400 auto-sampler at operating conditions described in (Kanani, 2004). All reagents were procured from Sigma.

- In the case of the plant and metabolite mix 1 samples, 100 μL of 20 mg/mL Methoxyamine HCL solution in pyridine was added to each sample and allowed to react for 90 mins followed by the addition of 100 μL MSTFA.
- In the case of pure metabolite samples, 30 instead of 100 μL MSTFA were used, balanced out by 70 μL of pyridine.
- In the case of the cysteine, arginine, histidine, tryptophan and metabolite mix 2 samples that were prepared without the addition of methanol-water solution and the subsequent drying, 100 μL of 2 μg/μL ribitol solution in pyridine and 300 μL of pyridine were initially added to each sample. The sample reacted with 100 μL of 20 mg/mL Methoxyamine HCL solution in pyridine for 30 mins followed by the addition of 500 μL MSTFA.

The exact derivatization times at which each sample was run along with the data acquired from each of the metabolomic profiles that were used to estimate the w_i^M weight values shown in Table 3-2 can be found in Appendix I, Table A1-2. Appendix I, Table A1-3

includes the retention times of the observed metabolite derivatives (see Table 3-1) at the given derivatization and GC-MS operating conditions. In Table 3-2, the weight values of:

- The metabolites 3,6-7,16-17,25 were based on the profiles of plant sample 1.
- The metabolites 1,8,10,12-13,20 were based on the metabolomic profiles of standard metabolite mix 1.
- The metabolites 2,5,14,18-19,22,24,26 were based on the metabolomic profiles of standard metabolite mix 2.
- The metabolites 4,21 were based on pure metabolite standards.

3.6.3 Data acquisition and analysis

Metabolite peak identification was based on (a) own library of standards, (b) MPL and the Public Repository for Metabolomic Mass Spectra - CSB.DB GOLM Metabolome database (Kopka et al., 2005) (in the rest of the text it will be referred to as CSB.DB) and (c) the commercially available NIST MS-library (Ausloos et al., 1999).

3.7 RESULTS-DISCUSSION

Similarly to any other high-throughput biomolecular profiling analysis to-date, GC-MS metabolomic profiling has been mainly used either to differentiate between various cellular states (Broeckling et al., 2005; Cook et al., 2004; Kanani, 2004; Noguchi et al., 2003; Sakai et al., 2004) and/or identify an environmental or genetic phenotype (Fiehn et al., 2000; Taylor et al., 2002; Weckwerth et al., 2004). When the objective is only the

Table 3-2: Estimated W_i^M weight values of the category-3 metabolite derivatives in the order shown in Table 3-1 for a particular set of GC-MS operating conditions and the indicated marker ion(s) (m/z); all utilized data are included in Appendix I Table A1-2. n/d: not consistently detected among the samples utilized for the estimation of the weight values.

	Amino Acid	Derivative 1		Derivative 2		Derivative 3	
	(M)	m/z	W_1^M	m/z	W_2^M	m/z	W_3^M
1	***Alanine***	116.0	**1.025**	188.2	**0.774**		
2	***Arginine***	142.2	**1.10**	174.2	**0.48**	257.2	**n/d**
3	***Asparagine***	231.3+258.0	**0.726**	188	**1.904**	405.3	**1.595**
4	***Aspartate***	160.0	**3.824**	232.2	**0.224**		
5	***Cysteine***	148.1+218.1	**n/d**	148.0	**12.67**	220.2	**0.37**
6	***Glutamate***	246.5	**1.014**	230.2 + 156.1	**0.988**		
7	***Glutamine***	156.1	**0.667**	227.3+317.2	**10.3**	155.1+301.3 + 344.3+227.0	**9.00**
8	***Glycine***	102.0	**9.397**	174.2	**0.773**		
9	***Histidine***	154.2 + 110.1	**n/d**	154.2 + 182.1	**n/d**	154.3 + 254.1	1.0
10	***iso-Leucine***	86.0	**2.55**	158.2	**0.92**	230.1	**n/d**
11	***Leucine***	170.0	**n/d**	158.2	**1.0**	230.1	**n/d**
12	***Lysine***	362.2+230.0	**n/d**	174	**1.005**	389.5+463.5	**2.124**
13	***Methionine***	176.1	**1.42**	248.3	**0.369**		
14	***Phenylalanine***	146	**1.30**	218.0	**0.48**		
15	***Proline***	142.1	**1.0**				
16	***Serine***	116.0	**2.97**	204.3	**0.299**	290	**7.87**
17	***Threonine***	219.0+130.0	**3.30**	292.3+218.3	**0.321**	290.2	**33.5**
18	***Tryptophan***	130.1	**n/d**	202.1	**1.0**	202.1	**n/d**
19	***Tyrosine***	179.1+268.1	**1.18**	179.1	**0.94**	218.1	**0.26**
20	***Valine***	72.0	**1.638**	218.0	**0.842**	188.0+216.0+ 172.0	**n/d**
21	***Allantoin***	374.2+259.2	**25.3**	331.3+431.2+ 446.2	**0.530**	518.5+428.4+ 188.3	**2.12**
22	***B Alanine***	117	**8.88**	102.1	**n/d**	248.3	**0.80**
23	***GABA***	102.1	**n/d**	174.2	**1.0**		
24	***Dopamine***	102.1	**4.16**	174.2	**0.73**		
25	***Homoserine***	146.1	**6.51**	218.3	**0.231**	290.3	**2.67**
26	***Ornithine***	142.2	**1.10**	174.2	**0.48**	257.2	**n/d**

former, profiles are compared as a whole with no particular interest in peak identity. In this case, based on the to-date published studies, it seems that each peak area in the metabolomic profiles has been typically considered independent of the others in multivariate statistical analysis, including peaks corresponding to derivatives of the same metabolite (Broeckling et al., 2005; Cook et al., 2004; Fiehn et al., 2000a; Noguchi et al., 2003; Sakai et al., 2004; Taylor et al., 2002; Weckwerth et al., 2004). When the objective is (also) the latter, peak identity is of interest.

Based on the reported results, it seems that in this case one of its derivative peak areas, (usually the largest) if multiple exist, has been typically used to represent a particular metabolite's concentration and derive biologically relevant conclusions (Broeckling et al., 2005; Cook et al., 2004; Fiehn et al., 2000a; Kanani, 2004; Noguchi et al., 2003; Sakai et al., 2004; Taylor et al., 2002; Weckwerth et al., 2004). Based, however, on earlier discussion in this report, regarding molecular categorization with respect to the derivatization reaction kinetics, both these practices could lead to erroneous conclusions in the context of GC-MS metabolomics.

To identify the extent of the bias introduced in the statistical analysis of category-3 metabolites if one of its derivative peak areas is selected to represent the metabolite's concentration in the original sample, and to validate the proposed normalization/correction strategy, multiple spectra of pure category-3 metabolite samples, of synthetic metabolite mixtures and two real plant samples were analyzed. Specifically, one of the two plant samples, plant sample 1, the standard metabolite mixes and pure category-3 metabolite standards underwent the repetitive measurement process

described in Step 3 of the presented strategy. Metabolomic profiles were acquired at different derivatization times spanning a period from 6h to 30h. Table 3-1 includes the derivatives of 26 category-3 metabolites (including all 20 protein amino acids) that were observed during this derivatization period, the latter being practically relevant in the presence of an auto-sampler and in lack of an automated derivatization system. Table 3-2 includes the weight values w_i^M for each of the 26 metabolites' derivatives as these were estimated according to the 4th step of the proposed strategy, based on the acquired peak area information listed in Appendix 1, Table A1-2 for the particular GC-MS operating conditions and the selected marker ions. Subsequently, as indicated in Step 5, the estimated weight coefficients were used in Eq. (3-6) to calculate the cumulative peak area of each known category-3 metabolite observed in the metabolomic profiles of plant sample 1 (direct result of the regression analysis) and plant sample 2, the relative peak areas from the latter being also listed in Appendix 1, Table A1-2. Table 3-3 shows the average peak area value for each measured category-3 metabolite derivative and the average estimated cumulative peak area for each metabolite along with their coefficient. After studying Tables 3-2, 3-3 and Appendix 1, Table A1-2, there are two significant observations that can be made regarding the values of the weight coefficients w_i^M:

- They varied in a range of two orders of magnitude, from ~0.1 to ~10;
- The largest weight coefficient did not always correspond to the largest derivative peak area of a particular metabolite.

Table 3-3 Measured relative (with respect to the internal standard ribitol) derivative and estimated cumulative peak areas of all category-3 metabolites listed in Table 3-2 that were observed in plant samples 1 and 2, averaged over metabolomic profiles acquired throughout the depicted derivatization periods (all utilized data are included in Appendix 1, Table A1-2). The cumulative peak areas were estimated using the values of the derivatives' weight coefficients shown in Table 3-2.

		Plant Sample 1 Derivatization Period (6-23 h)		***Plant Sample 2*** Derivatization Period (8-21 h)	
		Average RPA	Coefficient of Variation	Average RPA	Coefficient of Variation
Glutamate	Glutamate 3 TMS	0.2092	16%	0.2580	12%
	Pyroglutamate 2 TMS	0.2729	12%	0.3098	7%
	Cumulative	**0.4818**	**2%**	**0.5677**	**2%**
Asparagine	Asparagine 3 TMS	0.0206	26%	0.0271	20%
	Asparagine 4 TMS	0.0080	6%	0.0078	10%
	Asparagine 5 TMS (putative)	0.0041	47%	0.0031	39%
	Cumulative	**0.0368**	**1%**	**0.0394**	**4%**
Glutamine	Glutamine 3 TMS	0.0594	78%	0.1421	34%
	Glutamine 4 TMS	5.73E-04	137%	0.0035	58%
	Pyroglutamine 3 TMS (putative)	0.0106	40%	0.0082	49%
	Cumulative	**0.1410**	**2%**	**0.2047**	**10%**
Serine	Serine 2 TMS	0.0057	19%	0.0140	9%
	Serine 3 TMS	0.0170	17%	0.0169	20%
	Serine 4 TMS	8.93E-04	51%	9.74E-04	42%
	Cumulative	**0.0290**	**5%**	**0.0544**	**5%**
Threonine	Threonine 2 TMS	0.0049	14%	0.0080	7%
	Threonine 3 TMS	0.0168	11%	0.0133	15%
	Threonine 4 TMS	1.06E-04	64%	9.03E-05	53%
	Cumulative	**0.0250**	**4%**	**0.0338**	**3%**
Homoserine	Homoserine 2 TMS	3.78E-04	20%	5.47E-04	18%
	Homoserine 3 TMS	0.0025	22%	0.0013	34%
	Homoserine 4 TMS	3.13E-04	65%	2.09E-04	80%
	Cumulative	**0.0039**	**6%**	**0.0044**	**7%**

These observations further indicate that:

- Even a small category-3 derivative peak area could significantly contribute to the cumulative peak area and thereby should not be ignored, as it seems to be the current practice.
- Significant bias might be introduced in the analysis, if only one (often the largest) derivative peak area is selected to represent the metabolite of interest.

A measure of the latter bias can be provided from the difference between the coefficient of variation of the average individual derivative peak areas and the average cumulative peak area of the listed in Table 3-3 category-3 metabolites.

As discussed earlier in this chapter that only the cumulative peak area is directly proportional to the concentration of the metabolite in the original sample. The results in Table 3-3 validate this argument, considering that all acquired metabolomic profiles of plant samples 1 and 2 refer to the same biological sample measured at different derivatization times equal to or longer than the optimal derivatization time, as defined in the previous section. Therefore, barring changes in the GC-MS operating conditions, the peak area that is directly proportional to the concentration of a particular metabolite in the original sample should be of the same value in all profiles. Table 3-3 indicates that this is indeed true for the cumulative peak areas of the metabolites, their coefficient of variation being in average 3% and 5% in plant samples 1 and 2, respectively. On the other hand, the coefficient of variation of the individual derivative peak areas was in average 38% and 30%, respectively, the largest reaching a value of 137%.

This is a significant result indeed. First it validates the proposed methodology. Second it is the first time that the extent of the bias that could be introduced in the statistical analysis of the metabolomic data when, either the derivative peak areas of category-3 metabolites are all used as independent in the analyzed dataset or only one of them is selected as the representative of the concentration of the metabolite in the original sample, has been quantified.

Variation due only to the molecular characteristics of these metabolites and the mechanisms of the derivatization reaction could be attributed biological significance. Only the estimation of the cumulative peak area enables the accurate quantification of the change in a category-3 metabolite's concentration among various biological samples. If individual derivative peak areas of category-3 metabolites are compared, each may lead to a different value, and, even worse, different direction of change (see Table 3-3).

A significant "by-product" of the mass spectral analysis carried out for the validation of the proposed correction strategy was the identification of 15 derivatives of category-3 metabolites, which either had not been reported before in public databases (NIST, MPL, CSB.DB) (10), or matched reported peaks which have been currently assigned an unknown status in MPL (5) (See Table 3-1). This identification was made possible through the analysis of spectra of pure metabolite samples. One of the currently reported unknown peaks was identified as a chemical transformation derivative of glutamine-4-TMS. Moreover, pyroglutamate-2-TMS was validated to be a chemical transformation derivative of glutamate-3-TMS, as reported in (Gehrke et al., 1969; Leimer et al., 1977). This is underlined, because many recent studies have, however, treated this derivative as

independent of glutamate (Fiehn et al., 2000a; Roessner et al., 2000, Broeckling et al., 2005). Moreover, the previously reported (Halket, 1993; Halket et al., 2005) chemical transformation of arginine to ornithine derivatives in the presence of a silylating agent was also validated. The significance of these discoveries is quite high, considering that:

- A lot of effort in metabolomics is invested in the annotation of unknown peaks (Fiehn et al., 2000b; Halket et al., 2005),
- To-date statistical analyses might be biased due to dependency between peaks currently considered as independent in the metabolomic profiles' dataset,
- Variation in cumulative peak areas of known compounds with derivatization time, implies the presence of additional, still unidentified, derivative(s),
- Variation in unknown peak areas with derivatization time might provide clues regarding the chemical formula of the corresponding metabolite.

This data correction algorithm was used for ensuring accuracy and reproducibility for acquiring metabolomic profiles of systematically perturbed *A. thaliana* liquid culture network as described in the following chapter.

4 EXPERIMENTAL DESIGN APPROACH

In the traditional biological experiments to study a phenomenon, the key component of the experimental design is the initial hypothesis being proved or disproved by the series of experiments. The ability of high-throughput measurement techniques, to simultaneously measure all variables belonging to a class (in case of metabolomic analysis – metabolites), instead allow an experimental design approach which is based on comparatively a "limited" hypothesis such as: A biological system will respond at metabolic level in response to the given perturbation. In some sense the design of experiment is based on *"systematic discovery"* rather than a hypothesized invention. However with this freedom of not being restricted to a focused hypothesis answering a specific question also comes several challenges which have to be taken into account while designing experiments involving high-throughput analysis. Some of the important common challenges are listed below:

- The experiments should be designed in such a way that the observed changes in the molecular profiles should be attributed only to the applied perturbations
- To ensure enough experimental controls, biological and instrumental replicates without significantly compromising on the number of biological states that can be studied.
- Ensuring internal standards and appropriate data normalization and validation

techniques to monitor and remove the effect of biases introduced due to experimental and instrumental limitations.

These requirements are even more necessary when using novel high-throughput techniques like metabolomic analysis which is still in its early days of standardization. Keeping these criteria in mind a series of experiments were designed to achieve the specific objectives of the current study. The rational for the various design decisions and the experimental design set up are described in the following sections of this chapter.

4.1 MODEL SYSTEM SELECTION

Always, the selection of the system depends on the specific aims of the study. The specific aims of the current study were:

- To develop experimental and data analysis methods for using time-series metabolomic analysis to study complex biological systems where the ability to use metabolic flux analysis to study regulation is limited.

- To understand regulation of primary metabolism of plants by metabolomic analysis of systematically perturbed system.

The choice of plants was also governed by the fact that plants are complex eukaryotic organisms. Hence besides any functional insight that might be gained through the study, it is anticipated that analysis of multi-tissue organisms will also contribute in the development of systems biology principles that will have broader applicability than

studies in yeast or prokaryotic systems.

In this context, *Arabidopsis thaliana* (Columbia Ecotype) liquid cultures were chosen as the model system based on following reasons:

- *A. thaliana* is considered the model system of plant physiology, because of its short growth cycle and a small genome of 5 chromosomes,
- *A. thaliana* was the first plant genome to be fully sequenced.
- Liquid compared to soil cultures provide a controllable growth environment, ensuring that all plants in all experiments receive the same nutrients.
- In liquid cultures, as part of a perturbation any signalling molecule/growth hormone added to the media is uniformly distributed in a very short time, ensuring equal treatment received by all plants.
- Each liquid culture comprises of 50 - 80 *A. thaliana* seedlings. This provides a large biological population for each biological state, thereby reducing the effect of biological variability increasing, thereby, the confidence in the statistical significance of the acquired measurements. In the case of time-series analysis, this further helps in partially overcoming the trade off between number of replicates and number of timepoints (see further discussion in following section).
- Metabolism of plants is known to change rapidly and hence it is necessary to stop the metabolism by freezing the plants in liquid nitrogen. Since each liquid culture

comprises 50-80 plant samples, a large number of replicates are harvested in a very short time.

4.2 SELECTION OF PERTURBATIONS

In order to achieve the specific aims of the current analysis, the particular environmental perturbations were chosen based on the following criteria:

- Plants are expected to show a significant response to the perturbation at metabolic level
- The metabolic response of the perturbation can be obtained even in short period of time
- The response is expected to be in primary metabolism of plant
- Understanding the response of the system to the perturbation, in addition to providing important information for engineering desired traits in plants, also provides an example of the use of metabolomic analysis for studying a variety of commonly studied biological problems.

Based on these criteria perturbation following perturbations were chosen:

- Elevated CO_2 (in sucrose and glucose grown plants)
- Salt (NaCl) Stress
- Sugar (Trehalose) Signaling

- Hormone (Ethylene) Signaling

In order to develop the methodology for identifying interaction between different environmental stress, experiments were also conducted with the last three perturbation applied in combination to CO_2 stress. The motivation for the choice of each perturbation is described below.

4.2.1 Elevated CO_2

Increase in CO_2 was chosen as the common perturbation under all conditions, because:

- CO_2 being the primary carbon source of plants, plants are expected to respond to changes in the CO_2 level even for a short period of time.
- Improved CO_2 fixation can significantly increase the yield of industrial and agricultural output achieved from plants, thus any information gained would be extremely valuable
- Most of the prior studies of response of plants to elevated CO_2 have focused on long-term effects at a physiological level, hence very little is known about the short-term response which indicate metabolic pathways regulated by CO_2.
- CO_2 is expected to perturb the central carbon metabolism and amino acid biosynthesis networks. These networks have been well studied in plants and extensive information about their function exist both at the metabolic and genomic level. Further the majority of the involved metabolic pathways have been

well characterized in plants, while the regulation of these pathways has been extensively investigated at least in prokaryotic systems.

4.2.2 Salt (NaCl) Stress

All living organisms (bacteria (Verala et al., 2003), plants (Taiz, 2002)) are known to respond to salt and osmotic stress by changing concentration of osmoprotectants at the metabolic level. Thus measuring the response of plants just after giving the osmotic stress using metabolic profiling is expected to allow identification of metabolic mechanisms used by plants to protect against osmotic stress. Previous studies with NaCl stress to *A. thaliana* (Essah et al., 2003) adult plants could not sustain 50 mM NaCl stress for more than 4 days, suggesting a significant perturbation for the plant system. Higher concentration of 250 mM has also shown a very strong immediate response (few hours) at transcriptional level (Taii et al., 2004) and the plant was not able to survive for more than 2 days. Hence, 50 mM NaCl stress was chosen for *A. thaliana* plants to ensure a significant response to salt stress only.

4.2.3 Sugar (Trehalose) Signal

Sugar signaling regulates important physiological processes in plants like photosynthesis and nitrogen assimilation and is currently an important area of study in plants (Moore et al., 2003). Trehalose is a disaccharide like sucrose made from two glucose molecules. Trehalose is known to be an important osmoprotectant in many biological systems (Wingler, 2002). In yeast it is also known to regulate glycolysis flux. However, when genes encoding trehalose were first discovered in plants a decade ago, even their

presence was a surprise (Goddijen and Smiklens, 1998). Recent studies have indicated trehalose regulates starch redistribution between source tissues and sink tissues. In addition it is also known to cause starch accumulation in some cases. However detailed mechanism and the exact role of trehalose in regulation of starch accumulation is still not clear. Starch being an important product of photosynthesis, any new information is likely to provide important information to improve carbon fixation in engineered plants. This makes it an interesting perturbation for the current study independently and especially in combination with elevated CO_2 stress. 12 mM trehalose was chosen as in past experiments (Wingler et al., 2000) 10 mM trehalose present in the media (containing 50 mM sucrose) caused redistribution of starch (Wingler et al., 2000).

4.2.4 Plant Hormone (Ethylene) Signal

Hormones play an important regulatory role in plant physiology. Ethylene is an important plant hormone, and controls various plant physiological and developmental parameters like seed germination, seedling elongation, ripening, wounding, stress and defense response (Taiz, 2002). Recently, gene expression analysis of *A. thaliana* response to ethylene has identified as many as 245 genes whose expression changes considerably during just the first six hours after ethylene addition (De Paepe et al., 2004). Most of the study involving ethylene has either been at the gene expression level or physiological level, hence very little is known about the regulation of metabolic pathways by ethylene. Moreover studying the response of the plants to combined ethylene and elevated CO_2 treatment will enable conclusions to be derived about ethylene signaling and

photosynthesis.

4.3 TIME-SERIES VS SNAPSHOT ANALYSIS

Majority of high-throughput experimental designs can be classified into two categories:

- Comparison of Snapshots or two biological states with and without treatment/perturbation
- Comparison of Time-profiles, which compares a number of biological states at different times with and without the perturbation

The main advantages of snap-shot experimental design are:

- Less number of biological states are studied which allows more biological replicates for each stage. This allows a more statistically valid “snap-shot” of the biological system
- The data analysis methods for analysis of snap-shot experiments are well developed

The main advantages of using time-series approach in experiment design are:

- Most biological responses are inherently dynamic in nature which follow a signaling cascade at molecular level which can only be observed using time-series analysis.
- Response of biological systems to perturbations may be different at different time

scales, hence by comparing only two snap-shots, the complete understanding of pathways being perturbed would not be achieved.

- Time series also allows distinction between transient and sustained response which is not possible from snap-shot analysis.

- In case of metabolomic profiles, time-series metabolomic profile contain within them information about the net result of changes in metabolic fluxes which are directly related to enzymatic activity and hence closer to gene expression level. Thus time-series metabolomic analysis can allow identification of common regulatory networks and integration of profiles with transcriptomic analysis which is limited for snap-shot comparisons between only two biological states.

Thus in order to achieve the specific aims of the project, time-series analysis is preferred over snap-shot analysis.

Experiment design of time-series metabolomic analysis has to address following two issues:

Number of Timepoints Vs. Number of Bio-replicates: 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.

Previous time series metabolic profiling analysis to measure response of *A. thaliana* to elevated CO_2 (Kanani, 2004) following observations about time series experiment design were made which were incorporated in the current analysis:

- The harvesting times should be uniformly distributed during the day
- Large changes were observed from 12 to 24 hours, so intermediate time point and additional time point beyond 24 hours should be added.

Based on theses criterion, and availability of 20 liquid cultures for each experiment the number of bio-replicates and time points in the current experiment were chosen as follows:

- In each experiment, perturbations were applied in 12 day old *A. thaliana* seedlings which were in their young-adult (vegetative) state.
- In order to study pathways regulated by the perturbations, short term (few hours) response was chosen over long term (few days) response.

- In order to account for variability during their initial 12 days of growth, four liquid cultures were harvested as the 13 day 0h reference state in each experiment and were harvested before application of perturbations.

- The remaining 16 liquid-cultures were used to represent eight different time points with two bio-replicate cultures (each containing 60-80 seedlings ensuring a large number of plants thus avoiding bias due to a biological outlier/mutant).

- Immediately after the perturbations, the sampling frequency was kept higher as more change in expected immediately after the response.

- As per recommendations based on the previous experiment, a time point was added between 12 and 24h at 18h, and an additional time point was added beyond 24h at 30h.

- The final harvesting schedule chosen for the experiment were: 4 cultures at 13 day 0h serve as reference state in each experiment followed by 2 liquid cultures at 1h, 3h, 6h, 9h, 12h, 18h, 24h and 30h. The same harvesting protocol was followed for all the experiments.

Based on the criteria described from above a series of experiments were conducted as described in the following section.

4.4 PLANT GROWTH EXPERIMENTS

4.4.1 Liquid Cultures

Arabidopsis thaliana (Columbia Eco-type) liquid cultures were grown in 500 mL conical shake flasks. Each flask contained 200 mL autoclaved solution of B5 Gamborg media with minimal organics (Sigma, St. Louis) and 4g of sucrose (2% or 58.5 mM). The pH of the solution was adjusted to 5.7 and each flask opening was closed with Identi-Plug Plastic Foam Plugs (VWR) and aluminum foil prior to autoclaving the solution. 1 mL of 0.1% agar solution containing ~100 seeds of *A. thaliana* Columbia ecotype was added to each flask in bio-safety cabinet. The seeds were cleaned and sterilized as described in (Liu et al., 2005) and stored overnight at $4^{o}C$ prior to inoculation.

4.4.2 Experimental Setup

Each experimental setup comprised of 20 *A. thaliana* liquid cultures, which were grown on an orbital shaker platform at 150 rpm. The shaker was kept in a controlled environmental growth chamber (EGC Inc, Chagrin Falls, OH, USA, model M-40) at the University of Maryland, New Green House facility. The chamber allowed measurement and control of light, humidity, temperature and CO_2. The CO_2 level inside the growth chamber was measured using WMA-4 CO_2 Analyzer (PP Systems, Amesbury, MA).

4.4.3 Experiment Description

The experiment involved the growth of eight sets of 20 *A. thaliana* (Columbia ecotype) liquid cultures as shown in Figure 4-1. All sets were grown for 12 days in the controlled environment growth chambers under constant light intensity (80 - 100 $\mu E\ m^{-2}\ s^{-2}$), 60% relative humidity and $23^{o}C$ temperature.

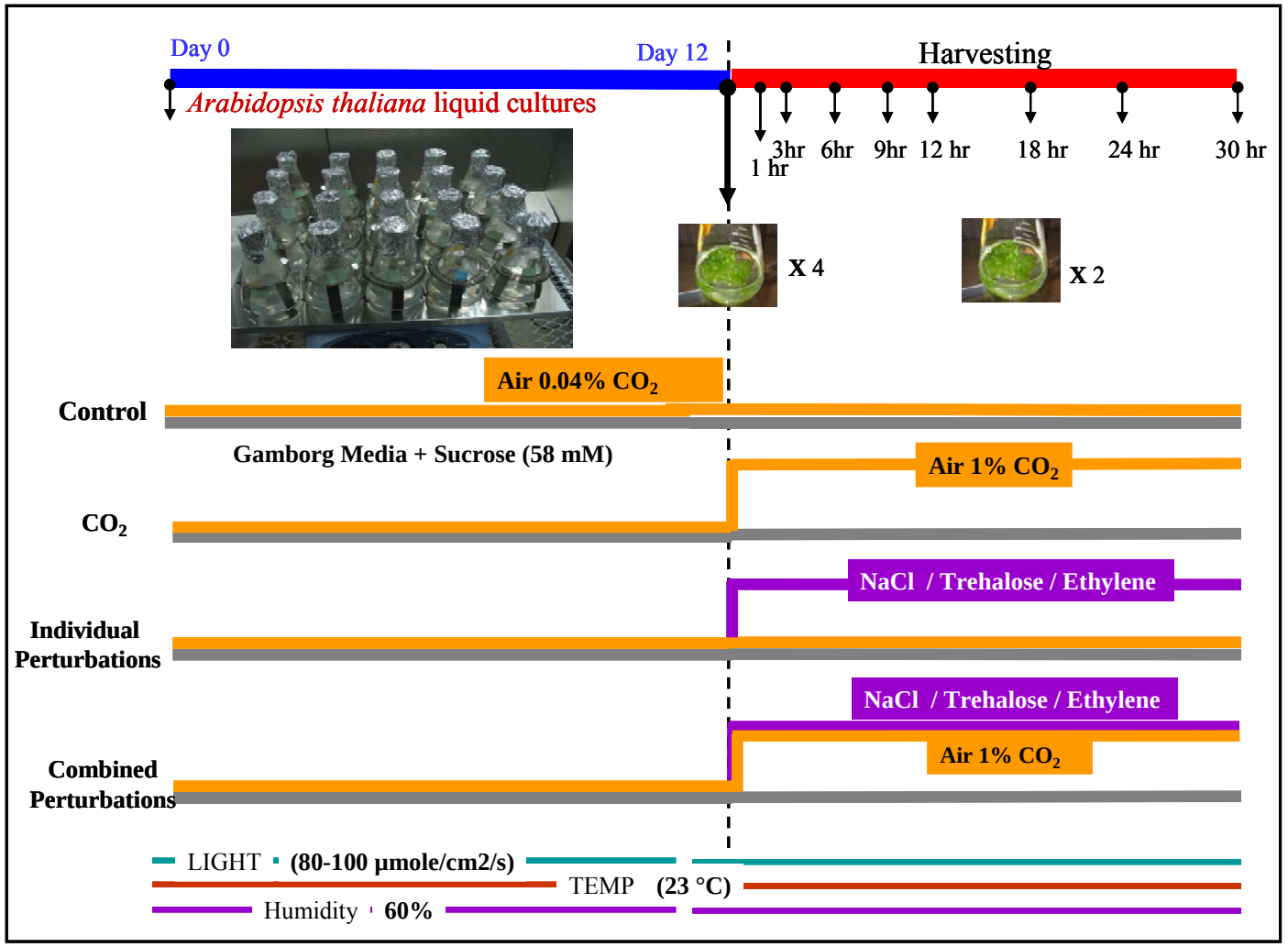


Figure 4-1: Graphical representation of the experimental conditions of the eight experiments performed as part of the project.

On the 13th day, 4 liquid cultures were harvested from both the control and perturbed sets. They were used as "reference markers" of the plant growth up to that stage (0h). After harvesting the reference marker liquid cultures, different conditions were maintained as shown in Figure 4-1 in eight different experiments as described below:

Experiment 1: Control experiment, Concentration of CO_2 was maintained at ambient concentration and media composition was also kept the same.

Experiment 2: Elevated CO_2 Stress - Concentration of CO_2 was increased to 10,000 ppm from ambient concentration. (Increase was achieved in less than 5 minutes).

Experiment 3: Salt Stress - 10 ml of 1 M NaCl solution was added to the media to achieve 50 mM final concentration.

Experiment 4: Trehalose Signal - 10 ml of 240 mM trehalose solution was added to the media to achieve 12 mM final concentration.

Experiment 5: Ethylene Signal - 1 ml of 2 mM 1-Aminocyclopropanecarboxylic acid (ACC – ethylene precursor) was added to the media to achieve 0.01 mM final concentration.

Experiment 6-8: Combined Stress Experiments, perturbations from experiment 3-5 were applied and simultaneously the CO_2 concentration was increased to 10,000 ppm of the air composition.

After applying the perturbations, 2 liquid cultures were harvested at each of the 1h, 3h, 6h, 9h, 12h, 18h, 24h and 30h time points in each experiment. Each harvested plant

culture was weighed, washed with de-ionized water to remove media, excess water removed with filter paper, wrapped in aluminum foil, frozen in liquid nitrogen and stored at –80^{o}C for further analysis. Each liquid culture spent about 2-5 minutes between removal from the growth chamber and freezing in liquid nitrogen.

4.5 GC-MS METABOLOMICS CONDITIONS

4.5.1 Grinding

Each of the frozen whole plant cultures was ground under liquid Nitrogen to a state of paste and then used to extract polar metabolites for the metabolomic profiling analysis. (The same ground plant was also used for a parallel transcriptomic analysis of the sample). Since no specific part of the plant was selected, the obtained metabolic profiles represent the average over the entire plant state of each liquid culture containing 50-80 *A. thaliana* seedlings.

4.5.2 Metabolite Extraction

Polar metabolite extracts were obtained from approximately 0.5 gm of the frozen liquid culture using methanol and water. The extraction protocol followed in this study has been developed by (Roessner et al., 2000; Kanani, 2004) which were subsequently optimized for optimum solvent volume for the current plant sample. Methanol (28 g/g plant), water (28 g/g plant) and ribitol (20 mg/gm of plant) were used, respectively, as the solvent to collectively extract the polar small molecules and the common internal standard for all classes of the observed metabolites. The methanol water extracts obtained from each

plant sample were divided into four equal aliquots which were vacuum-dried overnight at room temperature using SAVANT rotary vacuum drier and stored in -20 C freezer.

4.5.3 Derivatization of Polar Extracts

The dried plant extracts were subsequently derivatized to their trimethylsilyl (TMS)-Methoxime (Meox) derivatives. The derivatization conditions were 100 µL of 20 mg/mL methoxyamine hydrochloride solution in pyridine for 120 min followed by addition of 200 µL of N-methyl-trimethylsilyl-trifluoroacetamide (MSTFA) at room temperature.

The partial derivatization was identified as one of the problems in the previous analysis (Kanani, 2004). Hence optimization studies were carried out to optimize the derivatization time. Based on the analysis we found the number of identifiable peaks in a plant sample increased almost by 25% by increasing the MSTFA reaction time from 0.5 hrs to 4 hours. Hence the standard reaction time for derivatization was modified accordingly.

The derivatized samples were transferred to high recovery auto-sampler vials (National Scientific make) with Teflon inner liner for 4 hours after addition of MSTFA and transferred to auto-sampler.

4.5.4 Split Ratio Optimization

Since the GC-MS instrument used in the current analysis (Varian Inc, ion trap) was more sensitive as compared to instrument used in earlier study (Kanani, 2004), the new optimum split ratio needed to be determined. The approach for determination of the new

split ratio was developed in the previous study (Kanani, 2004) and involved ribitol calibration curves. Subsequently extracted plant samples were used to identify the largest peak area at different split ratios. Based on the obtained results 35:1 was chosen as the new split ratio. This ratio is much lower as compared to 75:1 split ratio used in the earlier analysis. This shows that apart from being more sensitive, the new GC-MS system has larger linear range, allowing more quantity of the sample to be injected in the system.

4.5.5 GC-MS Sample Runs

The derivatized samples were analyzed using Saturn 2000 GC-(ion trap) MS (Varian, CA) between 6-30 hours after addition of MSTFA, to ensure complete conversion of metabolites to their at least one derivative form without degradation. (See Chapter 3 for detailed discussion on selection of derivatization-injection time). The spectra were obtained at 25:1 split ratio with 1.5 μL injection volume, which was optimized to ensure linear range of operation for all metabolites. The plant extracts were derivatized in batches of 10 and two instrumental replicates were obtained for each plant sample extract. Therefore, at least 4 (2 biological replicates x 2 injections) spectra were available for each timepoint (at 0h up to 8 spectra were available). See (Kanani thesis) for detailed GC-MS operating conditions such as temperature programming, column conditions and flow rates.

4.5.6 Peak Identification

Metabolite peak identification was based on:

- Own library obtained using commercially available standards.

- The publicly available Max-Planck based curated TMS-derivative library [http://www.mpimp-golm.mpg.de/mms-library/index-e.html]

- The Public Repository for Metabolomic Mass Spectra - CSB.DB GOLM Metabolome database (Kopka et al., 2005).

- The commercially available NIST MS-library (Ausloos et al., 1999).

A comprehensive list of peaks was obtained by running initially samples in split-less mode increasing the signal strength 25 times the normal split mode. The higher sample causes saturation for high concentration peaks but at the same time brings the lowest concentration peak out of the noise region making it possible to identify them. Totally 550 peaks were identified from which about 140 had known structure, about 170 had at least a known functional group while the rest were complete unknowns. This approach significantly increased the total number of identified peaks and the contribution of known metabolites.

4.5.7 Peak Quantification

Each metabolite was quantified using peak area of its fragment ion/s using Varian Star software (version 6.5). The selection of the specific fragment ion/s for each metabolite was based on the criteria of:

- Specificity for the particular metabolite,

- Separation between co-eluting metabolites

- S/N optimization.

Specifically, only peaks with S/N≥5 and raw peak area ≥ 500 were considered in the analysis.

In order to ensure that the quantification is accurate, it is important to make sure that measurements are made in the linear range of operation of the machine. In current analysis, each set of experiment contains 20 plant samples. Each plant sample was analyzed two times, and each sample analysis contained 600 peak areas. Hence each experiment involved the calculation of approximately 24,000 peak areas. In the previous study, due to integration errors, peak areas were verified manually and corrected when required. However such manual verification was found to be a time consuming process, being the bottle-neck in the data analysis.

By utilizing and optimizing the additional integration control parameters of the new Varian Inc software and using a qualifying ion along with the marker ion, this integration errors and false identifications were reduced significantly from ~0.9% in the initial analysis to <0.1% for the current analysis. Further data filtering steps were created to identify abnormal standard deviation in injections of metabolites, or abnormal variation in a metabolite peak area, which would help identify such deviations. With these measures the requirement for manual verification of each peak area was avoided saving considerable time and resources.

4.5.8 Metabolomic Data Validation, Correction & Normalization

All known metabolites were classified into three categories, depending on their derivatization kinetics as described in Chapter 3 of this report. Category-1 consisted of metabolites forming a single derivative (e.g. alcohols & organic acids); Category-2 consisted of metabolites containing a ketone group (e.g. Glucose, Fructose, α-ketoglutarate) which at the methoximation step, react through parallel reactions to produce two syn and anti geometric isomers; Category-3 consisted of metabolites which contain at least one amine group and form multiple derivatives through series reaction (e.g. amino acids).

The following steps were followed for data normalization, validation and correction:

- Raw peak areas for all metabolite derivatives were normalized with peak area of internal standard ribitol to obtain relative peak area (RPA) of the metabolites.

- Data Validation was performed to ensure constant GC-MS conditions throughout the analysis as described in Chapter 3 of the report by estimating the ratio of peak areas of 6 known Category-2 metabolite derivative forms.

- For 16 Category-3 metabolites (for which 43 derivative forms were detected in total) "cumulative" peak areas representing the original metabolites were obtained using the data correction algorithm described in Chapter 3 of the report. The relative response ratios (w_i^M) values required to estimate the cumulative peak areas were obtained as described earlier in Chapter 3 of the report.

- Unknown metabolite derivative forms showing a coefficient of variation of more than 30% in replicate runs of plant extract at different derivatization times or containing marker ions with m/z 156 or m/z 174 in their mass-spectra (characteristic m/zs in amino acids, indicating presence of silylated amine group) were classified as unknown Category-3 metabolites and removed from further analysis.

- Finally a corrected RPA matrix consisting of: RPA of all known Category-1 metabolites; RPA of larger of the two derivative forms of known Category-2 metabolites; estimated Cumulative RPA of known Category-3 metabolites; and RPA of unknown Category-1&2 metabolite derivative forms; was prepared for further analysis as per the data correction algorithm described in Chapter 3. All RPA levels in the corrected RPA matrix are now expected to be directly proportional to the relative concentration of the original metabolite in the plant extract.

- In order to ensure consistency in metabolomic data and avoid mathematical artifacts peaks which were detected in blank control (consisting all derivatization reagents without plant extract); and peaks which are carried over (detected considerably in solvent GC-MS runs in between derivatized plant extract runs); were removed from further data analysis.

- The RPAs for each plant extract were obtained by taking arithmetic mean of the two instrumental replicates and for each time point by taking arithmetic mean of all instrumental & bio-replicates representing the given time point.

Finally Corrected RPA data sets for each experiments were normalized with metabolomic profiles of individual time zero controls (liquid cultures extracted at the end of 12 days) to obtain Normalized Peak Areas (NPA) by dividing the averaged RPA of a metabolite at time t hrs w. r. t. its averaged RPA at 0 hrs in the same data set. This normalization "corrects" for potential differences between the growth conditions of the two plant sets during 12 days before harvesting and improves the comparison between the various metabolites in each set.

For better comparison between the two plant sets, the time profile of a metabolite's concentration in the perturbed set was normalized with respect to its in the control and the logarithm with base 2 of this ratio was usually used to depict the difference in metabolite concentrations between the two sets. All mathematical operations described above were performed using an algorithm incorporated in Mathcad 2000 (Mathsoft, MA).

4.5.9 Metabolomic Data Analysis

Principal Component Analysis (PCA) and Pavlidis Template Matching (PTM) were carried out using algorithms incorporated in open source software MEV 3.1 (Saeed et al., 2003) which is freely available from http://www.tm4.org/mev.html. Paired-SAM (Tusher et. al., 2001) and Time-series analysis of the Metabolomic data was performed using the MiTimeS algorithm (Dutta et. al., 2007) software incorporated in Mathcad 2000 (Mathsoft, MA). The results obtained from the analysis of time-series metabolomic experiments of systematically perturbed *A. thaliana* liquid cultures are summarized in the next two chapters of this report.

5 INDIVIDUAL STRESS RESPONSES

In the 21st Century plants are re-emerging as the renewable, environment friendly source of energy, plastics, industrial chemicals and even pharmaceuticals. This industrial role coupled with the ever increasing demand for food and feed application is making it necessary for us to engineer plants which are (a) more efficient in light, carbon, nitrogen and nutrient fixation (b) more tolerant to adverse environmental conditions (abiotic stress) such as high salinity, flooding, temperature variations and drought conditions (c) resistance to external biotic stress such as worms, virus, bacteria, fungi. Engineering plants which have combination of traits to perform all the three category of function can significantly increase the agricultural output and make plants an attractive manufacturing host for the production of industrial chemicals.

In the last two decades a large amount of commercial agricultural research has focused on engineering genetically modified plants which are better able to combat biotic stress reducing the need for dependence on the chemical pesticides. These Genetically Modified plants have received very positive response in North and South American markets, accounting for more than 85% of the total agricultural output for some commercial crops (www.monsanto.com). The next generations of crops currently under various stages of development are however focused on improving resistance to the abiotic stresses and efficient utilization of nutrients. In order to successfully engineer plants for these applications we need to better understand:

- Regulation of primary metabolism of plants
- Strategies used by plants at molecular level in response to changes in the environment

To achieve this aim a series of experiments were designed which systematically perturbed the *Arabidopsis thaliana* liquid culture system with (a) elevated CO_2 levels (1%) in the growth environment (b) Salt (NaCl) stress in the liquid media (50 mM) (c) Trehalose sugar signal (12 mM) (d) Hormone Signal (Ethylene precursor ACC 0.01 MM) as described in Chapter 4 of this report. Their dynamic response at the metabolomic level was measured using the optimized GC-MS Metabolomic profiling algorithm described in Chapter 3. The results obtained from the statistical analysis of each individual stress and their biological significance is discussed in this chapter.

5.1 ELEVATED CO_2 RESPONSE

5.1.1 Review of Elevated CO_2 Stress Response in Plants

CO_2 is the main carbon and energy source of plants. One of the main roles of plants in global ecology is to maintain the carbon and nitrogen balance of the environment (Buchanan et. al., 2001). In this context and in light of global warming and increasing levels of CO_2 in last few decades response of elevated CO_2 on plant physiology has been extensively studied. A comprehensive review of the (a) carbon fixation mechanisms in plants, (b) known physiological effects of elevated CO_2 (c) detailed mechanisms of the pathways perturbed by elevated CO_2 and (d) the experiment design issues which can give

rise to contradictory results; was performed during an initial elevated CO_2 experiment study (part of Master's Thesis Research of this author) carried out under slightly different experimental conditions and the same is available from (Kanani, 2004). The main observations from the review were:

- Most of the experimental work on effects of elevated CO_2 was carried out in long-term effects (few weeks to years) in response to 700-1000 ppm CO_2. These studies were primarily governed by the aim to understand the physiological long term (adverse) effects of rise in environmental CO_2 levels rather than understanding regulation at molecular level.

- Most experiments made observations of the overall physiology like: root to shoot length ratio (found to increase), starch-sugar content (found to increase), total protein content (found to increase in absolute terms –but decrease on per gram dry weight basis), total lipid content (very few studies – showed no change), lignin (varying effect) and secondary metabolites.

The review of literature presented thus showed that the elevated CO_2, apart from affecting the Calvin cycle metabolite, also affects metabolites involved in carbohydrate synthesis, photorespiration, lipids, secondary metabolites and amino acids. The traditional analytical methods did not allow measurement of the change in all these metabolites simultaneously. In contrast high throughput method which can simultaneously measure changes in metabolites belonging to different classes will allow:

- Understanding the effect of elevated CO_2 on the metabolism of various plant sub systems.
- The extent of response of different plant sub systems in response to elevated CO_2.
- Understanding the interaction between different sub systems of the plants.

Hence a high throughput method for analysis of metabolic data would allow understanding of a holistic, complete response of the plant to elevated CO_2 level, as compared to the traditional methods which are focused more on a particular class of compounds. In deed during the Master's study the metabolomic analysis of the elevated CO_2 of *A. thaliana* liquid cultures showed:

- Inhibition of Nitrogen Storage due to competition for the reductive power between the CO_2 and Nitrogen fixation pathways (see (Kanani,2004) for details)
- Inhibition of photorespiration pathway
- Increase in precursors of structural carbohydrate and glycerolipids

In the current analysis, the experiment carried out during Master's thesis research was repeated with following differences:

- *A. thaliana* liquid cultures were grown in controlled environment growth chambers in University of Maryland Greenhouse, as compared to open growth chambers at The Institute of Genomic Research used in the previous experiment.

- In the current experiment, the elevated CO_2 stress was applied by increasing concentration of the growth chamber as compared to connecting to a cylinder in the previous experiment.

- Metabolomics was optimized for extraction derivatization and data-analysis.

The results obtained from the current analysis are described in the following sections.

5.1.2 Metabolomic profiling results

Polar Metabolomic Profiles (Roessner et al., 2000) were obtained using the optimized protocol and experimental conditions described in Chapter 3 and Chapter 4, respectively, of this report. One of the liquid cultures (harvested at 6h in control) weighed abnormally low (9.2 gm), about 50% of average of all sample (18.4 gm) and 60% of its bio-replicate (15.2 gm), and was also identified as an outlier in metabolomic and transcriptomic analysis (data not shown) and hence was removed from the analysis. In the acquired metabolomic profiles, 550, among which 147 known, metabolite methoxime (Meox)-trimethylsilyl(TMS)-derivative peaks were detected. After data correction, normalization and filtering, the set of metabolite peak areas that were finally considered in the analysis included 89 annotated and 216 unidentified. Among the 89 known metabolites, 61 correspond to metabolites forming only one TMS-derivative, 9 to one of the two geometric isomer derivatives of ketone-group containing metabolites (see Chapter 3), and

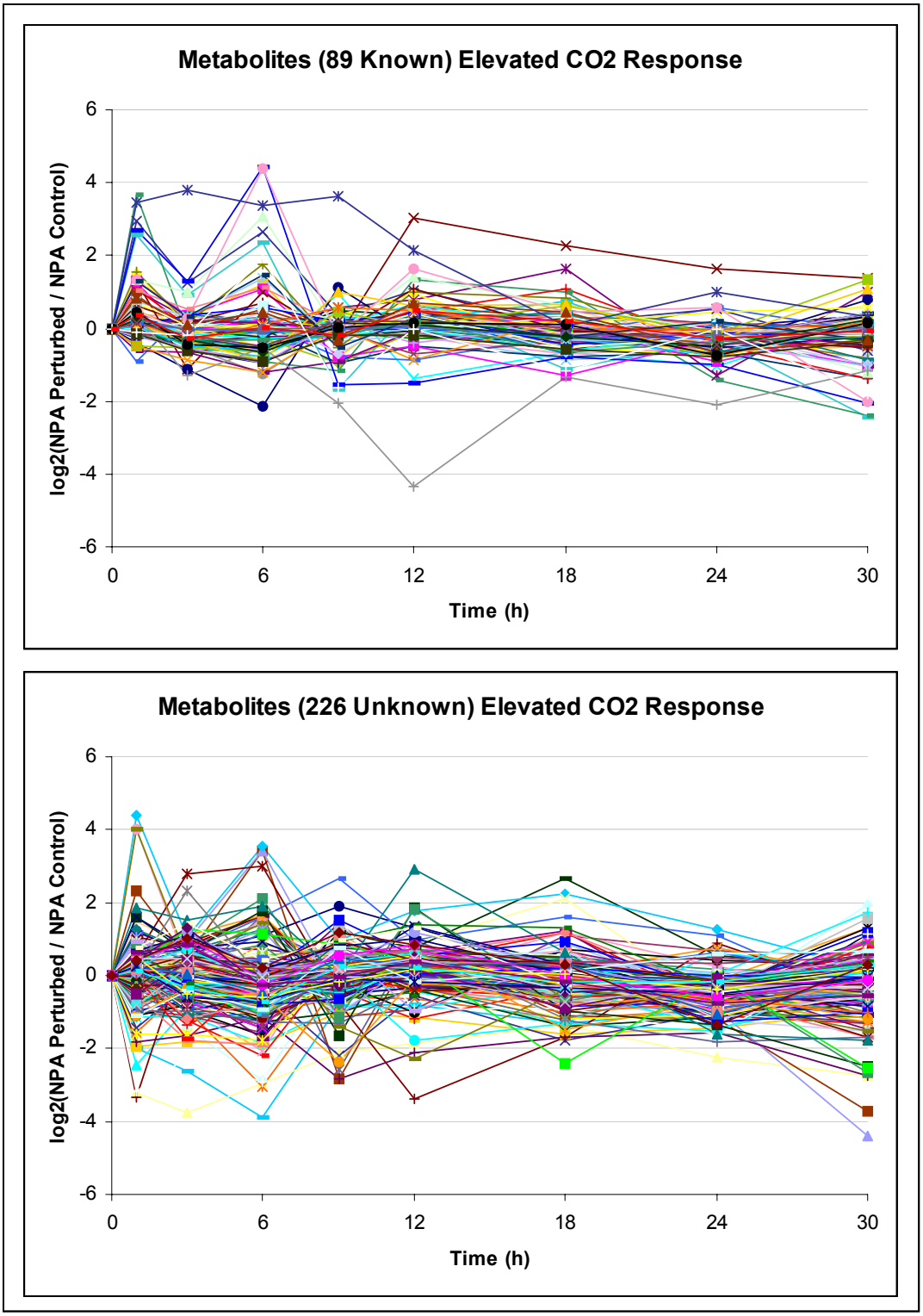


Figure 5-1: Log_2 Ratio Normalized Peak Area (NPA) profiles of Perturbed (elevated CO_2) to NPA Control profiles indicate up to 15 fold increase - decrease in metabolite levels.

19 to cumulative peak areas of amine-group containing metabolites. The coefficient of peak area variation between injections of the same sample and between biological replicates was 6.3% and 17%, respectively. Each timepoint (except control 6 h) was represented by at least 4 (2 biological x 2 instrumental replicate) spectra (at 0h, with 4 bio-replicates, 8 spectra were available). Data correction, filtering and normalization was done as described earlier in Chapter 3 to obtain normalized peak areas for control and perturbed experiments. The log ratio of this normalized peak profiles is shown in Figure 5-1. As can be seen from the figure up to 15 fold increase (most at 1h) as well as decrease in metabolites (most at 30h) is observed in response to elevated CO_2.

5.1.3 PCA Analysis

According to TIGR MeV Principal Component Analysis (PCA), the control metabolomic profiles can be clearly differentiated from their perturbed counterparts (Figure 5-2). This implies that the physiology of the plant liquid cultures is affected significantly by the applied perturbation at the metabolic levels, even during the first 30h of treatment. It is further interesting to note that even though elevated CO_2 modifies the metabolism significantly, the variation due to elevated CO_2 is mainly seen in the direction of principal component 3 which accounts for only 9% of the total variation, the largest variation is however with time which accounts for 45% of the total information content.

5.1.4 SAM and MiTimeS Results

Paired-SAM analysis (delta=1.35, 0% FDR) identified 28 (6 known) and 0, respectively, metabolites, whose average over time concentration had significantly decreased or

increased due to the applied perturbation (in the rest of the text, they will be referred to as negatively and positively significant, respectively) (Figure 5-3). It needs to be noted that, in the case that the new data correction and derivatization algorithm presented in Chapter 3 of the report had not been used, i.e. both derivative peak areas of the known ketone-group containing metabolites and all derivative peak areas of the amine-group containing metabolites had been included in the statistical analysis (361 peak areas in total), 14 (3 known) additional peak areas corresponding to amine-group containing metabolites would have been identified as negatively significant (Figure 5-4). This result reinforces the validity of the new algorithm and provides additional evidence supporting the need for correcting the metabolomic profiles from the derivatization biases (Kanani and Klapa, 2007).

MiTimeS analysis (Dutta et al., 2007), on the other hand, many more negatively and positively significant metabolites were identified at the individual timepoints (average 91 and 66, respectively) for the same significance threshold value (Figure 5-3). MiTimeS analysis (Dutta et al., 2007) also allows calculation of Significance Variability Score (SV Score) which represents the amount of fluctuations in the significance level for each metabolite on a scale of 0-2 as described earlier in Chapter 3. SV score of 0 represents the metabolite does not undergo change in its significance level and maintains the same significance level at all time points. In contrast SV score 2 represents that the metabolite changes its significance level from positively significant to negatively significant and vice-versa at each and every time point. The SV score distribution which plots the SV score against the fraction of the detected polar metabolites at each SV score is shown in Figure 5.5.

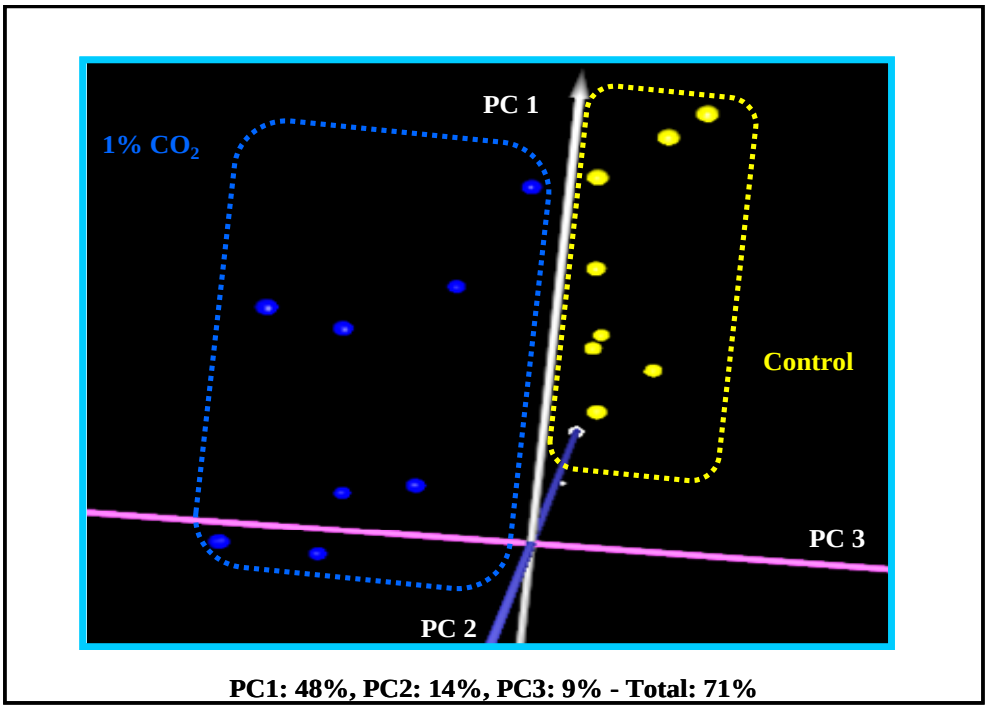


Figure 5-2: Principal Component Analysis (PCA) of Elevated CO_2 response using TIGR MEV 3.0 shows a significant difference in metabolism of *A. thaliana* liquid cultures even during the 1-30h.

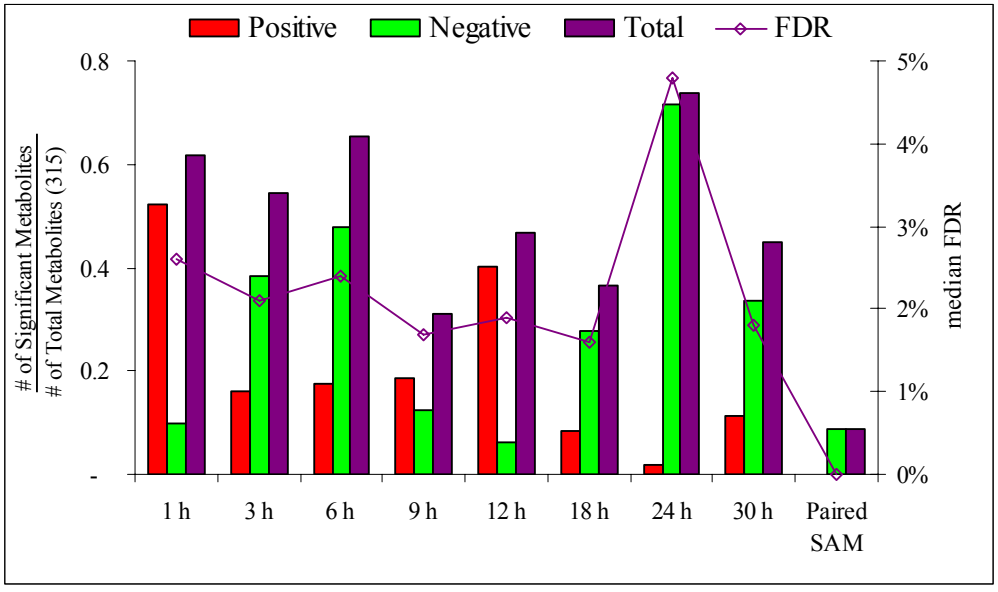


Figure 5-3: Number of Positively, Negatively and Total significant metabolites along with % median False Detection Rate (FDR) obtained for overall analysis and individual time points, using paired-SAM and MiTimeS analysis.

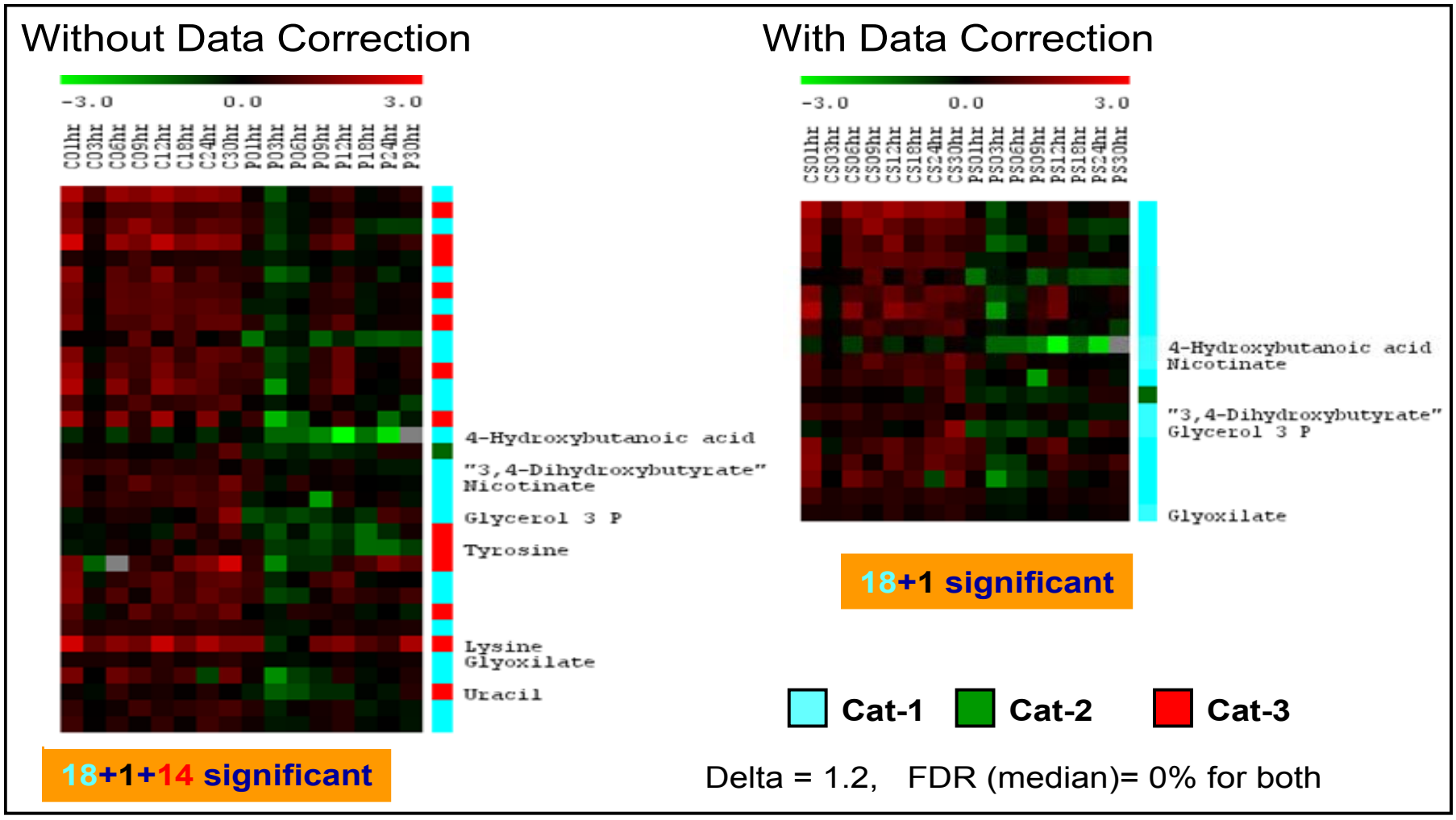


Figure 5-4: Paired-SAM Negatively significant metabolites with and without the use of data correction algorithm with takes care of biases in Category-3 metabolites (compounds containing amine groups). The comparison clearly shows in absence of data correction the results obtained can be biased due to variations in derivatization which may be assigned biological significance.

As can be seen from the figure, in response to elevated CO_2, only three (all unidentified) metabolites were observed in the same significance level (non-significant) at all timepoints (i.e. SV score = 0), while all others were observed as positively or negatively significant at one time point at least. This shows that there is no pathway which is not perturbed by elevated CO_2 at least during the short-term response.

According to Figure 5-3, CO_2 availability resulted in a drastic initial increase in the free polar metabolite pool sizes. The changes in the number of positively and negatively significant metabolites during the initial 6h of perturbation suggests (a) potential feedback inhibition of the pathways producing the measured free polar metabolites after the drastic initial increase in latter's concentration, and/or (b) initiation of the pools' consumption for the production of macromolecules and plant growth. The minimum and maximum total number of significant metabolites was observed at 9h and 24h, respectively, after the initiation of the perturbation. At 24h, more that 60% of the free polar metabolites were identified as negatively significant. The time point correlation network based on the positively significant metabolites (Figure 5-6A) indicates strong correlation between the time points up to 12h. The common known positively significant metabolites among these time points are mainly amino acids and other amine-containing compounds, intermediates of the one-half of the TCA cycle (citrate, iso-citrate, α-ketoglutarate), which is responsible for the production of glutamate-based amino acids, most of the triose- and hexose- phosphates and some fatty acids. The time point correlation network based on the negatively significant metabolites indicates strong correlation between the longer time points (18h, 24h, 30h).

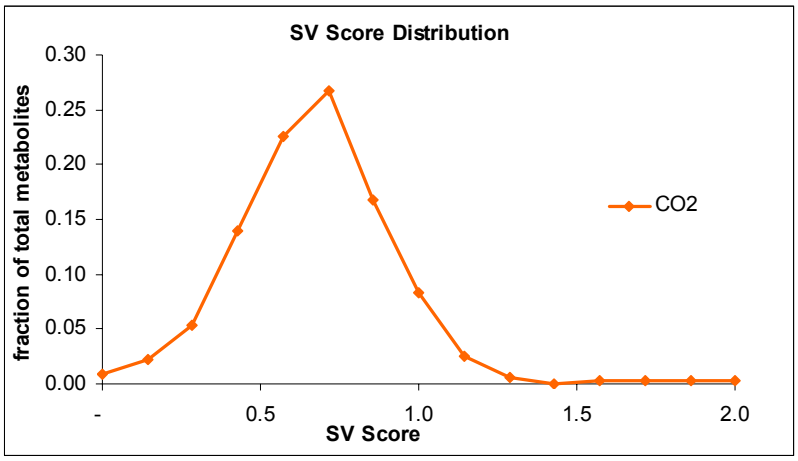


Figure 5-5: Significance Variability (SV) Score distribution of metabolites shows

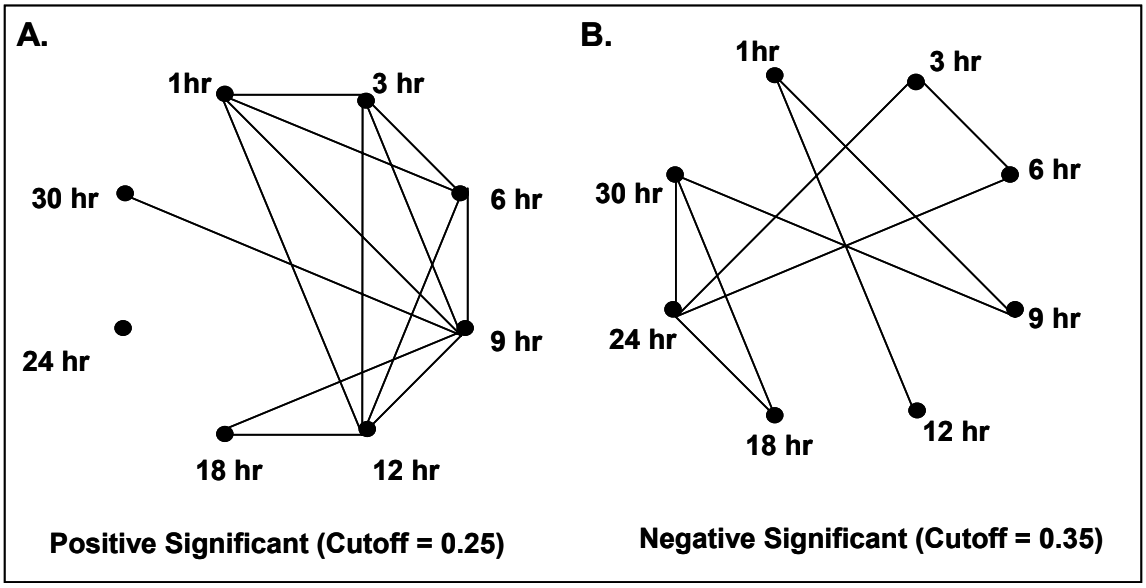


Figure 5-6: Time point correlation networks based on the common between time points (A) positively and (B) negatively significant metabolites. Two time points are connected, if their correlation coefficient is larger than the indicated threshold, the latter being selected in each case as the average of all between different timepoints correlation coefficients.

At these time points the number of positively significant genes is very low suggesting a continuous drain of the measured free polar metabolite pools without similar increase in their production rate. Most of the metabolites, which were identified as positively significant during the first 12h of the perturbation, belonged to the negatively

significance level at longer time points. Based on the number of metabolites that were significant at a particular time point only, the response of the system at 30 h of continuous perturbation can be singled out. This is in agreement with the observation that 24h of perturbation is a "turning point" in the response of the system to the applied perturbation.

In summary MiTimeS analysis of metabolomic data suggests a first turning point in the plant liquid culture's physiology between 9h and 12h and a second at 24h of continuous perturbation. After 24h the response of the cultures seems to be drastically changing giving rise to the unique profiles measured at 30h. The following sections discuss the most prominently observed changes in the physiology of the *A. thaliana* liquid cultures due to the applied perturbation in the context of specific pathways that are directly or indirectly related to carbon fixation.

5.1.5 Analysis of Individual Pathways

Photorespiration Pathway

Carbon and oxygen "compete" for Rubisco activity (Figure 5-7) as the Rubisco catalyses both the carboxylation and oxidation reaction. The pathway for regenerating Calvin cycle intermediate from the oxygenation product is known as Photorespiration pathway. This is an important pathway for plants as it wastes CO_2 and hence reduces the biomass yield. Changes in the CO_2/O_2 ratio have been shown to affect the flux distribution between the two pathways (Siedow and Day, 2001; Dey and Harborne, 1997; Coschigano et al., 1998) by: (a) rescuing mutants having genes deficient in the photorespiration NH3 recycle

pathway (b) *in vitro* enzymatic assays. The metabolomic analysis of the response of *A. thaliana* liquid cultures to elevated CO_2, showed significant changes in the metabolomic pools of all three detectable known organic acid intermediates of the photorespiration pathway, i.e. glycolate, glycerate and glyoxylate at (3,6,24h), (3,6,18,24h,30h and paired-SAM) and (3,6,18,24,30h and paired-SAM), respectively (Figure 5-7). This is the first time that the actual decrease in the metabolite pools of the photorespiration pathways has been quantified. This also demonstrates the effect of elevated CO_2 on the photorespiration pathway is similar to what has been observed in the soil grown plants.

Tri-Carboxylic Acid (TCA) Cycle Metabolites

Tri-Carboxylic acid Cycle (Figure 5-8) is the primary aerobic respiration pathway in the mitochondria for most eukaryotic systems (Tiaz, 2002). TCA cycle is responsible for the production of energy in the form of ATP and reducing power in the form of NADH and $FAdH_2$ by oxidation of pyruvate (derived from sucrose) into CO_2 (Tiaz, 2002). Apart from providing energy and reductive power TCA cycle also provides intermediates - α-ketoglutarate and oxaloacetate (OAA) which are used as carbon skeleton for various amino acids, purines, polyamides and various secondary metabolites (Tiaz, 2002). In response to elevated CO_2, as can be seen from Figure 5-8, most of the metabolite pools show a significant increase at 1h and significant decrease at 24h. This was in line with the over-all response of the free metabolite pools as the highest number of positively and negatively significant metabolites were observed at these time points. At other time points interestingly there were two distinct responses. During 3-12h, the right side of the TCA cycle from citrate -> α-ketoglutarate showed a significant increase at most of the

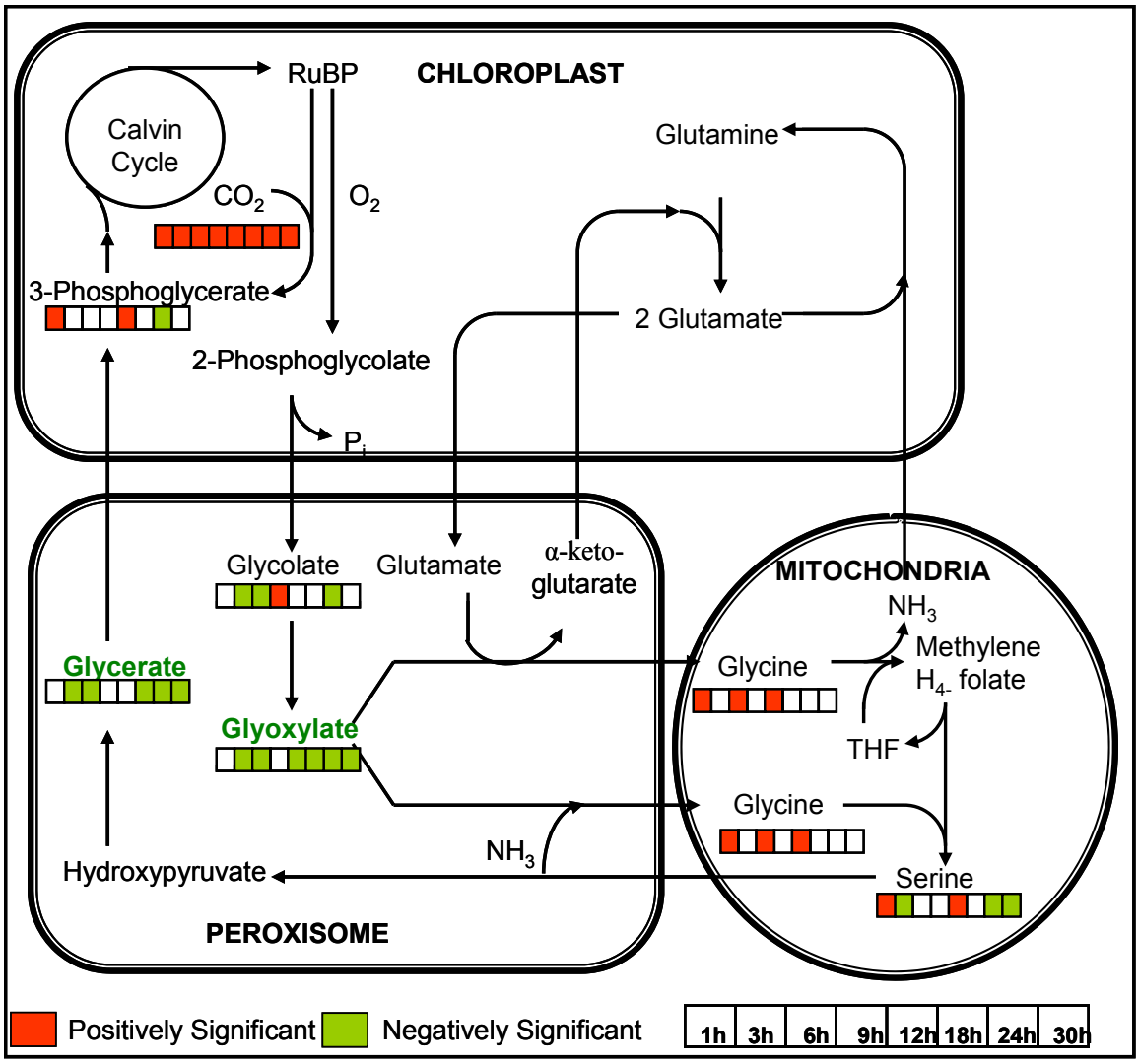


Figure 5-7: Observed effect of the applied perturbation on the physiology of photorespiration, at the metabolic levels. If a metabolite was identified as positively or negatively significant at a particular time point, the corresponding time point box under this metabolite's name is colored red or green, respectively. The significance level of the metabolite as identified by paired-SAM is indicated by the fonts' color of the metabolite's name (red or green for positively or negatively, respectively, significant metabolites).

time points, however the left side from succinate to malate showed a significant decrease. Similar contradictory changes were also observed at 24h, suggesting differential regulation of these metabolite pools in the same TCA cycle.

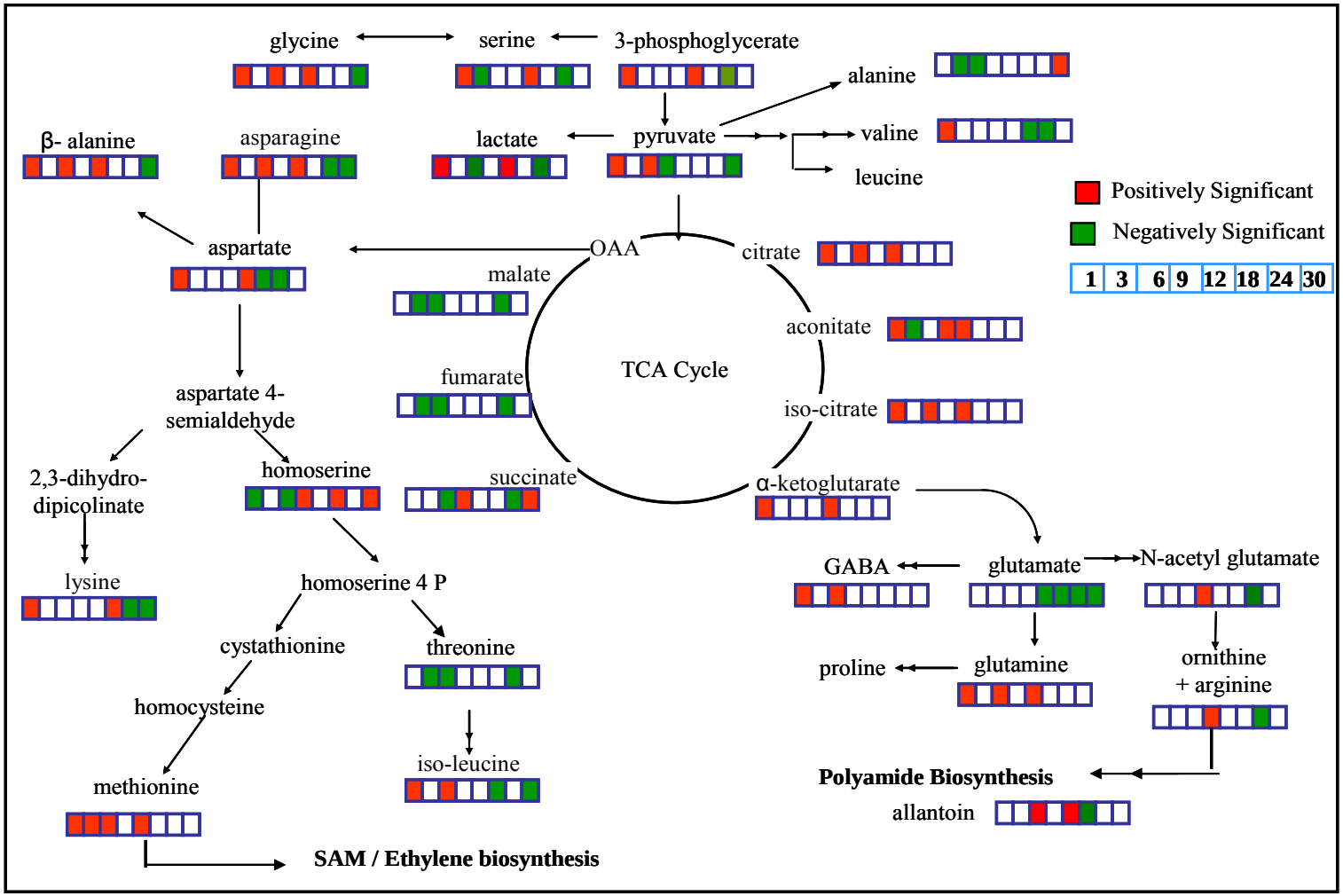


Figure 5-8: Observed effect of the applied perturbation on the physiology of the Tri Carboxylic Acid (TCA) cycle and amino acid biosynthesis at the metabolic level. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

Nitrogen Assimilation and Amino Acid Biosynthesis

Nitrate Reductase (NR) gene, is the first enzyme in the nitrogen assimilation pathway and light, nitrate ions and carbohydrates, have shown to induce expression of NR mRNA (Larios et al., 2001; Cheng et al., 1991; Cheng et al., 1992). It was speculated that the regulation by light is achieved through carbohydrates concentrations (Cheng et al., 1991). Elevated CO_2 study (Larrios et al., 2001) on cucumber plants indicated increase in the expression of the gene encoding nitrate reductase (NR), due to the applied stress. From the metabolomic data, three of the four nitrogen-storage amino acids (Coruzzi, 2000), glutamine, asparagine and aspartate, were identified as positively significant at most of the timepoints during the first 12h (Figure 5-8). Even though the concentration of the fourth amino acid, i.e. glutamate, remained unaltered for the first 9h of the perturbation, decreasing slightly at 12h, amine-containing metabolites derived from glutamate, i.e. 4-aminobutyrate (GABA), ornithine/arginine and N-acetyl-glutamate, were identified as positively significant at most of the time points. In addition, α-ketoglutarate required for the production of glutamate and glutamine along with its TCA cycle precursors citrate and isocitrate, were also positively significant at some/most of the timepoints during the first 12h (Figure 5-8).

Apart from the amino acids that are related to nitrogen assimilation and storage, valine, isoleucine, glycine, lysine and methionine were also identified as positively significant at some/most of the timepoints during the first 12h (Figure 5-8). Thus, the concentration of most of the known detected amino acids and metabolites produced by them showed a significant increase in response to elevated CO_2 stress during the first 12h. This was not,

however, the case for serine, threonine, and homoserine. The decrease in serine could be explained by the decrease in photorespiration pathway which was discussed earlier. It was interesting to note that both methionine and iso-leucine showed increase in their concentration during first 12 h even though homoserine and threonine, the intermediates to the pathway were negatively significant. The methionine pool size was identified as significantly increasing due to the applied perturbation during the first 12h. At subsequent time points, when most of the amino acids were identified as negatively significant, methionine belonged to the non-significant metabolites. The significant decrease in most of the amino acid pool sizes after 12h of perturbation could be because of either (a) decrease in the rate of CO_2 fixation through the Calvin cycle due to inhibition of RubisCo and/or (b) increase in the rate of these pools' depletion for the production of macromolecules for plant growth without equal increase in their production rate.

Sugar Metabolite Pools

In plants, Starch (a polymer of glucose) and sucrose are the major carbon storage molecules. Starch is usually used as a long-term storage where as sucrose is used for short term storage of carbon and for transport of carbon from source (leaf) to sink (roots, flower) tissues. Sucrose is also the most abundant sugar in *A. thaliana* liquid cultures thus representing a major portion of the sugar-metabolite pool. Both starch and sucrose concentrations are known to accumulate in response to long term CO_2 stress. In the present analysis however, in response to short term (1-day) exposure to elevated CO_2 levels, the concentration of sucrose significantly decreased at 3h, 6h, 24h and 30h with a

slight increase in concentration at 18h (Table 5-1). It is interesting to note that at 1h timepoint when most of the known metabolites show a significant increase in the metabolite pools, sucrose metabolite pools remain non-significant. This suggests there is very less carbon accumulation in the form of sucrose in response to elevated-CO_2 in the short term. In contrast, Fructose and Glucose obtained from sucrose were accumulated at (1h,12h,24h, 30h) and 30h, respectively, as shown in Table 5-1. The metabolomic analysis thus indicates a tendency towards fructose accumulation in response to elevated CO_2 in the short term.

Table 5-1: Significance level of Sugar, Sugar Phosphates and Sugar Alcohols at individual time points and paired-SAM analysis. Value of 1 in the table indicates metabolite levels were significantly increased in response to elevated CO_2 stress at the particular time-point or from paired-SAM analysis of overall response as indicated by the column heading. In contrast, value of -1 indicates a significant decrease.

		1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sucrose	sugar		-1	-1			1	-1	-1	
Trehalose	Sugar	1				1	1	-1	-1	
Maltose	Sugar	1					-1	-1		
Cellobiose	Sugar		1	1		1	-1			
Glucose	Sugar								1	
Fructose	Sugar	1				1		1	1	
Glucose 6 P	Phosph	1		-1		1		-1	-1	
Fructose 6 P	Phosph	1				1		-1	-1	
Inositol-1/2-Phosphate	Phosph	1		-1	1	1		-1		
Arabinose	sugar		-1	-1				-1		
Rhamnose	Sugar	-1			-1	-1			1	
Xylulose	Sugar		-1	-1						
Xylitol	alc	1		-1						
myo-Inositol	alc	1				1	-1		1	
Glycerol	alc		-1	-1			-1			
Glycerol 3 P	Phosph	-1	-1	-1	-1		-1		-1	-1

Maltose and cellobiose are metabolites produced from starch and cellulose breakdown respectively where as Trehalose competes with the starch and cellulose biosynthesis pathway for carbon substrates. All the three sugars (maltose, cellobiose, trehalose) like sucrose are disaccharides. In response to elevated CO_2 maltose shows significant increase at 1h as was observed for most of the metabolites, and significant decrease at 18h and 24h. In contrast cellobiose was found to be positively significant at 3h, 6h, 12h suggesting increase in cellulose breakdown. In the competing pathway trehalose concentration is increased significantly at 1h, 12h and 18h as shown in Table 5.1. However maltose and cellobiose are either negatively significant or non-significant beyond 18h, suggesting a decrease in starch and cellulose breakdown or probably accumulation of starch and cellulose which is in line with the overall changes which suggest increase in macro-molecule production in response to elevated CO2 at these time points. However these needs to be confirmed by actual measurement of starch/cellulose or by analyzing the corresponding transcriptomic fingerprint of the same response.

Apart from the role of sugars in carbon storage and transport, sugar metabolites are also used to build structural carbohydrates (oligosaccharides) required for the plant cell wall. Apart from cellulose which is the polymer of glucose, plant also produces hemi-cellulose which is a polymer made from a combination of hexose (glucose, fructose, galactose, and mannose) and pentose (xylulose, arabinose, rhamanose) sugars. As can be seen from Table 5.1 xylulose, rhamanose, arabinose and xylitol are negative or non-significant at most of the time points except 30h. In addition glycerol and glycerol-3-phosphate which are utilized for production of cell membranes constituents glycerolipids and

phospholipids were also negative significant at (3h, 6h, 18h) and (1-9h,18h,30h) respectively suggesting a possible drain for increased production of the cell membranes.

Secondary Metabolite Pools

The three aromatic amino acids: tryptophan, tyrosine and phenylalanine; are the precursors for many secondary metabolites which are important for a host of important physiological functions in plants. As shown in Figure 5-9, tryptophan produces important plant hormone Indole-Acetic Acid (IAA) also known as Auxin which was the first plant hormone discovered and which regulates growth of the plants. In addition it produces a family of indole alkanoids and indole glucosinolates. Phenylalanine is the other important secondary amino acid which is the precursor for the production of flavanoids, important plant cell wall constituents' lignin and syringin; and possibly regulatory compound salicylic acid important for regulating defense response to biotic stress in plants. The third aromatic amino acid tyrosine – is mainly used for production of alkanoids in higher plant, however in Arabidopsis, very few alkanoids are detected and the biosynthesis of alkanoids is still being studied.

Even though the choice of GC-MS for metabolomic analysis limits the extent to which secondary metabolism can be studied, a number of metabolites which are precursors to the secondary metabolite and the concentration of the three aromatic amino acids can be detected using GC-MS as shown in Figure 5-9. The analysis revealed that in *A. thaliana* liquid cultures, from the three aromatic amino acids, tryptophan pools was the smallest, close to or below the detection limit in most of the samples, however some of the metabolites derivatives derived from tryptophan had significantly higher metabolite

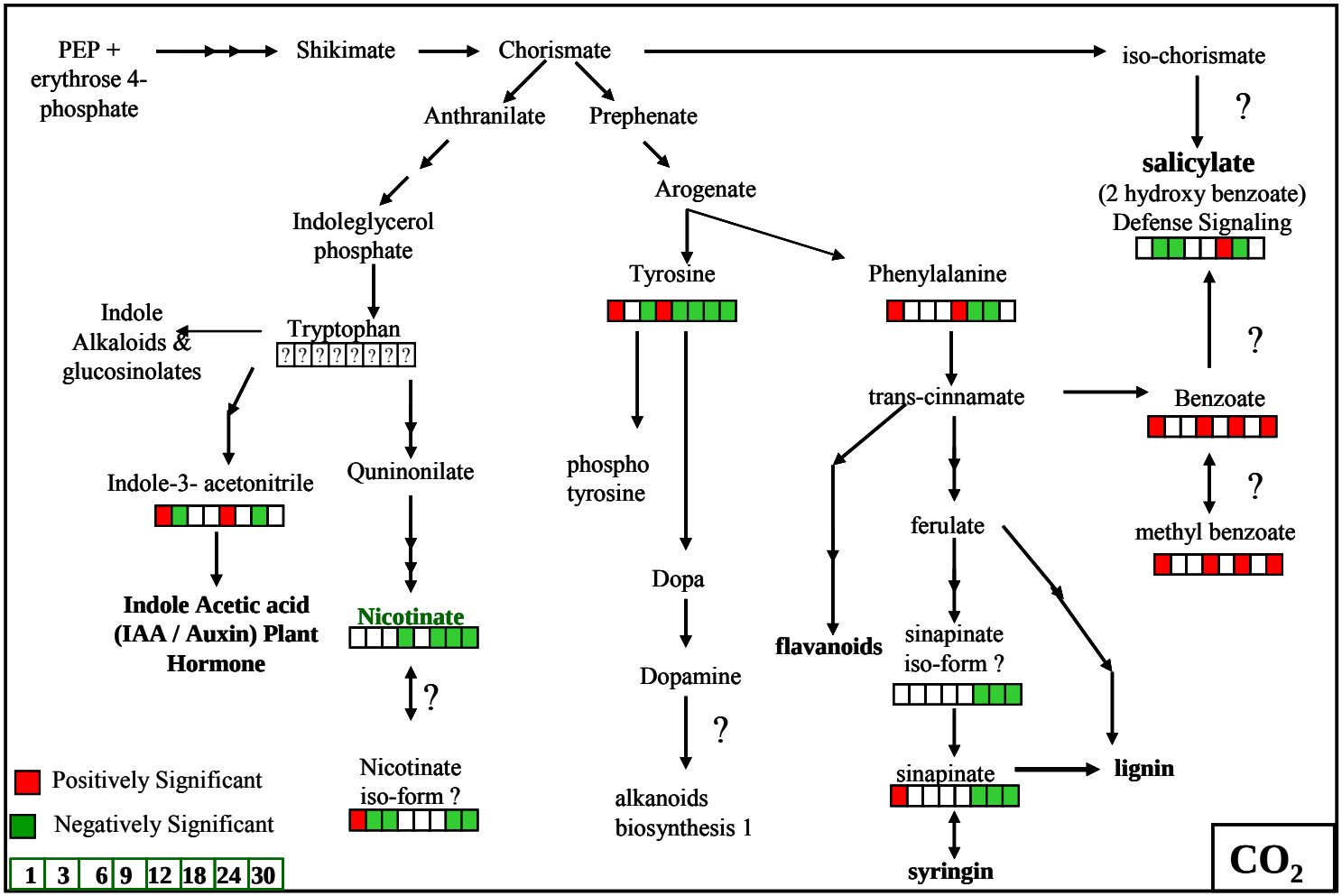


Figure 5-9: Response of Metabolite Pools in *Arabidopsis thaliana* secondary metabolism pathways. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

pools. In response to elevated CO_2, two metabolites, Nicotinate and an unidentified compound showing a very similar mass-spectra to nicotinate were identified as negatively or non-significant between 3-30h. From these two, nicotinate was also identified as negatively significant from paired-SAM analysis suggesting relative inhibition of nicotinate production. Another identified metabolite produced using tryptophan as a precursor was indole-3-acetamide which is in the auxin biosynthesis pathway which showed small fluctuations in metabolite pools in response to elevated CO_2 as shown in Figure 5-9.

From the other two aromatic amino acids, in response to elevated CO_2, tyrosine was positively significant at 1h and 9h, but negatively significant at 6h,12h,18h,24h and 30h time point; Phenylalanine was positively significant at 1h and 12h and negatively significant at 18h and 30h. This is in line with our earlier observation of significant increase in amino acid levels during first 9-12 hours followed by significant decrease at 12-30h. In addition to the secondary amino acids, metabolomic analysis also identified sinapinic acid and its derivative in the lignin/syringin biosynthesis pathway and salicylic acid and two related metabolites. In response to elevated CO_2 these metabolites showed significant increase at 1-2 timepoints during first 12h and significant decrease beyond 12 h as shown in Fig 5-9. The only exception in this pathways was however methyl benzoate, whose role and biosynthesis pathway in plants has still not been completely elucidated. Methyl benzoate in response to elevated CO_2, showed significant increase at 1h, 9h, 18h and 24h and no-significant at the rest of the timepoints.

Butanoate Metabolism

Butanoate metabolism is based on precursors from TCA cycle and is important for the production of Poly-hydroxybutyrates (PHBs) as well as intermediates for ketone bodies. PHBs are currently being commercialized as an important renewable, biodegradable alternative to petroleum based polymers. The pathway producing PHBs in plants however, have not yet been fully characterized and many genes and reactions involved in the same are still being studied. GC-MS metabolomic analysis identified several metabolites belonging to the known PHB pathways (2-hydroxyglutarate, 3-hydroxybutanoate, 4-hydroxybutanoate) as well as some related metabolites (3-hydroxy-3-methylglutarate, 2,3-dihydroxybutanoate, 3,4-dihydroxybutyrate) which may also be involved in Butanoate metabolism. As shown in Table 5.3, in response to elevated CO_2 all the metabolite pools belonging to this pathway showed significant decrease at time point beyond 1h with the sole exception of 3(or 2)-hydroxybutanoate which was positively significant at 1-12h and 30h. Two of the metabolites, 4-hydroxybutanoate and 3,4 dihydroxybutyrate were also identified as negatively significant from paired-SAM analysis. The results thus suggest elevated CO_2 significantly affects Butanoate metabolism and hence most probably also PHB production and its effect needs to be studied in further detail.

Lipid Metabolism

In the present metabolomic analysis experiments, the polar metabolomic profiles obtained by methanol-water extraction were studied. Most of the lipid metabolites are not extracted using this extraction method and the chloroform extraction is the preferred

Table 5-2: Significance level of metabolite pools belonging to Butanoate metabolism at individual time points and paired-SAM analysis. Positively and negatively significant metabolites are color-coded as described in the caption of Table 5-1.

		1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
2-hydroxyglutarate	acid			-1				-1		
3-Hydroxy-3-methylglutarate	acid	1	-1	-1			-1	-1		
2/3-hydroxybutanoate	acid	1	1	1	1	1			1	
4-hydroxybutanoate	acid		-1		-1	-1	-1	-1	-1	-1
3,4-dihydroxybutyric acid	acid		-1	-1			-1	-1	-1	-1
2,4-dihydroxybutanoic acid	acid	1	-1	-1				-1	-1	

method for studying lipid metabolism. However 5 fatty acids (Table 5-3), two of the four major Arabidopsis sterols and Tocopherol (Vitamin E) which is derived from the sterols were extracted in the methanol-water extraction and hence were studies in the current analysis. These lipids are important in determining the composition of the cell membranes. In addition campesterol is an intermediate of plant hormone Brassinosteroid which regulate important aspects of plant development. As can be seen from Table 5-3, all the detected fatty acids and Tocopherol showed significant increase at 1h as most metabolite pools increased in response to elevated CO_2. Similarly all FA and sterols were negatively/non significant at most of the time points beyond 18h except stearic acid. Stearic acid (C18, saturated FA) was positively significant at 1h, 6h, 9h, 18h and 30h and non-significant at the rest. Thus stearic acid showed an increased concentration of the metabolite pool even when most metabolite pools were negatively significant, suggesting an increased production in response to elevated CO_2.

Other Metabolite Pools

Metabolomic analysis of *A. thaliana* liquid culture polar extracts using GC-MS also identified several other metabolites shown in Table 5-4. Some of these metabolites the exact role and the biosynthesis of the metabolites is still not known. For e.g. erythretol

Table 5-3: Significance level of fatty acid and sterol metabolite pools at individual time points and paired-SAM analysis. Positively and negatively significant metabolites are color-coded as described in the caption of Table 5-1.

		1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Stearic Acid	FA	1		1	1		1		1	
Linolenic acid	FA	1					-1	-1		
alpha-Linolenic acid	FA	1	-1					-1		
Gama Linolenic Acid	FA	1					-1	-1	-1	
Icosanoic acid	FA	1	1					-1		
Stigmasterol	Sterol		-1	-1		1	-1		1	
CampeSterol	Sterol		-1	-1			-1	-1	-1	
Tocopherol	Pherol	1	-1	-1		1		-1	-1	

does not belong to any known *A. thaliana* pathway; and genes involved in biosynthesis and break-down of pyrole-2-carboxylic acid which is produced from proline degradation in some mammalian systems, are still not known. Even though the biosynthesis and regulatory role of these metabolites in Arabidopsis is still not understood, they do undergo significant change in response to elevated CO_2 as can be seen from Table 5-4. Citramalate is another such metabolite which is present in significant quantity and its role in plants is not well understood.

In addition metabolomic analysis also identified variation in Adenosine-Adenine, Ascorbate-Threonate concentration. Adenosine and Adenine are important plant metabolites and are involved in number of important cellular processes such as: Adenosine triphosphate (ATP) and Adenosine diphosphates (ADP) production, Cyclic AMP (CAMP) biosynthesis, Nucleotide / Purines production, methyl donor S-Adenosyl-Methionine (SAM) production and recycle. Unlike most metabolites in response to elevated CO_2, adenosine and adenine do not show increased production. In fact adenine was negatively significant at 3h, 6h, 12h, 24h, 30h

Table 5-4: Significance level of other know metabolite pools at individual time points and paired-SAM analysis. Positively and negatively significant metabolites are as described in Table 5-1.

		1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Erythritol	alc	1		-1			-1	-1		
Pyrrole-2-carboxylic acid	acid		-1	-1	-1			-1	-1	
Citramalate	acid		-1	-1		1		-1	-1	
Adenoise	sugar			-1	1	1		-1	-1	
Adenine (2TMS)	purine		-1	-1		-1		-1		
Threonate	acid	1		-1		1		-1	-1	
Ascorbate	acid	1		1		1			1	

This suggests increased consumption of adenine for biosynthesis of other metabolites/macro-molecules mentioned before. In contrast in response to elevated CO_2, Ascorbate (Vitamin C - an important plant anti-oxidant) metabolite pool showed increase at 1h, 6h, 12h and 30h. Threonate which is produced by oxidation of ascorbate however showed a slightly different variation in significance level and was found to be negatively significant at 6h, 24, 30h suggesting a decrease in ascorbate oxidation at this timepoints.

Unknown Metabolite Pools

In addition to the response of known metabolite pools to elevated CO_2 discussed in this section so far, GC-MS metabolomic analysis also identified additional 226 additional unknown metabolites. From these 226, it was possible to classify about 102 unknown metabolites into broad functional categories such as sugars, phospho-derivatives, secondary metabolites, organic and fatty acids (unknown amine containing metabolites were removed from the analysis during data correction and normalization). In some case the mass-spectra also hinted towards the possible metabolite structure by similarity to other known metabolites. The significance levels of these unknown metabolites are available from Appendix II – Table A2-1.

5.1.6 Conclusions

In summary, time-series metabolomic analysis of the short-term elevated CO_2 response of *A. thaliana* liquid cultures indicated following:

- In response to elevated CO_2, up to 12-15 fold increase (and decrease) in the metabolite concentrations were observed in the free-metabolite pools. All Metabolite (except 3 unknowns) pools showed significant increase or decrease in concentration at least at one time point. A large variation in number of metabolites showing positively/negatively significant level was observed.

- Elevated CO_2 levels, increased significantly concentration of the majority of free cellular metabolite pools during the first 1h of the treatment. These increased metabolite pools gradually decrease till 9h either due to feedback inhibition or due to increased utilization for production of macro-molecules. Between 9-12 hours and between 24-30 hours this trend is again reversed suggesting changes in carbon fixation pathways.

- Metabolite pools show large variations with time in response to elevated CO_2 giving rise to large variation in significance level of the metabolites. Most metabolite pools at one or other time point.

- Photorespiration is inhibited in response to elevated CO_2 stress.

- TCA cycle metabolites show two unique responses to elevated CO_2 separated by the conversion of α-ketoglutarate to succinate. This shows differential regulation

of parts of the same pathway, separated at the points where TCA cycle intermediates are exported for amino acid production.

- Amino acid metabolite pool increases during first 12h suggesting increased nitrogen assimilation. The amino acid metabolite pools decrease beyond 12h due to either inhibition of nitrogen assimilation or increased macro-molecule production. The exceptions were homoserine and methionine which did not show the decrease in metabolite pools beyond 12h; and tyrosine which did not show significant increase during 12h.

- In addition to primary metabolism, polar GC-MS metabolomic analysis also identified response of secondary metabolites, lipid, sterols and PHBs biosynthesis pathways identifying specific metabolites showing significant increase / decrease in response to elevated CO_2.

- GC-MS metabolomic analysis identified several metabolites which are detected in significant amounts and undergo significant change in response to elevated CO_2 however their roles in *A. thaliana* physiology is still not known and need to be further investigated.

5.2 SALT (NACL) STRESS RESPONSE

Salt or Osmotic stress is a common problem faced in commercial agriculture due to fluctuations in water levels, salt deposits from over-irrigation or the basic nature of the soil where the plants are being grown. Considerable amounts of resources have been devoted in public as well as private industry research for developing verities of plants with better salt tolerance strains. The increases acceptance of Genetically Modified (GM) plants in the Americas and the interest in plants for energy production has provided additional thrust to research in this area. In this section, the results of the first-ever time-series metabolomic analysis of salt stress response of whole plants is provided which is likely to further aid the current efforts to engineer more salt tolerant plants.

5.2.1 Review of Osmotic (Salt) Stress Response in Plants

Osmotic stress is a common stress encountered by most biological systems and hence the response to osmotic stress in last few decades have been studied in various systems such as plants, Bacteria, Yeast and Mammalian cell culture systems. It is an important defense mechanism in biological system necessary for their survival. All living organisms (bacteria (Verala et al., 2003), plants (Taiz, 2002)) are known to respond to osmotic stress by changing concentration of Osmoprotectants at the metabolic level. Specifically in case of *A. thaliana,* previous studies with NaCl stress (Essah et al., 2003; Taii et al., 2004) have shown that 50 mM and 250 mM salt stress could not be sustained by plants for more than 4 and 2 days respectively. In plants the NaCl salt stress affects the cellular physiology in two different ways:

- Increased salt concentration in the environment affects the osmotic balance within and outside the plant cells resulting in water transport out of the cells to maintain the water potential (Taiz, 2002).

- In addition to osmotic stress, presence of increased Na^+ and Cl^- ions also cause oxidative stress, the exact mechanism of which has still not been completely understood (Hamilton and Heckthorne, 2001)

The first stress condition of salt stress experienced by plant is also similar to conditions encountered during drought and cold stress conditions. Similarly the second condition is similar to oxidative stress created by hydrogen peroxide / ozone. Hence a number of transcriptional analysis experiments have also been directed towards identifying common regulatory elements between these stress responses. Previous analyses of osmotic stress in plants have revealed plants counter these two stress conditions in the short-term by:

- Increasing concentration of Osmoprotectants to balance the water potential stopping water loss out of the cell (Hayashi et al., 2000)

- Increasing production of anti-oxidants to counter oxidative stress (Munne-Bosch and Allegre, 2003)

- Regulating activity of transporter proteins to regulate ion influx/efflux between external environment-cell and vacuole-cell (Hamilton and Heckthorne, 2001).

These known stress response strategies of plants have motivated years of research in identifying the Osmoprotectants used by plants, regulation of their biosynthesis (Apse and Blumwald, 2002; Ronetein et al., 2002, Nuccio et al., 1999) and understanding

mechanism of their osmo-protection, some of which is still being actively investigated. In case of *A. thaliana* following Osmoprotectants have been identified:

- Proline (Hayashi et al., 2000)
- Glycine betaine (N,N,N- trimethyl Glycine) (Sakamoto and Murate, 2001)
- Raffinose family oligosaccharides (Galactinol) (Taji et al., 2002)
- Sucrose (Hamilton and Heckthorne, 2001)

In addition studies of osmotic stress response in other plants and study of helophytes which have the natural ability to with stand high osmotic stress conditions have also suggested:

- Inositol derivatives like onitol, pinitol (Vemon et al., 2007)
- Polyamines like spermine, spermidine, putrescine (Sannazaro et al., 2006; Tang, et al., 2007);
- Other Betaine compounds such as proline betaine (Trinchant et al., 2004) and β-alanine betaine (Bala et al., 2002)
- Polyols (Nelson et al., 1999)

In addition, trehalose is another such sugar which is known to be an important osmoprotactant in many eukaryotic and prokaryotic organisms. However osmotic stress response studies have indicated that trehalose does not have a significant role as osmoprotactant in plants (Wingler, 2000). The common antioxidants which have been

known to be utilized for oxidative stress response of Salt stress are (Hamilton and Heckthorne, 2001):

- Ascorbate (Vitamin C)
- A-tocopherol (Vitamin E)
- Glutathione
- Anti-oxidant enzyme (Catalase, Copper/Zn dimutase)

In recent years there have been several efforts to engineer more osmotolerant plants by introducing gene(s) involved in biosynthesis of Osmoprotectants or related regulatory enzymes (Apse and Blumwald, 2002; Ronetein et al., 2002, Nuccio et al., 1999) which have met with mixed success in achieving the desired goal. Better understanding of the stress response in plants at metabolic level will allow more efficient engineering strategies for introducing osmoprotection trait in commercial agricultural and bio-energy crops. The results obtained from the current analysis are described in the following sections.

5.2.2 Metabolomic profiling results

Polar Metabolomic Profiles were obtained using the optimized protocol and experimental conditions described in Chapter 3 and Chapter 4, respectively, of this report. One of the liquid cultures (harvested at 12h in NaCl treatment) weighed abnormally low (12.8 gm), about 53% of average of all sample (22.8 gm) and 50% of its bio-replicate (15.2 gm), and was also identified as an outlier in metabolomic analysis and hence was removed from

the analysis. In addition 2 out of 4 and 1 out of 2 *A. thaliana* liquid cultures representing 0h and 1h, respectively, were removed mid-way through the analysis due to media contamination and hence were not used in the current analysis. In the acquired metabolomic profiles, 550, among which 147 known, metabolite methoxime (Meox)-trimethylsilyl (TMS)-derivative peaks were detected. After data correction, normalization and filtering, the set of metabolite peak areas that were finally considered in the analysis included 82 annotated and 169 unidentified. Among the 82 known metabolites, 55 correspond to metabolites forming only one TMS-derivative, 9 to one of the two geometric isomer derivatives of ketone-group containing metabolites (see Chapter 3), and 18 to cumulative peak areas of amine-group containing metabolites. Each timepoint (except control 1h, 12h) was represented by at least 4 (2 biological x 2 instrumental replicate) spectra. Data correction, filtering and normalization were done as described earlier in Chapter 3 to obtain normalized peak areas for control and perturbed experiments. The log ratio of this normalized peak profiles is shown in Figure 5-10. As can be seen from the figure up to 30 fold increase as well as decrease in metabolites is observed in response to 50 mM NaCl stress in plants. From the 82 known metabolites 10 showed more than 4 fold increase at at-least one time point (Trehalose, Galactinol, Galacturonase, Xylulose, Glycerol, Indole-3-acetamide, Glutamine, Allantoin, N-acetylglutamate, β-alanine) where as 5 known metabolites showed more than 4 fold decrease at, at-least one time point (Glutamine, allantoin, adenosine, valine, nicotinate).

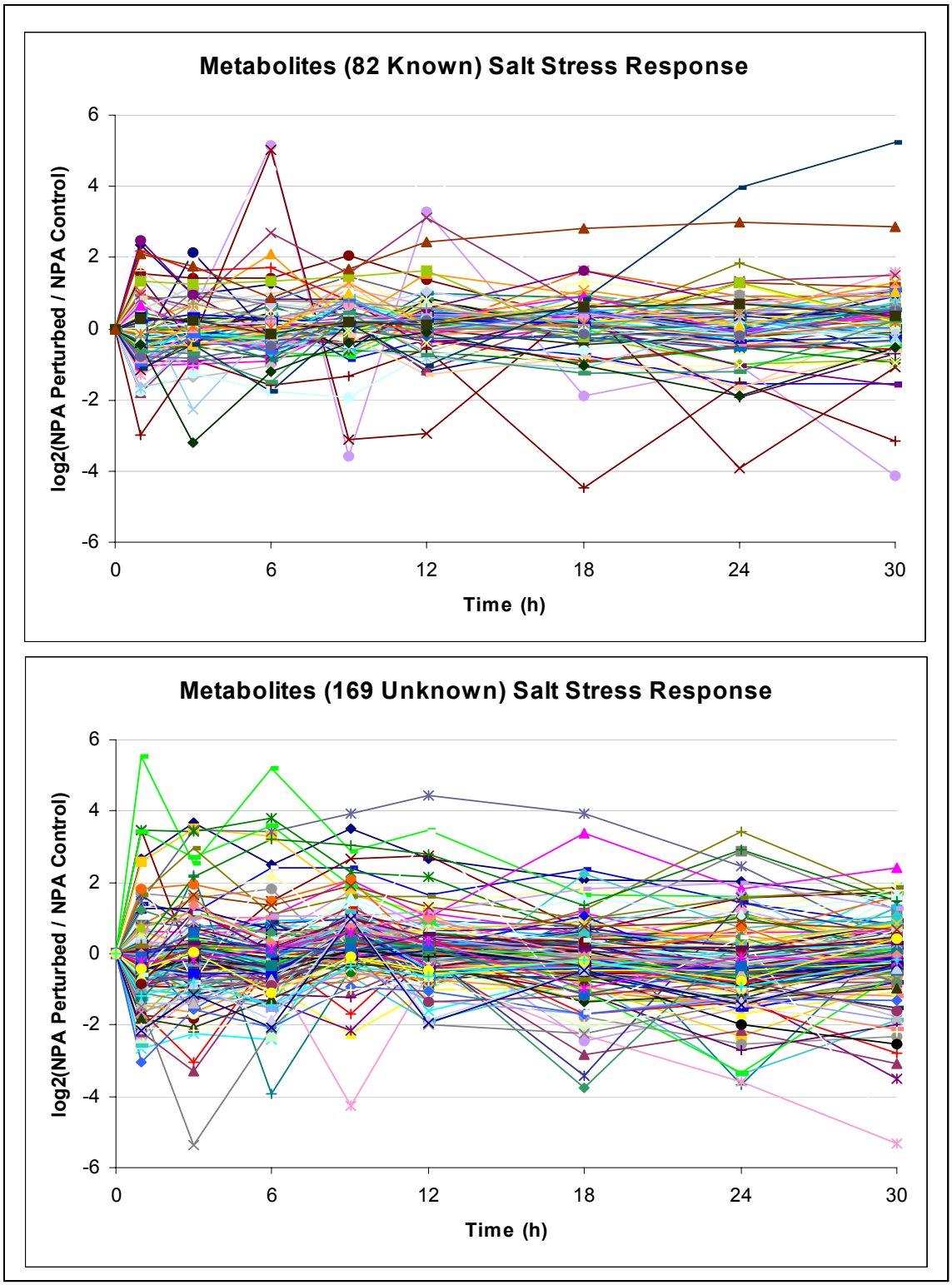


Figure 5-10: Log_2 Ratio Normalized Peak Area (NPA) profiles of Perturbed (NaCl Stress) to NPA Control profiles indicate up to 30 fold increase - decrease in metabolite levels.

5.2.3 PCA Analysis

According to TIGR MeV Principal Component Analysis (PCA), the control metabolomic profiles can be clearly differentiated from their perturbed counterparts (Figure 5-11). This implies that the physiology of the plant liquid cultures is affected significantly by the applied perturbation at the metabolic levels, even during the first 30h of treatment. It is further interesting to note that most of the variation in metabolism in response to NaCl stress is along principal component 1 which accounts for 45% of the total variation and is almost perpendicular to the time variation in the control samples which is mostly along principal component 2 which accounts for only 14% variation. PCA analysis thus clearly indicates that the salt stress creates a significant response in the *A. thaliana* metabolism.

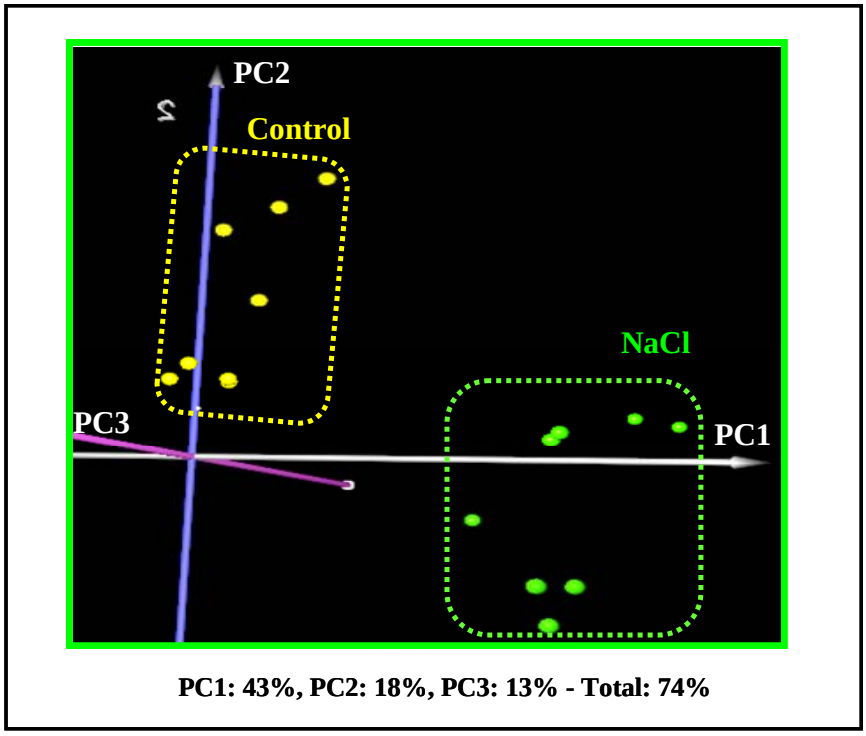


Figure 5-11: Principal Component Analysis (PCA) of Salt Stress response using TIGR MEV 3.0 shows a significant difference in metabolism of A. thaliana liquid cultures even during the 1-30h.

5.2.4 SAM and MiTimeS Results

Paired-SAM analysis identified (delta=1.5, 1.2% FDR) identified 32 (11 known) and 48 (11 known), respectively, positively and negatively significant metabolites in response to 50 mM NaCl stress (Figure 5-12). MiTimeS analysis (Dutta et al., 2007), on the other hand, identified many more positively and negatively significant metabolites at the individual timepoints (average 67 and 72, respectively) for the same significance threshold value and a slightly higher average median FDR (1.7%) (Figure 5-12). The significance level profiles over time along with the paired-SAM result for all known metabolites are discussed individually in subsequent sections in the context of their biology.

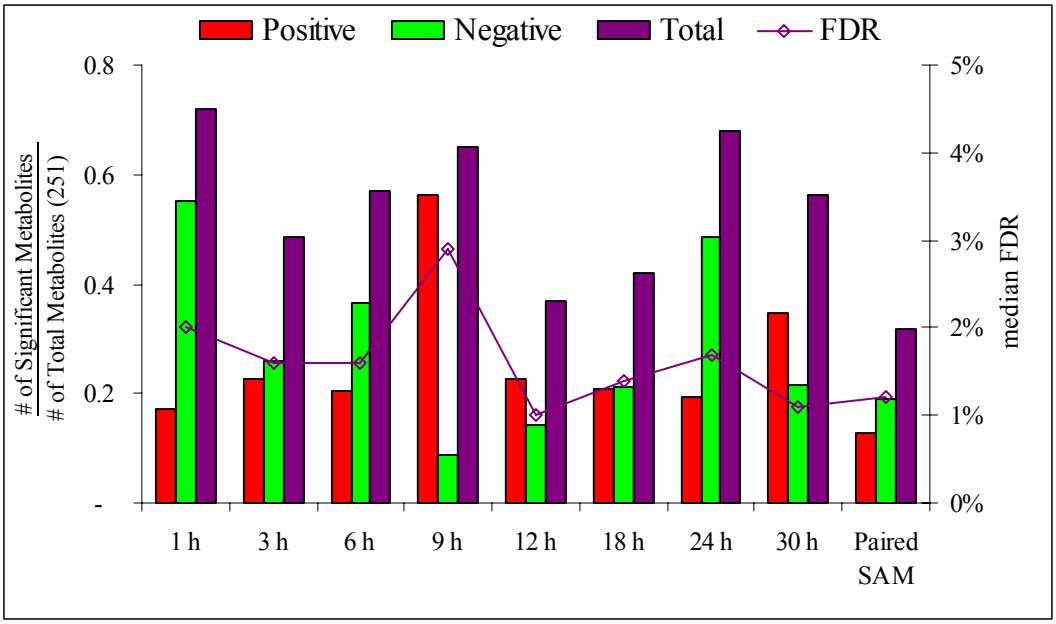


Figure 5-12: Number of Positively, Negatively and Total significant metabolites along with % median False Detection Rate (FDR) obtained for overall analysis and individual time points, using paired-SAM and MiTimeS analysis.

MiTimeS analysis (Dutta et al., 2007) also allowed calculation of Significance Variability Score (SV Score) which represents the amount of fluctuations in the significance level for each metabolite on a scale of 0-2 as described earlier in Chapter 3. SV score of 0 represents the metabolite does not undergo change in its significance level and maintains the same significance level at all time points. In contrast SV score 2 represents oscillatory metabolite, i.e. the metabolite changes its significance level from positively significant to negatively significant and vice-versa at each and every time point. The SV score distribution which plots the SV score against the fraction of the detected polar metabolites at each SV score is shown in Figure 5.13. As can be seen from the figure, in response to salt stress, 11 metabolites (4.4% of total) were observed in the same significance level at all timepoints (i.e. SV score = 0). From these 11, 2 were known metabolites: Xylulose was positively significant at all time points and Glycerate was negatively significant at all time points. The other 9 were unknown metabolites from which 5 and 4 metabolites were positively and negatively significant, respectively, at all time points. This again indicates none of the metabolite was non-significant at all time point, i.e. all the detected metabolites showed significant variations at least at one time point. The highest dynamics was shown by 2 metabolites (2-oxoglutarate and a unknown sugar) with SV score 1.29.

According to Figure 5-12, the highest numbers of total and negatively significant metabolites (~70%) were identified at 1h and 24h where as the highest number of positively significant metabolites were identified at 9h. These time points also showed maximum difference between number of positively and negatively significant metabolites, as for the rest of the time points they were more or less equal. A possible

explanation for these variations could be that at 1h, in response to salt stress the plant produces the necessary Osmoprotectants and proteins required to counter the salt stress. These create a significant drain on the metabolite pools giving rise to a large number of negatively significant metabolites at 1h (Figure 5-12). Gradually from 1h-9h, through improved carbon fixation or increased breakdown of carbon storage compounds like sucrose and starch, these metabolite pools are replenished and the production of Osmoprotectants and metabolite precursors for the necessary macromolecules is increased. This gives rise to gradual decrease in number of negatively significant metabolites and increase in number of positively significant metabolites (Figure 5-12). Beyond 9h, possibly, the plant has utilized the majority of stored carbon source or is unable to fix more CO_2 and hence the opposite phenomenon is observed from 12-24 h with the highest number of negatively significant metabolites (Figure 5-12).

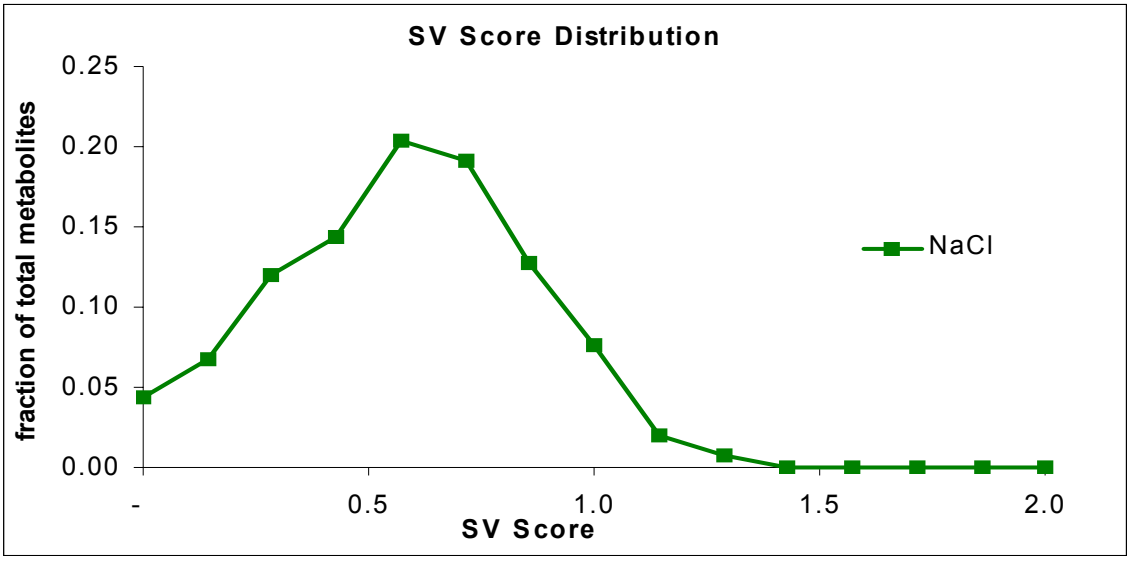


Figure 5-13: Significance Variability (SV) Score distribution of metabolites in response to NaCl stress shows the overall dynamics of the system. Note almost 5% of the total metabolites show the same significant level at all time points and hence has SV score 0.

The time point correlation network based on, both the positively and negatively significant metabolites (Figure 5-14A) indicate strong correlation between timepoints 1h-6h and between 18h-30h with the 9h time point showing the most unique response with least number of connections in both positively and negatively significant SCM networks. These suggests 9h to be the turning point at which, most probably the plant shifts from the immediate response to salt stress to a more long term strategy to resist salt stress. The detailed difference between these two responses will be discussed in more detail in the context of individual biological pathways.

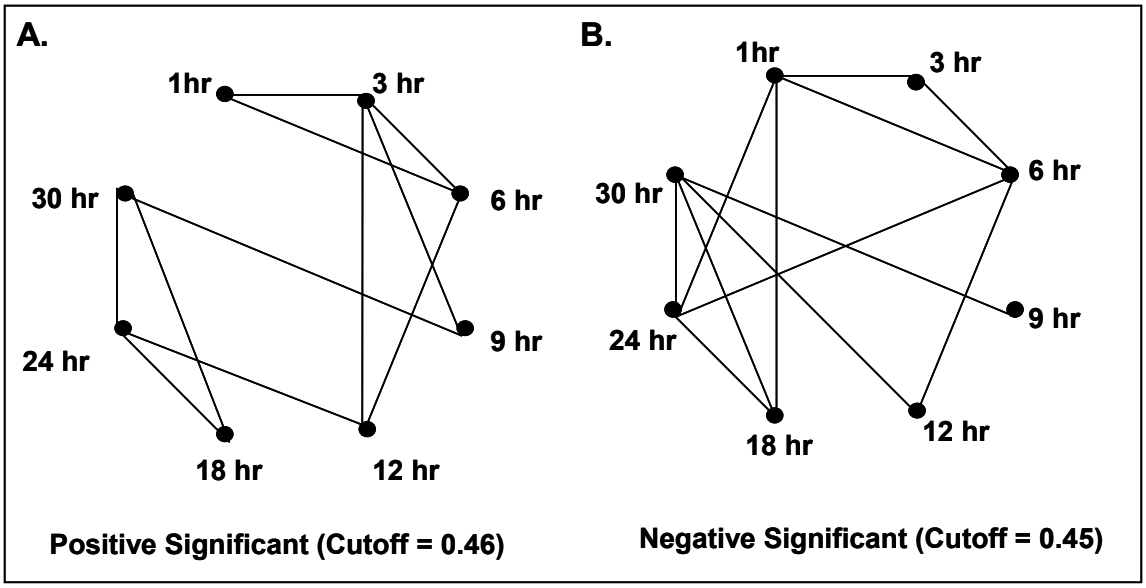


Figure 5-14: Time point correlation networks based on the common between time points (A) positively and (B) negatively significant metabolites in response to salt stress. Two time points are connected, if their correlation coefficient is larger than the indicated threshold, the latter being selected in each case as the average of all between different timepoints correlation coefficients.

5.2.5 Analysis of Individual Pathways

Photorespiration Pathway

Carbon and oxygen "compete" for Rubisco activity (Figure 5-15) as the Rubisco catalyses both the carboxylation and oxidation reaction. The pathway for regenerating Calvin cycle intermediate from the oxygenation product is known as Photorespiration pathway. Though effect of changes in the CO_2/O_2 ratio has been shown to affect the flux distribution between the two pathways (Siedow and Day, 2001; Dey and Harborne, 1997; Coschigano et al., 1998) however the effect of salt stress on photorespiration pathway has not been studied before. As can be seen from Figure 5-15, in response to salt stress two of the three organic acids, Glycolate and Glycerate were negatively significant at 1-12h, 1-30h and paired SAM respectively. The third organic acid glyoxilate, which also takes part in the glyoxilate pathways which is an alternative to TCA cycle for lipid respiration, was positively significant. This suggests for the first time, that photorespiration may be inhibited in response to salt stress. However the increase in glyoxilate would require more explanation since it was not possible to determine the change in glyoxilate level in photorespiration pathway and glyoxylate cycle pathway independently from metabolomic analysis alone.

Tri-Carboxylic Acid (TCA) Cycle Metabolites

Tri-Carboxylic acid Cycle (Figure 5-16) is the primary aerobic respiration pathway in the mitochondria for most eukaryotic systems (Tiaz, 2000). TCA cycle is responsible for the production of energy in the form of ATP and reducing power in the form of NADH and

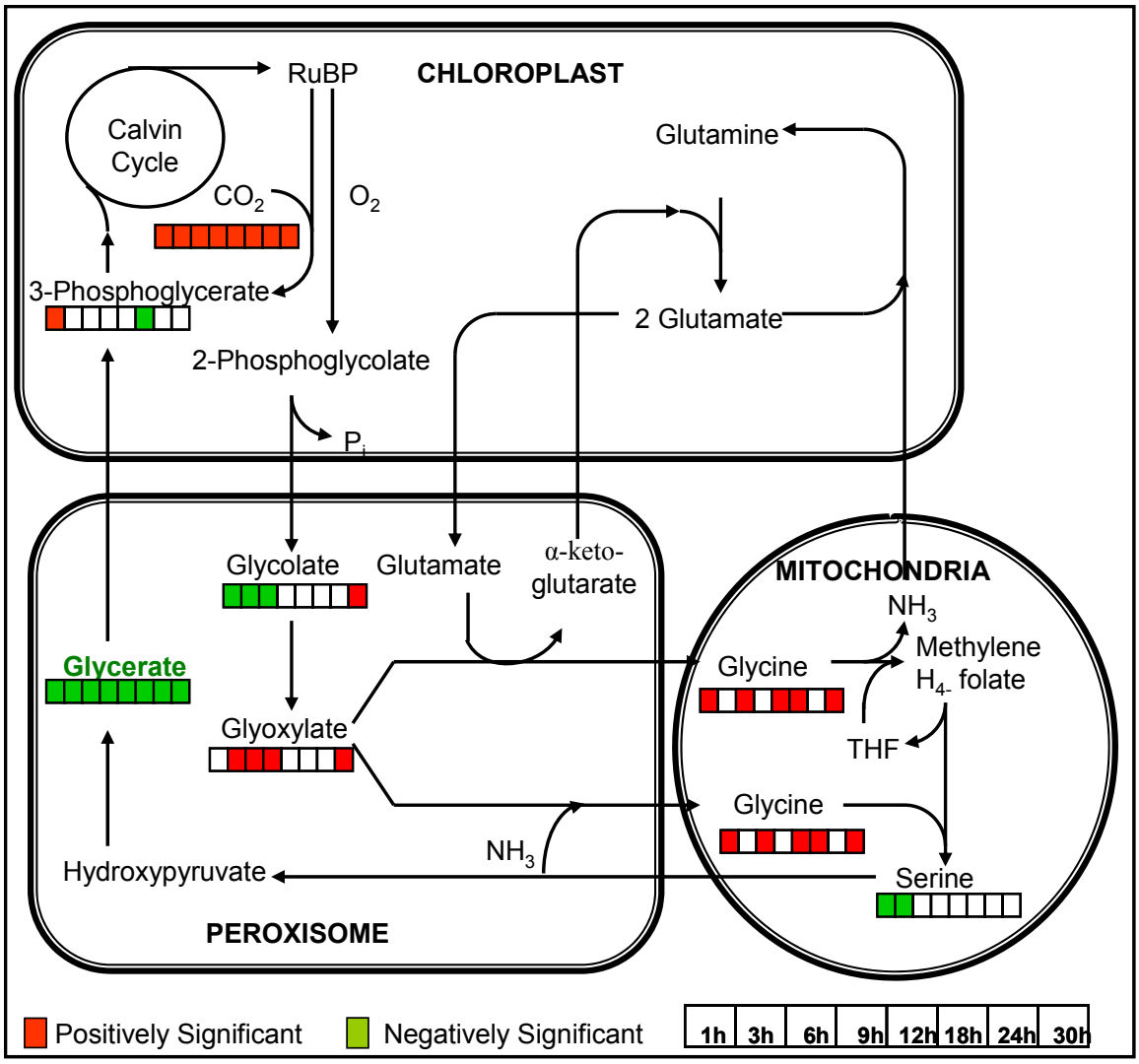


Figure 5-15: Observed effect of the salt stress on the physiology of photorespiration, at the metabolic levels. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

$FAdH_2$ by oxidation of pyruvate (derived from sucrose) into CO_2 (Tiaz, 2000). Apart from the osmotic stress effect of the salt stress, the salt stress is also known to cause oxidative stress which is known to significantly affect mitochondrial function. In fact, mitochondria are the first locations in cells for salt stress damage in plants (Hamilton and Heckthorne, 2001). However the exact mechanism of the same is still being studied (Hamilton and Heckthorne, 2001) and its effect on the TCA cycle in plants has not been

studied before. As can be seen from Figure 5-16, in response to salt stress, the TCA cycle intermediates from citrate to fumarate show significant increase in their concentration with the highest increase in concentration observed for aconitate, fumarate and iso-citrate, two of which were also found to be positively significant from paired-SAM analysis. In contrast, malate showed significant decrease in its concentration in response to the salt stress at all time points except 9h and also from overall analysis.

These results suggest increased activity in the TCA cycle to meet the increased ATP demand during salt stress. Prior metabolic flux analysis of *Corynobacterium glutamicum* under salt stress conditions also indicated an increased flux through glycolysis and TCA cycle pathway. The only exception of the metabolite pool of malate could be due to following reasons:

- Role of malate in maintaining ion balance between vacuole and cytoplasm which is known to balance Cl^- ions (Netting, 2001)
- Role of malate in regulating guard cell movements which regulate stomata closing, one of the response to stop water loss by plants (Netting, 2001)
- Possibility of activation of C4 / CAM pathways in which malate stored in vacuoles is converted to OAA and the carbon released is fixed by the Calvin Cycle. This pathway allows plants to close vacuoles and generate a very high concentration of carbon within the cells. This pathway also inhibits photorespiration as the CO_2 concentration generated using this pathway is much higher.

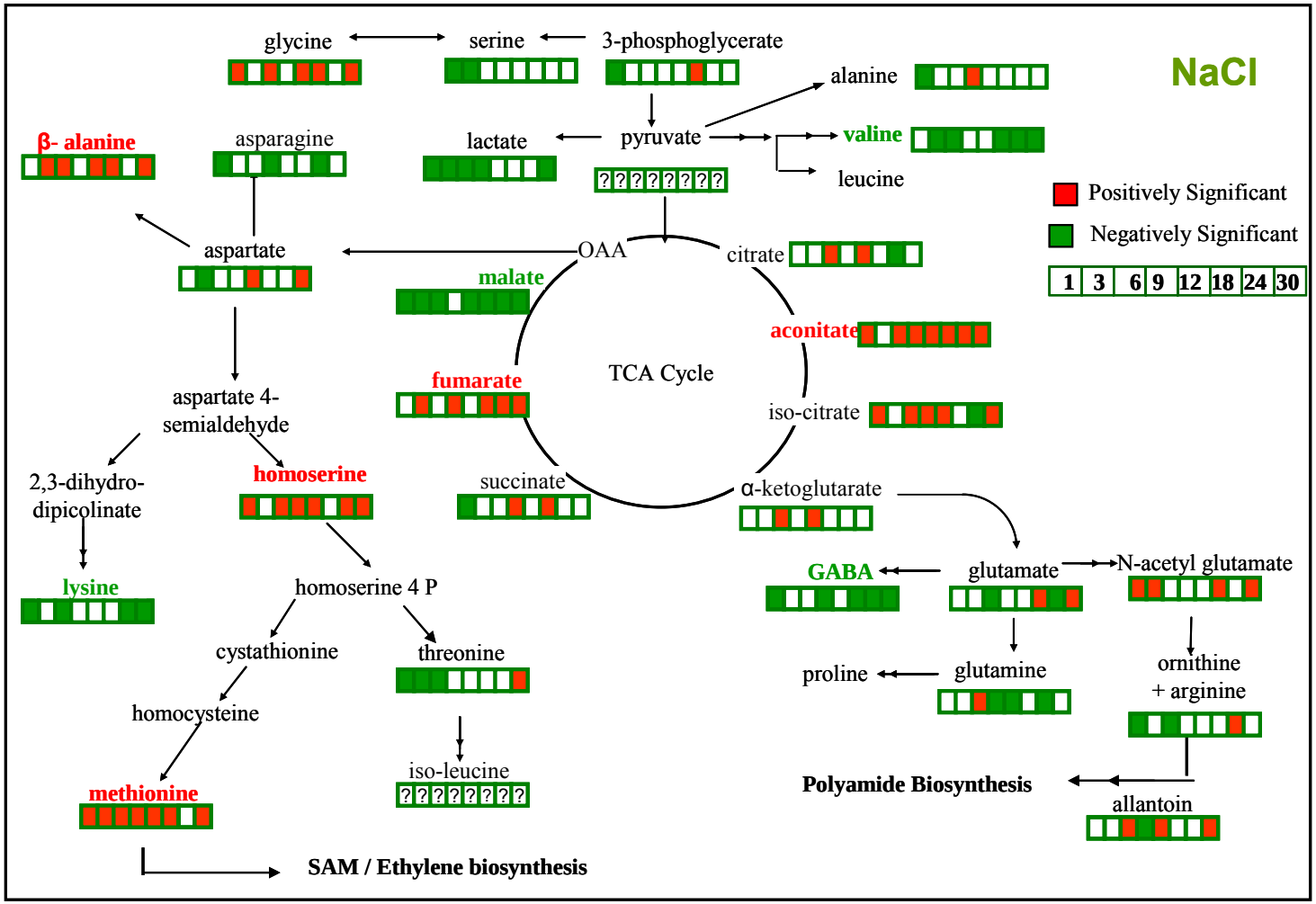


Figure 5-16: Observed effect of the applied perturbation on the physiology of the Tri Carboxylic Acid (TCA) cycle and amino acid biosynthesis at the metabolic level. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

Since the current metabolomic analysis used whole plant extracts, it is not possible to determine if the decrease in malate was due to reduced production from fumarate or due to the other reasons mentioned above. However there is still a strong possibility of increase in TCA cycle flux based on the rest of the metabolites. In addition the response also identified another unique aspect about TCA cycle. The conversion from citrate to iso-citrate and the reverse reaction are both catalyzed by the gene aconitate hydratase with aconitate as an intermediate. In response to elevated CO_2 however, aconitate shows a significantly more relative increase in concentration as compared to citrate and iso-citrate which are expected to be at near equilibrium concentrations. This can not be explained based on the known information about aconitate hydratase and needs further investigation.

Nitrogen Assimilation and Amino Acid Biosynthesis

Nitrate Reductase (NR) gene, is the first enzyme in the nitrogen assimilation pathway and light, nitrate ions and carbohydrates, have shown to induce expression of NR mRNA (Larios et al., 2001; Cheng et al., 1991; Cheng et al., 1992). Presence of salt stress is known to suppress NR expression an activity (Dluzniekwza et al., 2007; Debouba et al., 2006), however this suppression is also dependent on salt concentration and duration of stress. A number of heat shock, anti-oxidant and structural proteins are also induced in response to salt stress. Finally in some plant species and tress, polyamine compounds such as spermidine, spermine, citruline, etc are also expected to play the role of Osmoprotectants (Sannazaro et al., 2006; Tang, et al., 2007). The combined effect of this

is expected to drain the available organic/inorganic nitrogen pools in response to salt stress.

As can be seen from Figure 15.6, in response to salt stress from the four nitrogen storage and transport amino acids: aspartate, asparagine, glutamate and glutamine: asparagine and glutamine were negatively significant at four out of eight time points. Glutamine was also positively significant at 6h, which was also the time point when glutamate showed a significant decrease. Glutamate and aspartate were identified as positively significant as positively significant at later time points (18h,30h & 12h,30h) respectively. This together suggests as expected there is a slight decrease / no decrease during the first 9h and at later timepoints there may be slight increase / no change in the overall nitrogen pools as aspartate and glutamate show a moderate increase where as glutamine and asparagine show slight decrease during these time points.

From the other amino acids: homoserine, methionine, β-alanine showed significant increase at most of the time points and also from paired-SAM analysis of the overall-response. In addition, glycine, N-acetylglutamate and allantoin showed significant increase at 5, 4 and 3 time points respectively. These significant increases in amino acids can be explained based on known physiology as follows:

- The significant increase in homoserine and methionine in response to salt stress suggests an increased production of S-Adenosyl-Methionine (SAM) and Ethylene precursor 1-amino-cyclopropane-1-carboxylic acid. SAM is the most important methyl donor in plants and is especially need for biosynthesis of Glycine Betaine the main osmoprotactant in *A. thaliana*. Prior studies of stress response have also

shown increase in SAM methylation activity (Sanchez et al., 2004; Tabuchi et al., 2005). In addition, ethylene production is also known to be increased in response to salt stress (Cao et al., 2007).

- The significant increase in glycine confirms increased glycine-betaine levels, one of the two major known osmoprotectants in *A. thaliana.*

- The significant increase in β-alanine suggests that in *A. thaliana* liquid cultures, β-alanine betaine (which is produced by successive methylation of β-alanine by SAM) plays an important role as an osmoprotectant. This is the first time we show that β-alanine betaine contributes to *A. thaliana* stress response, as so far there is only study which has speculated this role in the plant family Plumbaginaceae (Bala et al., 2000).

- The significant increase in allantoin and N-acetylglutamate (at some time points more than 4 fold) suggests that polyamines which are produced using these intermediates also show a significant increase in response to salt stress. As discussed earlier, increased polyamine levels in trees and a number of other plant species have been shown to impart osmoprotection. The results from our metabolomic analysis indicate that the same also may be true for *A. thaliana* liquid cultures.

Along with this increase in amino acids, three amino acids: Lysine, Valine and GABA also showed a significant decrease at most of the time points and from paired-SAM analysis. This is also the first time that significant decrease in their metabolite pools in

response to salt stress has been identified. Another interesting observation from Figure 5-16 is that for each significant increase in the amino acid, we observe a significant decrease in the competing pathway. For e.g. homoserine-methionine show significant increase, however in its competing pathway lysine (and possibly iso-leucine whose concentration decreased in response to salt stress falling bellow detection limit at some timepoints and hence was removed during data normalization) showed a significant decrease in its concentration. Similarly: (a) β-alanine showed significant increase, but asparagine which competes for aspartate was negatively significant at 3 time points (b) N-acetylglutamate shows a significant increase but GABA (4-aminobutyric acid) shows a significant decrease at all time points (c) tyrosine shows significant decrease where as phenylalanine shows a moderate increase which is discussed later in secondary metabolism.

These examples suggest that in the immediate response to salt stress, production of specific amino acids is increased for production of osmoprotactants. However in many case, a significant part of this increase was possibly achieved by re-configuration of metabolic flux from the competing pathways. These also suggest nutrient limitation either for carbon or nitrogen in response to salt stress which makes re-distribution of flux a more efficient mechanism for production of osmoprotectants as compared to production from starch degradation, CO_2 fixation and NO_3^- fixation.

Sugar Metabolite Pools

In response to the salt stress several sugar and sugar based metabolites showed significant increase in their concentration as shown in Table 5-4. Specifically, Galactinol – which is

a combination of myo-Inositol and galactose, is known to be an osmoprotactant in various plants (Taji et al., 2002), showed significant increase at 7/8 time points and from overall analysis as well. Inositol, the precursor of galactinol also showed significant increase at 3 timepoints. Sucrose which is also known to be induced in response to stress showed significant increase at later time points beyond 6h confirming the similarity between the current and previous studies. In addition, since all known metabolites were monitored in the current analysis, we were also able to identify significant increase in Cellobiose at all time points beyond 6h and from overall paired SAM analysis as well suggesting possibly cellulose degradation in response to salt stress. Xylulose was another sugar which showed significant increase in response to salt stress, with more than four fold increase at several time points as described earlier. Xylitol which is produced from Xylulose also showed significant increase at 3 time points. Unlike other sugars whose main role is for carbon storage and assimilation or regulation by signaling, Xylulose with Arabinose are mainly used for producing hemi-cellulose family of compounds known as Xyloglucans and Arabinoglucans respectively. In the present analysis however we do not see an increase in arabinose, but specifically in Xylulose.

The change in hemi-cellulose, along with significant increase in cellobiose (produced from cellulose breakdown), suggests a novel long-term stress response strategy by plants in which the cell wall composition is changed to consist of more hemi-cellulose than cellulose. In addition, the hemi-cellulose composition would also undergo change to contain more Xyloglucans than Arabinoglucans. This is a novel observation, being made for the first time, and can provide an additional strategy for engineering more stress tolerant plants.

Table 5-5: Significance level of Sugar, Sugar Phosphates and Sugar Alcohols at individual time points and paired-SAM analysis in response to salt stress. Positively and negatively significant metabolites are as described in Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sucrose	sugar	P3840	-1			1		1		1	
Trehalose	Sugar	P3983			-1				1		
Cellobiose	Sugar	P4025			1	1	1	1	1	1	1
Glucose	Sugar	P2655							1	-1	
Fructose	Sugar	P2601							1		
Glucose 6 P	Phosph	P3548	-1	-1	-1			1			
Fructose 6 P	Phosph	P3513	-1				-1				
Inositol-1/2-Phosphate	Phosph	P3581	-1			1				1	
Arabinose	sugar	P2242	-1			1		1			
Rhamnose	Sugar	P2320	-1				-1	-1		-1	-1
Xylulose	Sugar	P2284	1	1	1	1	1	1	1	1	1
Xylitol	alc	P2230				1		1	1		
myo-Inositol	alc	P2910					1		1	1	
Galactinol	alc-sugar	P4270	1	1	1	1	1		1	1	1

Secondary Metabolite Pools

The three aromatic amino acids: tryptophan, tyrosine and phenylalanine; are the precursors for many secondary metabolites which are important for a host of important physiological functions in plants which were discussed earlier in Section 5.1.5. From the three amino acids, as with elevated CO_2 response, the tryptophan metabolite pools were very low and below the detection limit in the A. *thaliana* liquid culture plants. From the other two secondary amino acids, tyrosine showed a significant decrease in its metabolite pools at seven time points and also from paired SAM analysis as shown in Figure 5-17. In contrast, phenylalanine showed a significant increase at 4 time points. This suggests a partition of flux from tyrosine to phenylalanine derived products which play important role in stress response. This was further supported by the fact that benzoic acid and methyl benzoic acid, both derived from phenylalanine showed significant increase in

their metabolite pools at 4/8 and 6/8 timepoints as well as from paired-SAM analysis. Particularly between 12-30h, both benzoate and methyl benzoate showed significant increase. A related plant signaling molecule (2-hydroxybenzoate or Salicylic acid) also showed significant increase at 9h and 30h. In contrast, from the secondary metabolites produced from tryptophan, the concentration of nicotinate showed significant decrease at 6/8 time points and from paired-SAM analysis as well. The rest of metabolites, including plant hormone IAA's precursor 3-Indole-acetamide showed slight variation in from timepoint to timepoint. Thus it appears that in response to salt stress, phenylalanine and compounds derived from phenylalanine, showed significant increase in their metabolite negatively or non-significant between 3-30h. From these two, nicotinate was also identified as negatively significant from paired-SAM analysis suggesting relative pools, possibly with decreased flux towards tyrosine and tryptophan biosynthesis.

Butanoate Metabolism

Butanoate metabolism is based on precursors from TCA cycle and is important for the production of Poly-hydroxybutyrates (PHBs) as well as intermediates for ketone bodies. PHBs are currently being commercialized as an important renewable, biodegradable alternative to petroleum based polymers. The pathway producing PHBs in plants however, is still being studied. GC-MS metabolomic analysis identified several metabolites belonging to the known PHB pathways (2-hydroxyglutarate, 3-hydroxybutanoate, 4-hydroxybutanoate) as well as some related metabolites (3-hydroxy-3-methylglutarate, 2,3-dihydroxybutanoate, 3,4-dihydroxybutyrate) which may also be involved in Butanoate metabolism.

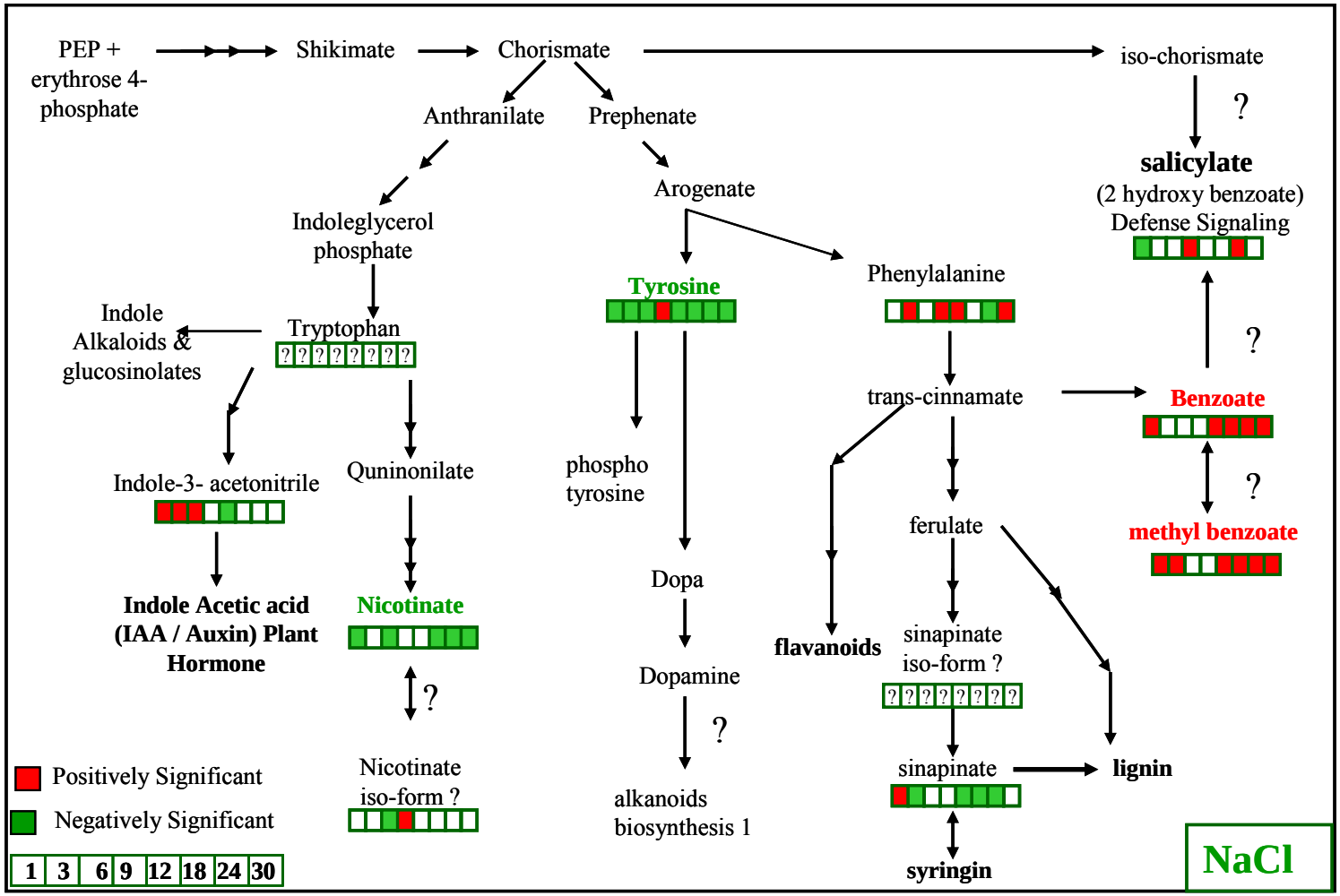


Figure 5-17: Response of metabolite pools in *A. thaliana* secondary metabolism pathways in response to 50 mM salt stress. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

As shown in Table 5-6, in response to salt stress, most metabolites with the exception of 3,4 hydroxybutyrate were negatively significant before 6h and beyond 24h. 3,4-hydroxybutyrate was the only metabolite which was negatively significant at 5/8 timepoints and from paired SAM.

Table 5-6: Significance level of metabolite pools belonging to Butanoate metabolism at individual time points and paired-SAM analysis in response to 50 mM salt stress. Positively and negatively significant metabolites are color-coded as described in the caption of Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
2-hydroxyglutarate	acid	P2184	-1		-1	1	-1		-1		
3-Hydroxy-3-methylglutarate	acid	P2197	-1				1	-1	-1		
4-hydroxybutanoate	acid	P1279		-1	1	-1	1			1	
3,4-dihydroxybutyric acid	acid	P1744	-1	-1	-1				-1	-1	-1
2,4-dihydroxybutanoic acid	acid	P1704	-1	-1	-1						

Lipid Metabolism

Lipid Metabolism showed a significant increase in response to salt stress during the first 12 hours. As can be seen from Table 5-7, all known FA and the sterols, tocopherol pools showed significant increase in their concentration during the first 12 hours. In addition the sterols and tocopherol showed significant increase in their metabolite pools also at 24 and 30h, as well as from overall paired-SAM analysis. This suggests an important role of steroids in NaCl salt stress response in Arabidopsis which has not been studied so far. From these metabolites only the tocopherol was known to be produced for its anti-oxidant property in response to the salt stress and it is the first time we also see a significant increase in other sterols and fatty acid pools during first 12 hours.

Other Metabolite Pools

The other metabolite pool group consists of metabolites (a) whose exact role and the

Table 5-7: Significance level of fatty acid and sterol metabolite pools at individual time points and paired-SAM analysis in response to salt stress. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Stearic Acid	FA	P3432		1	1	1			-1		
Linolenic acid	FA	P3461	1	1	1	1			-1		
alpha-Linolenic acid	FA	P3536	1			1			-1		
Gama Linolenic Acid	FA	P3509	1		1	1			-1		
Icosanoic acid	FA	P3747		1	1	1	1			1	
Stigmasterol	Sterol	**P5045**	1	1	1	1	1		1	1	1
CampeSterol	Sterol	**P5007**	1	1	1	1	1			1	1
Tocopherol	Pherol	**P4850**	-1	1		1	1		1	1	1

biosynthesis of the metabolites is still not known e.g. erythretol, pyrole-2-carboxylic acid, citramalate (b) isolated metabolite pairs like Adenosine-Adenine, Ascorbate-Threonate. From these metabolite erythretol shows a significant decrease in response to the salt stress. From Ascorbate-threonate, ascorbate is known to play a role in oxidative stress response and its metabolite pools decreased significantly falling below the detection limited in response to the salt stress, suggesting also a significant decrease. In the case of Adenosine-Adenine, which are involved in Adenosine triphosphate (ATP) and methyl donor S-Adenosyl-Methionine (SAM) metabolites, their concentration showed significant increase at 2/8 timepoints and 6/8 timepoints after a slight decrease in Adenosine concentration at 3h. This further supports increase in ATP, SAM production suggested by increase in TCA cycle and methionine biosynthesis pathway metabolites.

Unknown Metabolite Pools

The response of 90 unknown metabolites with at least some clue about their identity is shown in Appendix II (Table A2-2).

Table 5-8: Significance level of other known metabolite pools at individual time points and paired-SAM analysis in response to salt stress. Positively (1) and negatively (-1) significant metabolites are as described in Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sorbitol	alc	P2608	-1		-1	1					
Erythritol	alc	P1776	-1	-1	-1			-1	-1	-1	-1
Pyrrole-2-carboxylic acid	acid	P1716		1				-1	-1		
Citramalate	acid	P1868	-1			1		1	-1	1	
Adenoise	sugar	P4227		-1		1				1	
Adenine (2TMS)	purine	P3120	1	1		1		1	1	1	1
Threonate	acid	P2042	-1			1			-1		
Ascorbate	acid	P2948	?	?	?	?	?	?	?	?	

Some of the interesting biologically interesting observation are as follows:

- An unknown sugar metabolite showing a strong similarity to Maltitol shows significant increase in response to salt stress like Galactinol.

- An unknown FA showing close similarity to C-18 Fatty acids shows similar significance pattern as other FAs i.e. positively significant during first 12h and negatively significant at -24h.

From phosphor compounds, two unknown metabolites which are most probably Mannose-6-phosphate and ethanolamine phosphate show significant decrease in their metabolite pools at 7/8 time points and from paired-SAM analysis as well. Ethanolamine phosphate is known to be an important regulatory metabolite in the glycine betaine production pathway (Sakamoto and Murate, 2001) further supporting evidence of increase in its production.

5.2.6 Conclusions

In summary, time-series metabolomic analysis of the short-term 50 mM salt stress response of *A. thaliana* liquid cultures revealed following:

- In response to osmotic stress *A. thaliana* liquid cultures produce previously reported osmoprotectants like glycine betaine, sucrose and Raffinose-family oligosaccharides. In addition, they also produce β-alanine betaine and polyamines which have been known to be osmoprotectants in other plants.

- To counter the oxidative stress *A. thaliana* liquid cultures use tocopherol (Vitamin E) and ascorbate (Vitamin C). Where as in case of tocopherol a significant increase was observed in response to stress, in case of ascorbate, in contrast a significant decrease was observed, suggesting increased use of ascorbate without replenishing the supply by equivalent amount. In addition all the other known sterols were also found to increase significantly in response to the salt stress.

- TCA cycle metabolites showed a significant increase, specially fumarate, aconitate and iso-citrate, suggesting increased flux through TCA cycle to support increased demand for ATP.

- Malate was the only TCA cycle metabolite which showed a significant decrease in its concentration, giving rise to possibility of utilization of malate pools in vacuoles as either carbon source, regulation of guard cell movements or balancing Cl^- ions in vacuoles, which needs to be investigated further.

- Salt stress seems to have mild effect on the nitrogen storage and transport metabolite pools, suggesting not a significant change in nitrogen assimilation.

- Amino acids connected to the biosynthesis of osmoprotectants (β-alanine, homoserine-methionine, N-acetylglutamate, glycine) showed significant increase in their concentration. A large part of this increase is achieved by re-distribution of flux from competing pathways resulting in significant decrease of amino acids pools in those pathways (asparagine, lysine, GABA, serine). This also suggests possibility of limitation of organic carbon or nitrogen nutrients resulting in competition of resources.

- Another salt stress response strategy which was observed for the first time, is the possible re-configuration of the cell wall. Variations in sugar concentrations suggest possibility of increase in hemi-cellulose content, specifically xyloglucans, in the plant cell wall. In addition, at least in the short term, fatty acid concentrations also seem to be increasing.

The above results, suggest four novel strategies for engineering plants with higher osmotic stress tolerance which have not so far been tried out:

- Engineering plants to increase methionine and Adenosine biosynthesis increasing SAM production and/or increasing β-alanine betaine biosynthesis.

- Increasing hemi-cellulose xyloglucan production.

- Increasing increased carbon utilization for TCA cycle flux to increase ATP production, specifically increasing aconitate hydratase / iso-citrate lyase. A

possible strategy also could be to increase malate production outside the TCA cycle using genes such as malate synthase.

- Increasing sterol production by increasing flux through the HMG-CoA mevalonate pathway currently also being engineered in bacteria to increases anti-malaria drug Artemisinin production.

These four strategies independently or in combination with existing strategies to increase osmoprotectants proline, glycine betaine and polyamines, can significantly increase osmo-tolerance of plants. In addition, the present analysis also suggests a different mechanism for increasing methionine concentration in plants, a commercially important project for animal feed industry. The present analysis also shows the advantage of time-series metabolomic analysis to understand regulatory mechanisms in eukaryotic systems to develop novel strategies for engineering strains with desired traits.

5.3 TREHALOSE (SUGAR) SIGNAL RESPONSE

In plants, sugar concentrations play an important role in regulating biological processes at both molecular and physiological level. Trehalose, is a disaccharide of glucose, and is known to be an important osmoprotectants in most biological systems (Wingler, 2002). However in plants when genes involved in trehalose biosynthesis were discovered few years ago even their presence was a surprise to plant biochemists and physiologists (Goddijns and Smeekens, 1998). Several subsequent experiments in the current decade have shown that trehalose does not play a significant role as osmoprotectant (which was also seen in the osmotic stress response study described in earlier section). Instead trehalose has been known to regulate carbon storage and utilization between source and sink tissues (Eastmond and Graham, 2003; Wingler 2002). However the detailed analysis of how this is achieved and other processes regulated by trehalose has not been studied so far. In this section the results obtained by introducing 12 mM trehalose signal in the media of *Arabidopsis thaliana* liquid culture are presented which improve the existing knowledge of role o trehalose in regulating primary metabolism of plants. Since trehalose is known to regulate carbon assimilation and utilization in plants, any new insights would also be valuable for current efforts to engineer plants which are more efficient fixers of carbon dioxide.

5.3.1 Review of Trehalose (Sugar) Response in Plants

Trehalose is produced from UDP-Glucose and Glucose-6-Phosphate via Trehalose-6-Phosphate (T-6-P) as shown in Figure 5-18 (Wingler 2002). In spite of a plethora

trehalose synthesis genes present in *A. thaliana*(Layman, 2001), under normal conditions, very small amount of trehalose is accumulated due to very high activity of trehalose catabolism gene

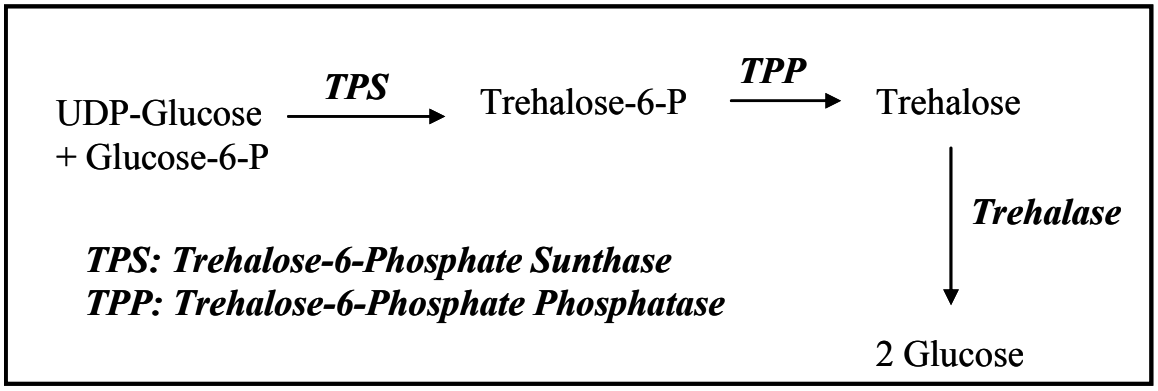


Figure 5-18: Trehalose metabolism in plants with key regulatory enzymes

trehalase (Wingler, 2002) which carries out hydrolysis of trehalose to two glucose molecules. After the discovery of these genes in Arabidopsis, subsequently similar genes have been found in most of the plants suggesting ubiquitous nature of trehalose in plant systems. However as mentioned earlier, unlike prokaryotic and certain mammalian system where trehalose is used for carbon storage and protection of proteins under stress (specially water and osmotic stress) (Wingler, 2002), in plants it was expected to play a role in carbon allocation (Wingler, 2002).

The initial hypothesis of the role of trehalose in plants was derived from yeast where it functions is to regulate sugar levels (Goddijns and Smeekens, 1998; Wingler, 2002). Specifically, trehalose was expected to increase carbohydrate storage where as its intermediate trehalose-6-phosphate (T6P) was expected to regulate flux of glucose into glycolysis by regulating hexokinase activity, as is the case is in yeast. However

subsequent experiments have shown that in plants, trehalose-6-phophate does not have the hexokinase regulatory activity. So far studies of trehalose activity have used (Wingler, 2002):

- Arabidopsis mutants in trehalose biosynthesis pathway
- Inhibitors of trehalase (validamycin A)
- Feeding studies feeding trehalose to plants

These experiments carried out mainly in last decade have shown (Wingler, 2002):

- Trehalose induces storage of carbohydrates in photosynthetic tissues
- In *Arabidopsis* trehalose induces expression of *ApL3* ADP-Glucose pyrophosphorylase gene which regulates starch biosynthesis in plants. These results in over accumulation of starch in leaves and reduction in root growth.
- In Barley, trehalose induced expression of regulatory genes involved in fructan biosynthesis from sucrose.
- Transgenic tobacco plants encoding *E-coli* trehalose phosphate synthase (TPS) gene which is expected to increase T6P levels show a more efficient photosynthetic activity per unit area.
- T6P accumulation due to over-expression of (TPS) or mutation in (TPS) results in shunted growth and leaf development.

- T6P is also supposed to regulate starch biosynthesis by post transcriptional redox regulation of *AGPase* involved in starch biosynthesis

Even though the results of the effects of trehalose and treahlsoe-6-phosphates have been now known for some time, the exact mode of this regulation in most cases is still not well understood. Further the effect of trehalose on the rest of the metabolism has not been investigated before. In the following sections, the results of time-series metabolomic analysis, the response of *A. thaliana* liquid cultures to trehalose feeding (12 mM) will be shown which gives a more system level understanding of regulation of metabolism by trehalose.

5.3.2 Metabolomic profiling results

Polar Metabolomic Profiles were obtained using the optimized protocol and experimental conditions described in Chapter 3 and Chapter 4, respectively, of this report. Specifically, in trehalose feeding experiments 2 out of 4 and 1 out of 2 *A. thaliana* liquid culture bio-replicates representing 0h and 1h, respectively, were removed mid-way through the analysis due to media contamination/irregular-growth and hence were not used in the current analysis. In the acquired metabolomic profiles, 550, among which 147 known, metabolite methoxime (Meox)-trimethylsilyl(TMS)-derivative peaks were detected. After data correction, normalization and filtering, the set of metabolite peak areas that were finally considered in the analysis included 89 annotated and 188 unidentified. Among the 89 known metabolites, 61 correspond to metabolites forming only one TMS-derivative, 9 to one of the two geometric isomer derivatives of ketone-group containing metabolites (see Chapter 3), and 19 to cumulative peak areas of amine-group containing metabolites.

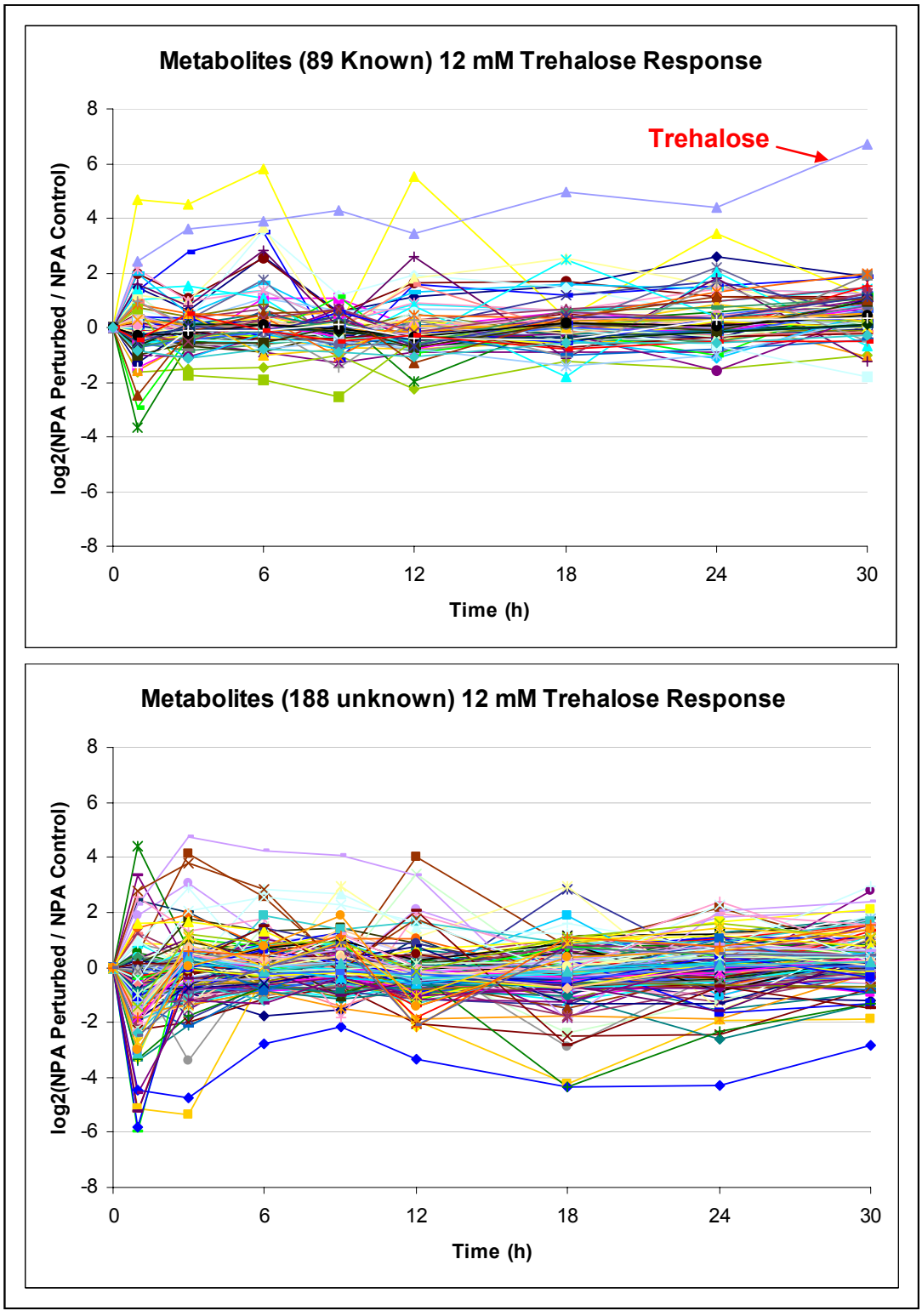


Figure 5-19: Log_2 Ratio Normalized Peak Area (NPA) profiles of Perturbed (Trehalose Signal) to NPA Control profiles indicate up to 20 fold increase - decrease in metabolite levels.

Each timepoint (except control 1h) was represented by at least 4 (2 biological x 2 instrumental replicate) spectra. Data correction, filtering and normalization were done as described earlier in Chapter 3 to obtain normalized peak areas for control and perturbed experiments. The log ratio of this normalized peak profiles is shown in Figure 5-19.

The highest variation in concentration observed was for intra-cellular trehalose. This is directly the result of exogenous trehalose supplied in the media. As can be seen from Figure 5-19, trehalose is represented by the blue line showing the highest increase in concentration in the known metabolite chart. The intra-cellular trehalose concentration increases significantly reaching 50 times the original concentration after 30h. Even though this result is expected result of a externally applied trehalose signal, it still confirms:

- External trehalose in transported into the cells by plants and its transport is not inhibited by internal trehalose concentration.
- In control plants, as discussed earlier from previous literature, trehalase activity ensures that trehalose concentration within the cell remains very low, however when trehalose is supplied exogenously, as can be seen from Figure 5-19, there is appreciable amount of trehalose accumulation, suggesting the trehalase activity is not increased significantly, if any, and this allows actual accumulation of trehalose and not just increased flux.
- Other than trehalose and another polyamine metabolite allantoin, most of the other known metabolites showed up-to only 4 fold increase and decrease (+/- 2 in

log_2 scale) in their relative concentration beyond 6 hours, suggesting smaller perturbations as compared to NaCl and CO_2 stress response.

5.3.3 PCA Analysis

According to TIGR MeV Principal Component Analysis (PCA), the control metabolomic profiles can be clearly differentiated from their perturbed counterparts (Figure 5-20). This implies that the physiology of the plant liquid cultures is affected significantly by the applied perturbation at the metabolic levels, even during the first 30h of treatment. It is further interesting to note that the trehalose treated plant group, has almost the opposite scores as compared to the control plants. Hence the variation with time and variation due to trehalose response are similar and are mostly accounted by principal component 1 which accounts for only 34% of the total variation.

5.3.4 SAM and MiTimeS Results

Paired-SAM analysis identified (delta=1.26, 0.8% FDR) identified 17 (10 known) and 23 (4 known), respectively, positively and negatively significant metabolites in response to 12 mM Trehalose signal (Figure 5-21). MiTimeS analysis (Dutta et al., 2007), on the other hand, identified many more positively and negatively significant metabolites at the individual timepoints (average 50 and 71, respectively) for the same significance threshold value and a slightly higher average median FDR (1.2%) (Figure 5-21). The significance level profiles over time along with the paired-SAM result for all known metabolites are discussed individually in subsequent sections in the context of their biology.

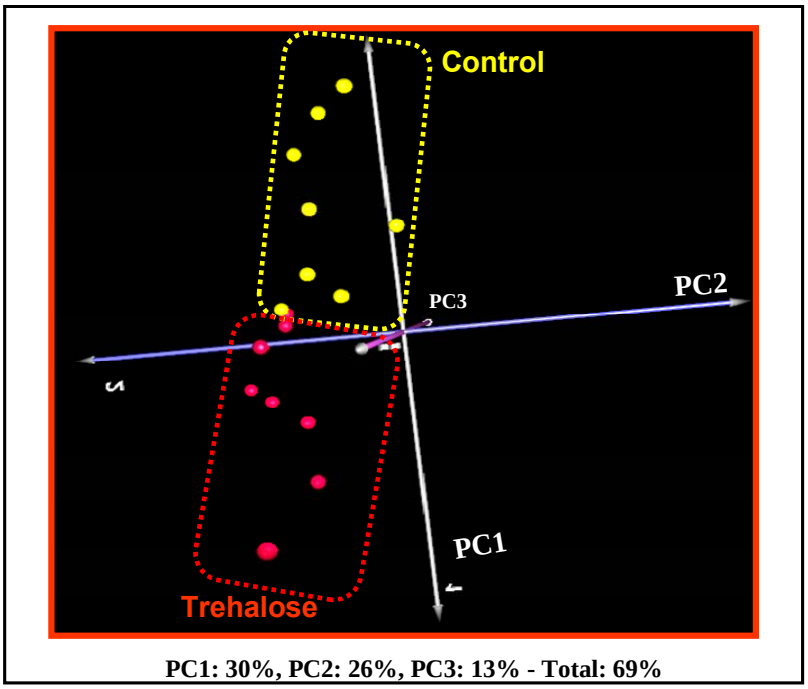


Figure 5-20: Principal Component Analysis (PCA) of Trehalose Signal response using TIGR MEV 3.0 shows a significant difference in metabolism of *A. thaliana* liquid cultures even during the 1-30h.

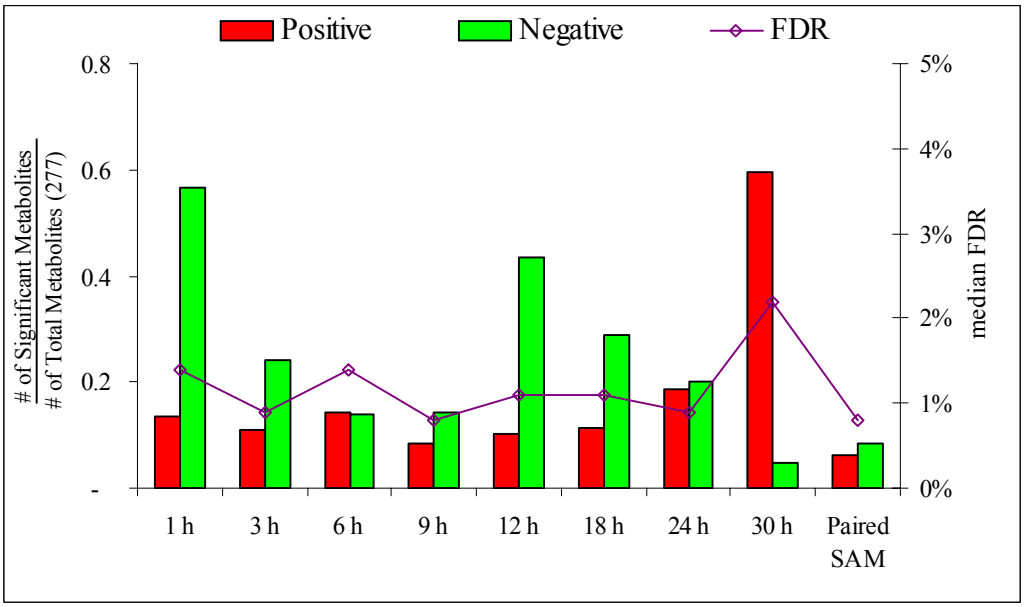


Figure 5-21: Number of Positively, Negatively and Total significant metabolites along with % median False Detection Rate (FDR), in response to Trehalose signal obtained for overall analysis and individual time points, using paired-SAM and MiTimeS analysis.

MiTimeS analysis (Dutta et al., 2007) also allowed calculation of Significance Variability Score (SV Score) which represents the amount of fluctuations in the significance level for each metabolite on a scale of 0-2 as described earlier in Chapter 3. SV score of 0 represents the metabolite does not undergo change in its significance level and maintains the same significance level at all time points. In contrast SV score 2 represents oscillatory metabolite, i.e. the metabolite changes its significance level from positively significant to negatively significant and vice-versa at each and every time point. The SV score distribution which plots the SV score against the fraction of the detected polar metabolites at each SV score is shown in Figure 5.22. As can be seen from the figure, in response to salt stress, 6 metabolites (2% of total) were observed in the same significance level at all timepoints (i.e. SV score = 0). From these 6, only 2 were known metabolites: Trehalose was positively significant at all time points and Tyrosine was negatively significant at all time points. The other 4 were unknown metabolites from which 3 and 1 metabolites were negatively and non-significant, respectively, at all time points. This also indicates only one metabolite was non-significant at all time point, i.e. almost all the detected metabolites showed significant variations at least at one time point. The highest dynamics was shown by only 1 metabolite Galactinol with SV score 1.29.

According to Figure 5-21, in response to trehalose stress the total number of significant metabolites is less than 40% of the total for most of the time points except 1h, 12h and 30h, where this was 70%, 53% and 64% respectively. Also at most of the time points, the number of negatively significant metabolites were higher or equal to positively significant metabolites except at 30h where almost 60% of metabolite pools showed a

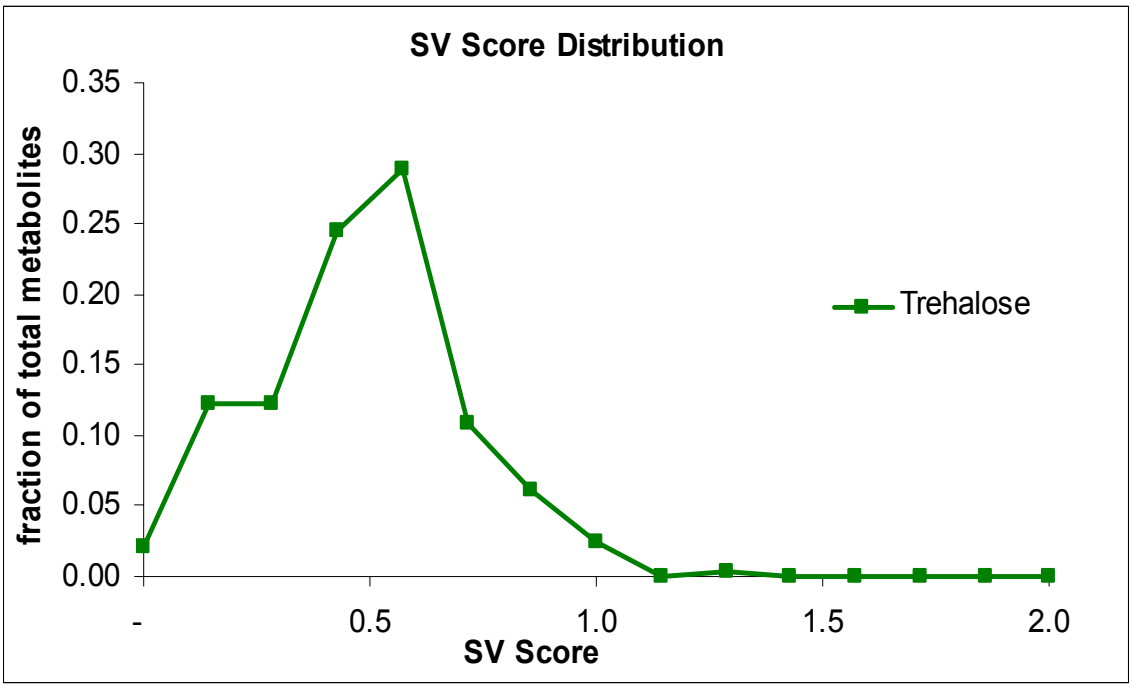


Figure 5-22: Significance Variability (SV) Score distribution of metabolites in response to 12 mM trehalose signal shows the overall dynamics of the system.

significant increase in comparison to 4% of the total metabolites showing significant decrease. This identifies 30h as having a unique response which could have been the result of very high trehalose concentration within the plant, as this is also the time point at which the trehalose concentration was the maximum. The result also indicates a possibility that trehalose might have different regulation of the metabolism at different concentration levels. At very high concentration, trehalose apart from having a regulatory role may also be utilized as a source of carbon increasing significantly majority of the polar metabolite pools. This however needs further investigation by either increasing the trehalose concentration in the medium or conducting the experiment for longer time.

The time point correlation network based on, both the positively and negatively significant metabolites (Figure 5-23) indicates strong correlation between consecutive

timepoints as well as between short (1h,3h,6h) and long time points (12h, 18h, 24h) except for 30h which showed the most unique response which was also indicated from the number of significant metabolites. The correlation within the time points suggests lower variability in the system which is also indicated by SV score distribution which had a lower mean and median as compared to NaCl and CO_2 stress. The biological significance of the response is discussed in the context of individual biological pathways in the following section.

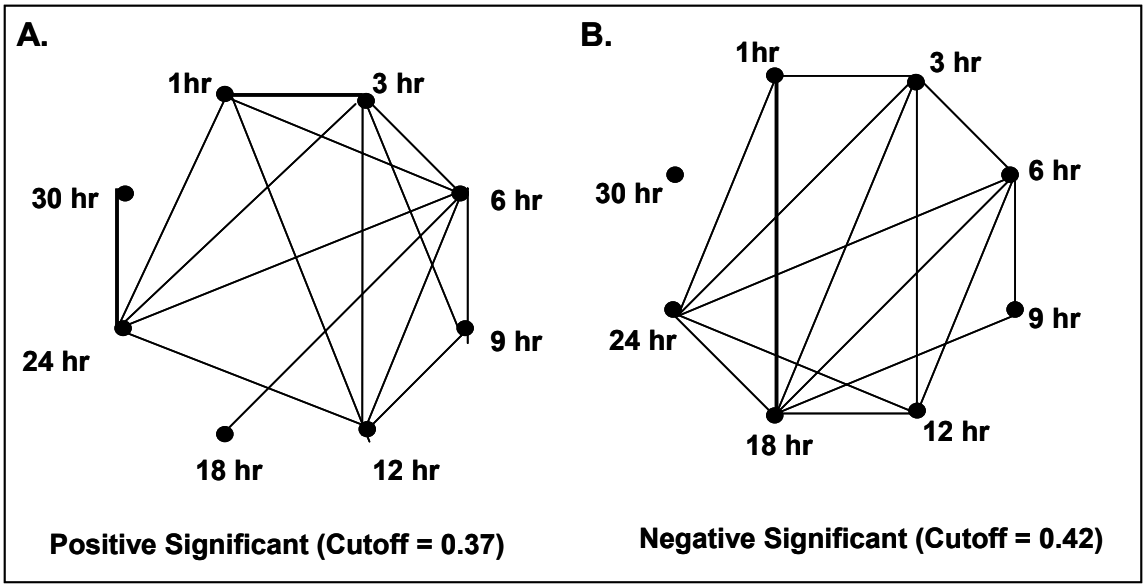


Figure 5-23: Time point correlation networks based on the common between time points (A) positively and (B) negatively significant metabolites in response to 12 mM trehalose signal. Two time points are connected, if their correlation coefficient is larger than the indicated threshold, the latter being selected in each case as the average of all between different timepoints correlation coefficients.

5.3.5 Analysis of Individual Pathways

Photorespiration Pathway

Carbon and oxygen "compete" for Rubisco activity (Figure 5-24) as the Rubisco catalyses both the carboxylation and oxidation reaction which form 3-phosphoglycerate and 2-Phosphoglycolate respectively. The pathway for regenerating Calvin cycle intermediate from the oxygenation product is known as Photorespiration pathway. As can be seen from Figure 5-24, in response to trehalose signal there was a mild decrease in the three organic acids in the pathway at 3/4 timepoints during first 24h. Only metabolite in the pathway showing significant response was Glycine which showed significant increase, however this was the result of overall increase in amino acid biosynthesis which is described later in this section. This suggests trehalose signal has no significant effect or very mild repression of the photorespiration pathway.

Tri-Carboxylic Acid (TCA) Cycle Metabolites

Tri-Carboxylic acid Cycle (Figure 5-25) is the primary aerobic respiration pathway in the mitochondria for most eukaryotic systems (Tiaz, 2000). TCA cycle is responsible for the production of energy in the form of ATP and reducing power in the form of NADH and $FAdH_2$ by oxidation of pyruvate (derived from sucrose) into CO_2 (Tiaz, 2000). As can be seen from Figure 5-25, trehalose does not have any significant overall effect on TCA cycle metabolites. Most of the metabolites are positively or negatively significant at 2/3 timepoints. However it is interesting to see that as observed earlier, even though citrate

and iso-citrate show similar variation in significance level, their intermediate aconitate does not follow the same trend suggesting additional regulatory elements for aconitate

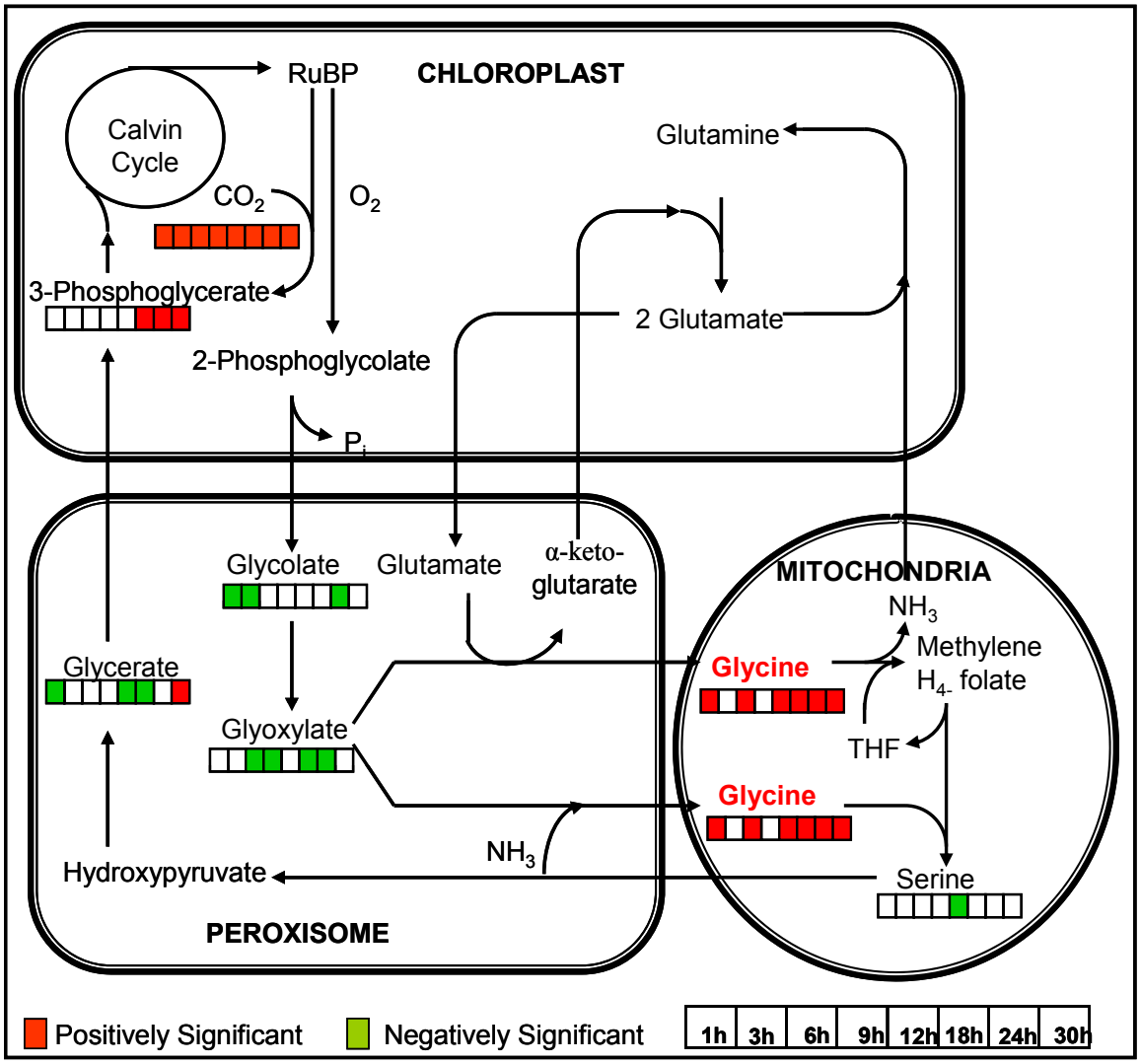


Figure 5-24: Observed effect of the trehalose signal on the physiology of photorespiration, at the metabolic levels. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7

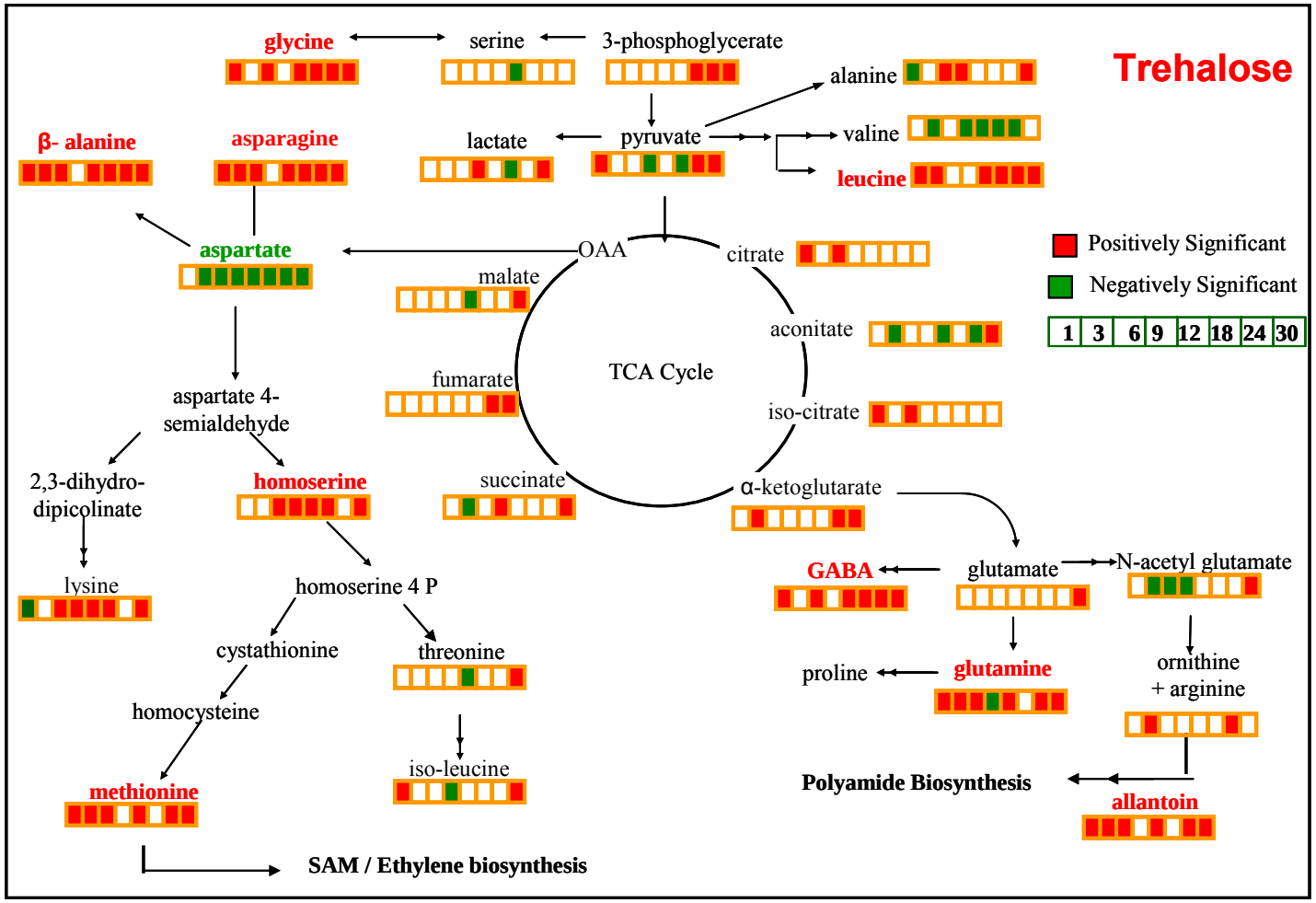


Figure 5-25: Observed effect of the trehalose signal on the physiology of the Tri Carboxylic Acid (TCA) cycle and amino acid biosynthesis at the metabolic level. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

Amino Acid Biosynthesis

Since trehalose was primarily known to affect carbohydrate metabolism, all the studies trying to understand the regulatory role of trehalose were focused on carbohydrate metabolism. However in the current experiment, metabolomic analysis showed in response to trehalose signal, most of the amino acids showed significant increase in their concentration. As can be seen from Figure 5-25, 9 amino acids showed significant increase at most of the time-points and also from paired-SAM analysis. This accounted for all the known metabolites showing significant increase along with intra-cellular trehalose concentration in response to trehalose signal. In addition to these 9 metabolites, lysine, alanine, (ornithine + arginine) and glutamate were also found to be positively significant at 5/8, 3/8, 2/8 and 1/8 timepoints respectively. Aspartate and tyrosine (Figure 5-26) were the only amino acids which were negatively significant at most of the time points and also from paired-SAM analysis. In addition valine, N-acetyl glutamate and serine also showed significant decrease in concentration at 5/8, 3/8 and 1/8 time points respectively. This suggests a significant increase in amino acid pools and most probably also in nitrogen assimilation in response to trehalose signal in plants.

Nitrate Reductase (NR) gene, is the first enzyme in the nitrogen assimilation pathway and light, nitrate ions and carbohydrates, have shown to induce expression of NR mRNA (Larrios et al., 2001; Cheng et al., 1991; Cheng et al., 1992)]. Previous studies of sucrose feeding in dark grown *A. thaliana* seedlings (Cheng et al., 1992) and elevated CO_2 stress response (Larrios et al., 2001) have shown induction of NR gene expression. This is the first time we show that trehalose signal also increases nitrogen assimilation. Further the

increase in amino acids in response to trehalose is much stronger and uniform as compared to the effect of elevated CO_2. This result thus demonstrates the advantage of using high-throughput techniques which allow us to identify phenomenon without *a priory* hypothesis.

Sugar Metabolite Pools

In response to trehalose signal, most sugar metabolite pools show a significant decrease at 1h as can be seen from Table 5-9. Increased intra-cellular trehalose levels are known to result in accumulation of starch. Accordingly in response to trehalose signal we see a significant decrease in sucrose at 1hr and 12 hr (which is the alternate carbon storage compound) and in maltose (which is formed by degradation of starch) at 1hr, 6hr, 12hr, 18hr and 24hr. Both these observation and lack of significant increase in other sugar levels indicate starch accumulation is increased. However suggested mechanism for trehalose induced starch accumulation has been increased production of starch by increased expression of gene ADP-glucose-phosphopyrolase. But the significant decrease in maltose concentration at 5 of the 8 time points suggest role of trehalose in not only increasing starch production (based on earlier studies) but also in decreasing starch breakdown. This is the first-time role of trehalose in regulating starch break-down for accumulating starch has been observed. It would be interesting to further study through future experiments if the increased production or the reduced break-down is the major cause behind starch accumulation as a result of increased trehalose concentrations in plants.

Table 5-9: Significance level of Sugar, Sugar Phosphates and Sugar Alcohols at individual time points and paired-SAM analysis in response to trehalose signal. Positively and negatively significant metabolites are as described in Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sucrose	sugar	P3840	-1				-1			1	
Trehalose	Sugar	P3983	1	1	1	1	1	1	1	1	1
Maltose	Sugar	P3923	-1		-1		-1	-1	-1		
Cellobiose	Sugar	P4025	-1							1	
Glucose	Sugar	P2655								1	
Fructose	Sugar	P2601								1	
Glucose 6 P	Phosph	P3548	-1				-1				
Fructose 6 P	Phosph	P3513			1		-1			1	
Inositol-1/2-Phosphate	Phosph	P3581	-1	-1			-1				
Arabinose	sugar	P2242	-1				-1			1	
Rhamnose	Sugar	P2320	-1				-1	-1		1	
Xylulose	Sugar	P2284	-1							1	
Xylitol	**alc**	**P2230**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**
myo-Inositol	alc	P2910	1							1	
Galactinol (9TMS); alpha-D-Gal-(1,3)-myo-I		P4270	1		1		1	-1	1	1	

In addition to regulation of starch accumulation, metabolomic analysis also indicated a significant decrease in Xylitol concentration in response to trehalose accumulation in plants. This could also indicate regulation of xyloglucans biosynthesis which is the major end use for xylitol by trehalose. The only sugar metabolite to show a significant increase was galactinol at 1h, 6h, 12h and 24h and myo-inositol at 1h. This is further significant as these metabolites show significant increase at 1h and 12h timepoints, at which more than 40-50% of the metabolite pools showed significant decrease. This further suggests trehalose may regulate Inositol based oligosaccharide biosynthesis. Finally at 30h majority of the sugar and sugar acid compounds showed a significant increase which was expected, as more than 60% of the total metabolite pools showed significant increase in response to the trehalose signal. Finally it was interesting to see that the concentration of glucose and fructose remain almost constant (non-significant) at all the time points except

30h. The constant glucose levels suggest that trehalose degradation by trehalase, producing glucose remained at almost the same levels, or alternatively the utilization of glucose is increased to the same extent as the increased trehalase activity.

Secondary Metabolite Pools

The three aromatic amino acids: tryptophan, tyrosine and phenylalanine; are the precursors for many secondary metabolites which are important for a host of important physiological functions in plants which were discussed earlier in Section 5.1.5. From the three amino acids, as with elevated CO_2 and NaCl stress response, the tryptophan metabolite pools were very low and below the detection limit in the A. *thaliana* liquid culture plants. From the other two secondary metabolite, tyrosine metabolite pools (Figure 5-26) showed significant decrease at all time points in response to trehalose signal where as the concentration of phenylalanine remained almost constant. From the other known secondary metabolites derived from tryptophan and phenylalanine, most of the metabolites showed slight decrease at 2 or 3 time points during first 18h with the exception of benzoate which showed a slight increase at 2 time points. Finally a number of secondary metabolites showed a significant increase at 30h, the time point at which there was a large increase in number of significant metabolites. Overall these results suggests except for significant decrease of tyrosine, (which seem to have been governed by amino acid biosynthesis rather than secondary metabolite production), trehalose has little of no significant role in regulating secondary metabolite biosynthesis.

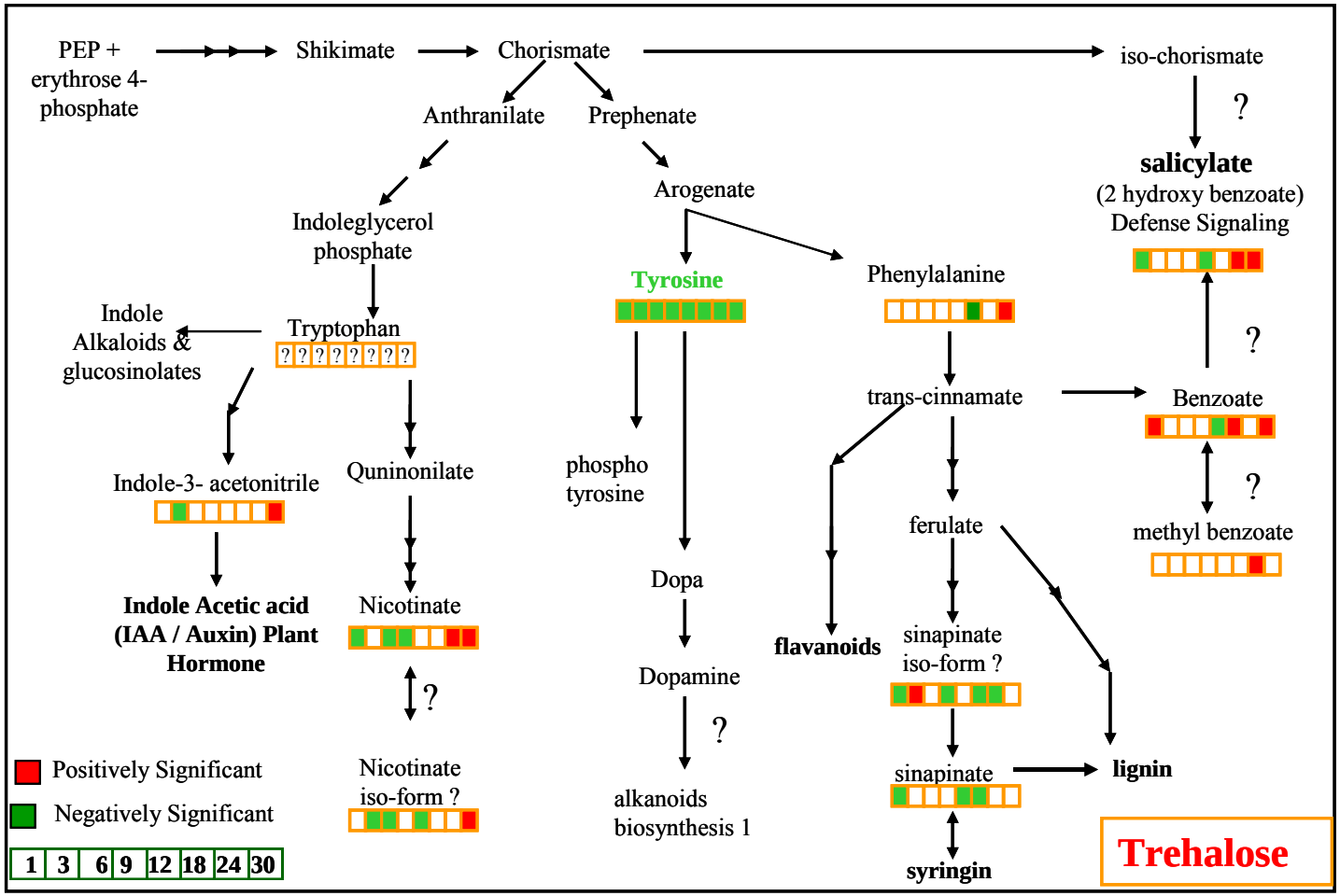


Figure 5-26: Response of metabolite pools in *A. thaliana* secondary metabolism pathways in response to 12 mM trehalose signal. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

Butanoate Metabolism

Butanoate metabolism is based on precursors from TCA cycle and is important for the production of Poly-hydroxybutyrates (PHBs) as well as intermediates for ketone bodies. The pathway producing PHBs in plants however is still under investigation. GC-MS metabolomic analysis identified several metabolites belonging to the known PHB pathways (2-hydroxyglutarate, 3-hydroxybutanoate, 4-hydroxybutanoate) as well as some related metabolites (3-hydroxy-3-methylglutarate, 2,3-dihydroxybutanoate, 3,4-dihydroxybutyrate) which may also be involved in Butanoate metabolism. From these metabolites as shown in Table 5-10, metabolite 3.4-dihydroxybutyrate undergoes significant decrease at 7/8 time points and also from over-all analysis. Further, most of the other metabolites in this pathway, except 4-hydroxybutanoate, were found to be negatively significant at 2 to 4 timepoints in response to the stress. At 30h however as a result of overall increase in metabolite pools, metabolites belonging to Butanoate metabolism pathway also showed a significant increase. In conclusion trehalose does seem to affect Butanoate metabolism significantly, reducing concentration of its metabolite pools at a number of time points, though the exact mechanism or the purpose of this regulation is not clear based on current analysis and needs further investigation.

Table 5-10: Significance level of metabolite pools belonging to Butanoate metabolism at individual time points and paired-SAM analysis in response to trehalose signal. Positively and negatively significant metabolites are color-coded as described in the caption of Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
2-hydroxyglutarate	acid	P2184	-1				-1			1	
3-Hydroxy-3-methylglutarate	acid	P2197	-1	-1				-1		1	
4-hydroxybutanoate	acid	P1279						1	1	1	
3,4-dihydroxybutyric acid	**acid**	**P1744**	**-1**	**-1**	**-1**		**-1**	**-1**	**-1**	**-1**	**-1**
2,4-dihydroxybutanoic acid	acid	P1704	-1	-1	-1		-1				

Lipid Metabolism

Trehalose had very small effect on the lipid metabolism. Most of the FA and sterols were negative significant at 12h and 1h the time points at which largest decrease in overall metabolite pools was observed as seen from Table 5-11. Similarly most of the lipids were positively significant at 30h when majority (60%) of the metabolite pools show a significant increase in concentration. Thus trehalose does not seem to affect lipid metabolism significantly and most of the changes in the pathway seems to be as a result of more global increase in free metabolite pools.

Table 5-11: Significance level of fatty acid and sterol metabolite pools at individual time points and paired-SAM analysis in response to trehalose signal. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Stearic Acid	FA	P3432							1	1	
Linolenic acid	FA	P3461					-1			1	
alpha-Linolenic acid	FA	P3536					-1		-1		
Gama Linolenic Acid	FA	P3509					-1			1	
Icosanoic acid	FA	P3747	-1				-1				
Stigmasterol	Sterol	**P5045**	-1				-1		1	1	
CampeSterol	Sterol	**P5007**	-1				-1			1	
Tocopherol	Pherol	**P4850**	-1				-1			1	

Other Metabolite Pools

The other metabolite pool group consists of metabolites (a) whose exact role and the biosynthesis of the metabolites is still not known e.g. erythretol, pyrole-2-carboxylic acid, citramalate (b) isolated metabolite pairs like Adenosine-Adenine, Ascorbate-Threonate. In response to trehalose signal, these metabolites show similar response to lipid metabolism i.e. metabolites show significant decrease at 1h and 12h but significant

increase at 24/30h timepoint. The only exception was ascorbate which was positively significant during first 3 hours.

Table 5-12: Significance level of other known metabolite pools at individual time points and paired-SAM analysis in response to trehalose signal. Positively (1) and negatively (-1) significant metabolites are as described in Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sorbitol	alc	P2608	-1	-1			-1			1	
Erythritol	alc	P1776		-1				-1		1	
Pyrrole-2-carboxylic acid	acid	P1716	-1			-1	-1			1	
Citramalate	acid	P1868	-1				-1				
3,4-dihydro-2(3H)-Furanone	SecM	P1841			1				1	1	
Adenoise	sugar	P4227	-1				-1		-1		
Adenine (2TMS)	purine	P3120					-1				
Threonate	acid	P2042	-1				-1			1	
Ascorbate	acid	P2948	1	1				-1	1	1	

Unknown Metabolite Pools

The response of 95 unknown metabolites to trehalose signal with at least some clue about their identity is shown in Appendix II (Table A2-3). They show a similar pattern as known metabolites, showing significant decrease at 1h and 12h and increase at 30h. However four unknown sugar metabolites showed a significant decrease in their concentration at most of the time points where as two higher weight sugar metabolites showed a significant increase, suggesting perturbations in the carbohydrate metabolites as was also observed from the known metabolite pools.

5.3.6 Conclusions

In summary, time-series metabolomic analysis of the short-term 12 mM trehalose response of *A. thaliana* liquid cultures revealed following:

- Exogenously applied Trehalose is accumulated inside the cells and its concentration gradually builds up increasing to 50 times the original concentration in plants at the end of 30h. In control wild type conditions in which most of trehalose produced is broken down by trehalase so very small quantity of trehalose is accumulated. Hence trehalose accumulation within the cells suggests, trehalase activity is not regulated by trehalose concentration or beyond a certain concentration, trehalase activity is not sufficient to breakdown majority of trehalose produced or transported.

- Metabolomic analysis also indicates possibility of starch accumulation. The known mechanism by which trehalose regulates starch accumulation in plants is by activating genes regulating starch biosynthesis. Metabolomic analysis however, also indicated an additional mechanism of regulating starch accumulation by inhibiting break down of starch to malate and subsequently to glucose.

- Trehalose did not effect the concentration of glucose (produced by trehalose hydrolysis by *trehalase* enzyme) and fructose during the 1-24 hours.

- Trehalose significantly decreased xylitol concentration and showed significant increase in Galactinol concentration. Thus in addition to previously observed effect of trehalose increasing fructan biosynthesis (Muller et al., 2000), trehalose may also increase xyloglucan and Raffinose family oligosaccharides.

- Trehalose has very small effect (possibly secondary) on TCA cycle, Photorespiration, Secondary metabolism, Lipids and Butanoate metabolism.

- The most significant effect of Trehalose signal is observed in amino acids productions. In response to trehalose, 12 amino acids showed significant increase (from which 9 were also observed from paired SAM). In contrast only 4 amino acids showed decrease in concentration of which only 1 (aspartate) was negatively significant at most time points.

- Trehalose increased 60% of the total metabolite pools at 30h when intracellular trehalose concentration was the highest (50 times the original). This suggests possibility of a different response at very high trehalose concentration.

The above results, demonstrate the advantage of using high-throughput analysis which allowed identification three new regulatory roles of trehalose which would not have been detected simultaneously based on traditional analysis approach. Specifically the regulation of amino acid biosynthesis by trehalose is an important result which can help in engineering efforts to produce plants which have a better ability to assimilate nitrogen and plants which can produce more amino acids there by improving the nutritional quality.

5.4 ETHYLENE RESPONSE

5.4.1 Review of Ethylene Response

Ethylene is an important gaseous plant hormone. In plants, ethylene regulates: seed germination, seedling elongation, hook formation, pathogen/disease response, wounding response, leaf abscission and ripening (Taiz, 2002). Due to importance of ethylene in these important physiological and developmental events, ethylene has been studied in great detail for almost a century (Taiz, 2002). Throughout these period focus of the research has focused on understanding (Benavante et al., 2005, De Paepe et al.,2005) ethylene biosynthesis, identifying ethylene binding sites, dissecting the ethylene signaling pathways, identifying genes and transcription factors induced by ethylene and identifying interaction between ethylene and other hormone regulatory networks (Van loon 2006) Thus most of the research on ethylene, except for understanding regulation of ethylene biosynthesis pathway, has been to understand the change in gene expressions and change in physiology or development. However changes in gene expression, if encoding enzymes in metabolic pathway, are also likely to affect the metabolism of plants. Similarly changes in physiology (like shoot elongation or leaf abscission) also require complementing changes in metabolism of the plants. Thus metabolism of plants is also likely to be significantly effected; however no studies have been made so far to study the response of ethylene at the metabolic level. In this section, the results from the first metabolomic analysis (time-series or not) of the response of *Arabidopsis thaliana* liquid cultures to ethylene signal are presented.

Ethylene is an important plant hormone regulating plant development and morphology (Guo and Ecker, 2004). Ethylene is produced in the plants through the oxidation of 1-aminocyclopropane-1-carboxylate (ACC) by oxygen in the presence of ascorbate and iron; ACC oxidase is activated by CO_2 (Godzinski 2004). As described earlier in Chapter 4 of this report, ethylene signal was introduced to the plants by adding ACC to the liquid media so that its final concentration was 0.01 MM. ACC is known to get readily converted to ethylene and has been a common choice for many studies involving ethylene response. Once ethylene is produced by ACC oxidase, the ethylene signal is sensed and transmitted through a signaling cascade in *A. thaliana* and higher plants involving a large number of genes (Chen 2005), whose regulation and specific role in transduction has been studied for a long time, and still many aspects of the same are currently under study. The response of ethylene in plants has also been studied before using DNA microarrays analysis (De Paepe et al., 2004. The transcriptomic analysis indicated even short treatment of ethylene (1-6h) created significant differences at the gene expression level. However as discussed earlier, the effect of these changes at metabolic level has not been studied before. The results from our study are described in the following sections.

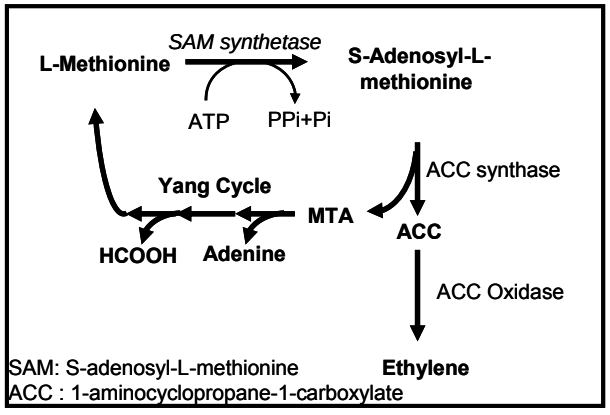


Figure 5-27: Ethylene Biosynthesis Pathway from Methionine.

5.4.2 Metabolomic profiling Results

Polar Metabolomic Profiles were obtained using the optimized protocol and experimental conditions described in Chapter 3 and Chapter 4, respectively, of this report. In the acquired metabolomic profiles, 550, among which 147 known, metabolite methoxime (Meox)-trimethylsilyl(TMS)-derivative peaks were detected. After data correction, normalization and filtering, the set of metabolite peak areas that were finally considered in the analysis included 87 annotated and 199 unidentified. Among the 87 known metabolites, 60 correspond to metabolites forming only one TMS-derivative, 9 to one of the two geometric isomer derivatives of ketone-group containing metabolites (see Chapter 3), and 18 to cumulative peak areas of amine-group containing metabolites. Each timepoint was represented by 4 (2 biological x 2 instrumental replicate) spectra. Data correction, filtering and normalization were done as described earlier in Chapter 3 to obtain normalized peak areas for control and perturbed experiments. The log ratio of this normalized peak profiles is shown in Figure 5-28. In response to ethylene, the known metabolites showed up to 8 fold increase or decrease (+/-3 log2 ratio). Maltose showed the highest significant increase at all the time points. Alanine showed the largest decrease in concentration during the first 9 hours. Most of the other known metabolites showed 4 fold increase or decrease at most of the time points. In contrast, the unknown metabolites showed a much larger variation in their metabolites showing variation of up to 30 fold increase or decrease from original concentration during first-12 hours and up to 10 fold after 12-hours. The figure also clearly showed a change in metabolism from 9-12 hours suggesting a shift in the response to ethylene around this time points.

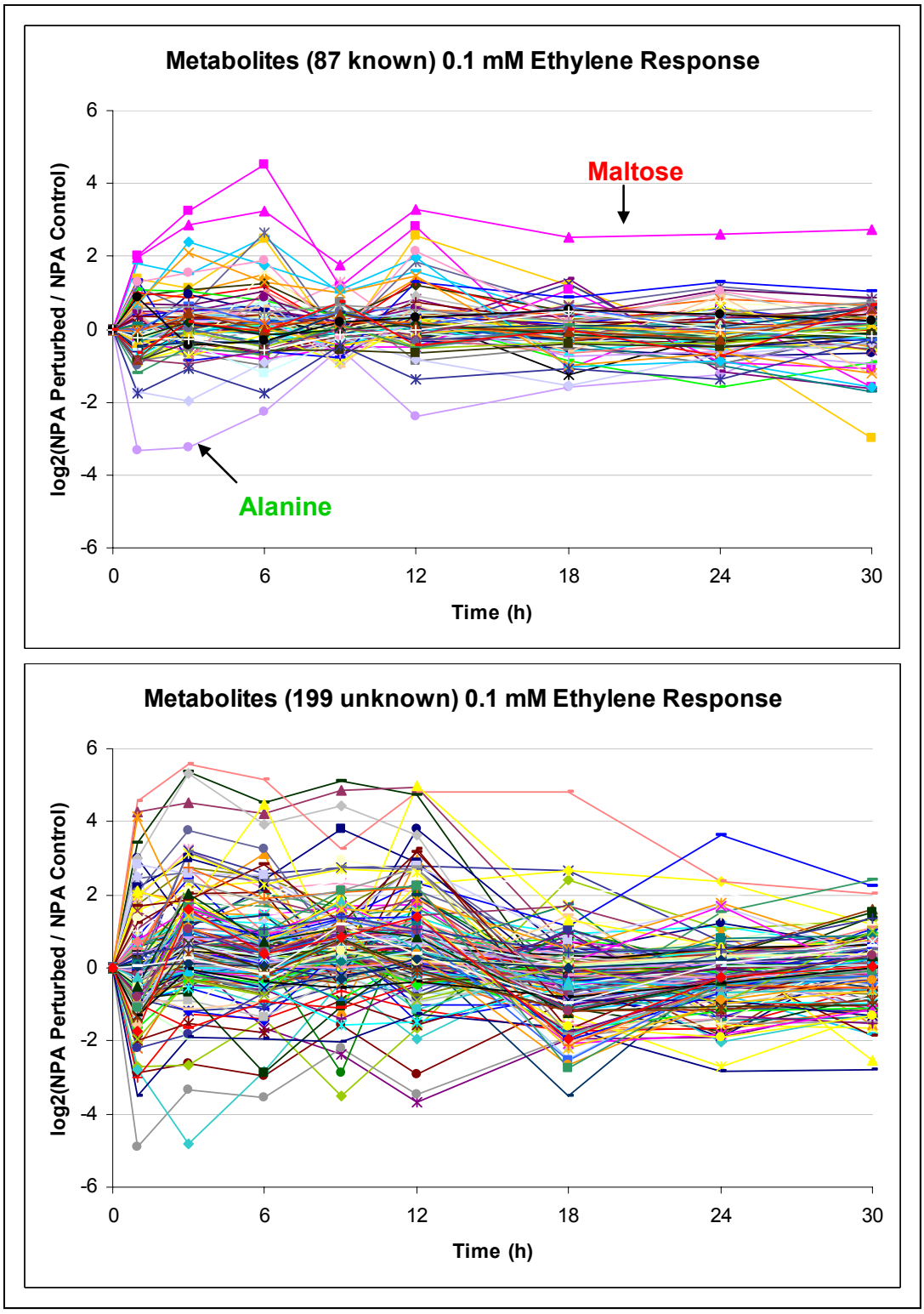


Figure 5-28: Log_2 Ratio Normalized Peak Area (NPA) profiles of Perturbed (Ethylene Signal) to NPA Control profiles indicate up to 20 fold increase - decrease in metabolite levels.

5.4.3 PCA Analysis

According to TIGR MeV Principal Component Analysis (PCA), the control metabolomic profiles can be clearly differentiated from their perturbed counterparts (Figure 5-29). However the most dramatic feature of the principal component analysis is the separation between of the first 12 hour timepoints in response to ethylene. As can be seen from the figure, the *A. thaliana* liquid cultures undergo a dramatic change in their metabolism even at 1h which is sustained up to 12h. At the end of 12h however the metabolism shifts back again closer to the control plants, but still showing a distinct separation from the control time points. These variations in ethylene response will be discussed in greater details in their biological context later in this section.

5.4.4 SAM and MiTimeS Results

Paired-SAM analysis identified (delta=1.45, 0.7% FDR) identified 43 (15 known) and 19 (6 known), respectively, positively and negatively significant metabolites in response to 0.01 MM ACC - Ethylene signal (Figure 5-30). MiTimeS analysis (Dutta et al., 2007), on the other hand, identified many more positively and negatively significant metabolites at the individual timepoints (average 90 and 52, respectively) for the same significance threshold value and a slightly higher average median FDR (0.96%) (Figure 5-30). The significance level profiles over time along with the paired-SAM result for all known metabolites are discussed individually in subsequent sections in the context of their biology.

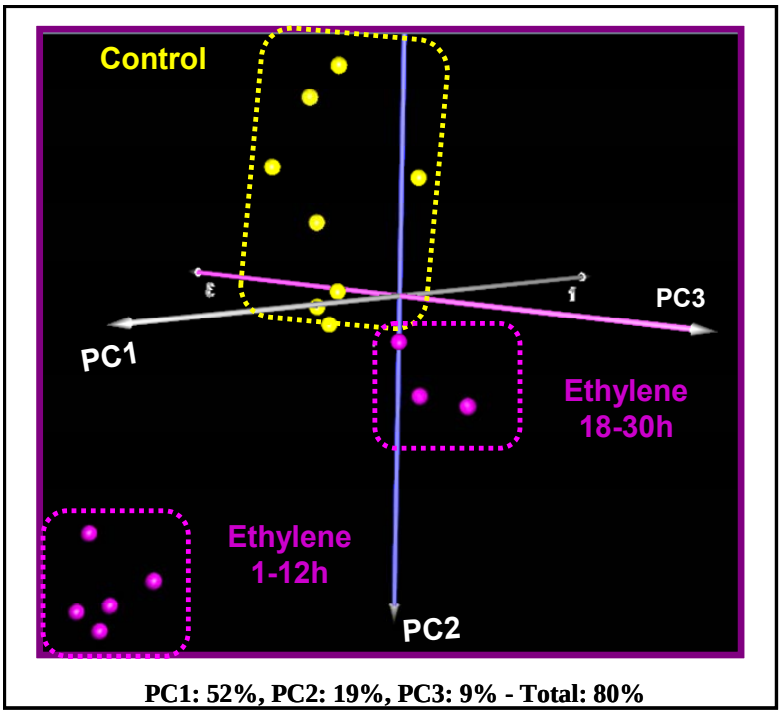


Figure 5-29: Principal Component Analysis (PCA) of Ethylene response using TIGR MEV 3.0 shows a significant difference in metabolism of *A. thaliana* liquid cultures even during the 1-30h.

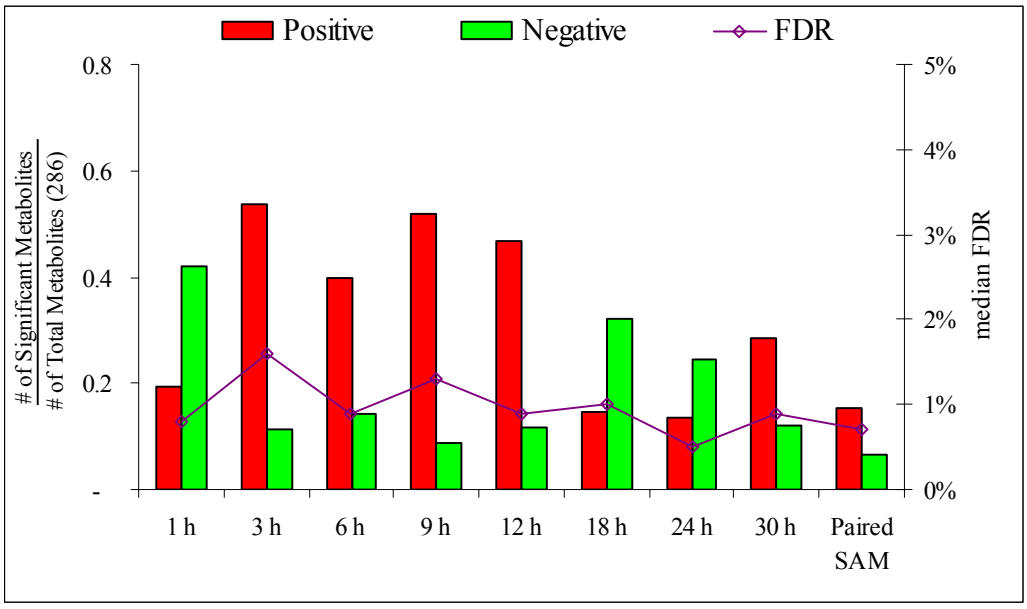


Figure 5-30: Number of Positively, Negatively and Total significant metabolites along with % median False Detection Rate (FDR), in response to ethylene signal obtained for overall analysis and individual time points, using paired-SAM and MiTimeS analysis.

According to Figure 5-30, in response to ethylene stress almost 50% - 60% of the metabolite pools show a significant change in their metabolism during the first 12 hours. During 3-12h time period the number of positively significant metabolites are much more as compared to negatively significant metabolites. However beyond 12h, the number of significant metabolites drops gradually from almost 60% of the metabolite pools to 40% of the total metabolite pools at 24h and 30h. Further at 18 and 24h, we also see more negatively, rather than positively, significant metabolites. This correlates to the significantly different response observed during first 12h from the PCA analysis.

MiTimeS analysis (Dutta et al., 2007) also allowed calculation of Significance Variability Score (SV Score) which represents the amount of fluctuations in the significance level for each metabolite on a scale of 0-2 as described earlier in Chapter 3. SV score of 0 represents the metabolite does not undergo change in its significance level and maintains the same significance level at all time points. In contrast SV score 2 represents oscillatory metabolite, i.e. the metabolite changes its significance level from positively significant to negatively significant and vice-versa at each and every time point. The SV score distribution which plots the SV score against the fraction of the detected polar metabolites at each SV score is shown in Figure 5.31. As can be seen from the figure, in response to salt stress, 8 metabolites (3% of total) were observed in the same significance level at all timepoints (i.e. SV score = 0). From these 8, only 3 were known metabolites: Maltose and Stigmasterol was positively significant at all time points and Threonine was negatively significant at all time points. The other 5 were unknown metabolites from which 2 and 3 metabolites were positively and negatively significant, respectively, at all time points. This also indicates none of the metabolites were non-significant at all time

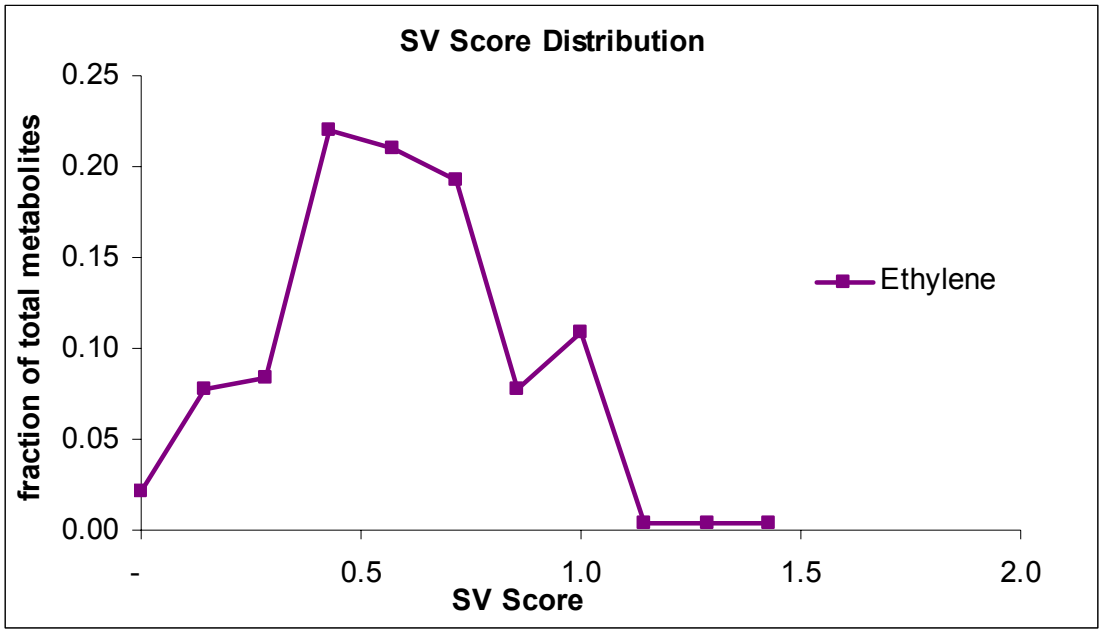


Figure 5-31: Significance Variability (SV) Score distribution of metabolites in response to ethylene signal shows the overall dynamics of the system.

points i.e. almost all the detected metabolites showed significant variations at least at one time point in response to ethylene. The metabolites which showed the highest dynamics was Glycine with SV score 1.29, followed by Adenine (SV score 1.14) and pyruvate (SV score 1).

The time point correlation network based on, both the positively and negatively significant metabolites (Figure 5-32) indicates strong correlation between 1h to 12h timepoints for both positively and negatively significant metabolites. In case of positively significant metabolites, there is only one connection between the initial 12h time points and longer time points suggesting a big change in metabolism which was also observed from PCA analysis variation of significant genes between time points. The negatively significant metabolites however showed more correlations between the time points.

However in both the networks 18h and 30h showed the least connections indicating most unique response at these time points.

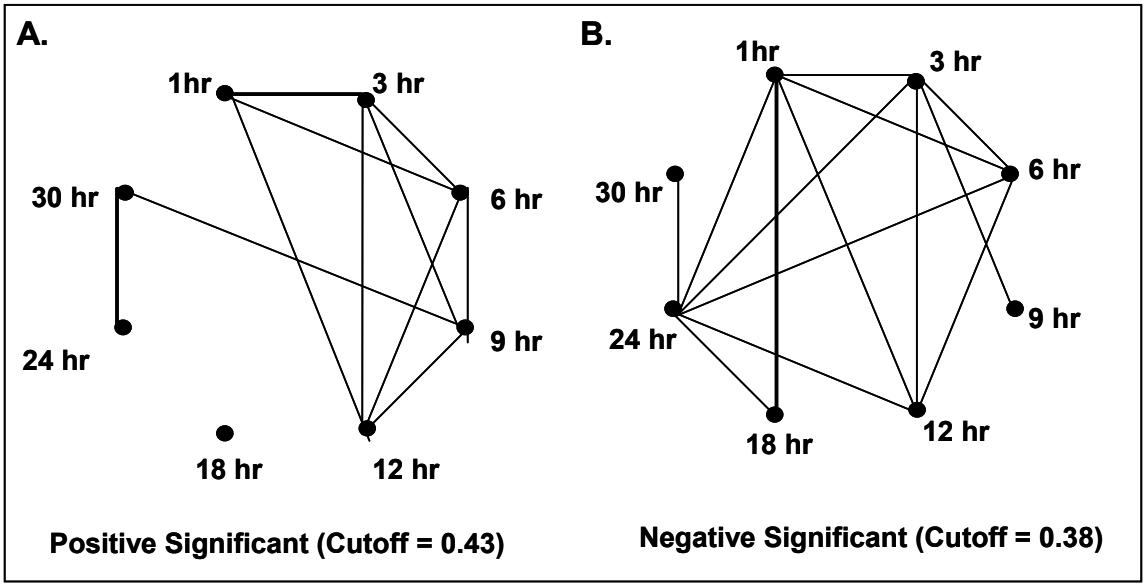


Figure 5-32: Time point correlation networks based on the common between time points (A) positively and (B) negatively significant metabolites in response to ethylene signal. Two time points are connected, if their correlation coefficient is larger than the indicated threshold, the latter being selected in each case as the average of all between different timepoints correlation coefficients.

5.4.5 Analysis of Individual Pathways

Tri-Carboxylic Acid (TCA) Cycle Metabolites

Tri-Carboxylic acid Cycle (Figure 5-33) is the primary aerobic respiration pathway in the mitochondria for most eukaryotic systems (Tiaz, 2000). TCA cycle is responsible for the production of energy in the form of ATP and reducing power in the form of NADH and $FAdH_2$ by oxidation of pyruvate (derived from sucrose) into CO_2 (Tiaz, 2000). Finally, TCA cycle also provides the carbon backbone in the form of α-ketoglutarate and oxaloacetate for amino acids biosynthesis.

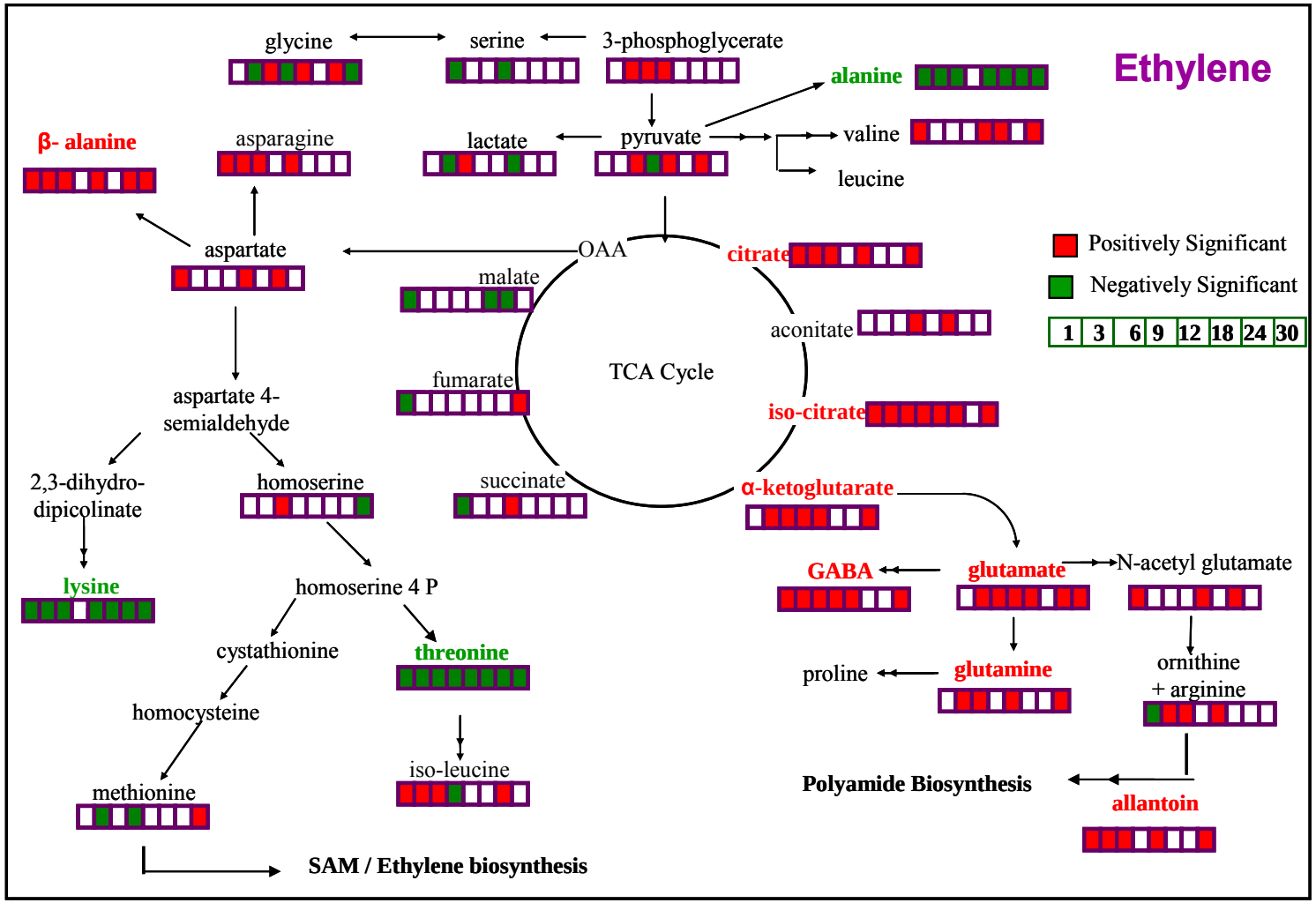


Figure 5-33: Observed effect of the ethylene signal on the physiology of the Tri Carboxylic Acid (TCA) cycle and amino acid biosynthesis at the metabolic level. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

As can be seen from Figure 5-33, ethylene has a very different response on two parts of TCA cycle. All the metabolites from citrate to α-ketoglutarate are positively significant at most of the time points and paired-SAM analysis. In addition, amino acids glutamate, glutamine, GABA and allantoin derived from α-ketoglutarate also showed significant increase in their pools at most of the time points and from paired-SAM analysis. Where as metabolites in the other half of the TCA cycle, which convert α-ketoglutarate to oxaloacetate did not show much change except at 1 or 2 time points. This again suggests that the flux through the first part of the TCA cycle in response to ethylene is governed mainly by the need for carbon skeleton for amino acid biosynthesis rather than the overall TCA cycle goal of ATP and NADH production. These results for the first time show that Ethylene in addition to regulating physiological development and gene expressions also regulates concentration of metabolite pools of the TCA cycle intermediates at the metabolic level.

Amino Acid Biosynthesis

In response to ethylene, as can be seen from Figure 5-35, all the nitrogen storage and transport amino acids: glutamate, glutamine, aspartate and asparagine; showed significant increase in their metabolite pools. Particularly, glutamate and glutamine were also found to be positively significant from overall response using paired-SAM analysis. These results suggest nitrogen assimilation is significantly increased in response to ethylene signal. In addition valine, iso-leucine, ornithine + arginine, non-protein amino acids – 4-aminobutyric acid (gaba) and β-alanine and polyamine allantoin showed significant increase in their concentration. In contrast four amino acids, alanine, threonine, lysine

and tyrosine (discussed later during secondary metabolism) showed a significant decrease in response to ethylene at most of the time points as well as for overall response from paired-SAM analysis. Serine showed a slight decrease in its concentration at two time points, where as glycine, methionine and homoserine fluctuated between positively and negatively significant levels. These results for the first time show the role of ethylene in increasing nitrogen assimilation and differential regulation of various amino acid biosynthesis pathways.

Sugar Metabolite Pools

In response to ethylene signal Maltose was the only sugar metabolite which showed significant increase (up to 30 fold) in its relative concentration at all timepoints and also from overall paired-SAM analysis. In addition to maltose, Inositol also showed significant increase at a number of time points. Sucrose, sugar phosphates showed a slight decrease at 1h and also at 6h for sucrose. Hexose sugars Glucose and Fructose were also insignificant at most of the time points and showed a slight decrease at 18h and increase at 24h. Hence the most prominent effect of ethylene on sugar metabolite pools was the significant increase in maltose concentration.

The main source of maltose biosynthesis is from starch breakdown. The 30 fold increase in maltose concentration thus clearly indicates ethylene enhances breakdown of starch in plants. This is the first time the regulation of starch breakdown by ethylene in vegetative growth phase of plants has been shown.

Table 5-13: Significance level of Sugar, Sugar Phosphates and Sugar Alcohols at individual time points and paired-SAM analysis in response to ethylene signal. Positively and negatively significant metabolites are as described in Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sucrose	sugar	P3840	-1		-1						
Trehalose	Sugar	P3983	-1	-1				1		1	
Maltose	Sugar	P3923	1	1	1	1	1	1	1	1	1
Cellobiose	Sugar	P4025	-1						-1		
Glucose	Sugar	P2655			-1			1	-1		
Fructose	Sugar	P2601						1	-1		
Glucose 6 P	Phosph	P3548	-1			-1		-1			
Fructose 6 P	Phosph	P3513	-1				1	-1			
Inositol-1/2-Phosphate	Phosph	P3581	-1			1				1	
Arabinose	sugar	P2242									
Rhamnose	Sugar	P2320		1		1		1			
Xylulose	Sugar	P2284	-1			1			-1		
Xylitol	alc	P2230			-1			1			
myo-Inositol	alc	P2910	1	1	1		1			1	

Secondary Metabolite Pools

The three aromatic amino acids: tryptophan, tyrosine and phenylalanine; are the precursors for many secondary metabolites which are important for a host of important physiological functions in plants which were discussed earlier in Section 5.1.5. From the three amino acids, as with elevated CO_2, NaCl and trehalose stress response, the tryptophan metabolite pools were very low and below the detection limit in the *A. thaliana* liquid culture plants. From the other two secondary amino acids, tyrosine metabolite pools (Figure 5-34) showed significant decrease at all time points (except 9h) and from paired-SAM analysis. Where as phenylalanine concentrations increased significantly between 3-12 hour time points. At these time points, which also showed the highest amount of positively significant metabolites, along with phenylalanine, secondary metabolites derived from it – namely sinapinic acid and its derivative also showed

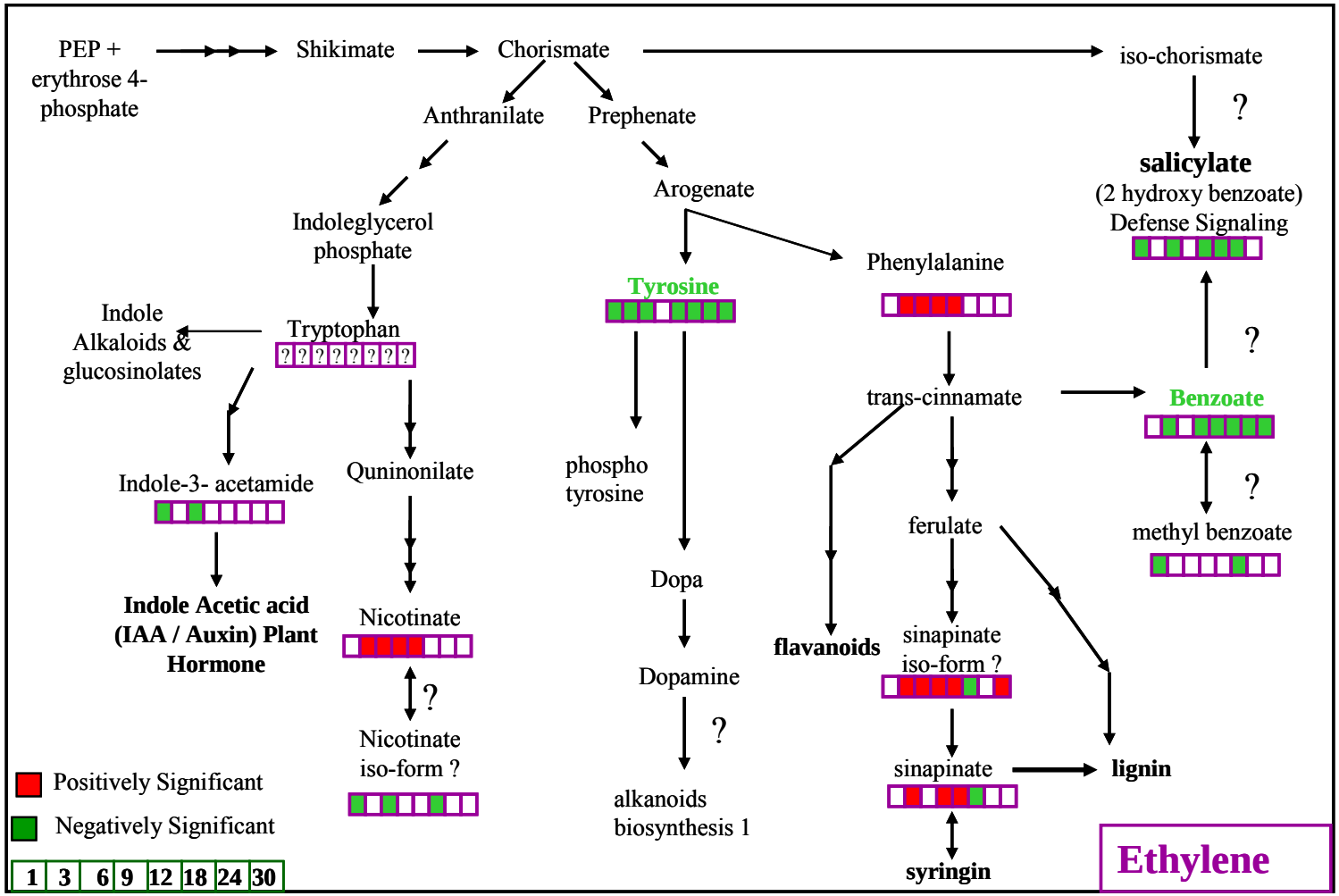


Figure 5-34: Response of metabolite pools in *A. thaliana* secondary metabolism pathways in response to ethylene signal. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

significant increase in their relative concentration. These metabolites are part of the lignin / Syringin biosynthesis pathway, suggesting increase in its synthesis in response to ethylene signal in the first 12 hours. Similar result was also seen in nicotinic acid, which is produced starting from tryptophan and has important cellular functions such as nucleotide biosynthesis.

Ethylene responsive elements are also known to interact with other plant hormones to co-ordinate disease and wounding response. The interaction between their signaling pathways has been a subject of significant study in recent times. From our analysis, ethylene significantly reduces the concentration of the Auxin intermediate Indole-3-acetamide at 1h and 6h, however at rest of the time points it is non-significant. A stronger inhibiting effect of ethylene is seen however of Salicylic Acid and its intermediate Benzoic acid which were found to be negatively significant at (1h, 6h, 12h, 18h, 24h) and (3h, 9h, 12h, 18h, 24h, 30h also from paired-SAM), respectively, as can be also seen from Figure 5-34. The increase in both lignin and nucleotide intermediate at the same timepoints 3h-12h suggest possibility of increase in cell division of plant growth in response to ethylene signal. However beyond 12h, there seems to be a dramatic change as both these metabolites were non-significant at time points beyond 12h.

Calvin Cycle and Photorespiration Pathway

Carbon and oxygen "compete" for Rubisco activity (Figure 5-24) as the Rubisco catalyses both the carboxylation and oxidation reaction which form 3-phosphoglycerate and 2-Phosphoglycolate respectively. The pathway for regenerating Calvin cycle intermediate from the oxygenation product is known as Photorespiration pathway. As can

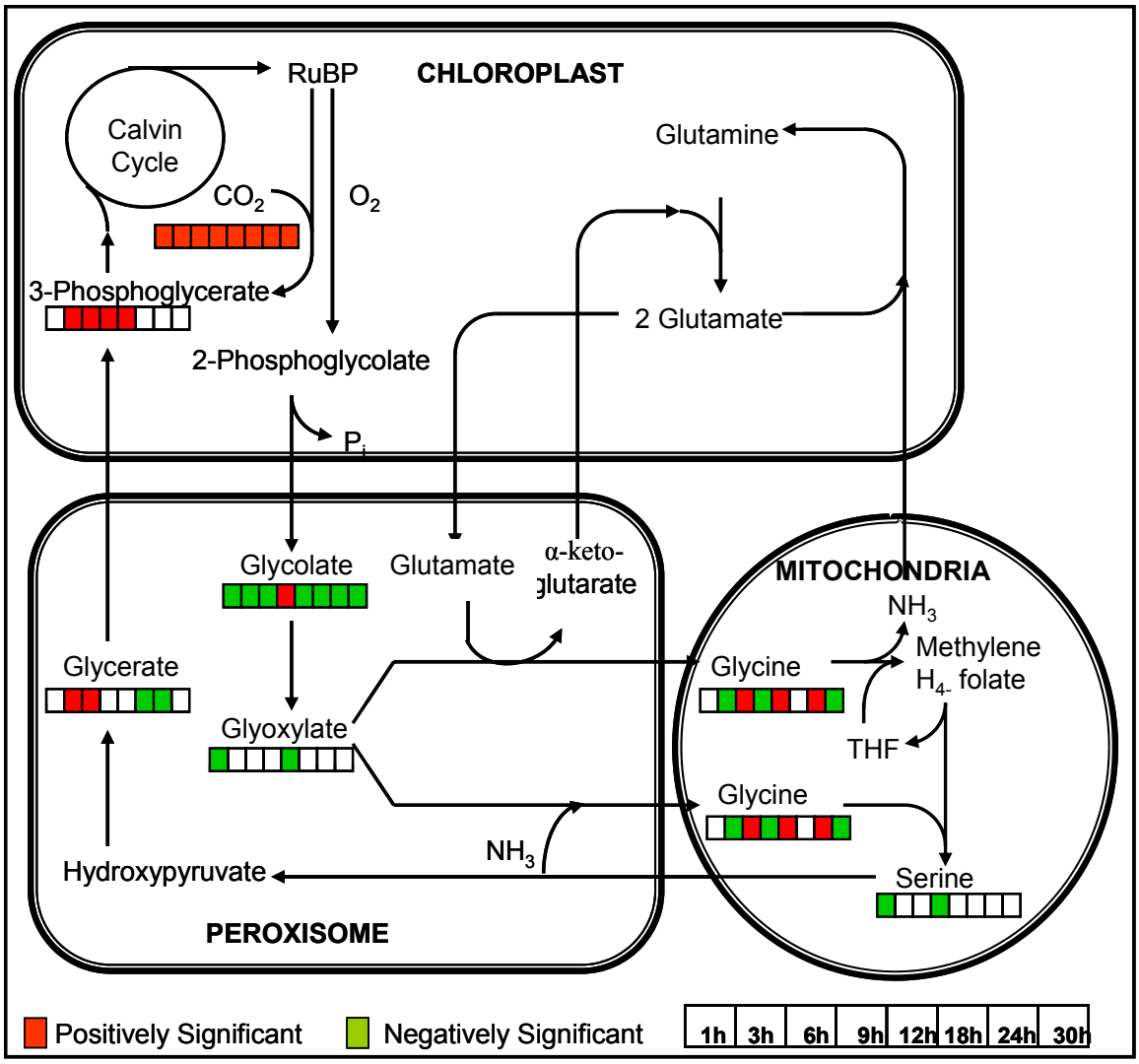


Figure 5-35: Observed effect of the ethylene signal on the physiology of photorespiration, at the metabolic levels. Positively and negatively significant metabolites are color-coded as described in the caption of Figure 5-7.

be seen from Figure 5-24, in response to ethylene signal from the three organic acids in these pathway: glycerate, glycolate and glyoxylate; glycolate shows significant decrease at 7/8 time points, however is also positively significant at 9h. Glyoxylate shows slight decrease at 1h and 12h, where as glycerate is positively significant at 3-6h but is later negatively significant at 18-24h. Due to the variation in response of metabolite intermediates at different time points it is difficult to determine the effect of ethylene on

photorespiration, however significant decrease in glycolate at 7/8 time points suggests possible inhibition of the photorespiration. The competing Calvin cycle, however, seemed to be significantly increasing during first 12h as 3-phosphoglycerate of the triose-phosphate pool was positively significant at these time points.

Lipid Metabolism

Ethylene significantly increased the concentration of all five known fatty acids during the first 12h as shown in Table 5-14 below. Especially the increase was seen more in the unsaturated forms of (C18) stearic acid (linolenic and γ-linolenic acid). These seem to further support increased cell division and growth in response to ethylene signal during first 12 hours.

In sterol biosynthesis, all the three know sterols/pherol were positively significant at 3h and 9h. However Stigmasterol was the only one which was positively significant at all eight time points and also from overall paired-SAM analysis. Usually Brassinosteroid is a plant hormone which is usually associated with regulation of cell division and growth.

Table 5-14: Significance level of fatty acid and sterol metabolite pools at individual time points and paired-SAM analysis in response to ethylene signal. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Stearic Acid	FA	P3432			1					1	
Linolenic acid	FA	P3461	-1	1	1	1	1	-1			
alpha-Linolenic acid	FA	P3536	-1		1			-1			
Gama Linolenic Acid	FA	P3509	-1	1	1	1	1			1	
Icosanoic acid	FA	P3747		1	1			-1			
Stigmasterol	Sterol	**P5045**	1	1	1	1	1	1	1	1	1
CampeSterol	Sterol	**P5007**		1		1					
Tocopherol	Pherol	**P4850**	-1	1		1					

Campesterol is the important plant sterol which is involved in Brassinosteroid production. However recent studies (He et al., 2002) have shown that other plant sterols like Stigmasterol and sitosterol are also involved in regulation of cell division and cellular growth. From the results of the current experiment shows that ethylene increase production of Stigmasterol (and possibly also sitosterol which was not identified) but not of campesterol which is involved in Brassinosteroid pathway. These suggests a possible mechanism by which ethylene regulates specific aspects of cell division and growth using phytosterols rather than another plant hormone Brassinosteroid. This further supports a recent comparison of transcriptomic response of young Arabidopsis seedlings to all plant hormones (He 2005) which concluded that different plant hormones regulate the same physiological phenomenon (in this case cell division) through different signaling and regulatory cascades. This further supports the assertion in a previous paper that the phytosterols can independently regulate cell division and growth like Brassinosteroid.

Butanoate Metabolism

Butanoate metabolism is based on precursors from TCA cycle and is important for the production of Poly-hydroxybutyrates (PHBs) as well as intermediates for ketone bodies. The pathway producing PHBs in plants however is still under investigation. GC-MS metabolomic analysis identified several metabolites belonging to the known PHB pathways (2-hydroxyglutarate, 3-hydroxybutanoate, 4-hydroxybutanoate) as well as some related metabolites (3-hydroxy-3-methylglutarate, 2,3-dihydroxybutanoate, 3,4-dihydroxybutyrate) which may also be involved in Butanoate metabolism. From these metabolites as shown in Table 5-15, metabolite 3.4-dihydroxybutyrate undergoes

significant decrease at 7/8 time points and also from over-all analysis in response to ethylene signal. Further, 2,4 dihydroxybutanoate was also negatively significant during first 12 hours. The rest of the metabolites showed variation in their significant levels with time as can be seen from Table 5-15.

Table 5-15: Significance level of metabolite pools belonging to Butanoate metabolism at individual time points and paired-SAM analysis in response to ethylene signal. Positively and negatively significant metabolites are color-coded as described in the caption of Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
2-hydroxyglutarate	acid	P2184	-1	1		1			-1		
3-Hydroxy-3-methylglutarate	acid	P2197		1	1		1	-1			
4-hydroxybutanoate	acid	P1279				-1			1	1	
3,4-dihydroxybutyric acid	acid	P1744	-1	-1	-1			-1	-1	-1	-1
2,4-dihydroxybutanoic acid	acid	P1704	-1	-1	-1	-1					

Other Metabolite Pools

The other metabolite pool group consists of metabolites (a) whose exact role and the biosynthesis of the metabolites is still not known e.g. erythretol, pyrole-2-carboxylic acid, citramalate (b) isolated metabolite pairs like Adenosine-Adenine, Ascorbate-Threonate (c) Phosphate ions. As shown in Table 5-16, in response to ethylene signal, pyrole-2-carboxylic acid showed significant decrease in concentration at 1h, 9h and 18-30h. In addition significant increase in major phosphate pools – phosphoric acid and methyl phosphoric acid (which could be a result of methanol extraction) – showed significant increase in response to ethylene signal. This suggests ethylene also up-regulates phosphorus uptake.

Unknown Metabolite Pools

The response of 91 unknown metabolites to ethylene signal with at least some clue about

Table 5-16: Significance level of other known metabolite pools at individual time points and paired-SAM analysis in response to ethylene signal. Positively (1) and negatively (-1) significant metabolites are as described in Table 5-1.

			1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Sorbitol	alc	P2608	-1			1					
Erythritol	alc	P1776		1		1					
Pyrrole-2-carboxylic acid	acid	P1716	-1			-1		-1	-1	-1	
Citramalate	acid	P1868	-1			1					
3,4-dihydro-2(3H)-Furanone	SecM	P1841	1	1	1	1	1				
Adenoise	sugar	P4227		1		1			-1		
Adenine (2TMS)	purine	P3120	-1		-1	1	-1				
Threonate	acid	P2042	-1	1					-1		
Ascorbate	acid	P2948		1	1		1			-1	
Phosphoric Acid	Phosph	P1465	1	1	1		1		1	1	1
MonomethylPhosphate	Phosph	P1337	1	1	1		1			1	1

their identity is shown in Appendix 2 (Table A2-4). They show a similar pattern as known metabolites, showing significant increase in number of metabolites, especially secondary metabolites during first 12 h. Also organo-phosphate pools especially of unknown metabolites matching Inositol phosphate and ethanol amine phosphate (both involved in phospholipids formation) increased significantly at most of the time points and also from paired-SAM analysis. In addition for sugars having mass spectra matching (hence suggesting similar structures) turanose, maltose, trehalose, iso maltose showed significant increase in response to ethylene stress at most of the time points. Among metabolites showing significant decrease were, some unknown sugar metabolites and lactones and an unknown proline like compound which shows a strong similarity to proline mass spectra suggesting a proline derivative like methyl proline.

5.4.6 Conclusions

In summary, time-series metabolomic analysis of the short-term ethylene response of *A. thaliana* liquid cultures revealed following:

- Ethylene significantly increases starch degradation at all time points as suggested by significant increase in maltose concentration (up to 30 fold).

- Ethylene significantly increases concentrations of first half of the TCA cycle intermediates (citrate, iso-citrate, α-ketoglutarate) and a number of amino acids at most of the time points during first 30h. The carbon available from starch seems to have been used primarily to increase TCA cycle intermediates, amino acids for protein, and secondary metabolite production possibly to support faster cell division and plant growth.

- The increase in precursors for protein (amino acids), lignin (sinapinic acid), and nucleotides (nicotinate) biosynthesis, unsaturated lipids, phosphor-lipid precursors especially during first 12 hours, also suggest the possibility of increased rate of cell division and plant growth in response to ethylene signal specially during first 12 hours. However these observation needs to be further investigated with more experiments.

- Ethylene achieves the regulation of increased cell division and cellular growth by regulating Stigmasterol (and possibly other phytosterols) rather than using the Brassinosteroid regulation machinery.

All of the above are novel findings, being reported for the first time. They demonstrate at metabolic level changes taking place in plants to achieve the increased cell division or growth at the macro-molecular level. In addition to adding to our understanding of

ethylene stress response, the analysis also indicated regulatory nodes in the *A. thaliana* liquid culture metabolic network as demonstrated by following examples:

- In response to ethylene, starch breakdown increases, however most of the carbon is stored in the form of maltose which shows the highest increase (30 fold) rather than fructose / glucose or their phosphates, suggesting maltose catabolism as one of the regulating nodes.

- In response to ethylene, TCA cycle intermediates citrate, iso-citrate and α-ketoglutarate show a significant increase in their pools. However the rest of the TCA cycle metabolites do not show similar increase. Instead the increased flux is used for production of glutamate and other amino acids derived from α-ketoglutarate. This suggests α-ketoglutarate to be an important node which regulates the flux towards further TCA cycle reactions or amino acid biosynthesis.

From the discussion in this chapter, time-series metabolomic analysis of the response of *A. thaliana* liquid cultures to four stresses: elevated CO_2, salt stress, trehalose (sugar) signal and ethylene signal; have provided significantly new information about the regulation of primary metabolism in plants. These analyses have also demonstrated clearly the importance of time-series and high-throughput nature of our study. In the next chapter, compassion of the individual stress responses and analysis of combined stresses (of the last three with elevated CO_2) will demonstrate the advantages of multiple combined systematic perturbations of the same biological system.

6 TIME-SERIES METABOLOMIC ANALYSIS OF MULTIPLE STRESS RESPONSES

In complex biological system, simultaneously many biological processes are taking place. These processes are tightly controlled and interact and regulate one another. In the case of a plant which is an example of such a complex system, simultaneously there might be carbon fixation, nitrogen fixation, plant defense response taking place along with many developmental and regulatory processes. These processes are usually controlled tightly to either maximize the utilization of available nutrient and sources of energy OR to maximize growth of the plant OR the most likely case a combination of these and many other parameters which could change depending on the developmental stage of the plant and changing environmental conditions. In order to achieve this desired function, a complex regulatory network exist involving cross-talk or cross-regulation between these simultaneously occurring biological processes. Due to this the response of the biological system to a given perturbation is very seldom expected to be a linear independent response.

In spite of the non-linearity of regulation of biological processes, most research, especially in high-throughput metabolomic analysis so far, focuses on response of the system to one perturbation at a time. However some of the most interesting and some time exquisitely beautiful regulatory mechanisms are seen in interaction between different biological processes. From a practical stand point also, and efforts to engineer a

complex system like plants, is expected to require more than one genetic change of the system. Already we are close to the commercial release of the next generation of genetically modified crops which combine along with insect protection other traits such as better water tolerance. Thus it is really important to study the effects in the biological systems when more than one perturbations are applied simultaneously using the available high-throughput measurement techniques. With this motivation in mind, as described in Chapter 4 we carried out combined perturbations where we increased the CO_2 level to 1% in the growth chamber and simultaneously also modified the media composition with (a) Salt (b) Trehalose Signal (c) Ethylene Signal and measured the response of the system at metabolic level in a dynamic manner using GC-MS metabolomic analysis. In this chapter, we first compare the effect of individual perturbations which were discussed in the previous chapter to understand the similarities and differences between these perturbations followed by results obtained from the combined perturbation.

6.1 COMPARISON OF INDIVIDUAL PERTURBATION

In order to study the regulation of primary metabolism of *Arabidopsis thaliana* liquid cultures four different perturbations were applied which involved a primary plant carbon source, environmental abiotic stress, sugar signal and hormone signal. The response of the system to these stresses at metabolic level was studied using metabolomic analysis and significant metabolites at each timepoint were identified using MiTimeS analysis (Dutta et al., 2007) and discussed in detail individually in the previous chapter. In these section their results are compared with each other and later in the chapter the effect of their combination will be discussed.

6.1.1 Principal Component Analysis

Comparison of principal component analysis (PCA) of the dynamic metabolomic profiles in response to four perturbations is shown in Figure 6-1. Following observations can be made from the comparison of PCA analysis for individual experiments:

- As can be seen form the Figure 6-1 all the four stresses show a clear separation from the control experiment.

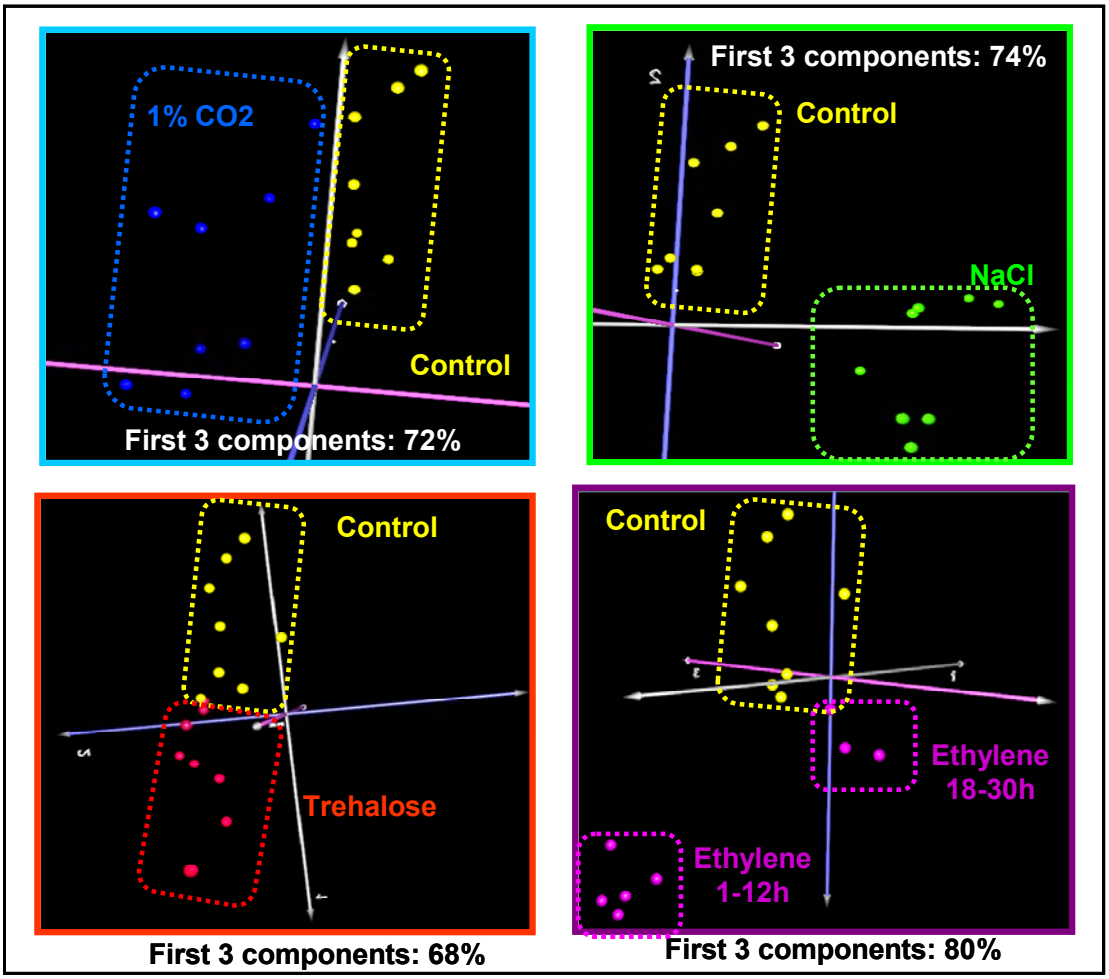


Figure 6-1: Comparison of principal component analysis of four individual perturbations with respect to control.

- The first 12h of ethylene signal and Salt (NaCl) stress show the largest difference in the metabolism.

- The Salt (NaCl) moves the metabolism almost in a perpendicular manner suggesting a dramatic change in the metabolic profiles.

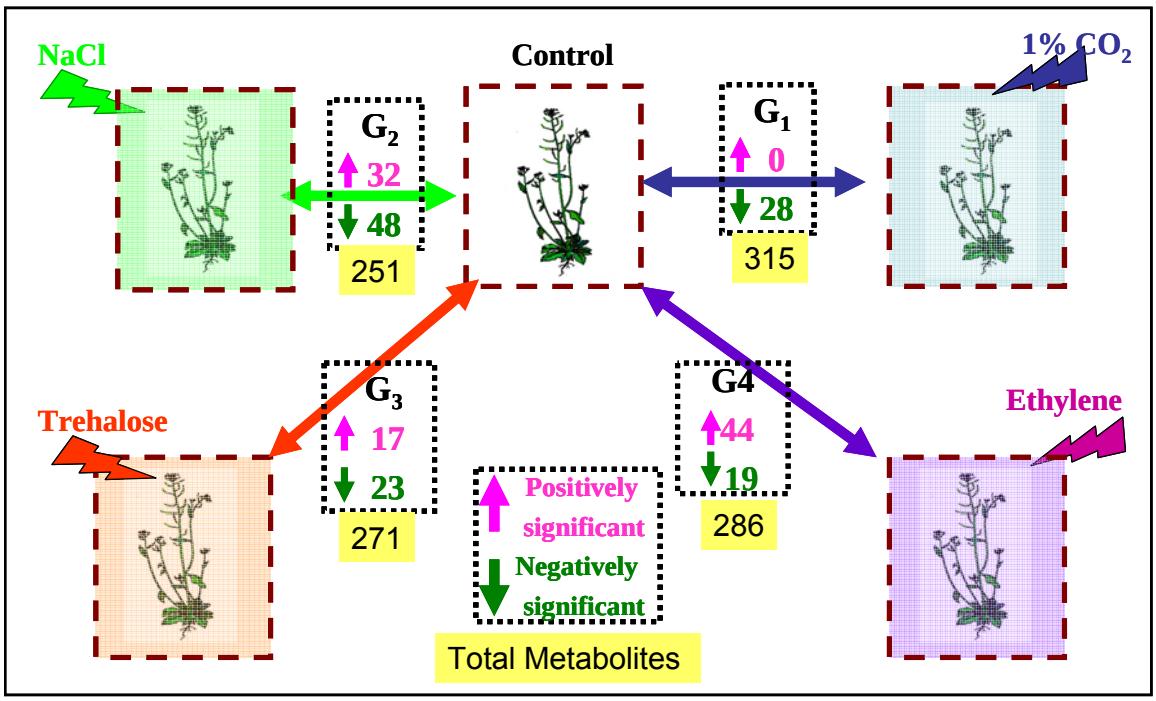
- Where as trehalose signal creates a shift in metabolomic profiles mainly along Principal Component 1 which has the maximum contribution, Elevated CO_2 stress moves it along Principal Component 3 which accounts for the smallest contribution of the three principal components.

Based on the PCA analysis and the distance between the control and the perturbed experiments, 1-12h of ethylene has the strongest response, followed by Salt (NaCl) stress, Trehalose Signal and elevated CO_2 in that order.

6.1.2 Significant Metabolites

PCA analysis of metabolomic profiles identified a significant change in the metabolic profiles of *A. thaliana* liquid cultures in response to the four perturbations. The metabolites which were responsible for these significant changes were identified using Paired Significance Analysis of Microarrays (paired-SAM) for the overall analysis for all the time points and at individual timepoints using MiTimeS analysis (Dutta et al., 2007). Figure 6.2 shows the comparison of the significance level from paired-SAM and individual time points for these perturbations. Since the total number of metabolites varied from each stress comparison, in MiTimeS analysis results all significant metabolites were normalized with the total metabolites in each comparison.

A. Paired-SAM Analysis Results



B. MiTimeS Analysis

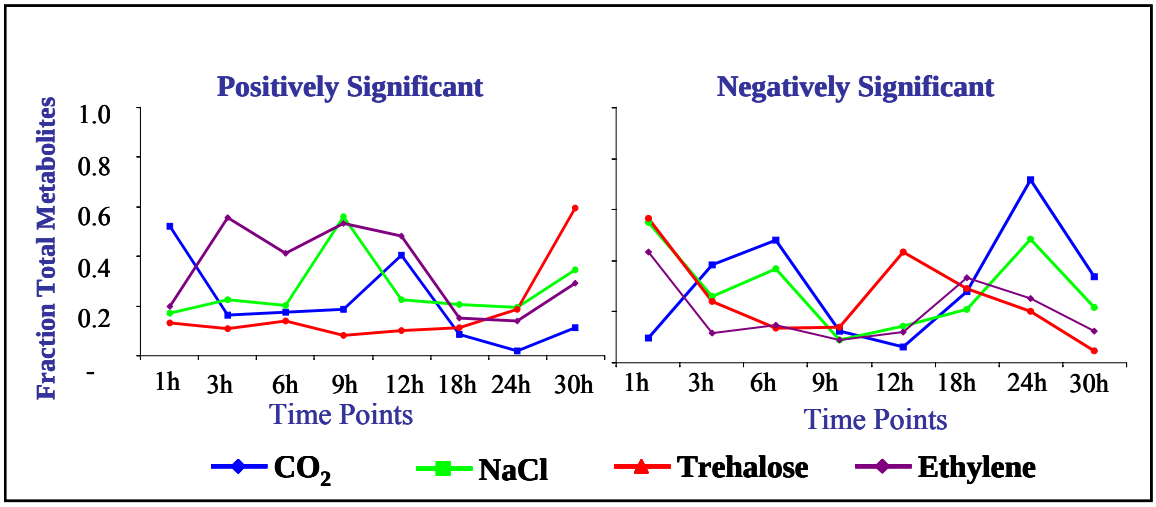


Figure 6-2: Comparison of significant metabolites for four individual perturbations from (A) paired-SAM and (B) MiTimeS analysis.

The comparisons of paired-SAM and individual time-point significance results indicate:

- The number of metabolites showing a significant change in response to perturbation at individual time points is much higher as compared to those identified from the overall analysis using paired-SAM analysis.

- Ethylene showed the highest number of positively significant metabolites from overall-analysis where as Salt stress showed the highest number of negatively significant (and also total) metabolites.

- Metabolomic profiles show a large dynamics / variability in response to stress as up to 70-80% of the metabolites show a positive / negative significance at a given time point.

- Ethylene shows the highest sustained significant increase in metabolite pools during first 12 hours suggesting ethylene signal perturbs a large number of metabolite pathway. This also confirms the large change observed in these time points from PCA analysis.

- In contrast trehalose shows smallest positively significant metabolite pools except at 30h suggesting a more specific and focused response.

- All perturbations show a much higher number of metabolite pools showing significant decrease (<50%) at 1h as compared to increase. The only exception to this was elevated CO_2, which was the only perturbation which showed a large number of metabolites (~60%) pools showing significant increase at 1h.

- For all perturbations, the number positively significant metabolite increase from 24h - 30h while those for negatively significant metabolites show decrease during the same time. This suggests, possibly a developmental change taking place at the beginning of 14^{th} day which trumps the effects of stress response and results in overall increase in metabolite pools.

6.1.3 Significance Variability (SV) Score Distribution

In order to study the dynamics of the response and variability in the significance level Significance Variability (SV) scores were calculated. Figure 6-3 shows the comparison of SV score distributions in response to individual perturbations. In the distribution, SV score of 0 indicates a metabolite which maintains the same significance level at all time points while the SV score of 2 indicates a metabolite which changes state from positively significant to negatively significant and vice versa at each consecutive timepoint. The closer the distributions to SV score 0 or the smaller the mean and median, the smaller the variability in the response. The comparison of SV score distribution shows:

- Elevated CO_2 response shows highest SV score distribution i.e. most variability in significance and lowest fraction of metabolites with SV score 0.

- NaCl shows the lowest SV score distribution i.e. least variation in significance level of metabolites and highest fraction of metabolites with SV score 0.

- Highest Recorded SV score for all distributions was 1.2 to 1.4 and most of the metabolites showed a score below 1 which indicates an average change from significant to non-significant at each consecutive time points.

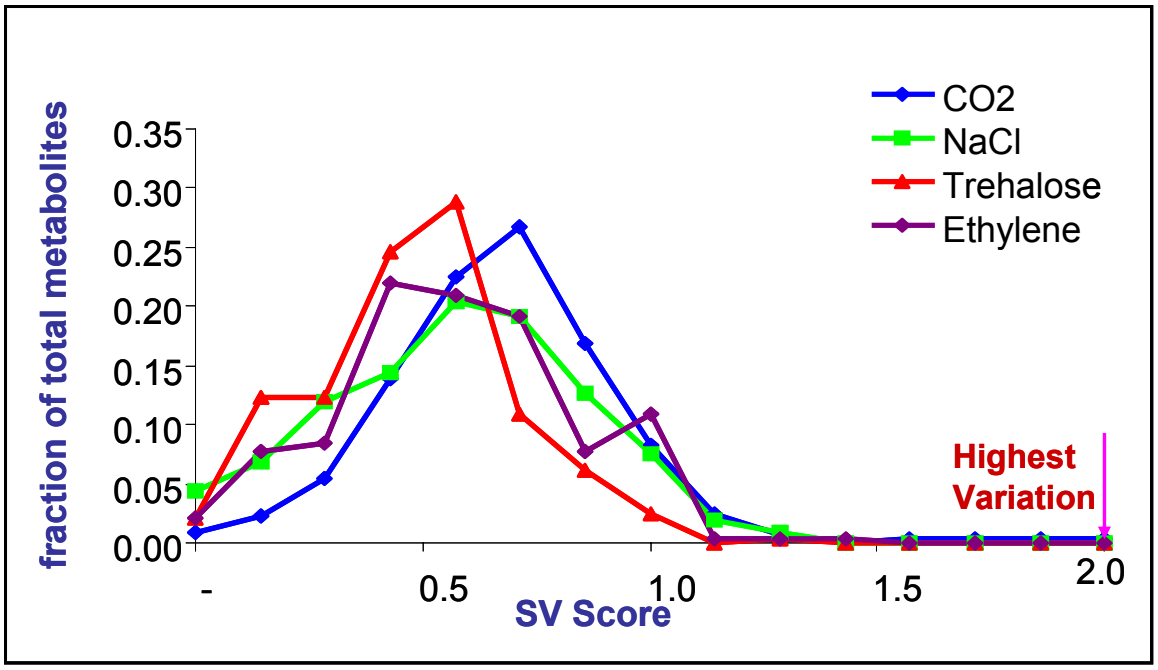


Figure 6-3: Comparison of Significance Variability (SV) Score distribution of four individual perturbations. SV score 0 represents no change in significance state, SV score 2 represents highest possible variability in significance state.

6.1.4 Significance Correlation Network

From the variation in number of significant metabolites and SV score distribution it is clear that the response of metabolic pools to different perturbations shows different dynamics. In order to identify which time-points show similarity in their response a Significance Correlation Matrix was used which is a part of MiTimeS software suite (Dutta et al., 2007). The results were converted to a graphical form in the form of Significance Correlation Network for positively and negatively significant metabolites in response to each stress. The comparison of these networks for positively and negatively significant metabolites is presented in Figure 6-4 and Figure 6-5 respectively. A connection between two timepoints represents a stronger than the average (cut off)

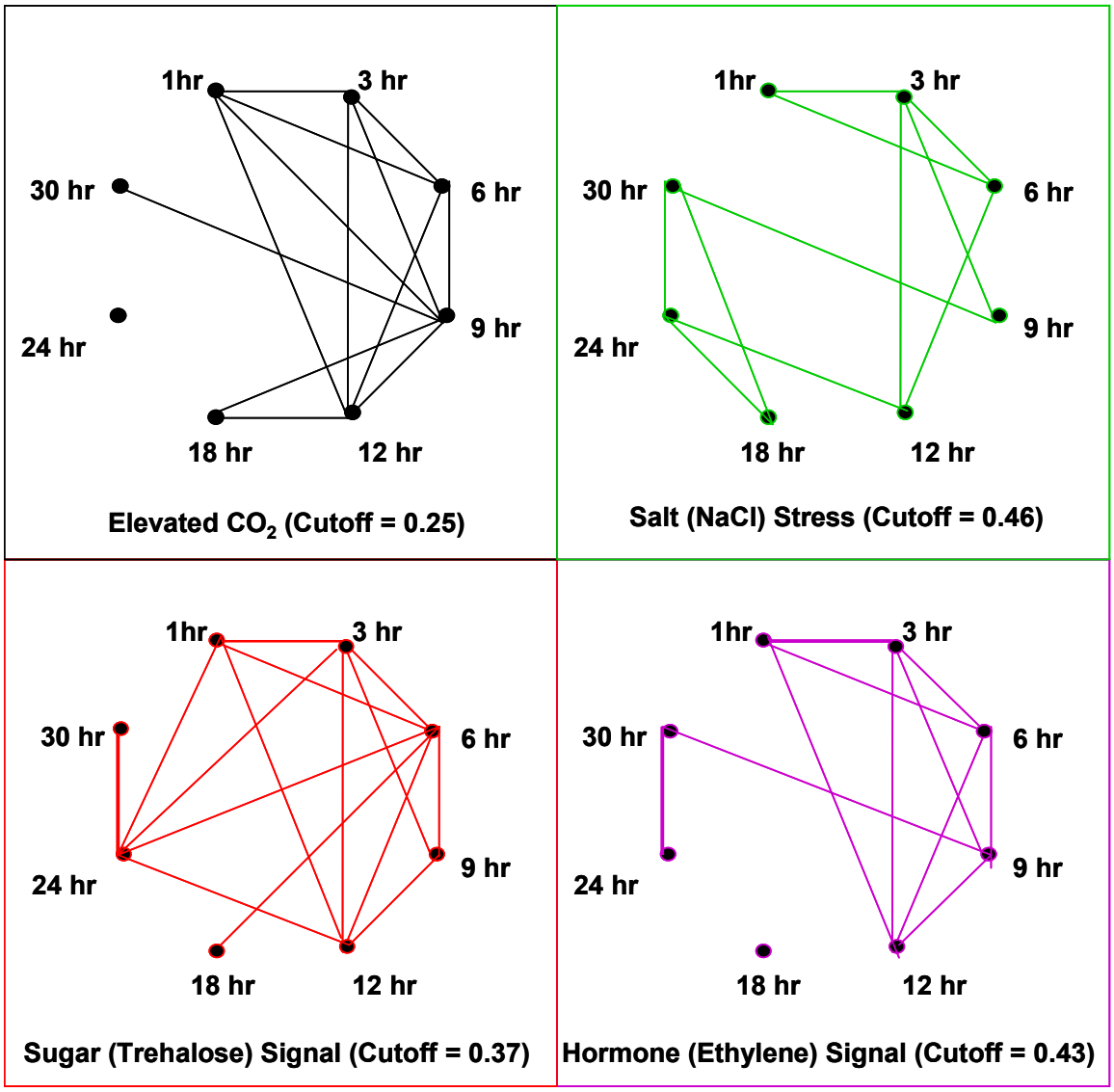


Figure 6-4: Comparison of Significance Correlation Matrix (SCM) Network for Positively Significant metabolites for four individual perturbations. A connection between two timepoints represents a stronger than the average (the cutoff) correlation between timepoints in the given experiment.

correlation between timepoints in the given experiment. A highly connected network indicates strong correlation and sustained responses to stress. A non-connected timepoint in contrast indicates a unique or transient response to the stress at that time point.

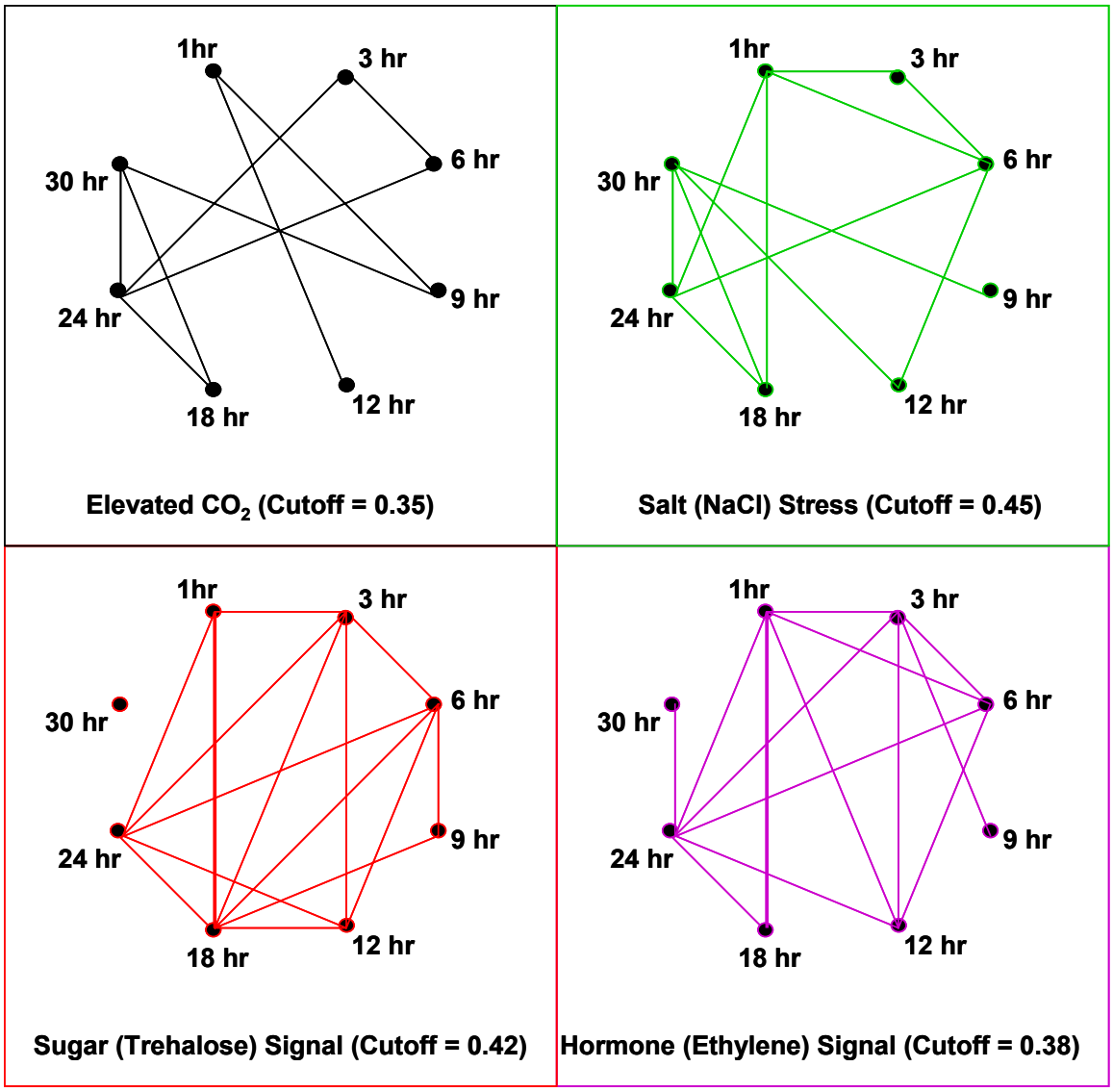


Figure 6-5: Comparison of Significance Correlation Matrix (SCM) Network for Negatively Significant metabolites for four individual perturbations. A connection between two timepoints represents a stronger than the average (the cutoff) correlation between timepoints in the given experiment.

Comparison of positively and negatively significant SCM networks suggests:

- Most of the SCM networks showed strong correlation between timepoints during first 12h and especially for the first 6h. This suggests that most system level changes in response to stress are conserved for at least 6 to 12 hours.

- None of the SCM networks show a link between all consecutive timepoints. The closest to this scenario is the trehalose negatively significant network which showed links between all consecutive time points except between 9 and 12h and 24 and 30h.

- Among consecutive time points, 12h and 18h were disconnected in 7 out of 8 networks (trehalose negatively significant being the exception), suggesting a shift in the response of the system during these time frame.

- Among all the perturbation, salt stress showed the highest average correlation (cut off) for both positively and negatively significant metabolites but fewer connections, suggesting an even stronger correlation between the connections in Salt Stress SCMs.

- In contrast, among all perturbations CO_2 showed the lowest mean correlation values for both positively and negatively significant metabolites and also fewer connections suggesting a more dynamic system.

- Trehalose with about average mean correlation values showed the highest number of links for both positively and negatively significant metabolites.

6.1.5 Conclusions

The comparison of the results from Principal Component Analysis, Paired-SAM analysis, Variation in Significant metabolites, Significance Variability (SV) Scores and

Significance Correlation Matrix (SCM) allowed following conclusions to be made about the response of *A. thaliana* liquid cultures to four individual perturbations:

- The strongest response in metabolism was shown for Salt Stress (NaCl) and first 12h of Ethylene Signal, followed by Trehalose and Elevated CO2.
- The dynamics showed two common patterns:
    - Strong correlation in response between timepoints from 1h-6h or 1h-12h.
    - Low correlation between response at 12h and 18h at all time points, suggesting a shift in the response in between these timepoints.
    - Increase in metabolite pools (more positively significant metabolites and less negatively significant metabolites) between 24 to 30h.
- The stronger the response of the system to the stress, the more conserved was the response showing lower dynamics or variability between timepoints. Hence NaCl stress showed the least variability and the Elevated CO_2 stress showed the highest variability.

Thus comparison of the same system allows additional general conclusions to be made about the dynamic regulations in the system in addition to the information about the specific response of the system to each individual stress. The effects on variability and dynamics of the system response, when these stresses with widely varying dynamic effects are combined, have been discussed in the following sections.

6.2 COMBINED SALT (NACL) AND ELEVATED CO_2 PERTURBATIONS

In the previous chapter, the effect of individual perturbations on the metabolism of plants to four individually applied stresses was studied using time-series metabolomic analysis. To further gain insight into the response of the system and regulation of plant primary metabolism, for the first, time-series metabolomic analysis was used to study effect of combined perturbations. The unique experimental design, not only allows us to understand the effect of the combined stress, but because the individual and combined stress were applied at the same biological growth phase of the plants, also allow us to study the interaction between the two perturbations. As discussed in previous section of this Chapter, the two stresses 50 mM NaCl and 1% elevated CO2 stress have very different effects on the system which were almost perpendicular to each other as observed from Principal component analysis. In this section, the results from the combined stress and its comparison with individual stress are reported.

6.2.1 Principal Component Analysis

The principal component analysis of the combined stress metabolomic profiles and control experiment is shown in Figure 6-6, along with the comparison with the individual stress. As can be seen from Figure 6-6, the combined stress also shows a perpendicular response showing a separation along principal component 1 which accounts for the highest variation in the system. However unlike NaCl stress it also shows a significant spread along principal component 2, which is the direction of the time variation in the

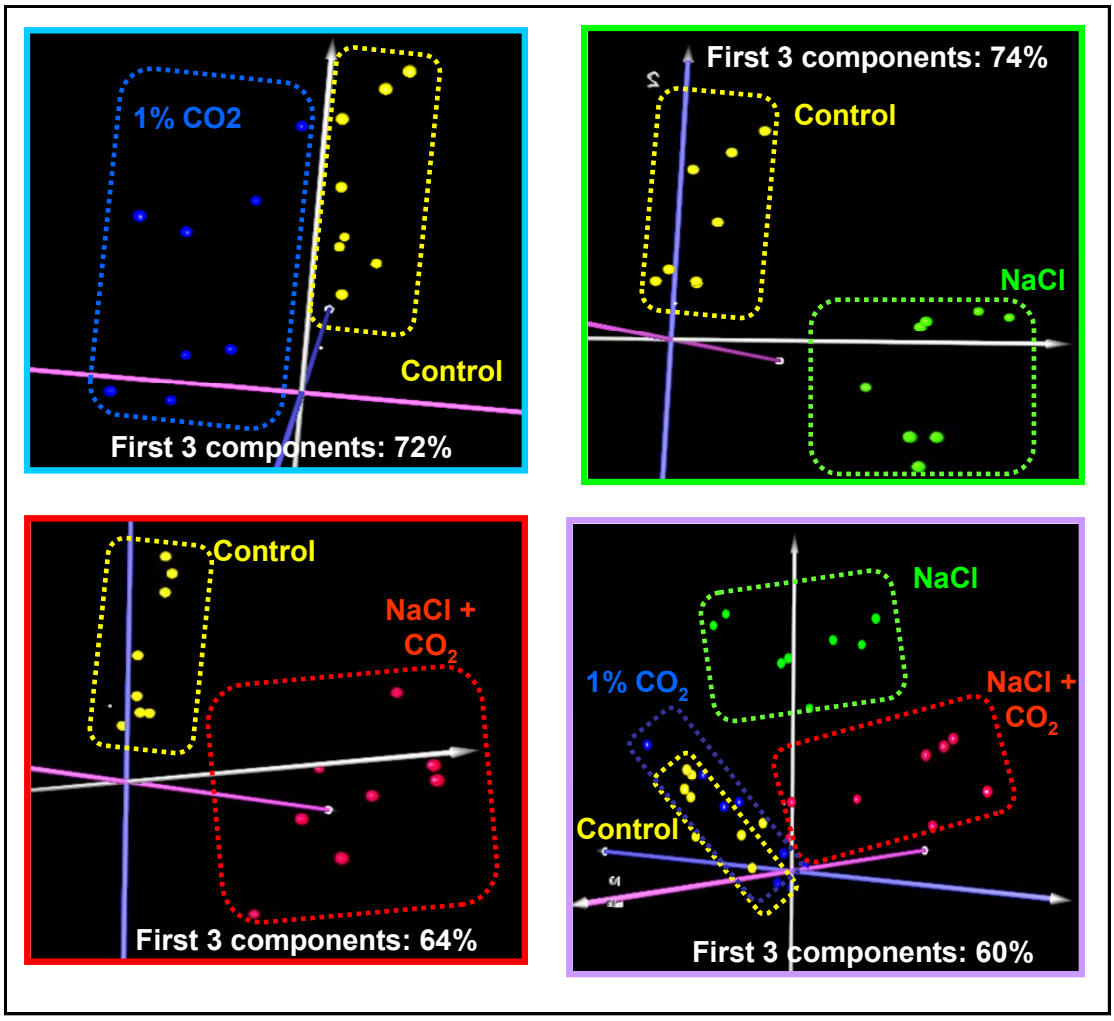


Figure 6-6: Comparison of Principal Component Analysis (PCA) of elevated CO_2, Salt (NaCl) and the combined perturbations.

control experiments. The combined PCA of all four stresses in Figure 6-6 clearly shows:

- The combined stress has unique response separate from the individual response.
- The combined stress shows a shift of metabolism in the direction of salt stress but is closer to the control stress than only salt stress.

Thus PCA analysis thus clearly indicates a unique response of the system which seem to be reducing the strong effect of the salt stress in the plants.

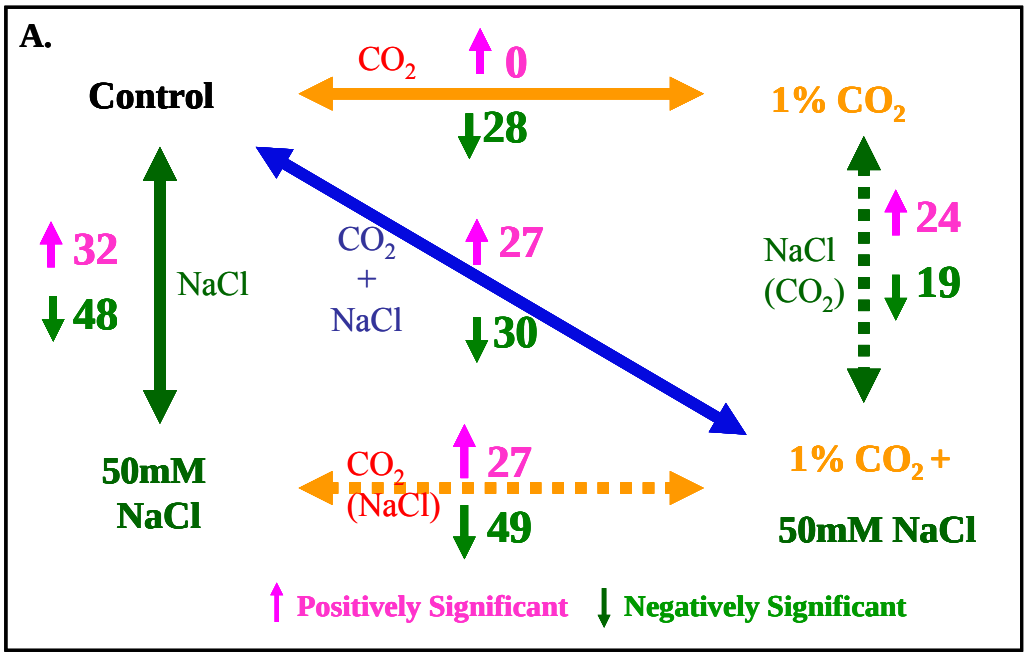
6.2.2 Significant Metabolites

Paired-SAM analysis and MiTimeS algorithm were used for identifying the significant metabolites from overall analysis and individual time points as before. The effect of the combined stress could be identified by comparing the time profiles of the control experiment with the NaCl and CO_2 combined stress experiment. However in addition there were two more comparisons possible as shown in Figure 6-7. Thus comparisons between the control experiment, individual stress experiments and the combined stress experiment as shown in Figure 6-7 allow us to identify:

- Elevated CO_2 response
- Salt (NaCl) Stress response
- Elevated CO_2 + NaCl Combined Stress Response
- Elevated CO_2 response in presence of NaCl stress referred to as CO_2 (NaCl)
- NaCl response in presence of elevated CO_2 conditions referred to as NaCl (CO_2)

The comparison of the significant metabolites for each of these responses from paired-SAM and MiTimeS algorithm are shown in Figure 6-7. The comparison shows following general features about the response:

- The combined stress has less significant metabolites as compared to NaCl stress but more significant metabolites as compared to elevated CO_2 stress.
- CO_2 (NaCl) has more significant metabolites as compared to CO_2 alone from



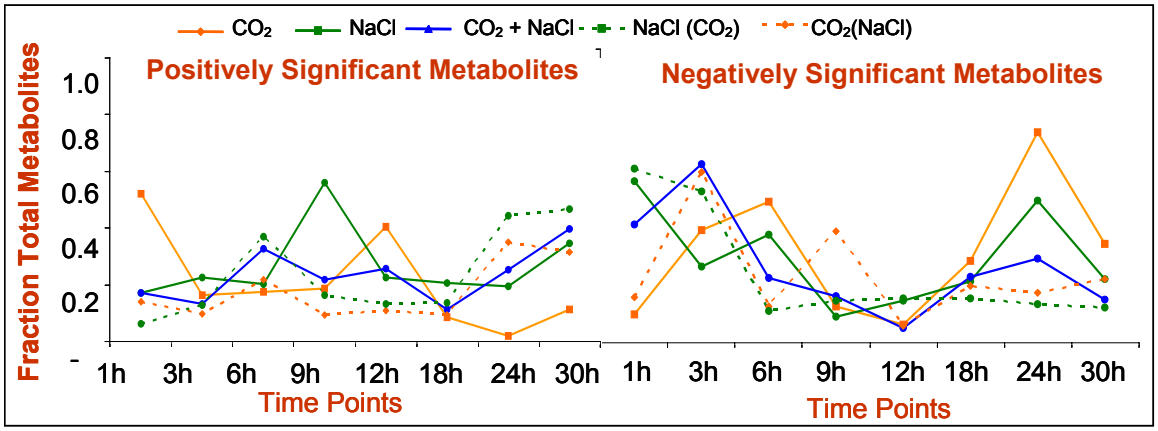


Figure 6-7: Comparison of positively and negatively significant metabolites from overall analysis and elevated CO_2, Salt (NaCl) Stress, Combined Stress, CO_2 Stress in presence of NaCl stress – CO_2 (NaCl) and NaCl Stress in presence of CO_2 stress – NaCl(CO_2) from (A) paired-SAM (B) MiTimeS analysis.

paired SAM analysis. This indicates elevated CO_2 perturbation on its own but in presence of NaCl perturbs more biochemical pathways than individually. Converse is true for NaCl (CO_2) which shows a decrease from NaCl response.

Similarly from Figure 6-7 (B) following observations can be made about the significant metabolites at each individual time points:

- The significant increase in more than 60% of metabolite pools observed in response to elevated CO_2 effect alone was not observed in CO_2 effect in presence of NaCl or the combined response.

- The number of significant metabolites in the combined stress showed lower variation in significant metabolites as compared to the individual CO_2 and NaCl stress. For e.g. for positively significant metabolite the large increase in number of positively significant metabolites in response to elevated CO_2 at 12h and in response to NaCl at 9h was not observed in response to the combined stress.

- Finally the profile for variation in number of significant metabolites for NaCl(CO_2) and CO_2 (NaCl) showed similar trends as the combined stress (CO_2+NaCl) profile rather than the response of the two stresses individually.

These results show that the variability in the system in response to individual stress is reduced significantly in case of combined stress. Also even though the CO_2 and NaCl stress were perpendicular stresses, i.e. showing very different regulation of the biochemical network, when applied in combination, they significantly altered response of each other. If these two stresses were indeed independent, a more closer similarity would have been observed between CO_2 (NaCl) and CO_2 and similarly between NaCl (CO_2) and NaCl however this was not the case based on the observations above of the significant

metabolites. This clearly suggests both the stresses interact and moderate the other stress's response.

6.2.3 Significance Variability (SV) Score Distribution

In order to better understand this effect, the SV score distribution of the metabolites in response to the five comparisons were plotted as shown in Figure 6-8. All three plots clearly indicate that the combination of stress significantly reduces the variability. Specifically the plots show:

- CO_2 (NaCl) i.e. CO_2 effect in presence of NaCl had SV score distribution with similar profile but shifted closer towards 0, suggesting lower variability, with higher number of metabolites with SV score 0, lower mean and median as compared to elevated CO_2 stress only.

- NaCl (CO_2) i.e. NaCl effect in presence of elevated CO_2 had SV score distribution with similar profile but shifted closer to 0 with a significantly lower mode, mean and median as compared to NaCl stress only.

- The combined stress NaCl + CO_2 also shows a similar shift, though to a lesser extent, moving towards left closer to zero.

These observations reinforce the previous observations from PCA (Figure 6-6) and variation of significant metabolites (Figure 6-7) analysis that the combined stress reduces the effects of individual stress and brings system closer to control conditions. However they are based only on the statistical analysis.

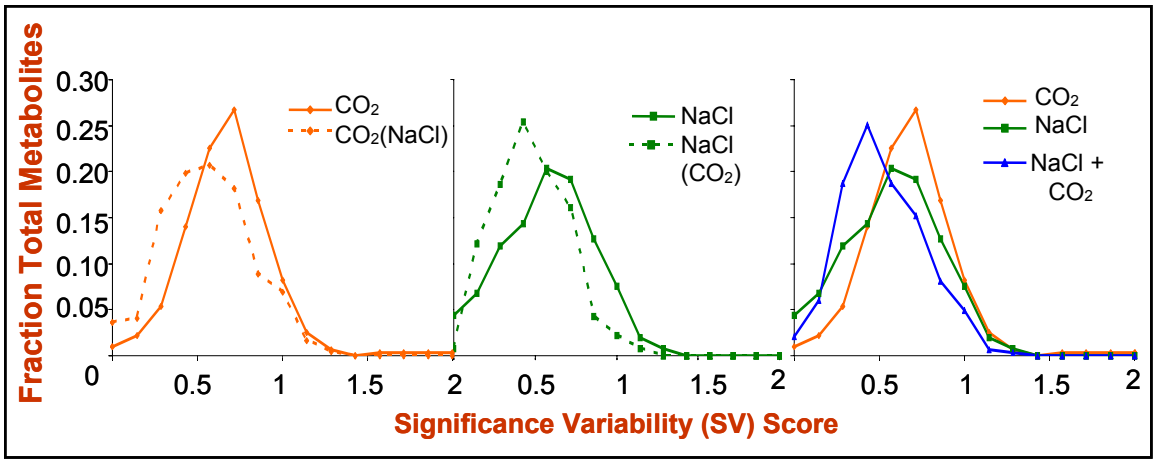


Figure 6-8: Comparison of Significance Variability (SV) Score distribution for elevated CO_2, Salt (stress), Combined stress, Salt Stress Response in presence of CO_2 stress NaCl (CO_2) and CO_2 stress in presence of Salt Stress – CO_2 (NaCl). SV score 0 represents no change in significance state, SV score 2 represents highest possible variability in significance state.

In order to understand this better the response of the system will be analyzed in the context of TCA cycle and Amino acid biosynthesis pathways which was perturbed significantly in response to both the perturbations.

6.2.4 TCA cycle and Amino Acid Biosynthesis in Combined Stress

The effect of elevated CO_2 and NaCl stress on these pathways has been discussed in detail in Chapter 5 of this report. The most important observation in response to elevated CO_2 in these pathways were:

- Elevated CO_2 significantly increased the concentration of most amino acids during first 12 hours, but decreased concentration at later time points for most amino acids.

- Elevated CO_2 had difference response on two parts of the TCA cycle, from Citrate to α-ketoglutarate, and the other between succinate to malate.

Similarly, the most important observation in response to Salt Stress in these pathways were:

- Salt stress increased concentration of amino acids involved in production of osmoprotectants along with some other amino acids but decreased concentration of several amino acids in the competing pathway.

- Salt stress significantly increased concentration of most TCA cycle intermediates except Malate.

The significance level of all known metabolites in this pathway in response to both these stress are reproduced along with the response of these metabolites pools to the combined stress in Figure 6-9 for easy comparison. Comparison of the individual perturbations with the combined perturbation immediately shows:

- The increase in TCA cycle intermediates from aconitate to fumarate observed in response to NaCl is not observed in the combined stress.

- Malate which was negatively significant at almost all time points in NaCl stress is non-significant at most time points in the combined stress.

- A number of amino acids which showed significant increase (homoserine, methionine, β-alanine and glycine) in response to Salt stress showed a much lower increase (methionine , β-alanine) in response to the combined stress, with

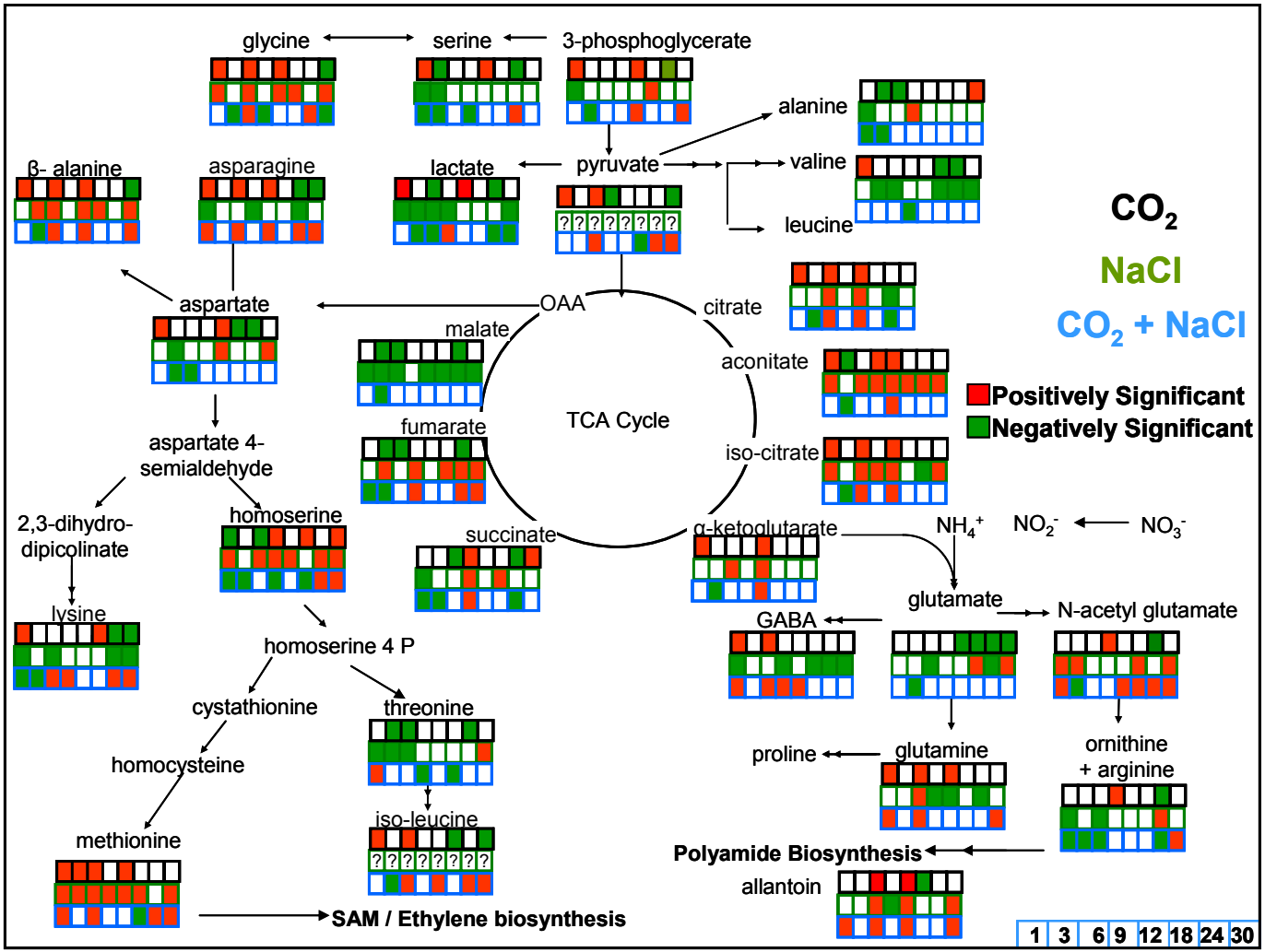


Figure 6-9: Comparison of Significance levels of individual metabolites in the TCA cycle and amino acid biosynthesis pathway at individual time points in response to Elevated CO_2, Salt Stress and the Combined Elevated CO_2 and Salt Stress Perturbations.

some even becoming negatively significant (homoserine, glycine) at some of the timepoints.

- Amino acids which showed significant decrease in metabolite pools at most time points in response to NaCl stress (asparagine, lysine, gaba, valine)were instead non-significant (valine) and even positively significant (aspargine, lysine and gaba) at those timepoints in response to the combined stress.

The above observations based on Figure 6-9 clearly show that most of the metabolite pools showing significant change in response to Slat stress the stronger of the two individual stress did not show the same effect in the combined stress, with some metabolites even showing opposite trend. In addition, some of the amino acids glycine and homoserine which show a significant increase in response to elevated CO_2 as well as Salt stress during the first 12 hours, showed significant decrease at these time points in response to the combined stress. These observations suggest significant changes in metabolomic pools in response to NaCl stress are altered in presence of elevated CO_2 levels in the combined stress. Thus CO_2 in presence of NaCl stress seem to be creating an opposite effect of salt stress so that the overall variability of the system is reduced. In order to see this "opposite" effects of elevated CO2 response to that of individual Salt stress, the significance level of metabolites for these comparisons is shown in Figure 6-10. As can be seen from the figure 13 of the total 26 metabolite pools in these pathways showed opposite significance level in

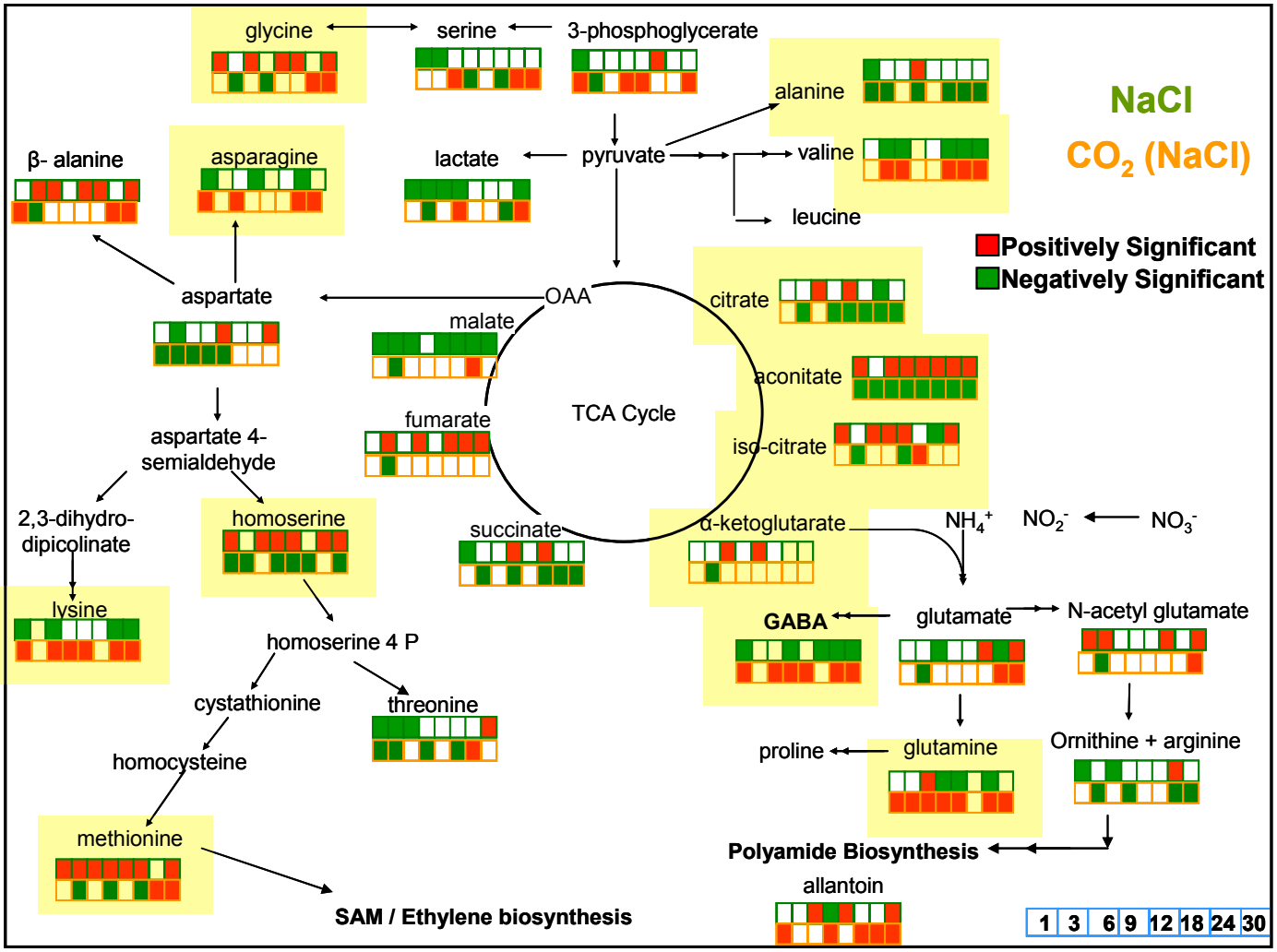


Figure 6-10: Comparison of Significance levels of individual metabolites in the TCA cycle and amino acid biosynthesis pathway at individual time points in response to Salt Stress and response to Elevated CO_2 in presence of Salt Stress.

response to elevated CO_2 in presence of NaCl as compared to NaCl stress alone. These results clearly show that in the combined stress, elevated CO_2 seem to relieve the strong effect of salt stress reducing its impact by using the available additional CO_2 (and possibly corresponding increase in nitrate reductase providing additional nitrogen) to produce the osmoprotectants rather than re-configuring available metabolite pools (indicated by opposite response of competing amino acids) in Salt stress. Similar observations were also made for several other metabolites though the same are not shown in the current report.

This is an important observation which provides an additional direction for engineering plants with better osmotic stress tolerance by improving their carbon or nitrogen fixation or reducing starch accumulation in favor of more metabolite pools, rather than simply increasing concentration of osmoprotectants. In order to however validate this further more experiments are needed looking at long term physiological and molecular effects of the combined stress to ensure this short-term response can be sustained over a longer period of time.

The above result from the systems biology perspective also clearly shows the non-linearity of the response of plant system suggested at the beginning of this chapter. Seemingly perpendicular or independent stresses also alter each other significantly when applied together creating a unique response. Thus it is not the strength or independence, but the biological role of each perturbation which seem to regulate the response of the combined stress. One more similar example of response of *A. thaliana* liquid cultures to

moderate but less dynamic sugar stress (trehalose) and elevated CO_2 are combined is presented in the next section.

6.3 COMBINED TREHALOSE SIGNAL AND ELEVATED CO_2 PERTURBATIONS

The combined effect of elevated CO_2 and Salt stress showed a significantly altered response which reduced the strong effects of NaCl stress. In this section the effect of combination of Trehalose and elevated CO_2 stress are presented to further understand the common elements of the response of the system to combined stress by similar comparisons and statistical analysis as was described in the previous section.

6.3.1 Principal Component Analysis

The Principal Component Analysis (PCA) of the combined stress metabolomic profiles and control experiment is shown in Figure 6-11, along with the comparison with the individual stresses. The comparison of the four stresses allow following observations:

- Individual application of Trehalose and Elevated CO_2 perturbations, move the overall metabolism in principal component 1 and principal component 3 respectively, however in both cases the time variation of the experiments of control, elevated CO_2 and trehalose signal experiments is along principal component 2.

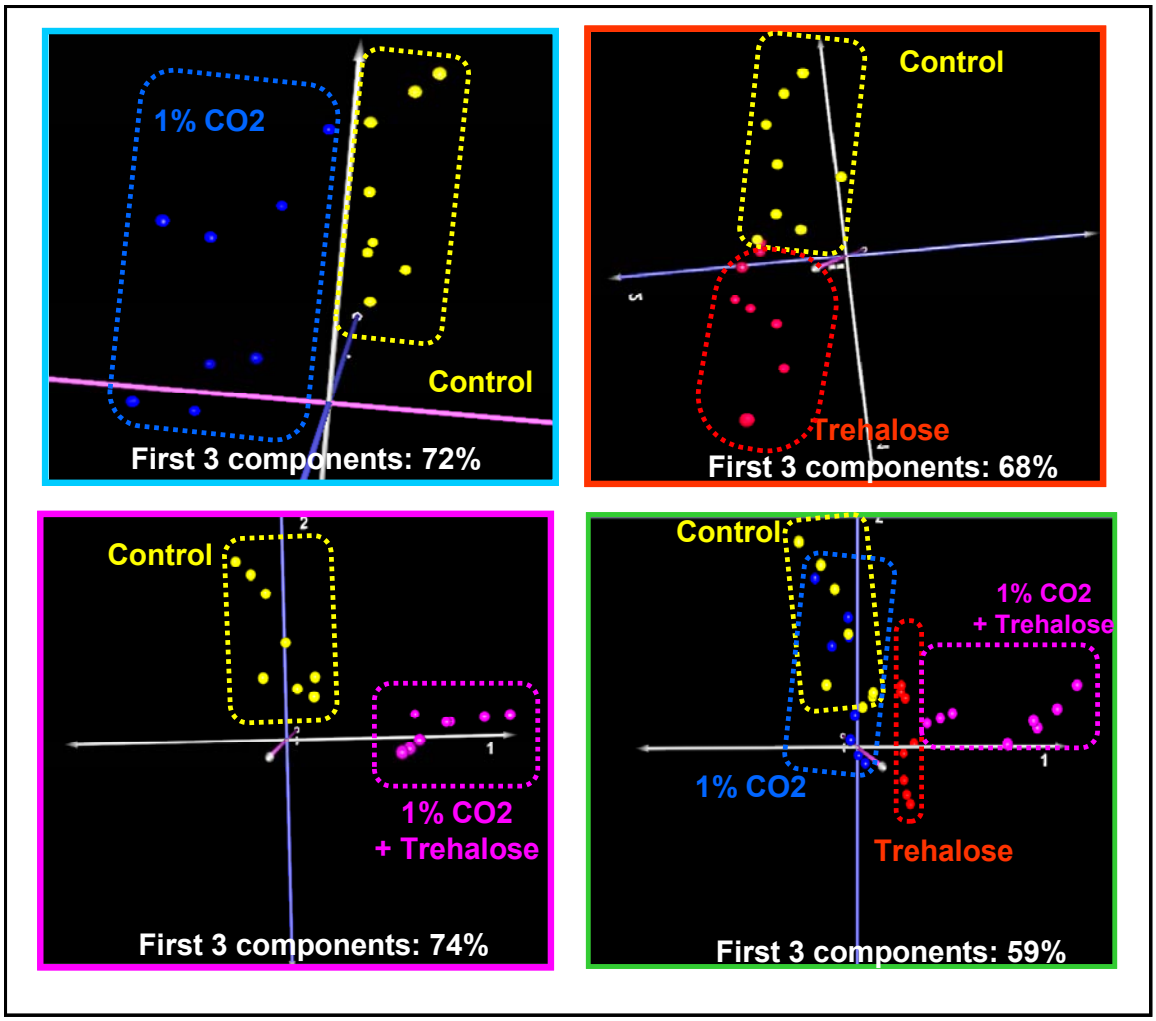


Figure 6-11: Comparison of Principal Component Analysis (PCA) of Elevated CO_2, Trehalose and the Combined perturbations.

- PCA analysis of control, individual perturbations and the combined perturbation shows similar response as individual plots. The Control and CO_2 experiments are close to each other and are not separated completely. Trehalose stress shows a clear separation from the rest of the stress moving down parallel to the control stress. The combined stress showed maximum variation and was perpendicular similar to NaCl stress.

- The combined elevated CO_2 + Trehalose perturbation shows a strong response similar to NaCl with a clear separation along principal component 1 which accounts for the most variation of the system.

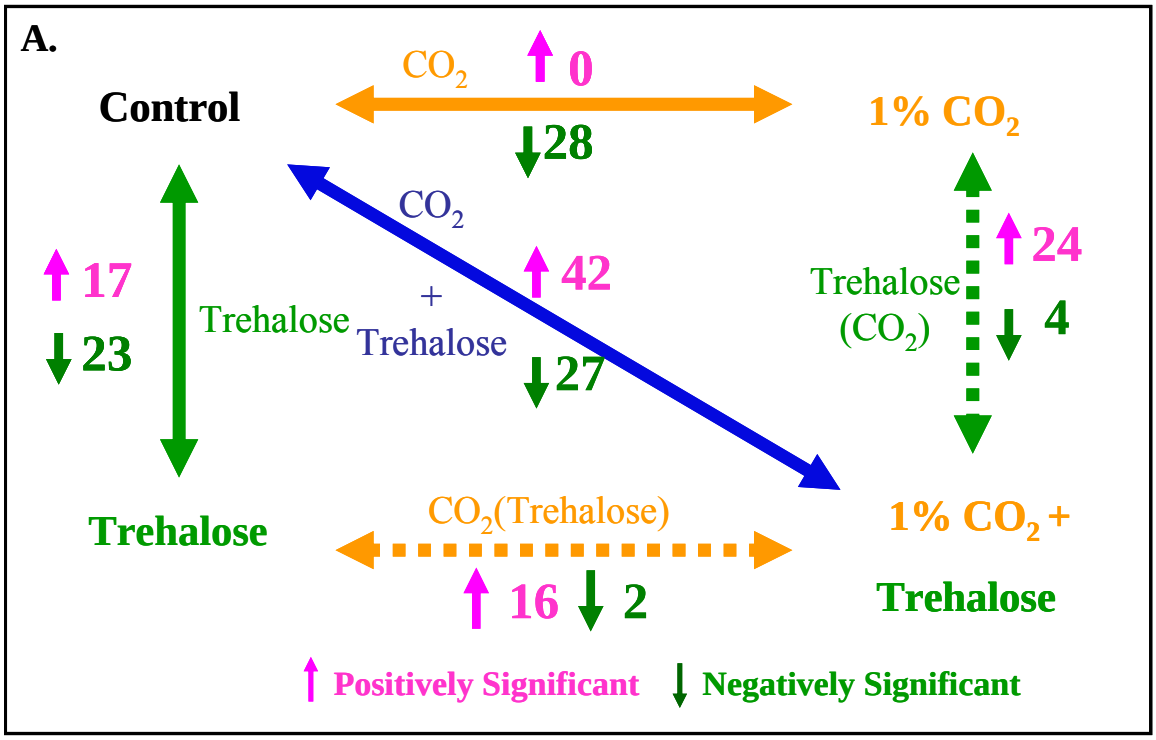
- Unlike combined NaCl and CO2 stress, the combined Trehalose and elevated Stress shows a stronger effect and moves trehalose stress away from the control stress along principal component 1.

The principal analysis thus itself shows significant difference of effect of combined stress in NaCl and trehalose. It seems to enhance the effect of trehalose rather than reduce the effect of NaCl. This will be analyzed in further detail from significance metabolite analysis and biological pathways in subsequent sections.

6.3.2 Significant Metabolites

Paired-SAM analysis and MiTimeS algorithm were used for identifying the significant metabolites from overall analysis and individual time points as before. The effect of the combined stress could be identified by comparing the time profiles of the control experiment with the Trehalose Signal and CO_2 combined stress experiment. However in addition there were two more comparisons possible like NaCl stress and as shown in Figure 6-12. Thus data analysis strategy for comparison between the control experiment, individual stress experiments and the combined stress experiment as shown in Figure 6-12 allow us to identify:

- Elevated CO_2 response



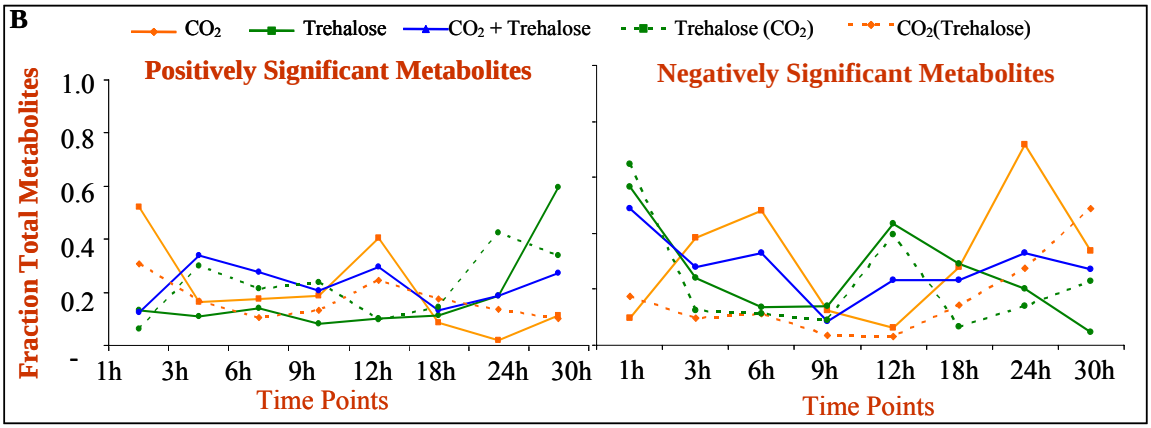


Figure 6-12: Comparison of positively and negatively significant metabolites from overall analysis and Elevated CO_2, Trehalose Signal, Combined Stress, Elevated CO_2 Stress in presence of Trehalose – CO_2 (Trehalose) and Trehalose Stress in presence of CO_2 stress – Trehalose(CO_2) from (A) paired-SAM (B) MiTimeS analysis.

- Trehalose Signal response

- Elevated CO_2 + Trehalose Signal Combined Stress Response

- Elevated CO_2 response in presence of Trehalose Signal referred to as CO_2 (Trehalose)
- Trehalose response in presence of elevated CO_2 conditions referred to as Trehalose (CO_2)

The comparison of the significant metabolites for each of these responses from paired-SAM and MiTimeS algorithm are shown in Figure 6-12. The paired-SAM comparison (Figure 6-12A) shows following general features about the response:

- The combined stress showed the highest number of significant metabolites as compared to individual stresses.
- Both interaction between the stress CO2 (Trehalose) and Trehalose (CO2) show a significantly higher number of positively significant metabolite as compared to negatively significant metabolites.

Similarly from Figure 6-12 (B) following observations can be made about the variation of significant metabolites at each individual time points:

- The combined stress shows many more significant metabolites as compared to trehalose stress at all time points except at 30h.
- Large increase in number of positively significant metabolites at 12h and 30h in elevated CO_2 and trehalose respectively is not observed in the combined stress, thus reducing variability.

- CO_2 (Trehalose) also showed a higher number of positively significant metabolites at 1h which was also observed in the individual perturbation.

- The large increase in number of negatively significant metabolites at 12 h in trehalose and 24h in elevated CO2 was not observed in the combined stress reducing variability in significance.

These results show that the variability in the system in response to individual stress is reduced significantly specially when compared to elevated CO_2 stress. As compared to trehalose stress the number of positively significant metabolites increased significantly. This seems to suggest both the stresses interact and enhance response to trehalose and reduce dynamics of elevated CO_2 stress.

6.3.3 Significance Variability (SV) Score Distribution

In order to better understand this effect, the SV score distribution of the metabolites in response to the five comparisons were plotted as shown in Figure 6-13. All three plots clearly indicate that the combination of stress significantly reduces the variability. Specifically the plots show:

- CO_2 (Trehalose) i.e. CO_2 effect in presence of Trehalose had SV score distribution with similar profile but shifted closer towards 0, suggesting lower variability, with higher number of metabolites with SV score 0 (highest number with SV score zero in all comparisons, 15% of total metabolite pools), lower mean and median as compared to elevated CO_2 stress only.

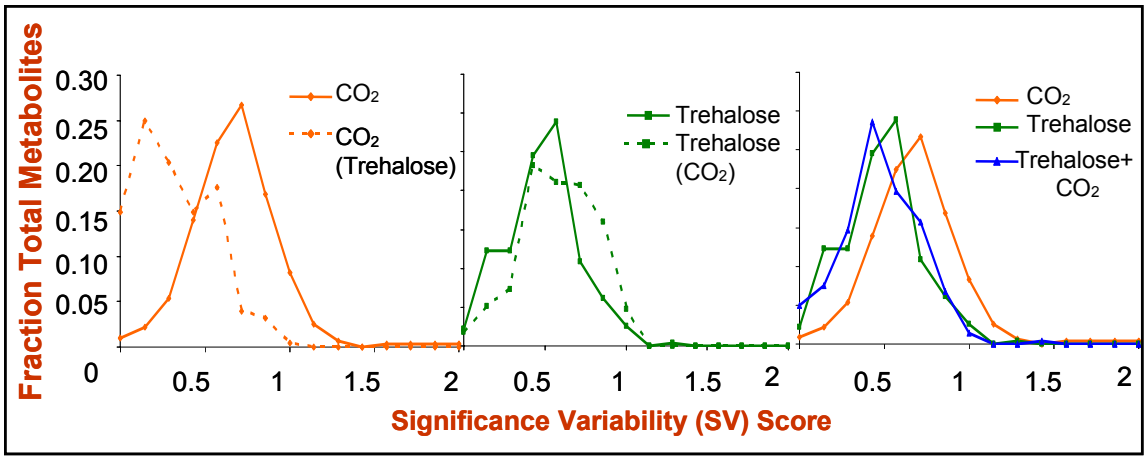


Figure 6-13: Comparison of Significance Variability (SV) Score distribution for Elevated CO_2, Trehalose, Combined stress, Trehalose Response in presence of Elevated CO_2-Trehalose(CO_2) and Elevated CO_2 in presence of Trehalose–CO_2(Trehalose). SV score 0 represents no change in significance state, SV score 2 represents highest possible variability in significance state.

- Trehalose (CO_2) i.e. Trehalose effect in presence of elevated CO_2 had SV score distribution almost similar with a slight decrease in mean.

- The combined stress Trehalose + CO_2 also shows a similar shift, though to a lesser extent when compared to trehalose, moving towards left closer to zero

These observations reinforce the previous observations from variation of significant metabolites (Figure 6-12) analysis that the combined stress reduces the variability in significance or dynamics of the metabolite profiles specially when compared to elevated CO_2.

6.3.4 TCA cycle and Amino Acid Biosynthesis in Combined Stress

The effect of elevated CO_2 and Trehalose stress on these pathways has been discussed in detail in Chapter 5 of this report. The most important observation in response to elevated CO_2 in these pathways were:

- Elevated CO_2 significantly increased the concentration of most amino acids during first 12 hours, but decreased concentration at later time points for most amino acids.

- Elevated CO_2 had difference response on two parts of the TCA cycle, from Citrate to α-ketoglutarate, and the other between succinate to malate.

Similarly, the most important observations in response to Trehalose response in these pathways were:

- Trehalose significantly increased concentration of most of the amino acids.

- Trehalose had very small effects on TCA cycle intermediates.

The significance level of all known metabolites in this pathway in response to both these stresses are reproduced along with the response of these metabolites pools to the combined stress in Figure 6-14 for easy comparison. Comparison of the individual perturbations with the combined perturbation immediately shows:

- Amino acids which showed a significant increase in response to trehalose also showed significant increase in response to the combined perturbation.

- Aspartate and Valine the amino acids which were negatively significant in response to trehalose showed a lower decrease and were also positively significant at few time points, further suggesting increase in amino acid productions.

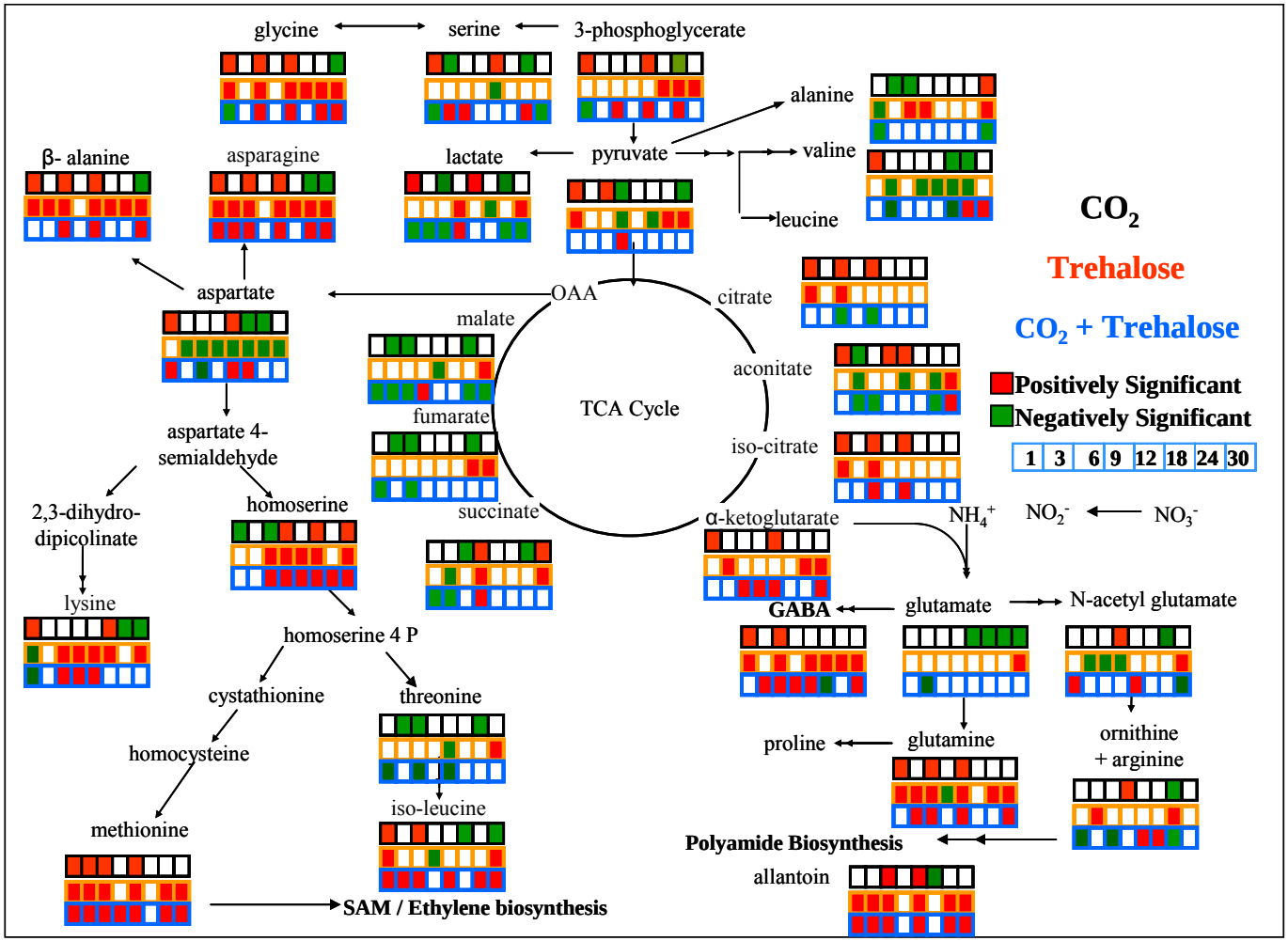


Figure 6-14: Comparison of Significance levels of individual metabolites in the TCA cycle and amino acid biosynthesis pathway at individual timepoints in response to Elevated CO_2, Trehalose Signal and the Combined Elevated CO_2 and Trehalose Signal Perturbations.

- TCA cycle metabolites were found to be more negatively significant in response to combined stress, suggesting possibly a reduction in respiration. Malate showed the largest decrease among all TCA cycle metabolites, while α-ketoglutarate, the precursor for major amino acids was more positively significant.

- Lactate showed a significant decrease suggesting decrease in anaerobic respiration in response to the combined stress which was not observed in individual stress.

In addition to similarity in TCA cycle and amino acids biosynthesis pathway similar trend was also shown in other pathways specially the sugar metabolites. Two metabolites which showed a significant increase as compared to individual stress in the combined stress were intracellular trehalose and galactinol. Figure 6-15 shows the relative peak area of trehalose in the four experiments. The figure showed a significant increase in trehalose concentration in response to the combined stress. Thus CO_2 in presence of trehalose increases trehalose uptake rate thus increasing its intracellular concentration. The higher concentration could also be the main reason behind a stronger trehalose response in the combined stress. The trehalose profile further shows after reaching the highest trehalose concentration at 9h, the trehalose concentration does not increase but decreases slightly over time suggesting possibly an increase in trehalase enzyme activity at these high trehalose concentrations.

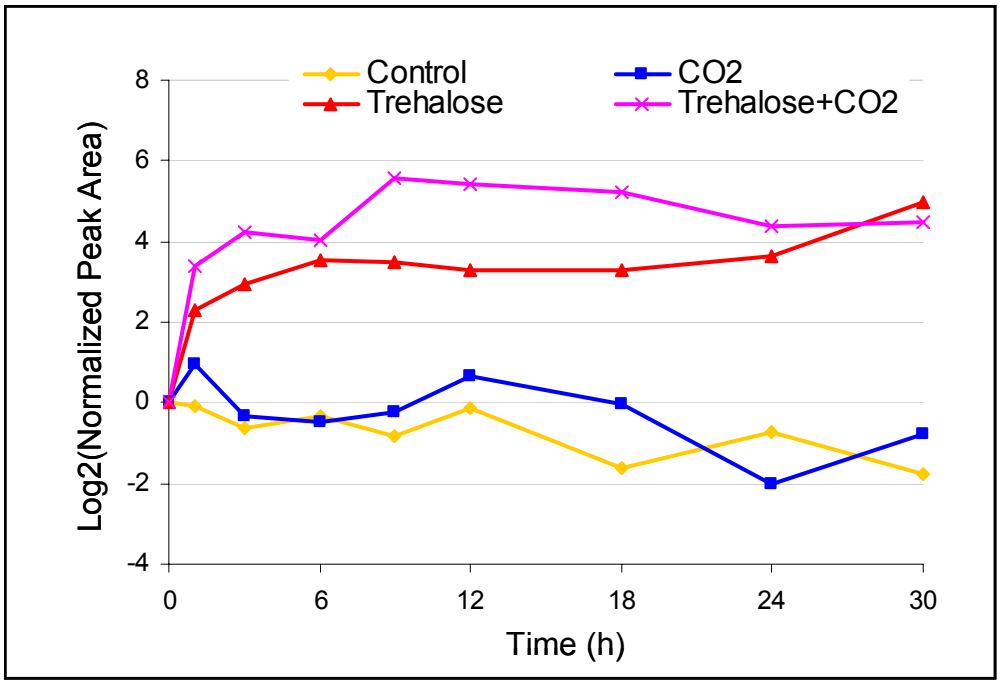


Figure 6-15: Intracellular Trehalose concentration in Control, Elevated CO2, Trehalose and the combined Trehalose+CO2 perturbation suggests, a significant increase in intracellular concentration of trehalose in the combined stress accumulating up to 30 times the original concentration.

6.3.5 Comparison of Combined Stresses

Comparing the effect of individual and combined stress with elevated CO_2 stress suggests the same elevated CO_2 stress can play different regulatory roles depending on the presence of the other stress. In case of salt stress, elevated CO_2 decrease the perturbation due to salt stress by showing an opposite regulation of primary metabolism as the individual stress. In contrast, in trehalose stress the CO_2 significantly magnifies the trehalose response by showing a larger change for the combined stress in the same direction as the combined stress. The three elevated CO_2 stresses: only Elevated CO_2, CO_2 in presence of NaCl and CO_2 in presence of Trehalose show completely different significance profiles, SV score distribution and biology with no significant overlap between them. These results, from the systems biology perspective also clearly

demonstrate the non-linear nature of the biological processes and their regulation in complex biological systems such as plants. In addition they provide a frame-work for designing future combined perturbations studies and their multivariate data analysis.

6.4 RECONSTRUCTION OF BIOCHEMICAL NETWORK

One of the important motivations for Metabolic Engineering and Systems Biology studies is to re-construct the regulatory networks which are present in complex biological systems. Detailed regulatory networks of biological systems can then be exploited for:

- Developing strategies to identify and correct these networks whenever they are disturbed causing a disease state.
- Developing strategies for engineering systems to add or remove regulatory elements to develop the desired traits (such as over production of metabolites)

Dynamic metabolomic profiles obtained in response to systematically perturbed network contain within them the results of changes taking place in transcriptomic and proteomic levels in response to the perturbations. These profiles are expected to show a strong correlation when:

- Metabolites are regulated by a common regulatory elements
- Metabolites are in near equilibrium reactions
- Metabolites belong to the same pathway whose enzymatic activities are regulated.

Thus systematically measuring correlation of metabolites can allow re-construction of biochemical regulatory networks governing the response of the system to the perturbations. In case of systematically perturbed biological system, the variation in correlation between metabolites between different experiment can even help identification of gene/protein responsible for disrupting the co-regulatory networks.

An immediate practical advantage of such re-construction of metabolomic networks is identification of novel pathways in plants which are not yet identified. This is especially important in plants, since as compared to microbial systems and even mammalian systems, information about several pathways is putative or derived from other biological systems. In-spite of these advantages, there have been very few studies so far with limited success at re-construction of the pathways due to non-availability of dynamic profiles in systematically perturbed network.

In order to re-construct the biochemical network using dynamic metabolomic profile obtained from different perturbations, Pavlidis Template Matching (PTM) was used. Specifically, methionine time profile was used as a template to match metabolite profiles which showed strong correlation in control, elevated CO2, salt stress, trehalose signal and ethylene signal perturbations, representing 40 different metabolic states under different conditions (time and/or environment). Identifying correlations in profiles in response to all these perturbations reduces the possibility of a "chance" correlation and is expected to show a match with a high confidence (or very p-values). In order to see which profiles

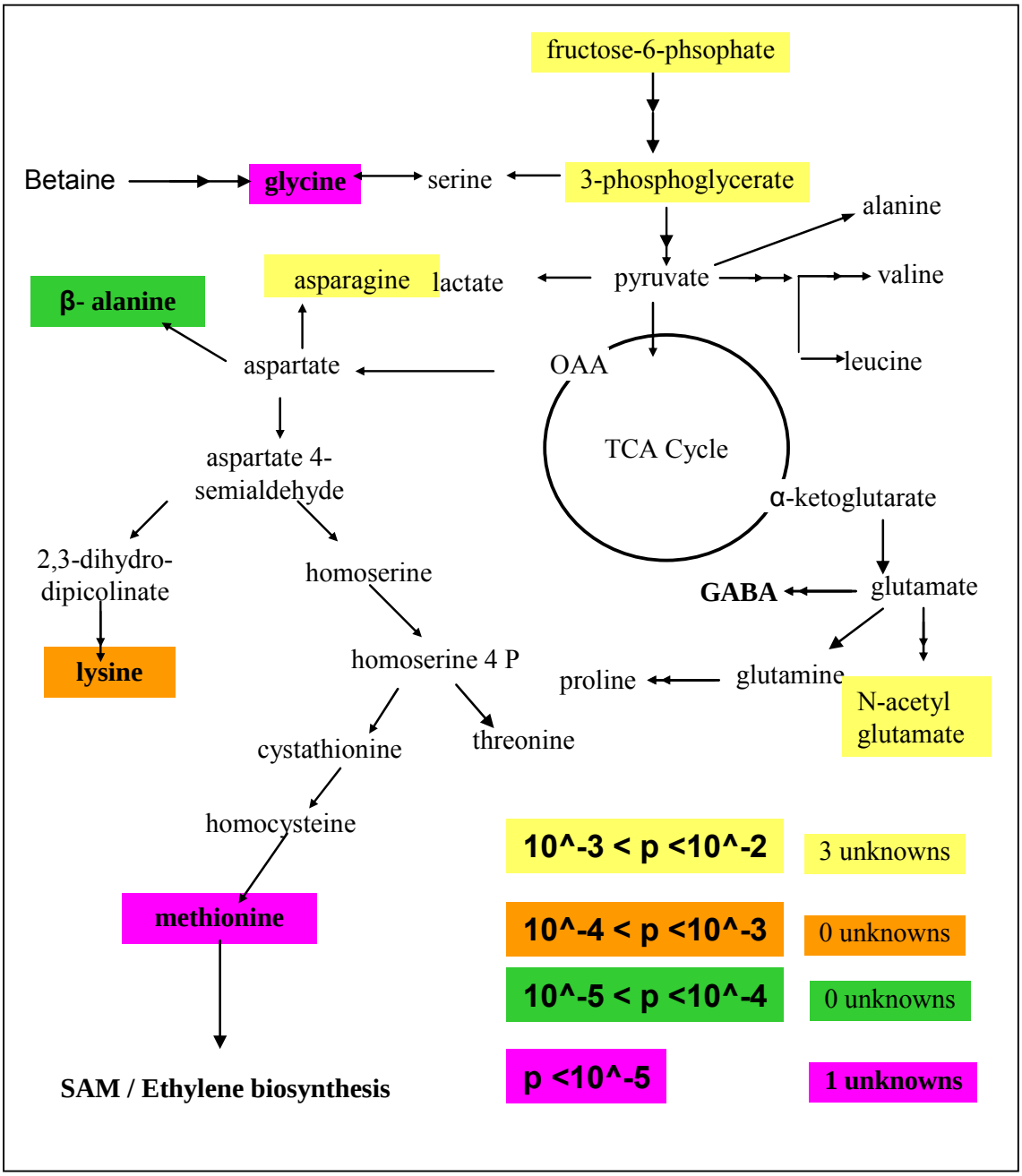


Figure 6-16: Results from Pavlidis Template Matching (PTM) for the metabolomic time profiles obtained from five different experiments with Pearson Correlation distance matrix. PTM results at different p values using methionine as a template. The analysis indicates presence of a novel biochemical pathway or co-regulatory elements for methionine, glycine and an unknown metabolite.

were most correlated, PTM analysis was used with successively increasing stringency (decreasing p-values). The results from these PTM analyses are shown in Figure 6-16.

As can be seen from the Figure 6-16, p-value cut off of 0.01 identified 11 metabolites matching the methionine pathway. These metabolite matches were all across the pathways and with 4 unknown metabolites. As the stringency of match was increased and p-value cutoff was decreased to 0.001, only four metabolites (glycine, lysine, β-alanine and an unknown) showed a match to methionine at this stringency. From these lysine, β-alanine and methionine are bio-chemically connected as they are all derived from aspartate. Further increasing the stringency to p-value cut off to 0.00001, showed a match with only glycine and an unknown metabolite. The profile of these metabolites is shown in Figure 6-17 below. As can be seen from the figure all these metabolite show a similar pattern of variation in response to perturbations though the absolute values of response are different. These pattern is conserved across all four perturbations even when these metabolites show up to 8 times increase in their original concentration at 0h.

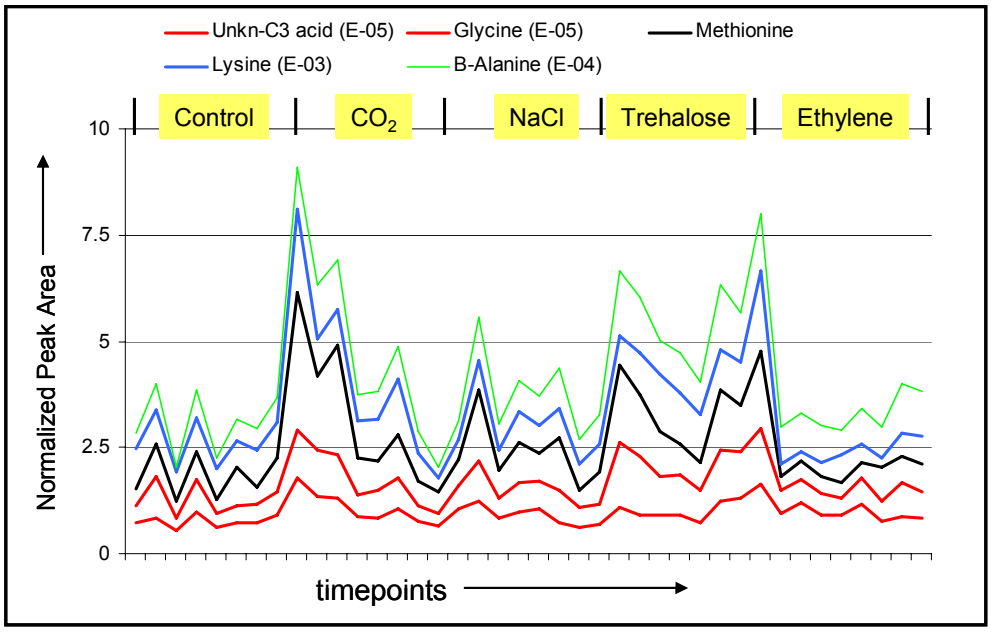


Figure 6-17: Time profiles of metabolites showing strong correlations to methionine during all perturbations using PTM analysis (see Figure 6-15).

These correlations suggest a strong possibility of co-regulation between methionine and glycine which can not be explained easily based on the existing knowledge as their biochemical pathways are expected to be independent. A further study of the biochemical pathways involved in their production suggested two possible links between these two metabolites (glycine and methionine) shown in Figure 6-18.

As can be seen from the figure, based on the current knowledge, methionine is produced by methylation of homocysteine. This methylation step requires a methyl donor. The Arabidopsis genome contains a number of genes which produce enzymes catalyzing this methylation steps. The known methyl donors for this reaction are S-Adenosylmethionine (which is produced from methionine itself and is a methyl donor in a number of other reactions) and 5-methyltetrahydropteroyltriglutamate. However there is an additional methyl donor-gene system common to several biological system but whose presence or gene (5-methyltetrahydrofolate-homocysteine S-methyltransferase or vitamin B12 dependent S-methyltransferase) involved has not yet been identified in Arabidopsis as yet, which uses 5-methyltetrahydrofolate as the methyl donor. The de-methylated product of the methyl donor is known to be the substrate for the gene producing glycine from serine as shown in Figure 6-18. The gene converting the product – 5,10-methyleneltetrahydrofolate to the methyl donor of methionine biosynthesis has also been identified in Arabidopsis. These pathway suggest the methyl group required for methionine biosynthesis effectively is received from the demethylation of serine to form glycine. Thus the amount of glycine produced regulates the methyl groups available for methionine production. In addition it also suggests a common regulatory element – folic acid which provides the main skeleton for the methyl donors.

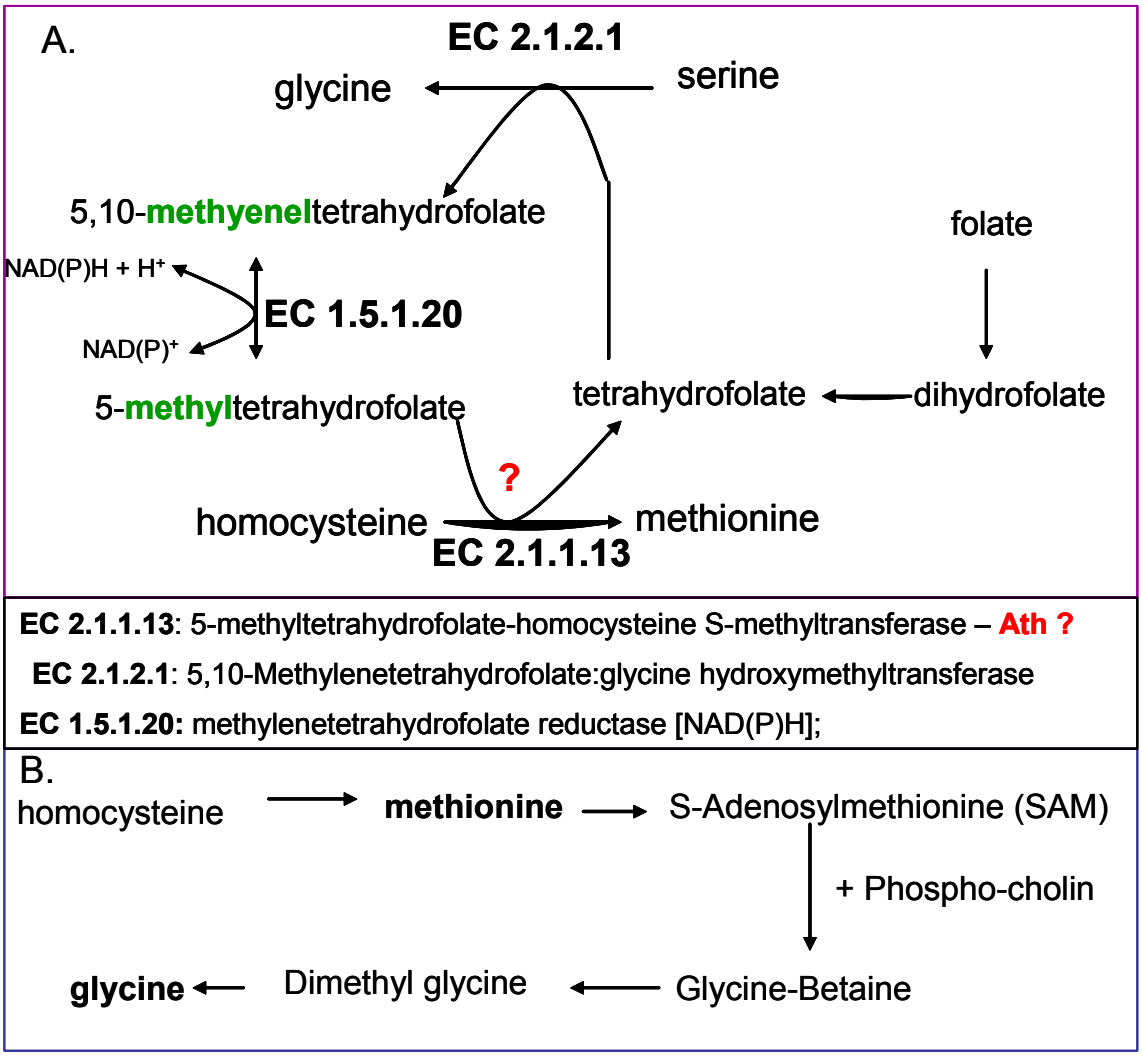


Figure 6-18: Two possible co-regulatory elements between methionine and glycine (A). The possibility of presence of an enzyme in Arabidopsis EC 2.1.1.13 links methionine and glycine through tetrahydrofolate. (B) S-Adenosylmethionine (SAM) produced from methionine regulates Glycine Betaine production, which could also be converted to glycine thus being the common regulator.

The other possible known pathway which connects the two enzymes is one of the major use of methionine is to produce SAM which along with phospho-cholin through a series of steps produces glycine betaine which is one of the important osmoprotectant in plants. Glycine-betain is also known to undergo successive methylation to produce glycine. However the second pathway involves many genes and is known to be differentially regulated specially during slat stress conditions so is less likely to be the case. Hence

most possible explanation of the presence of strong correlation between the glycine and methionine could be the presence of reaction catalyzed by B12 dependent S-methyl-transfarase gene. The above results also indicate that this is also the major methionine producing pathway if present, since it is regulating the overall methionine concentration. In an effort to identify this gene, the methionine and glycine profile will also be compared with the transcriptomic profiles from the same experiment to identify possible targets for genes catalyzing this reaction.

Thus PTM analysis using methionine as the template provides an example of the ability to use the dynamic profiles for re-construction of the metabolic regulatory networks. Similar analysis with malate as a template showed the strongest correlation (p-value 1E-09) with the citramalate (2-methylmalate) profile. Arabidopsis does not have a gene to directly convert malate into citramalate or vice-versa however genes catalyzing such reaction have been previously reported in bacterial systems. These are just two examples of the ability of using systematically perturbed time-series profiles for re-constructing the biochemical networks which can provide important regulatory information or metabolic engineering targets (in this case high corn methionine – an important nutrient for the feed industry) in plants.

7 SUMMARY OF RESULTS - FUTURE WORK

The late 19th and the beginning of the 20th century were marked by a series of discoveries which allowed us to identify the structure of fundamental unit of matter – an atom. The discoveries identified the constituents of atom, the atomic structure and forces governing (regulating) their motion (behavior). This knowledge gained about these systems over a period of two-three decades inspired a host of technologies from nuclear energy and weapons to medical imaging and finally culminating into the computer revolution. Our ability to measure, understand and model the constituents and response of the atoms and molecules was key to the development of these technologies over the entire century.

In almost a comparable way, we are today in the midst of another revolution which is based on the biological systems. The OMICs revolution of the 1990s which allowed successful sequencing of many organisms was followed by a number of high-throughput techniques to measure various cellular levels of the biological system. Once again, our ability to measure, understand and model the regulation of constituents (gene, protein and metabolites being the most important ones) of the biological systems will be the key to utilize these systems for developing technologies which benefit the mankind (which always is the hope of the scientists). The recent environmental changes and likely shortage of energy; difficult to control diseases like cancer, AIDs and possible threat of a Avian flue pandemic have also raised the stakes for increasing our understanding of the biological systems. Already our ability to precisely monitor the cellular level of microbes using techniques in the metabolic engineering tool box like: metabolic flux analysis and

In Silico modeling has allowed rational engineering of microbes to produce monomers (1,3-propanediol), animal feed amino acids (lysine), biopolymers (PHBs) and bulk drugs (Artemisinin) from renewable biological resources. However the use of these techniques for *systematic identification* of targets for engineering plants, or other complex eukaryotic systems, has met with only limited success. In addition these techniques are not high-throughput in nature, especially for complex biological system, limiting the ability to design experiments for integrating it with other cellular fingerprints obtained in the high-throughput manner. Thus a highthroughput alternative to these techniques is needed to monitor the response of the biological systems at the metabolic level.

7.1 METABOLOMIC ANALYSIS

Metabolomic analysis, with its ability to measure hundreds of small metabolites, provided a highthroughput alternative for obtaining the metabolic fingerprint of the biological system. Since metabolites are the ultimate products of changes taking place in gene expression and protein activity; and are the most commonly used source of cell-to-cell communications; our ability to measure them in an accurate manner will be of prime importance in achieving the goal of building new technologies utilizing biological systems.

7.1.1 Method Development

The main aim of my PhD thesis project was to develop a set of experimental tools and data analysis methodologies for using metabolomic analysis to understand regulation in a complex biological system. Being the newest of the OMICs techniques, metabolomic

analysis is still in its early days of development and standardization. During the course of the project an optimized metabolomic profiling protocol was developed using the most common platform (GC-MS). During method development, parameters which need to be optimized for any GC-MS metabolomic study were systematically identified and procedures and criteria for their optimization were prepared.

7.1.2 Data Normalization Methodology

There were additional challenges, inherent to the analysis, which could not be addressed by optimizing experimental condition. To overcome this limitation of GC-MS metabolomic analysis following algorithms were developed:

- A data validation algorithm to ensure constant GC-MS conditions for all the samples analyzed which is a pre-requisite for metabolomic analysis.

- A data correction algorithm which removed biases introduced during sample preparation (derivatization) in metabolomic analysis without jeopardizing the highthroughput nature of the analysis. The problem was a critical limitation for analyzing biological samples using GC-MS for more than 3 decades.

- A data validation and normalization algorithm was finally developed incorporating these algorithms, internal standard normalization, filtering outliers and normalization with biological reference standards. This was the first comprehensive data normalization approach for metabolomic data which takes into account all sources of variability and ensures that the biological conclusions derived from the GC-MS metabolomic profiles is free of experimental biases.

This data normalization strategy is being used not only in plants, but also for metabolomic analysis for (a) yeast (b) mammalian cell cultures (c) brain tissue samples and (d) zebra fish in our group.

7.1.3 Time-series Metabolomic Data Analysis

Most initial metabolomic analysis studies were driven by snap-shot analysis. This was one of the earliest systematically perturbed time-series metabolomic analyses. In order to systematically identify significant metabolites from overall analysis and at individual time-points, methods used in transcriptomic analysis, paired-SAM and MiTimeS respectively, were studied and incorporated for metabolomic analyses. The current project is the first example of time-series metabolomic analysis in which significant metabolites were identified at individual time points using multivariate statistical techniques, rather than using just fold change of different responses as criteria to determine significance in time-series metabolomic analysis.

Using these techniques, a data analysis strategy was worked out for studying the effect of simultaneously identified perturbations and comparison of the same with individual perturbations. The analysis allowed along with identification of effect of combined perturbation, interaction between the regulatory effects of individual perturbations.

In order to reduce time of data analysis, A Mathcad algorithm which incorporated all the data normalization, filtering and data correction steps was developed. In addition, paired-SAM and MiTimeS algorithm were also incorporated in Mathcad reducing significantly

the time required for data analysis which is currently the major bottle neck for metabolomic analysis.

7.1.4 Reconstruction of Metabolic Regulatory Network

In response to perturbation in a biochemical network, the response of metabolites within the network is expected to be correlated if (a) metabolites are regulated by a common regulatory element (b) metabolites are in near equilibrium reactions (c) metabolites belong to the same pathway whose enzymatic activities are regulated. Based on this assumption, carrying out Pavlidis Template Matching in dynamic metabolomic profiles in response to perturbations it was possible to re-construct the metabolic regulatory network. Such reconstruction identified possibility of novel metabolic pathways, which are known to exist in other organisms were thought be absent in plants. This is the first known example of reconstruction of metabolic network and identification novel biochemical pathways / co-regulatory elements in plants using time-series metabolomic analysis data from a systematically perturbed network.

Thus, the import contributions of the research performed in the current PhD project can be summarized as:

- Optimized methodology for GC-MS metabolomics and procedure for developing the same in a new lab.
- Data validation, data correction and data normalization technique which is likely to benefit all researchers using GC-MS metabolomic analysis.

- Incorporated data analysis methods for time-series metabolomic analysis from methods used from transcriptomic analysis.

- Developed a data analysis strategy to compare simultaneously applied perturbations to identify interaction between the stresses.

- A methodology to re-construct biochemical networks and identify novel biochemical pathways or co-regulatory elements using time-series data.

7.2 REGULATION OF PLANT PRIMARY METABOLISM

Plants play an important role in the food chain on earth due to their ability to harness solar energy to reduce carbon dioxide and nitrate ions into organic carbon and nitrogen compounds respectively. Engineering the primary metabolism of complex biological systems like plants is a very difficult but equally important task for a renewable source of energy and chemicals. Incomplete information (as compared to microbes) about their metabolism and lack of steady state make it difficult to use metabolic flux analysis to study regulation in plants. Instead in the present analysis, high-throughput time-series metabolomic analysis was used to study the regulation of *Arabidopsis thaliana* (plant model system) liquid cultures primary metabolism. The system was systematically perturbed with a number of perturbations and their response was monitored using dynamic metabolomic profiling analysis. Summary of the insights about primary metabolism obtained from the same are described below.

7.2.1 Common Observations about Metabolic Regulations

From comparison of all the perturbations applied in this project a number of common perturbations were easily made about the overall regulation of the free small metabolite pools:

- In response to perturbations the maximum increase/decrease in metabolite pools was up to 30 times their original concentrations.
- In response to stress at any given point of time, up to 70% of the metabolite pools showed significant variations. This indicates perturbations can have global effects on the metabolome. In addition the number of metabolites showing increase or decrease can very greatly between time-points in response to the stress.
- In response to each stress, at least 95% of the metabolites showed a significant change at least at 1 time point.
- In response to each stress, less than 10% of the metabolites maintained same significance level at all timepoints, the rest changed significance level at least at one of the time points.
- Metabolomic analysis identified a number of metabolites whose position in the biochemical pathways were not known. This suggests possibility of new pathways not known still in plants.

7.2.2 CO_2 stress

Elevated CO_2 levels in plants significantly alter the metabolism of plants. Some of the main changes taking place are summarized below:

- Increased metabolite levels significantly for majority of metabolites at 1 h, the only stress to do so. CO_2 stress also showed the highest dynamics of significance level of metabolites.
- Inhibited photorespiration and decreased concentration of organic acid pools of the pathway.
- Increased nitrogen assimilation and amino acid pools during first 12 hours, but the concentration of the amino acid pools decreased beyond 12h.

7.2.3 Salt (NaCl) Stress

50 mM Salt (NaCl) stress in plants causes oxidative and osmotic stress which is fatal in *A. thaliana* 4 days from the application of stress. Time-series metabolomic analysis of salt stress showed:

- *A. thaliana* liquid cultures produce their known osmoprotectants like glycine betaine, sucrose and Raffinose-family oligosaccharides (Galactinol).
- *A. thaliana* liquid cultures also produce β-alanine betaine and polyamines which have been known to be osmoprotectants in woody plants.

- *thaliana* liquid cultures use tocopherol (Vitamin E) and ascorbate (Vitamin C) to counter oxidative stress. All the other known sterols also show increase.

- TCA cycle metabolites showed a significant increase, specially fumarate, aconitate and iso-citrate, suggesting increased flux through TCA cycle. This was observed for the first time in plants.

- In contrast Malate also a TCA cycle metabolite, showed a significant decrease suggesting regulation of malate pools by its role in regulation of guard cell movements or balancing Cl^- ions in vacuoles. These needs further investigation.

- Salt stress seems to have mild effect on the nitrogen storage and transport metabolite pools, suggesting not a significant change in nitrogen assimilation.

- Amino acids connected to the biosynthesis of osmoprotectants (β-alanine, homoserine-methionine, N-acetylglutamate, glycine) showed significant increase in their concentration suggests increased flux through their metabolite pools.

- Increase in AAs involved in osmoprotectant is achieved by re-distribution of flux from competing pathways resulting in significant decrease in their pools. This suggests possibility of limitation of organic carbon or nitrogen.

- In response to stress possibly hemi-cellulose content is increased, specifically xyloglucans, in the plant cell wall. This is a novel effect of salt stress response.

The above results, suggest four novel strategies for engineering plants with higher osmotic stress tolerance:

- Engineering plants to increase methionine and adenosine biosynthesis increasing SAM production and/or increasing β-alanine betaine biosynthesis.

- Increasing hemi-cellulose xyloglucan production.

- Increasing carbon utilization for TCA cycle flux to increase ATP production,

- Increasing sterol/ tocopherol production.

These four strategies independently or in combination with existing strategies to increase osmoprotectants proline, glycine betaine and polyamines, can significantly increase osmo-tolerance of plants.

7.2.4 Sugar (Trehalose) Signal

- Exogenously applied Trehalose is accumulated inside the cells and its concentration gradually builds up increasing to 50 times the original concentration in plants at the end of 30h. In presence of the high concentration of trehalose, the trehalase activity is not sufficient to breakdown majority of trehalose produced or transported.

- Metabolomic analysis also indicates possibility of starch accumulation by decreasing starch breakdown in contrast to current theory that accumulation is achieved by increasing gene expression in starch biosynthesis.

- Trehalose did not effect the concentration of glucose (produced by trehalose hydrolysis by *trehalase* enzyme) and fructose during the 1-24 hours.

- Trehalose significantly decreased xylitol concentration and showed significant increase in Galactinol concentration. Trehalose hence may also increase xyloglucan and Raffinose family oligosaccharides production.

- Trehalose has very small effect (possibly secondary) on TCA cycle, Photorespiration, Secondary metabolism, Lipids and Butanoate metabolism.

- In response to trehalose, 12 amino acids showed significant increase (from which 9 were also observed from paired SAM). In contrast only 4 amino acids showed decrease in concentration of which only 1 (aspartate) was negatively significant at most time points. This effect of trehalose on AA production is observed for the first time.

- Trehalose increased 60% of the total metabolite pools at 30h when intracellular trehalose concentration was the highest (50 times the original). This suggests possibility of a different regulatory response at very high trehalose concentration.

The regulation of amino acid biosynthesis by trehalose is an important result which can help in engineering efforts to produce plants which have a better ability to assimilate nitrogen and plants which can produce more amino acids there by improving the nutritional quality.

7.2.5 Hormone (Ethylene) Signal

Ethylene is a plant hormone which regulated several developmental and stress response in plants, however its effect on metabolism have not been studied so far. The first ever time-series metabolomic study of ethylene response:

- Up to 30 fold increase in maltose suggest increased starch degradation observed for the first time.
- Significant increases in first half of the TCA cycle intermediates (citrate, iso-citrate, α-ketoglutarate) and glutamate derived amino acids.
- Increase in precursors for protein (amino acids), lignin (sinapinic acid), and nucleotides (nicotinate) biosynthesis, unsaturated lipids, phosphor-lipid precursors especially during first 12 hours – which show the strongest response of all stresses.
- Regulation of increased cell division and cellular growth by ethylene is achieved by regulating Stigmasterol (and possibly other phytosterols) rather than using the Brassinosteroid regulation machinery.

All of the above are novel findings, being reported for the first time. They demonstrate at metabolic level changes taking place in plants to achieve the increased cell division or growth at the macro-molecular level. In addition to adding to our understanding of ethylene stress response, the analysis also indicated regulatory nodes in the *A. thaliana* liquid culture metabolic network as demonstrated by following examples:

- In response to ethylene, starch breakdown increases, however most of the carbon is stored in the form of maltose which shows the highest increase (30 fold) rather than fructose / glucose or their phosphates, suggesting maltose catabolism as one of the regulating nodes.

- In response to ethylene, TCA cycle intermediates citrate, iso-citrate and α-ketoglutarate show a significant increase in their pools. However the rest of the TCA cycle metabolites do not show similar increase. Instead the increased flux is used for production of glutamate and other amino acids derived from α-ketoglutarate. This suggests α-ketoglutarate to be an important node which regulates the flux towards further TCA cycle reactions or amino acid biosynthesis.

7.2.6 Novel Biochemical Pathways

In addition to specific information about response of primary metabolism pathways in response to stress, the large number of dynamic perturbations also allowed identification of novel biochemical pathways of co-regulation elements based on strong correlation between the metabolites. Using Pattern matching algorithm (PTM) with the time profile of methionine as a template, showed a very strong correlation (p-value 1E-05) between methionine and glycine. Similar strong correlation was also observed between malate and citramalate (p-value 1E-09) with malate as a template. The correlation between these metabolite pairs was much stronger than the correlation between the known co-regulated metabolites. This correlation was observed not only in one of the experiments but across a number of perturbations. Currently there are no known biochemical pathways or common regulatory elements between malate and citramalate in plants, but such a

pathway does exist in some bacteria. The result from metabolic network reconstruction thus suggests possibility of the existence of the pathway also in plants.

7.3 FUTURE WORK

Metabolomic analysis is rapidly emerging as a technique of choice to obtain metabolic fingerprint in many different biological systems. In spite of the ability of the metabolomic analysis to add a wealth of information about the biological systems being studied, there is still room to improve the biological knowledge which can be generated using metabolomic analysis. One of the obvious, but difficult, steps in this direction is to build better metabolite libraries which allow identification of at least 80-90% of metabolites being studied. This would not only add a number of pathways which are still not being studied, but also improve information content of the existing pathways by adding more known metabolites in the pathway. Such an exercise is also likely to identify more metabolites whose biochemical role or position in the biochemical networks is still not known pointing future research direction.

The additional advantage of knowing most of the metabolites in the sample are it would then become possible to prepare a synthetic sample which consists of all the known metabolites. The advantages of having such a synthetic sample are:

- It will significantly improve data normalization strategy and allow cross comparison between data generated in different labs. This would be critical for metabolomic applications in the field of medical diagnostics.

- It will allow us to obtain absolute concentrations of the metabolites being studied in a high-throughput manner.

- Ability to measure absolute concentrations will also allow us to design data analysis strategies which take into account three orders of magnitude difference between metabolite pools.

- Better understanding of all known metabolites with ability to measure their dynamic concentrations will allow us to obtain dynamic flux information in a high-throughput manner.

In the current analysis also, a synthetic sample consisting of 30 known metabolites was prepared which was the starting point for identification of the problems of derivatization which resulted in the data correction algorithm.

In addition to creating a synthetic sample the other steps which can significantly improve the interpretation of the results are as listed below:

- Developing software which combines multiple data analysis steps of metabolomic analysis into single software with ability to integrate the results in biochemical pathway reducing significantly the time it takes to go from GC-MS runs to biological results.

- Creation of a data base which summarizes past results involving individual metabolites. This would make it easier to interpret result as to what changes in the high-throughput analysis mean biologically.

- Extending the metabolome to include important biological energetic state molecules like ATP/ADP, reductive state NADP/NADPH and nucleotides.

- If possible parallel highthroughput analysis to also measure macromolecules like starch, total protein and lipid contents in a high-throughput manner or develop a total metabolite extraction technique which breakdown the macromolecules into their monomers which can then be analyzed separately in a highthroughput manner.

These changes would allow us to get a really global view of all the changes taking place in the biological system. Once we have acquired this ability or some of the important parts of it we can also improve biological sampling looking at individual cell types or even compartments within the cell. Already advances in the field of laser micro-dissection have allowed researchers to look at transcriptomic data for individual cell types of roots in plants (Galbraith and Birnbaum, 2006) and the same also been shown to work with metabolomic analysis (Schad et al., 2005). These results can then be integrated with information available from transcriptomic, proteomic and ionomics fingerprints. This would allow us to develop highly complex compartmentalized models for changes taking place in the biological system taking us one step closer to engineer them for the benefit of all life forms on this beautiful plant we call home.

APPENDICES

Appendix I: Data Correction Algorithm Additional Tables

Table A1-1: Composition of the standard metabolite mix 1. Actual weights of each metabolite in the 600 μL solution are provided in column 2 (first published in Kanani and Klapa, 2007).

	Metabolite	**Concentration (μg/mL)**	**Amount in 600 μL Solution (μg)**
1	Alanine	2.5	1.5
2	Asparagine	83.3	50.0
3	Aspartic acid	11.1	6.7
4	Citric Acid	111.2	66.7
5	Fructose	83.3	50.0
6	Fumarate	11.7	7.0
7	Galactose	16.7	10.0
8	Glucose	116.7	70.0
9	Glutamic acid	133.3	80.0
10	Glutamine	50.0	30.0
11	Glycine	8.3	5.0
12	iso-Leucine	0.2	0.1
13	Lactose	1.7	1.0
14	Leucine	0.3	0.2
15	Lysine	3.3	2.0
16	Malic acid	83.3	50.0
17	Maltose	1.7	1.0
18	Mannitol	0.8	0.5
19	Methionine	0.8	0.5
20	nor-Leucine	13.3	8.0
21	Phenylalanine	0.8	0.5
22	Proline	16.7	10.0
23	**Ribitol (Internal Standard)**	**33.3**	**20.0**
24	Serine	8.3	5.0
25	Succinic acid	12.4	7.4
26	Sucrose	83.3	50.0
27	Threonine	5.0	3.0
28	Valine	8.3	5.0

Table A1-2: Measured relative (with respect to the internal standard ribitol) peak areas of the derivative forms of the category-3 metabolites listed in Tables 3-1 &3-2 in the metabolomic profiles used for the estimation of the w_1^M weight values listed in Table 3-2 (first published in Kanani and Klapa, 2007).

		Plant Sample 1-RPA						
	Derivatization Time →	***6 hr***	***11 hr***	***12 hr***	***17 hr***	***18 hr***	***23 hr***	***C Values***
Glutamate	Glutamate 3 TMS	0.251072	0.237657	0.217070	0.196430	0.191739	0.161459	
	Pyroglutamate 2 TMS	0.230808	0.258363	0.251117	0.280638	0.289289	0.326998	
	Cumulative	**0.4826**	**0.4962**	**0.4682**	**0.4765**	**0.4802**	**0.4868**	**0.4819**
Asparagine	Asparagine 3 TMS	0.027924	0.024085	0.022217	0.018717	0.017888	0.012792	
	Asparagine 4 TMS	7.543E-03	7.517E-03	7.928E-03	8.411E-03	8.381E-03	8.488E-03	
	Asparagine 5 TMS (putative)	1.380E-03	3.208E-03	3.318E-03	4.598E-03	5.188E-03	7.011E-03	
	Cumulative	**0.0368**	**0.0369**	**0.0365**	**0.0369**	**0.0372**	**0.0366**	**0.0368**
Glutamine	Glutamine 3 TMS	0.135769	0.079174	0.070707	0.035678	0.029325	0.005884	
	Glutamine 4 TMS	2.075E-03	5.912E-04	5.638E-04	2.03E-04	n.d.	n.d.	
	Pyroglutamine 3 TMS (putative)	3.248E-03	8.745E-03	0.010159	0.012590	0.013769	0.015121	
	Cumulative	**0.1412**	**0.1376**	**0.1444**	**0.1392**	**0.1435**	**0.1400**	**0.1411**
Serine	Serine 2 TMS	0.007493	0.006191	0.005855	0.004677	0.005350	0.004558	
	Serine 3 TMS	0.021300	0.015568	0.017809	0.018195	0.012529	0.016312	
	Serine 4 TMS	2.88E-04	6.270E-04	6.76E-04	1.025E-03	1.153E-03	1.587E-03	
	Cumulative	**0.0309**	**0.0280**	**0.0280**	**0.0274**	**0.0287**	**0.0309**	**0.0291**
Threonine	Threonine 2 TMS	6.056E-03	4.895E-03	4.965E-03	4.110E-03	4.787E-03	4.290E-03	
	Threonine 3 TMS	0.018956	0.015755	0.016772	0.018428	0.013826	0.017302	
	Threonine 4 TMS	n. d.	7.85E-05	8.18E-05	1.42E-04	1.37E-04	1.98E-04	
	Cumulative	**0.0261**	**0.0238**	**0.0245**	**0.0242**	**0.0248**	**0.0263**	**0.0250**
Homoserine	Homoserine 2 TMS	4.847E-04	3.745E-04	4.470E-04	3.108E-04	3.469E-04	3.024E-04	
	Homoserine 3 TMS	3.408E-03	2.327E-03	2.691E-03	2.701E-03	1.708E-03	2.385E-03	
	Homoserine 4 TMS	n. d.	2.43E-04	2.34E-04	3.83E-04	4.17E-04	6.01E-04	
	Cumulative	**3.943E-03**	**3.624E-03**	**4.156E-03**	**3.670E-03**	**3.766E-03**	**4.124E-03**	**3.943E-03**

		Plant Sample 2 –RPA					
	Derivatization Time →	***8 hr***	***9 hr***	***14 hr***	***15 hr***	***20 hr***	***21 hr***
Glutamate	Glutamate 3 TMS	0.289357	0.293374	0.261782	0.255368	0.229210	0.218927
	Pyroglutamate 2 TMS	0.279567	0.284105	0.313765	0.315481	0.331679	0.334087
	Cumulative	**0.5696**	**0.5782**	**0.5754**	**0.5706**	**0.5601**	**0.5521**
Asparagine	Asparagine 3 TMS	0.032668	0.033411	0.027955	0.025636	0.021702	0.020932
	Asparagine 4 TMS	6.415E-03	7.382E-03	8.145E-03	8.133E-03	8.741E-03	7.858E-03
	Asparagine 5 TMS (putative)	1.606E-03	1.840E-03	3.181E-03	3.318E-03	4.338E-03	4.465E-03
	Cumulative	**0.0385**	**0.0412**	**0.0409**	**0.0394**	**0.0393**	**0.0373**
Glutamine	Glutamine 3 TMS	0.192906	0.203068	0.143653	0.132696	0.093441	0.086970
	Glutamine 4 TMS	5.550E-03	6.313E-03	3.332E-03	3.037E-03	1.594E-03	1.324E-03
	Pyroglutamine-3-TMS (putative)	2.487E-03	4.748E-03	8.254E-03	8.587E-03	0.012801	0.012199
	Cumulative	**0.2082**	**0.2432**	**0.2044**	**0.1971**	**0.1939**	**0.1814**
Serine	Serine 2 TMS	0.016386	0.013592	0.014349	0.012940	0.013015	0.013913
	Serine 3 TMS	0.014420	0.023790	0.016139	0.016093	0.016631	0.014616
	Serine 4 TMS	4.972E-04	5.877E-04	8.927E-04	9.636E-04	1.373E-03	1.529E-03
	Cumulative	**0.0569**	**0.0521**	**0.0545**	**0.0508**	**0.0544**	**0.0577**
Threonine	Threonine 2 TMS	9.091E-03	7.413E-03	8.188E-03	7.588E-03	7.916E-03	7.897E-03
	Threonine 3 TMS	0.011811	0.017305	0.012697	0.012903	0.013443	0.011824
	Threonine 4 TMS	n.d.	8.00E-05	9.95E-05	1.05E-04	1.26E-04	1.31E-04
	Cumulative	**0.0338**	**0.0327**	**0.0344**	**0.0327**	**0.0347**	**0.0342**
Homoserine	Homoserine 2 TMS	6.956E-04	6.306E-04	5.353E-04	4.300E-04	4.797E-04	5.086E-04
	Homoserine 3 TMS	1.073E-03	2.178E-03	1.210E-03	1.084E-03	1.207E-03	1.004E-03
	Homoserine 4 TMS	0.00E+00	0.00E+00	2.41E-04	3.02E-04	3.70E-04	3.38E-04
	Cumulative	**4.776E-03**	**4.608E-03**	**4.409E-03**	**3.855E-03**	**4.390E-03**	**4.445E-03**

		Standard Metabolite Mix1 – RPA								
Derivatization Time -->		***6***	***7***	***9***	***10***	***14***	***15***	***17***	***18***	***C Values***
Alanine	Alanine NO	0.033705	0.031266	0.031235	0.033505	0.024798	0.027107	0.023076	0.025426	
	Alanine NNO	5.026E-03	5.728E-03	9.281E-03	9.862E-03	0.013905	0.014379	0.017820	0.018751	
	Cumulative	**0.03844**	**0.03648**	**0.03920**	**0.04198**	**0.03618**	**0.03891**	**0.03745**	**0.04057**	**0.038732**
Glycine	Glycine N O	0.02105	0.02007	0.02182	0.02070	0.02319	0.02120	0.02781	0.02680	
	Glycine NNO	0.72804	0.72662	0.70297	0.71781	0.67659	0.69549	0.62924	0.65167	
	Cumulative	**0.76059**	**0.75025**	**0.74842**	**0.74941**	**0.74091**	**0.73687**	**0.74770**	**0.75557**	**0.749086**
isoLeucine	isoLeucine O	1.689E-03	1.475E-03	1.526E-03	1.424E-03	1.637E-03	1.172E-03	1.306E-03	1.371E-03	
	isoLeucine NO	0.01692	0.01659	0.01579	0.01754	0.01500	0.01666	0.01489	0.01592	
	Cumulative	**0.01988**	**0.01903**	**0.01842**	**0.01977**	**0.01797**	**0.01831**	**0.01702**	**0.01814**	**0.018613**
Lysine	Lysine NNNO	0.06040	0.05917	0.05770	0.05861	0.05454	0.05636	0.05305	0.05375	
	Lysine NNNNO	n.d.	n.d.	1.149E-03	1.640E-03	1.968E-03	1.982E-03	2.859E-03	3.225E-03	
	Cumulative	**0.06040**	**0.05917**	**0.06043**	**0.06238**	**0.05899**	**0.06085**	**0.05939**	**0.06087**	**0.060400**
Methionine	Methionine N O	0.01786	0.01773	0.01697	0.01675	0.01650	0.01696	0.01457	0.01389	
	Methionine NNO	8.944E-03	7.429E-03	0.01327	8.703E-03	9.172E-03	8.257E-03	9.106E-03	7.777E-03	
	Cumulative	**0.02866**	**0.02791**	**0.02896**	**0.02700**	**0.02681**	**0.02713**	**0.02405**	**0.02260**	**0.026801**
Valine	Valine O	0.02708	0.02674	0.02927	0.03026	0.02946	0.02649	0.02946	0.02463	
	Valine NO	0.09162	0.09102	0.08668	0.08856	0.08164	0.08834	0.07891	0.08520	
	Cumulative	**0.1217**	**0.1207**	**0.1211**	**0.1243**	**0.1171**	**0.1180**	**0.1148**	**0.1123**	**0.11870**

		Standard Metabolite Mix2- RPA									
Derivatization Time -->		***6***	***8***	***10***	***12***	***14***	***16***	***18***	***20***	***22***	***C Value***
B Alanine	B Alanine NNO	15.2052	15.9998	14.5027	16.9377	15.2167	15.6929	12.8514	15.4956	14.7220	
	B Alanine O	0.4060	0.3413	0.4135	0.1890	0.3697	0.3473	0.5662	0.4221	0.4116	
	Cumulative	**15.8151**	**15.8786**	**15.3175**	**15.2793**	**15.5019**	**15.6854**	**15.3475**	**16.1912**	**15.4768**	**15.6112**
Dopamine	Dopamine NOO	0.3132	0.2421	0.2989	0.1722	0.2684	0.2752	0.4074	0.2924	0.3045	
	Dopamine NNOO	3.6453	3.9818	3.7981	4.5306	3.8138	3.8572	3.1544	3.7618	3.6517	
	Cumulative	**3.9576**	**3.9066**	**4.0093**	**4.0151**	**3.8938**	**3.9537**	**3.9924**	**3.9559**	**3.9261**	**3.9585**
Phenylalanine	Phenylalanine O	1.2511	1.2399	1.2708	1.1471	1.2631	1.2961	1.4528	1.3148	1.3454	
	Phenylalanine NO	0.7122	0.7373	0.6171	0.9357	0.6722	0.6665	0.1132	0.5630	0.5240	
	Cumulative	**1.9631**	**1.9606**	**1.9432**	**1.9351**	**1.9596**	**1.9996**	**1.9384**	**1.9744**	**1.9955**	**1.9633**
Tyrosine	Tyrosine O	0.3300	0.2500	0.3200	0.1700	0.2900	0.2900	0.4400	0.3100	0.3200	
	Tyrosine NO	1.6205	1.6723	1.7182	1.8711	1.7497	1.6662	1.5229	1.7727	1.7961	
	Tyrosine NOO	0.0310	0.0712	0.0716	0.0424	0.0775	0.1019	0.0281	0.0107	0.0891	
	Cumulative	**1.9221**	**1.8866**	**2.0127**	**1.9712**	**2.0084**	**1.9363**	**1.9598**	**2.0362**	**2.0906**	**1.9815**

Derivatization Time -->		***8***	***10***	***12***	***14***	***16***	***20***	***22***			***C Value***
Cysteine	Cysteine NSO	0.1635	0.1727	0.1277	0.1772	0.1647	0.1716	0.1729			
	Cysteine NNO	2.869	2.2392	3.5237	2.4538	2.803	2.0037	2.0115			
	Cumulative	**3.1331**	**3.0166**	**2.9217**	**3.1530**	**3.1239**	**2.9155**	**2.9349**			**3.0325**

Derivatization Time -->		***8***	***10***	***13***	***16***	***19***	***22***	***25***			***C Value***
Ornithine	Ornithine $N_2N_2N_5O$	1.256	1.355	1.383	1.374	1.392	1.400	1.466			
	Ornithine $N_5N_5N_2O$	0.552	0.711	0.668	0.566	0.593	0.542	0.566			
	Cumulative	**1.647**	**1.832**	**1.842**	**1.783**	**1.816**	**1.800**	**1.884**			**1.8080**

		Pure Metabolite Standards – RPA						
Derivatization Time -->		***6***	***6.5***	***13***	***13.5***	***27***	***27.5***	***C Value***
Aspartate	Aspartate OO	2.108E-03	1.883E-03	2.418E-03	2.492E-03	0.018432	0.016652	
	Aspartate NOO	0.30668	0.29906	0.29415	0.30321	0.05375	0.02309	
	Cumulative	**0.0760**	**0.0757**	**0.0750**	**0.0767**	**0.0818**	**0.0694**	**0.0760**
Derivatization Time -->		***7***	***11***	***15***	***19***	***23***		***C Value***
Allantoin	Allantoin NNN	7.532E-03	4.482E-03	7.887E-03	3.986E-03	2.368E-03		
	Allantoin NNNN	0.462107	0.529965	0.600697	0.709147	0.680483		
	Allantoin NNNNN	6.212E-03	5.406E-03	4.733E-03	0.012829	0.024902		
	Cumulative	**0.44864**	**0.40573**	**0.52793**	**0.50390**	**0.47335**		**0.47585**

Table A1-3: Retention times (RT) for the derivative forms of the 26 category-3 metabolites listed in Table 3-1; they correspond to the GC-MS operating conditions at which all data discussed in the paper were acquired (first published in Kanani and Klapa, 2007).

Amino Acid	RT Derivative 1 Min	RT Derivative 2 Min	RT Derivative 3 Min
Alanine	7.96	14.83	
Arginine	25.19	25.39	29.30
Asparagine	24.77	20.30	28.13
Aspartate	19.48	20.30	
Cysteine	16.56	20.89	21.29
Glutamate	22.64	23.00	
Glutamine	27.01	23.45	21.78
Glycine	9.06	13.54	
Histidine	32.00	31.75	31.02
Iso-Leucine	12.04	13.38	19.35
Leucine	11.21	12.74	18.27
Lysine	23.62	27.25	30.93
Methionine	21.11	26.03	
Phenylalanine	24.00	23.73	
Proline	14.45		
Serine	14.20	15.43	20.04
Threonine	14.81	15.83	22.01
Tryptophan	39.01	37.75	36.08
Tyrosine	26.15	30.44	37.59
Valine	9.30	11.10	11.729
Allantoin	35.91	29.34	27.59
β-Alanine	8.11	11.42	17.06
GABA	14.727	19.694	
Dopamine	26.97	30.95	
Homoserine	16.90	17.75	22.47
Ornithine	25.19	25.39	29.30

Appendix- II Significance Level Of Unknown Metabolites

Table A2-1: Significance level of unknown metabolite pools at individual time points and paired-SAM analysis in response to elevated CO_2 stress. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAI
acid	Hydroxy-pyruvic acid/Ara Col leaf 2 / 4-keto Glucose diMeox	P1906		-1	-1				-1		
acid	Glyceric Acid like unknown	P2793		-1	-1				-1		
acid	Ara Col Leaf 012 or 2-keto gulonic acid Methoxime	P1361		-1			-1		-1	-1	
acid	2 Pentenoic Acid 1 TMS	P0769	-1	-1	-1	-1	-1	-1	-1	-1	-1
acid	unknown similar to Malic Acid 3 TMS / à Keto isovaleric acid	P1677		-1	-1		-1		-1	-1	-1
acid	Ascorbic Acid derivative in Tomato leaf ?	P2322		-1		-1		-1	-1	-1	-1
acid	ARTH unknown 1534-21-1 / 3-ketovalaric acid	P1570	-1			-1		-1	-1	-1	-1
acid	unknown similar to Citramalate ? (Citramaleate ?)	P1655	1	-1	-1				-1	-1	
acid	2,3 dihydroxybutanedioic /pentanoic acid / galactonic acid	P2077	1	-1	-1		1		-1		
acid	2-keto-gluconic acid / Potato 015 unknown sugar	P2427	1				1		-1		
acid	Propenoic acid, 2,3,3-tris[(trimethylsilyl)oxy]-, trimethylsilyl	P2893	1	1	1		1				
acid	acid ester	P1391	1	1		1	1	1		1	
acid	small mol. Wt. acid (C3 acid ?)	P0479	1					-1			
acid	small mol. Wt. acid (C3 acid ?)	P0421	1					-1	-1		
acid	Propane, 2-methyl-1,2-bis(trimethylsiloxy)- /3-Methyl-1,3-bis(	P0802						1	-1		
acid	Ketone / Aldehyde (Tartaric acid - Glyceraldehyde)	P0871	1	1	1		1				
alcohol	like Galactinol	P2835	-1	-1	-1	-1	-1			-1	-1
Alcohol	Coniferyl alcol (2TMS)	P4691	1		-1				-1	-1	
FA	unknown FA	P3188	1		1	1	1		-1		
FA	Heptadecanoic acid 1 TMS	P3263	1	1	1	1		1			
FA	9,12-Octadecadienoic acid, tert-butyldimethylsilyl ester, (Z,Z)	P3635		1				-1	-1		
FA	Docosanoic acid, trimethylsilyl ester	P4040	1	1				-1	-1		
FA?	methylanthraquinone, O,O'-bis(trimethylsilyl	P3295	1	-1	-1				-1	-1	
oxime	Trehose / Erythrose Methoxiamine OR Glycolic acid	P1091			-1						
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAI**
SecM - S	Scopolin, tetra(trimethylsilyl)- ?	P1750		-1	-1	-1			-1	-1	
secM	biphenol 2 TMS	P2234	-1	-1	-1	-1	-1		-1	-1	-1
SecM	unknown trimethylsillyl estrone ? Aromatic ring	P1790	-1	-1	-1			-1	-1	-1	-1
SecM	unknown secondary metabolite	P2113	-1		-1	-1		-1	-1	-1	-1
SecM	unknown	P2224		-1	-1			-1	-1	-1	
SecM	1H-Indole-2,3-dione	P3559		-1	-1				-1	-1	
SecM	2-Piperidinecarboxamide, 1-butyl-N-(2,6-dimethylphenyl)-	P4718		1	-1		1		-1	-1	
SecM	unknown	P3636	1					-1	-1	-1	
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P3620	1	-1	-1				-1		
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P5165	1	-1	-1				-1		
Similar to	Proline like compound - methyl proline ?	P1629		-1	-1				-1		
SecM	unknown secondary metabolite	P1801		-1	-1			-1			
SecM	Secondary Metabolite	P3252		-1	-1			-1			
SecM	Secondary Metabolite	P3656		-1	-1	-1					
secM	unknown secondary metabolite large mol	P4833	1		-1	1	1	1	-1		
SecM	no 73 peak - non hydroxy/carboxy/amine TMS compound	P2350	1		-1	1	1	1	-1		
SecM	similar to (1-methyl-1-phenylethoxy)-	P5082	1	1			1	-1	-1		
SecM	unknown	P5073			-1		1		-1		
SecM	Benzamide, 3,4-dimethoxy- derivative	P3154		1		1	1		-1		
SecM	Secondary amine	P3521				1	1	1	-1		
SecM	secondary aromatic metablite	P5050	1	1			1		-1		
SecM	1'H-Cholest-2-eno[3,2-b]indole, 5'-methoxy-, (5?)-	P4646	1				1		-1		
SecM	4-hydroxybenzoic acid	P2391		1	-1	1					
SecM	5-Aminocarboxy-4,6-dihydroxypyrimidine	P2435			1	1			-1		
SecM	Benzoic acid derivative	P1923							-1		
SecM	non-TMS compound	P5001		1	1		1	1	1	1	
SecM	secondary metabolite (imidazole derivative ?)	P3211	1			1	1	1		1	
SecM	Benzoic acid, 5-methoxy-2-oxy-TMS	P5105	1	-1	-1			-1	-1		

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAI
Phospho	Sugar Phosphate (MPIMP-ID:221004-21-1)	P3438	1				1		-1		
Phospho	Sugar Phosphate	P3681	1		1	-1					
Phospho	Triose (2,3 Propanediol) phosphate	P3372	1				1	1	-1	-1	
Phospho	Mannose Phosphate	P3502	1				1		-1	-1	
Phospho	Ethanolamine - phosphate	P2643	1				1		-1		
Phospho	Inositol phosphate	P4125	1		1		1		-1	-1	
Phospho	myo-Inositol-1 or 2- Phosphate	P3705	1		1	1			1	1	
Phospho	could be other inositol phosphate	P3585	1	-1	-1		1		-1	-1	
Phospho	Sugar Phosphate	P3611	1		-1		1				
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAI**
Sugar	027 / POTATO TUBER 019 / Glucose impurity	P2957	1	1	1	1	1			1	
Sugar	similar to fructose / ketogluconic acid non-meox	P2412	1		1	1			1	1	
Sugar	d-Turanose / di-tri sachharide	P3288	1	1	1		1	-1	-1		
Sugar Am	Adenosine, N-(trimethylsilyl)-2',3'-bis-O-(trimethylsilyl)-, 5'-[bi	P4882	1	1	1		1		-1		
Sugar	MPIMP-ID:196004-45-1 / Melizitose / Maltose / iso Maltose	P2922	1	1	1			-1		1	
Sugar	P3867 Meox 1 ?	P3867	1	1	1		1		-1		
Sugar	Sugar like Trehalose	P4631	1		-1	1	1		-1		
Sugar	Trehalose ?	P4784	1	1			1	-1	-1		
Sugar	unknwn Sugar / Ploem Cmax 013	P2054	1			-1	-1	1		1	
Sugar	D-Gal-(1,6)-D-Glc (8TMS)]	P4510	1	1			1		-1		
Sugar	unknown similar to Fructose 5 TMS	P2457		-1	1				1	1	
Sugar		P4295	1	1		1	-1				
Sugar	Maltose Methox ?	P3991		1	1	1		-1			
Sugar		P4814		1		1	1				
Sugar	Isomaltose/Melibiose/Cellibiose (900 Reverse match)	P4140		1			1			1	
Sugar	(8TMS); alpha-D-Gal-(1,6)-D-Glc (8TMS)]/Mannose	P3400	1	-1	-1		1	-1	-1		
Sugar	(11TMS)) / alpha-D-Glc-(1,3)-beta-D-Fru-(2,1)-alpha-D-Glc	P2980	1				1	-1	-1	-1	
Sugar	Nigrose	P3816		-1	-1		1	1	-1		
Sugar	di / tri saccharide / Trehalose/Maltose/Lactose	P3013	1	-1	-1		1		-1		
Sugar (fu	1/Ribofuranose/Xylofuranose/Arabinofuranose/Phloem C	P2083	1			1			-1	-1	
Sugar	Ara Col Leaf 094 / MPIMP-ID:250001-20-1 /Lactose	P3677	1	-1	-1		1				
Sugar	Ath Sugar - MPIMP-ID:213-21-1	P3726	1			-1	1				
Sugar	Galactinol (9TMS); alpha-D-Gal-(1,3)-myo-Inositol (9TMS)]	P4312	1				1		-1		
Sugar	mono / di / saccharide	P3046				1		1			
Sugar	sugar/Pentonic acid (5TMS)	P2445		-1	-1	-1	1	-1	-1	-1	
Sugar Alc	Maltitol	P4024			-1		1	-1	-1	-1	
Sugar	Galacto-lactone or sugar acid	P2203	1	-1	-1				-1	-1	
Sugar	P3867 Meox 2	P3890	1	-1	-1				-1	-1	
Sugar	Arabidopsis Col Leaf 046 / Altrose / Galactofuranose /	P2509	1					-1	-1	-1	
Sugar	Gulose / Potato Tuber 032	P4167	1		-1				-1	-1	
Sugar	2-keto-L gulonic acid methoxime / Xylofuranose/Ribofuranos	P2000	1	-1	-1				-1		
Sugar	Glucose derivative unknown ?	P1324	1	-1	-1						
Sugar	Matches Sucrose - quite large peak / Trehalose /Melizitos	P4692	1		-1				-1		
Sugar		P4756		1				-1			
Sugar	Gluconic Acid 4 TMS lactone ?	P2823	-1	-1	-1	-1			-1	-1	-1
Sugar	Raffinose	P4455	-1		-1	-1		-1	-1	-1	-1
Sugar	Glucose derivative ?	P1540	-1	-1	-1				-1	-1	-1
Sugar	di / tri saccharide	P3033		-1	-1				-1	-1	
Sugar		P3387		-1	-1				-1		
Sugar	Cellobiose ?	P3106		-1	-1				-1		
Sugar	alpha-D-Glc	P4053		-1	-1				-1		

Table A2-2: Significance level of unknown metabolite pools at individual time points and paired-SAM analysis in response to salt stress. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
acid	Glyceric Acid like unknown	P2793	-1	1		1	1		1	1	
acid	Ara Col Leaf 012 or 2-keto gulonic acid Methoxime	P1361	1	1	1	1	1	1	1	1	1
acid	unknown similar to Malic Acid 3 TMS / à Keto isovaleric acid	P1677	-1		-1		-1		-1	-1	-1
acid	Ascorbic Acid derivative in Tomato leaf ?	P2322	-1	-1		-1		-1	-1	-1	-1
acid	unknown similar to Citramalate ? (Citramaleate ?)	P1655	-1		-1		-1		-1		
acid	2,3 dihydroxybutanedioic /pentanoic acid / galactonic acid	P2077				1			-1		
acid	2-keto-gluconic acid / Potato 015 unknown sugar	P2427	-1		-1				-1	-1	-1
acid	Propenoic acid, 2,3,3-tris[(trimethylsilyl)oxy]-, trimethylsilyl	P2893	1	1	1		1		-1		
acid	acid ester	P1391	1			1		1			
acid	Propane, 2-methyl-1,2-bis(trimethylsiloxy)- /3-Methyl-1,3-bis(	P0802		-1	-1	-1			-1	-1	-1
alcohol	like Galactinol	P2835	-1	-1	-1				1	-1	
Alcohol	Coniferyl alcol (2TMS)	P4691				1		-1			
FA	unknown FA	P3188	-1		-1	1			-1		
FA	Heptadecanoic acid 1 TMS	P3263							-1	1	
FA	(Z,Z)-9,12-Octadecadienoic acid TMS ester	P3635	1	1	1	1			-1		
FA	Docosanoic acid, trimethylsilyl ester	P4040	-1	1		1			-1		
FA?	TMS[1Docosanol(1TMS)]1,6Dihydroxy3methylanthraquinon	P3295	-1			1				1	
oxime	Trehose / Erythrose Methoxiamine OR Glycolic acid	P1091	-1		-1		-1	-1	-1	-1	-1
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAM**
SecM - Sc	Scopolin, tetra(trimethylsilyl)- ?	P1750	-1					-1	-1	-1	-1
SecM	unknown trimethylsillyl estrone ? Aromatic ring	P1790	1	1	1	1			-1		
SecM	unknown	P2224	-1		-1	1			-1		
SecM	1H-Indole-2,3-dione	P3559	-1	-1	-1	1					
SecM	2-Piperidinecarboxamide, 1-butyl-N-(2,6-dimethylphenyl)-	P4718	-1	1		1					
SecM	unknown	P3636	-1		-1			-1	-1	-1	-1
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P3620			-1	1			-1	1	
Similar to	Proline like compound - methyl proline ?	P1629	-1	-1	-1	1					
SecM	Secondary Metabolite	P3252	1	1	1	1			-1	1	
secM	unknown secondary metabolite large mol	P4833	-1	-1	-1		-1	-1	-1	-1	-1
SecM	no 73 peak - non hydroxy/carboxy/amine TMS compound	P2350	-1			1		1		1	
SecM	similar to (1-methyl-1-phenylethoxy)-	P5082				1		-1			
SecM	unknown	P5073	-1	-1	-1	1	-1		-1	-1	-1
SecM	Benzamide, 3,4-dimethoxy- derivative	P3154	-1			1			-1	1	
SecM	Secondary amine	P3521	-1	-1	-1	1			-1		
SecM	secondary aromatic metablite	P5050		1				-1	-1		
SecM	1'H-Cholest-2-eno[3,2-b]indole, 5'-methoxy-, (5?)-	P4646	-1				-1	-1	-1	-1	-1
SecM	4-hydroxybenzoic acid	P2391	-1	1		1			-1		
SecM	5-Aminocarboxy-4,6-dihydroxypyrimidine	P2435				1		1	-1	1	
SecM	Benzoic acid derivative	P1923	1	1	1			-1	-1	1	
SecM	non-TMS compound	P5001		1	1	1	1	1	1	1	1
SecM	secondary metabolite (imidazole derivative ?)	P3211							-1		
SecM	Benzoic acid, 5-methoxy-2-oxy-TMS	P5105	-1		-1				-1		

Continued..

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Phospho	Sugar Phosphate (MPIMP-ID:221004-21-1)	P3438	-1	-1		1			-1		
Phospho	Sugar Phosphate	P3681	-1					1		1	
Phospho	Triose (2,3 Propanediol) phosphate	P3372	-1		-1	1		1			
Phospho	Mannose Phosphate	P3502	-1	-1	-1				-1	-1	-1
Phospho	Ethanolamine - phosphate	P2643	-1	-1	-1	-1			-1	-1	-1
Phospho	Inositol phosphate	P4125	-1			1	1	1			
Phospho	myo-Inositol-1 or 2- Phosphate	P3705	-1			1				1	
Phospho	could be other inositol phosphate	P3585		-1		1	1	1			
Phospho	Sugar Phosphate	P3611	-1				1			1	
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAM**
Sugar	Digalactosylglycerol (9TMS); 882; Melibiose (8TMS);	P4510		1	1	1	1	1	1	1	1
Sugar Alc	Maltitol	P4024	1		1	1	1	1	1	1	1
Sugar		P4756		1	1	1	1	1	1	1	1
Sugar	Mannonic Acid/MPIMP-ID:203003-31-1TOMATOLEAF027	P2957							1		
Sugar	similar to fructose / ketogluconic acid non-meox	P2412	-1					1	1		
Sugar	MPIMP-ID:196004-45-1 / Melizitose / Maltose / iso Maltose	P2922	1				-1	1	1	-1	
Sugar	P3867 Meox 1 ?	P3867					1		1	-1	
Sugar	Sugar like Trehalose	P4631	-1	-1	-1	1	-1	-1	-1		
Sugar	Trehalose ?	P4784				1		-1	-1		
Sugar	unknown similar to Fructose 5 TMS	P2457	-1						1		
Sugar		P4295	-1	-1	-1	1			-1		
Sugar		P4814	-1	-1	-1	1			-1		
Sugar	Isomaltose/Melibiose/Cellibiose (900 Reverse match)	P4140	-1	1					1	1	
Sugar	(8TMS); alpha-D-Gal-(1,6)-D-Glc (8TMS)]/Mannose	P3400				1				1	
Sugar	(11TMS)) / alpha-D-Glc-(1,3)-beta-D-Fru-(2,1)-alpha-D-Glc	P2980		-1				-1	-1	-1	-1
Sugar	Nigrose	P3816				1		1	-1		
Sugar	di / tri saccharide / Trehalose/Maltose/Lactose	P3013		-1		1				1	
Sugar (fu	1/Ribofuranose/Xylofuranose/Arabinofuranose/Phloem C	P2083	-1	1		1			-1		
Sugar	Ara Col Leaf 094 / MPIMP-ID:250001-20-1 /Lactose	P3677	-1			1			-1		
Sugar	Galactinol (9TMS); alpha-D-Gal-(1,3)-myo-Inositol (9TMS)]	P4312	-1		-1				-1		
Sugar	mono / di / saccharide	P3046	-1		-1	1			1		
Sugar	sugar/Pentonic acid (5TMS)	P2445	-1	-1	-1	-1		-1		-1	-1
Sugar	Galacto-lactone or sugar acid	P2203	-1		-1	1			-1	1	
Sugar	P3867 Meox 2	P3890	-1			1				1	
Sugar	Arabidopsis Col Leaf 046 / Altrose / Galactofuranose /	P2509	-1		-1		-1	-1	-1	-1	-1
Sugar	Gulose / Potato Tuber 032	P4167	1	1	1	1				1	
Sugar	2-keto-L gulonic acid methoxime / Xylofuranose/Ribofuranos	P2000	-1		-1		-1	1	1	1	
Sugar	Matches Sucrose - quite large peak / Trehalose /Melizitos	P4692				1	-1				
Sugar	Gluconic Acid 4 TMS lactone ?	P2823	-1	-1	-1		-1	-1	-1	-1	-1
Sugar	Raffinose	P4455	-1	-1	-1			-1	-1	-1	-1
Sugar	Glucose derivative ?	P1540				1		-1	-1		
Sugar	di / tri saccharide	P3033	-1	-1	-1	1			-1		
Sugar		P3387	-1		-1	1			-1		
Sugar	Cellobiose ?	P3106		-1		1				1	
Sugar	Melezitose (11TMS)	P4053	-1			1		1		1	

Table A2-3: Significance level of unknown metabolite pools at individual time points and paired-SAM analysis in response to Trehalose stress. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAI
acid	Hydroxy-pyruvic acid/Ara Col leaf 2 / 4-keto Glucose diMeox	P1906	-1				-1				
acid	Glyceric Acid like unknown	P2793	-1							1	
acid	Ara Col Leaf 012 or 2-keto gulonic acid Methoxime	P1361	1		1			1	1	1	1
acid	Ascorbic Acid derivative in Tomato leaf ?	P2322	-1			-1		-1	-1		
acid	ARTH unknown 1534-21-1 / 3-ketovalaric acid	P1570				-1		-1	1	1	
acid	unknown similar to Citramalate ? (Citramaleate ?)	P1655	-1				-1				
acid	2,3 dihydroxybutanedioic /pentanoic acid / galactonic acid	P2077					-1			1	
acid	2-keto-gluconic acid / Potato 015 unknown sugar	P2427		-1			-1			1	
acid	Propenoic acid, 2,3,3-tris[(trimethylsilyl)oxy]-, trimethylsilyl	P2893	1		1			1	1	1	
acid	Propane, 2-methyl-1,2-bis(trimethylsiloxy)- /3-Methyl-1,3-bis(	P0802	1					1			
acid	Ketone / Aldehyde (Tartaric acid - Glyceraldehyde)	P0871									
alcohol	like Galactinol	P2835	-1	-1	-1	-1	-1				
Alcohol	Coniferyl alcol (2TMS)	P4691	-1				-1			1	
FA	unknown FA	P3188									
FA	Heptadecanoic acid 1 TMS	P3263								1	
FA	(Z,Z)-	P3635		1						1	
FA	Docosanoic acid, trimethylsilyl ester	P4040	-1							1	
FA?	methylanthraquinone, O,O'-bis(trimethylsilyl	P3295	-1				-1			1	
oxime	Trehose / Erythrose Methoxiamine OR Glycolic acid	P1091					-1				
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAI**
SecM - S	Scopolin, tetra(trimethylsilyl)- ?	P1750	-1					1	1		
secM	biphenol 2 TMS	P2234	-1		-1		-1				
SecM	unknown trimethylsillyl estrone ? Aromatic ring	P1790	-1						1	1	
SecM	unknown	P2224	-1	-1			-1				
SecM	1H-Indole-2,3-dione	P3559	-1	-1	-1	-1	-1	-1	-1		-1
Similar to	Proline like compound - methyl proline ?	P1629	-1	-1	-1		-1	-1	-1		-1
SecM	2-Piperidinecarboxamide, 1-butyl-N-(2,6-dimethylphenyl)-	P4718	-1						1		
SecM	unknown	P3636	-1				-1	-1			
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P3620	-1	-1	-1		-1		-1		
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P5165	-1				-1			1	
SecM	Secondary Metabolite	P3252	-1	-1				-1	-1	1	
secM	unknown secondary metabolite large mol	P4833	-1	-1						1	
SecM	no 73 peak - non hydroxy/carboxy/amine TMS compound	P2350		1				1		1	
SecM	similar to (1-methyl-1-phenylethoxy)-	P5082	-1							1	
SecM	unknown	P5073	-1				-1			1	
SecM	Benzamide, 3,4-dimethoxy- derivative	P3154								1	
SecM	secondary aromatic metablite	P5050	-1							1	
SecM	1'H-Cholest-2-eno[3,2-b]indole, 5'-methoxy-, (5?)-	P4646	-1				-1	-1		1	
SecM	4-hydroxybenzoic acid	P2391		1		1				1	
SecM	5-Aminocarboxy-4,6-dihydroxypyrimidine	P2435			1	1					
SecM	non-TMS compound	P5001	-1				-1	1	1	1	
SecM	secondary metabolite (imidazole derivative ?)	P3211					-1			1	

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SA
Phospho	Sugar Phosphate (MPIMP-ID:221004-21-1)	P3438	-1				-1				
Phospho	Sugar Phosphate	P3681	-1	1					1	1	
Phospho	Triose (2,3 Propanediol) phosphate	P3372								1	
Phospho	Mannose Phosphate	P3502	-1			-1	-1			1	
Phospho	Ethanolamine - phosphate	P2643				-1				1	
Phospho	Inositol phosphate	P4125	-1							1	
Phospho	myo-Inositol-1 or 2- Phosphate	P3705	-1				-1		1	1	
Phospho	could be other inositol phosphate	P3585	-1	-1							
Phospho	Sugar Phosphate	P3611							1	1	
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SA**
Sugar	unknwn Sugar / Ploem Cmax 013	P2054	-1	-1	-1	-1			-1		-1
Sugar	di / tri saccharide	P3033	-1	-1	-1		-1	-1	-1		-1
Sugar		P3387	-1	-1	-1		-1	-1	-1	-1	-1
Sugar	Raffinose	P4455	-1	-1	-1	-1	-1	-1		-1	-1
Sugar	Cellobiose ?	P3106	-1	-1			-1	1			
Sugar	D-Gal-(1,6)-D-Glc (8TMS)]	P4510		1	1	1	1		1	1	
Sugar	Unknown Sugar	P4756		1	1	1	1		1	1	1
Sugar	similar to fructose / ketogluconic acid non-meox	P2412					-1			1	
Sugar	d-Turanose / di-tri sachharide	P3288	1	1				-1		1	
Sugar	MPIMP-ID:196004-45-1/Melizitose/Maltose/iso Maltose	P2922	1			-1				1	
Sugar	Mannonic acid / Tomato Leaf 027/ Potato Tuber 019	P2957							-1		
Sugar	P3867 Meox 1 ?	P3867	-1						-1		
Sugar	Sugar like Trehalose	P4631	-1				-1	-1			
Sugar	Trehalose ?	P4784	-1							1	
Sugar	unknown similar to Fructose 5 TMS	P2457	-1				-1				
Sugar		P4295	-1	-1	-1		-1	-1			
Sugar	Maltose Methox ?	P3991	-1					-1			
Sugar		P4814	-1			1			1	1	
Sugar	Isomaltose/Melibiose/Cellibiose (900 Reverse match)	P4140	-1							1	
Sugar	(8TMS); alpha-D-Gal-(1,6)-D-Glc (8TMS)]/Mannose	P3400								1	
Sugar	Melezitose (11TMS)) / alpha-D-Glc-(1,3)-beta-D-Fru-(2,1)-al	P2980	1					-1			
Sugar	Nigrose	P3816					-1	1		1	
Sugar	di / tri saccharide / Trehalose/Maltose/Lactose	P3013		-1			-1	-1			
Sugar (fu	Ribofuranose/Xylofuranose/Arabinofuranose/Phloem C Max	P2083					-1	-1	-1		
Sugar	Ara Col Leaf 094 / MPIMP-ID:250001-20-1 /Lactose	P3677	-1							1	
Sugar	Ath Sugar - MPIMP-ID:213-21-1	P3726	-1						1	1	
Sugar	Galactinol (9TMS); alpha-D-Gal-(1,3)-myo-Inositol (9TMS)]	P4312	-1							1	
Sugar	mono / di / saccharide	P3046		-1	1	1					
Sugar	Galactonic acid /2-keto-gluconic acid / Potato 015 unknown s	P2445	-1	-1				-1		1	
Sugar Alc	Maltitol	P4024	-1								
Sugar	Galacto-lactone or sugar acid	P2203	-1				-1			1	
Sugar	P3867 Meox 2	P3890	-1				-1		-1		
Sugar	Arabidopsis Col Leaf 046 / Altrose / Galactofuranose /	P2509	1	1				-1		1	
Sugar	Gulose / Potato Tuber 032	P4167	-1				-1			1	
Sugar	2-keto-L gulonic acid methoxime / Xylofuranose/Ribofuranos	P2000			-1					1	
Sugar	Matches Sucrose - quite large peak / Trehalose /Melizitos	P4692	-1				-1			1	
Sugar	Gluconic Acid 4 TMS lactone ?	P2823	-1					-1		1	
Sugar	Glucose derivative ?	P1540	-1				-1				
Sugar	alpha-D-Glc	P4053	-1				-1			1	

Table A2-4: Significance level of unknown metabolite pools at individual time points and paired-SAM analysis in response to Ethylene stress. Positively (1) and negatively (-1) significant metabolites are color-coded as described in the caption of Table 5-1.

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
acid	Hydroxy-pyruvic acid/Ara Col leaf 2 / 4-keto Glucose diMeox	P1906		1		1					
acid	Glyceric Acid like unknown	P2793		1				-1	-1		
acid	Ara Col Leaf 012 or 2-keto gulonic acid Methoxime	P1361	1		1	1		1			
acid	Ascorbic Acid derivative in Tomato leaf ?	P2322		1	1	1	1				
acid	ARTH unknown 1534-21-1 / 3-ketovalaric acid	P1570	1	1	1	1	1			1	1
acid	unknown similar to Citramalate ? (Citramaleate ?)	P1655	-1		-1		1				
acid	2,3 dihydroxybutanedioic /pentanoic acid / galactonic acid	P2077		1		1		1			
acid	2-keto-gluconic acid / Potato 015 unknown sugar	P2427					1				
acid	Propenoic acid, 2,3,3-tris[(trimethylsilyl)oxy]-, trimethylsilyl	P2893		1	1		1			1	
acid	acid ester	P1391	-1		-1					1	
acid	Propane, 2-methyl-1,2-bis(trimethylsiloxy)- /3-Methyl-1,3-bis(	P0802	1			1	1	1			
acid	Ketone / Aldehyde (Tartaric acid - Glyceraldehyde)	P0871									
alcohol	like Galactinol	P2835						1	-1		
Alcohol	Coniferyl alcol (2TMS)	P4691	-1			1	1			1	
FA	unknown FA	P3188			1	1	1				
FA	Heptadecanoic acid 1 TMS	P3263	-1		1						
FA	(Z,Z)-	P3635	-1	1	1	1	1	-1			
FA	Docosanoic acid, trimethylsilyl ester	P4040		1	1	1	1	-1			
FA?	methylanthraquinone, O,O'-bis(trimethylsilyl	P3295	-1			1			-1		
oxime	Trehose / Erythrose Methoxiamine OR Glycolic acid	P1091	-1	1			-1	-1	-1		
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAM**
SecM - S(	Scopolin, tetra(trimethylsilyl)- ?	P1750		1	1	1	1		-1		
secM	biphenol 2 TMS	P2234	1	1	1	1	1	-1		1	
SecM	unknown trimethylsillyl estrone ? Aromatic ring	P1790	1	1	1	1	1				
SecM	unknown	P2224	-1	1		1					
SecM	1H-Indole-2,3-dione	P3559				1					
SecM	2-Piperidinecarboxamide, 1-butyl-N-(2,6-dimethylphenyl)-	P4718	-1	1	1	1	1	-1			
SecM	unknown	P3636		1	1	1					
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P3620					-1				
SecM	Cholest-2-eno[2,3-b]naphthalene, 5'-amino-	P5165	-1			-1		-1	-1		
Similar to	Proline like compound - methyl proline ?	P1629	-1	-1	-1	-1	-1	-1	-1	-1	-1
SecM	unknown secondary metabolite	P1801	###	###	###	###	###	###	#N/A	###	##
SecM	Secondary Metabolite	P3252	1	1	1	1	1			1	1
SecM	Secondary Metabolite	P3656	###	###	###	###	###	###	#N/A	###	##
secM	unknown secondary metabolite large mol	P4833	-1			1			-1		
SecM	no 73 peak - non hydroxy/carboxy/amine TMS compound	P2350	1	1	1	1	1	1	1	1	1
SecM	similar to (1-methyl-1-phenylethoxy)-	P5082	-1	1		1	1	-1		1	
SecM	unknown	P5073	-1	1		1		-1	-1		
SecM	Benzamide, 3,4-dimethoxy- derivative	P3154		1	1	1	1				
SecM	Secondary amine	P3521	###	###	###	###	###	###	#N/A	###	##
SecM	secondary aromatic metablite	P5050		1	1	1	1	-1		-1	
SecM	1'H-Cholest-2-eno[3,2-b]indole, 5'-methoxy-, (5?)-	P4646	-1	1	1	1	1			1	
SecM	4-hydroxybenzoic acid	P2391		1		1		1		1	
SecM	5-Aminocarboxy-4,6-dihydroxypyrimidine	P2435		-1		1			1	-1	
SecM	Benzoic acid derivative	P1923		1	1	1	1	1	1	1	1
SecM	non-TMS compound	P5001				-1	1			1	
SecM	secondary metabolite (imidazole derivative ?)	P3211			1		1			-1	
SecM	Benzoic acid, 5-methoxy-2-oxy-TMS	P5105		1	1	1	1		1	1	1

Category	Best Match	Peak ID	1 hr	3 hr	6 hr	9 hr	12 hr	18 hr	24 hr	30 hr	SAM
Phospho	Sugar Phosphate (MPIMP-ID:221004-21-1)	P3438		1	1	1	1			-1	
Phospho	Sugar Phosphate	P3681		1	1		1				
Phospho	Triose (2,3 Propanediol) phosphate	P3372		1	1		1			1	
Phospho	Mannose Phosphate	P3502	-1	1			1	-1			
Phospho	Ethanolamine - phosphate	P2643	1	1			1		1	1	1
Phospho	Inositol phosphate	P4125		1	1		1		1	1	1
Phospho	myo-Inositol-1 or 2- Phosphate	P3705				1			1		
Phospho	could be other inositol phosphate	P3585	-1	-1	-1			1			
Phospho	Sugar Phosphate	P3611		1			1				
Category	**Best Match**	**Peak ID**	**1 hr**	**3 hr**	**6 hr**	**9 hr**	**12 hr**	**18 hr**	**24 hr**	**30 hr**	**SAM**
Sugar	027 / POTATO TUBER 019 / Glucose impurity	P2957	-1	-1	-1			1	-1		
Sugar	similar to fructose / ketogluconic acid non-meox	P2412		1				1	-1		
Sugar	d-Turanose / di-tri sachharide	P3288	1	1	1	1	1		1	1	1
Sugar	MPIMP-ID:196004-45-1 / Melizitose / Maltose / iso Maltose	P2922	1	1	1		1	1		1	
Sugar	P3867 Meox 1 ?	P3867	-1	-1	-1	-1			-1	-1	-1
Sugar	Sugar like Trehalose	P4631	-1			1		-1	-1		
Sugar	Trehalose ?	P4784	1	1	1	1	1			1	1
Sugar	unknwn Sugar / Ploem Cmax 013	P2054	1	1	1		1	1	1	1	1
Sugar	D-Gal-(1,6)-D-Glc (8TMS)]	P4510	-1	1		1	1	-1	-1		
Sugar	unknown similar to Fructose 5 TMS	P2457						1	-1	1	
Sugar		P4295	-1			1		-1	-1		
Sugar		P4814		1		1					
Sugar	Isomaltose/Melibiose/Cellibiose (900 Reverse match)	P4140	-1	1			1				
Sugar	(8TMS); alpha-D-Gal-(1,6)-D-Glc (8TMS)]/Mannose	P3400		1		1					
Sugar	(11TMS)) / alpha-D-Glc-(1,3)-beta-D-Fru-(2,1)-alpha-D-Glc	P2980	1	1	1	1	1	1		1	1
Sugar	Nigrose	P3816	-1	-1				1			
Sugar	di / tri saccharide / Trehalose/Maltose/Lactose	P3013				1					
Sugar (fu	1/Ribofuranose/Xylofuranose/Arabinofuranose/Phloem C	P2083	-1	1	1	1					
Sugar	Ara Col Leaf 094 / MPIMP-ID:250001-20-1 /Lactose	P3677	-1			1	-1		-1		
Sugar	Ath Sugar - MPIMP-ID:213-21-1	P3726							1	-1	
Sugar	Galactinol (9TMS); alpha-D-Gal-(1,3)-myo-Inositol (9TMS)]	P4312	-1	1			1	-1			
Sugar	mono / di / saccharide	P3046	-1	-1	-1		-1		-1	-1	-1
Sugar	sugar/Pentonic acid (5TMS)	P2445		1			1			1	
Sugar Alc	Maltitol	P4024		1		1	1				
Sugar	Galacto-lactone or sugar acid	P2203	-1		-1		-1		-1		
Sugar	P3867 Meox 2	P3890	-1		-1	1	-1				
Sugar	Arabidopsis Col Leaf 046 / Altrose / Galactofuranose /	P2509	1	1	1	1	1			1	1
Sugar	Gulose / Potato Tuber 032	P4167	-1		-1			-1		1	
Sugar	2-keto-L gulonic acid methoxime / Xylofuranose/Ribofuranos	P2000					1	1		-1	
Sugar	Glucose derivative unknown ?	P1324							1		
Sugar	Matches Sucrose - quite large peak / Trehalose /Melizitos	P4692	-1	1		1			1		
Sugar	Gluconic Acid 4 TMS lactone ?	P2823	-1	-1	-1		-1	-1	-1	-1	-1
Sugar	Raffinose	P4455	-1		-1			-1	-1	-1	-1
Sugar	Glucose derivative ?	P1540	-1			1		-1	-1		
Sugar	di / tri saccharide	P3033	-1			1	-1		-1		
Sugar		P3387		1		1		-1	-1		
Sugar	Cellobiose ?	P3106	-1	-1	-1	-1	-1		-1	-1	-1
Sugar	alpha-D-Glc	P4053	-1		-1	1	-1				

REFERENCES

Adams-Phillips L, Barry C, Kannan P, Leclercq J, Bouzayen M, Giovannoni J: Evidence that CTR1-mediated ethylene signal transduction in tomato is encoded by a multi-gene family whose members display distinct regulatory features. Plant Mol Bio 2004, 54:387–404.

Ajjawi I, Shintani D: Engineered plants with elevated vitamin E: A neutraceutical success story. Trends Biotechnol 2004, 3:104-107.

Apse MP, Blumwald E. Engineering salt tolerance in plants. Curr Opin Biotechnol. 2002, 13(2):146-50.

Arabidopsis functional genomics at TIGR [http://atarrays.tigr.org/]

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

Ausloos P, Clifton CL, Lias SG, Mikaya AI, Stein SE, Tchekhovskoi DV, Sparkman OD, Zaikin V, Zhu D: The critical evaluation of a comprehensive mass spectral library. J Am Soc Mass Spectrom 1999, 10:287-299.

Bassi PK, Spencer MS: Effect of Carbon Dioxide and Light on Ethylene Production in Intact Sunflower Plants. Plant Physiol 1982, 69:1222-1225.

Benavente LM, Alonso JM. Molecular mechanisms of ethylene signaling in Arabidopsis. J Plant Physiol. 2006, 163(12):1247-58. Mol Biosyst 2006, 2(3-4):165-73.

Bloom AJ, Smart D, Nguyen D, Searls P: Nitrogen assimilation and growth of wheat under elevated carbon dioxide. Proc Natl Acad Sci USA 2002, 99:1730-35.

Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorlach J: Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. Plant Cell 2001, 13:1499-1510.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001, 29:365-71.

Broeckling CD, Huhman DV, Farag MA, Smith JT, May GD, Mendes P, Dixon RA, Sumner LW: Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. J Expt Bot 2005, 56:323-336.

Buchanan B, Gruissem W, Jones R: Biochemistry & Molecular Biology of Plants. Rockville: American Society of Plant Physiologists; 2001.

Burja AM, Dhamwichukorn S, Wright PC: Cyanobacterial postgenomic research and systems biology. Trends Biotechnol 2003, 21(11):504-11.

Butte AJ, Tamayo D, Slonim TR, Golub IS, Kohane, P: Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proc Natl Acad Sci USA 2000, 97:12182-12186.

Cao WH, Liu J, He XJ, Mu RL, Zhou HL, Chen SY, Zhang JS. Modulation of ethylene responses affects plant salt-stress responses. Plant Physiol 2007, 143(2):707-19.

Castrillo JI, Hayes A, Mohammed S, Gaskell SJ, Oliver SG: An optimized protocol for metabolome analysis in yeast using direct infusion electron-spray mass spectrometry. Phytochemistry 2003, 62(6):929-37.

Chang IF, Szick-Miranda K, Pan S, Bailey-Serres J: Proteomic characterization of evolutionarily conserved and variable proteins of Arabidopsis cytosolic ribosomes. Plant Physiol 2005, 137:848-862.

Chen YF, Etheridge N, Schaller GE: Ethylene signal transduction. Ann Bot (Lond) 2005, 95:901-915.

Cheng CL, Acedo GN, Cristinsin M., Conkling MA: Sucrose mimics the light induction of Arabidopsis nitrate reductase gene transcription. Proc Natl Acad Sci USA 1992: 89:1861-1864.

Cheng CL, Acedo GN, Dewdney J, Goodman HM, Conkling MA: Differential Expression of the Two Arabidopsis Nitrate Reductase Genes. Plant Physiol 1991, 96:275-279.

Cheng SH, Moore B, Seemann JR: Effects of short- and long-term elevated CO2 on the expression of ribulose-1,5-bisphosphate carboxylase/oxygenase genes and carbohydrate accumulation in leaves of *Arabidopsis thaliana* (L.). Plant Physiol 1998, 116:715-723.

Colón LA, Baird LJ: Detectors in Modern Gas Chromatography. In: Grob, R. L., Barry, E. F. (Ed.s), Modern Practice of Gas Chromatography 4th Edition, John Wiley & Sons Inc., Hoboken, New Jersy, USA, 2004, pp 277-338.

Cook D, Fowler S, Fiehn O & Thomashow MF: A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci USA 2004, 101:10583-89

Corruzi J, Last R: Amino Acids. In Biochemistry & Molecular Biology of Plants, Edited by Buchanan B, Gruissem W, Jones R. Rockville: American Society of Plant Physiologists; 2001: 358-411.

Coschigano K, Schultz C, Melo-Oliveria R, Lim J, Coruzzi G: Arabidopsis gls Mutants and Distinct Fd-GOGAT Genes: Implications for Photorespiration and Primary Nitrogen Assimilation. Plant Cell 1998, 10:741-752.

De Paepe A, Van der Straeten D. Ethylene biosynthesis and signaling: an overview. Vitam Horm 2005, 72:399-430.

De Paepe A, Vuylsteke M, Van Hummelen P, Zabeau M, Van Der Straeten D: Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in Arabidopsis. Plant J 2004, 39(4):537-59.

Debouba M, Gouia H, Suzuki A, Ghorbel MH. NaCl stress effects on enzymes involved in nitrogen assimilation pathway in tomato "Lycopersicon esculentum" seedlings. J Plant Physiol 2006, 163(12):1247-58.

Deng C, Zhang X, Zhu W, Qian J: Gas chromatography-mass spectrometry with solid-phase microextraction method for determination of methyl salicylate and other volatile compounds in leaves of Lycopersicon esculentum. Anal Bioanal Chem. 2004, 378(2):518-22

Denis DT, Blakely SD: Carbohydrate Metabolism. In Biochemistry & Molecular Biology of Plants, Edited by Buchanan, B, Gruissem W, Jones R. Rockville: American Society of Plant Physiologists; 2001: 630-675.

Dey P, Harborne J: Plant Biochemistry, San Diego: Academic Press Inc.; 1997.

Dhawan KR, Bassi PK, Spencer MS: Effects of Carbon Dioxide on Ethylene Production and Action in Intact Sunflower Plants. Plant Physiol 1981, 68:831-834.

Dixon RA, Lamb CJ, Masoud S, Sewalt VJ, Paiva NL: Metabolic engineering: prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses--a review. Gene 1996, 179:61-71.

Dluzniewska P, Gessler A, Dietrich H, Schnitzler JP, Teuber M, Rennenberg H. Nitrogen uptake and metabolism in Populus x canescens as affected by salinity. New Phytol. 2007, 173(2):279-93.

Dutta B, Snyder R, Klapa MI: Significance Analysis of Time-Series Transcriptomic Data: A methodology that enables the identification and further exploration of the differentially expressed genes at each time-point. Biotech Bioeng 2007, DOI 10.1002/bit.21432.

Eastmond PJ, Graham IA. Trehalose metabolism: a regulatory role for trehalose-6-phosphate? Curr Opin Plant Biol 2003, 6(3):231-5.

Eastmond PJ, Li Y, Graham IA.Is trehalose-6-phosphate a regulator of sugar metabolism in plants? J Exp Bot 2003, 54(382):533-7.

Essah PA, Davenport R, Tester M: Sodium influx and accumulation in Arabidopsis. Plant Physiol 2003, 133(1):307-18.

Evershed, R: Advances in Silylation. In: Blau, K., Halket, J. (Ed.s), Handbook of Derivatives for Chromatography 2nd Edition, John Wiley & Sons Ltd, West Sussex, England 1993, pp 51-100.

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L: Metabolite profiling for plant functional genomics. Nature Biotech 2000a, 18:1157-1168.

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

Fiehn O: Integrated studies in plant biology using multi parallel techniques. Current opinions in biotechnology 2000, 112:82-86.

Fogler SH: Elements of Chemical Reaction Engineering, 3rd Edition, Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA 2002.

Galbraith DW, Birnbaum K. Global studies of cell type-specific gene expression in plants. Annu Rev Plant Biol 2006, 57:451-75.

Gamborg OL, Murashige T, Thorpe TA, Vasil IK: Plant tissue culture media. In Vitro 1976, 12:473-478.

Gao Z, Chen YF, Randlett MD, Zhao XC, Findell JL, Kieber JJ, Schaller GE: Localization of the Raf-like kinase CTR1 to the endoplasmic reticulum of Arabidopsis through participation in ethylene receptor signalling complexes. J Biol Chem 2003, 278:34725–34732.

Gehrke CW, Leimer K: Trimethylsilylation of amino acids derivatization and chromatography. J Chromatogr 1971, A57: 219-238.

Gehrke CW, Nakamoto H, Zumwalt RW: Gas-liquid chromatography of protein amino acid trimethylsilyl derivatives. J Chromatogr 1969, A45: 24-51.

Gemesi LI, Kapas M, Szeberenyi S: Application of LC-MS analysis to the characterisation of the in vitro and in vivo metabolite profiles of RGH-1756 in the rat. J Pharm Biomed Anal 2001, 24(5-6):877-85.

Ghassemian M, Lutes J, Tepperman JM, Chang HS, Zhu T, Wang X, Quail PH, Markus Lange B: Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis. Arch Biochem Biophys 2006, 448:45-59.

Goddijn O, Smeekens S. Sensing trehalose biosynthesis in plants. Plant J 1998, 14(2):143-6.

Gopaul SV, Farrell K, Abbott FS: Gas chromatography/negative ion chemical ionization mass spectrometry and liquid chromatography/electrospray ionization tandem mass spectrometry quantitative profiling of N-acetylcysteine conjugates of valproic acid in urine: application in drug metabolism studies in humans. J Mass Spectrom 2000, 35(6):698-704.

Grodzinski B, Woodrow L, Leonardos ED, Dixon M, Tsujita MJ: Plant responses to short- and long-term exposures to high carbon dioxide levels in closed environments. Adv Space Res 1996, 18:203-211.

Gullberg J, Jonsson P, Nordstrom A, Sjostrom M, Moritz T: 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 2004. 15: 331(2):283-95.

Guo H, Ecker JR: The ethylene signaling pathway: new insights. Curr Opin Plant Biol 2004, 7:40-49.

Halket JM, Waterman D, Przyborowska AM, Patel RK, Fraser PD, Bramley PM: Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Expt Bot 2005, 56:219-43.

Halket, JM: Derivatives for Gas Chromatography-Mass Spectrometry. In: Blau, K., Halket, J. (Ed.s), Handbook of Derivatives for Chromatography 2nd Edition, John Wiley & Sons Ltd, West Sussex, England, 1993, pp 297-322.

Hamilton EW 3rd, Heckathorn SA. Mitochondrial adaptations to NaCl. Complex I is protected by anti-oxidants and small heat shock proteins, whereas complex II is protected by proline and betaine. Plant Physiol 2001, 126(3):1266-74.

Hawkesford MJ: Transporter gene families in plants: the sulphate transporter gene family — redundancy or specialization? Physiologia Plantarum 2003, 117:155-163.

Hayashi F, Ichino T, Osanai M, Wada K. Oscillation and regulation of proline content by P5CS and ProDH gene expressions in the light/dark cycles in Arabidopsis thaliana L. Plant Cell Physiol. 2000, 41(10):1096-101.

He JX, Gendron JM, Yang Y, Li J, Wang ZY. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in Arabidopsis. Proc Natl Acad Sci USA 2002, 99(15):10185-90.

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

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J,

Papenbrock J, Saito K: Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem 2005, 280:25590-25595.

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M,Fujiwara T, Saito K: Integration of transcriptomics and metabolomics for understanding of globalresponses to nutritional stresses in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 2004, 101(27):10205-10.

Hua J, Sakai H, Nourizadeh S, Chen QG, Bleecker AB, Ecker JR, Meyerowitz EM: Ein4 and ERS2 are members of the putative ethylene receptor family in Arabidopsis. Plant Cell 1998, 10:1321–1332.

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

Hwang D, Rust AG, Ramsey S, Smith JJ, Leslie DM, Weston AD, Atauri P, Aitchison JD, Hood L, Siegel AF, Bolouri H: A data integration methodology for systems biology. Proc Natl Acad Sci USA 2005, 102:17296-17301.

Hwang D, Smith JJ, Leslie DM, Weston AD, Rust AG, Ramsey S, Atauri P, Siegel AF, Bolouri H, Aitchison JD, Hood L: A data integration methodology for systems biology: Experimental verification. Proc Natl Acad Sci USA 2005, 102:17302-17307.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001, 292:929-934.

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

Jiye A, Trygg J, Gullberg J, Johansson A, Jonson P, Antti H, Marklund SL, Moritz. T: Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome. Anal Chem 2005, 77:8086-94.

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

Kanani H, Klapa MI: Data normalization strategy for quantitative high-throughput metabolomic profiling analysis using Gas Chromatography – Mass Spectrometry. Metabolic Eng 2007, 9(1):39-51.

Kanani H: Time series metabolic profiling analysis of the short term *Arabidopsis thaliana* response to elevated CO2 using gas chromatography mass spectrometry. 2004, Master's Thesis. University of Maryland, College Park, MD.

Kataoka T, Watanabe-Takahashi A, Hayashi N, Ohnishi M, Mimura T, Buchner P, Hawkesford MJ, Yamaya T, Takahashi H: Vacuolar sulfate transporters are essential determinants controlling internal distribution of sulfate in Arabidopsis. Plant Cell 2004a, 16:2693–2704.

Kataoka T, Hayashi N, Yamaya T, Takahashi H: Root-to-shoot transport of sulfate in Arabidopsis. Evidence for the role of SULTR3;5 as a component of low-affinity sulfate transport system in the root vasculature. Plant Phys 2004b, 136:4198–4204.

Katona Zs.F, Sass P, Molnar-Perl: Simultaneous determination of sugar, sugar alcohols, acids and amino acids in apricots by gas chromatography-mass spectrometry. J of Chromat 1999, 847:91-102.

Katz JE, Dumlao DS, Clarke S, Hau J: A new technique (COMSPARI) to facilitate the identification of minor compounds in complex mixtures by GC/MS and LC/MS: tools for the visualization of matched datasets. J Am Soc Mass Spectrom 2004, 15(4):580-4.

KEGG: Kyoto Encyclopedia of Genes and Genomes [www.kegg.com]

Kieber JJ, Rothenberg M, Roman G, Feldman KA, Ecker JR: CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the Raf family of protein kinases. Cell 1992, 72:427–441.

Kim H, Snesrud EC, Haas B, Cheung F, Town CD, Quackenbush J: Gene expression analyses of Arabidopsis chromosome 2 using a genomic DNA amplicon microarray. Genome Res 2003, 13:327-340.

Kitson FG, Larsen B, McEwen CN: Gas Chromatography and mass spectrometry: A practical guide, Academic Press, New York, 1996.

Klapa MI, Quackenbush J: The Quest for the Mechanisms of Life. Biotech Bioeng 2003, 84:739-742.

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

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

Kopka, J, Schauer, Krueger S, Birkemeyer C, Usadel, B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D: GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 2005, 21: 1635-8.

Kopriva S: Regulation of sulfate assimilation in Arabidopsis and beyond. Ann Bot (Lond) 2006, 97:479-95.

Kose F, Weckwerth W, Linke T, Fiehn O: Visualizing plant metabolomic correlation network using clique-metabolite metrices. Bioinformatics 2001, 17: 1198-1208.

Krebbers E, Seurinck J, Herdies L, Cashmore AR, Timko MP: Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant Mol Biol 1988, 11:745–759.

Kursteiner O, Dupuis I, Kuhlemeier C: The pyruvate decarboxylase1 gene of Arabidopsis is required during anoxia but not other environmental stresses. Plant Physiol 2003, 132:968-78.

Laine RA, Sweeley CC: Analysis of Trimethylsilyl O-Methyloximes of Carbohydrates by Combined Gas-Liquid Chromatography-Mass Spectrometry. Anal Biochem 1971, 43:533-538.

Larios B, Aguera E, de la Haba P, Perez-Vicente R, Maldonado JM: A short-term exposure of cucumber plants to rising atmospheric CO2 increases leaf carbohydrate content and enhances nitrate reductase expression and activity. Planta 2001, 212:305-312.

Leimer K, Rice RH, Gehrke CW: Complete mass spectra of N-trifluoroacetyl-n-butyl esters of amino acids. J Chromatogr 1977, A141: 355-375.

Leyman B, Van Dijck P, Thevelein JM. An unexpected plethora of trehalose biosynthesis genes in Arabidopsis thaliana. Trends Plant Sci 2001, 6(11):510-3.

Liu F, Vantoai T, Moy LP, Bock G, Linford LD, Quackenbush J: Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. Plant Physiol 2005, 137:1115-1129.

Mayr M, Siow R, Chung YL, Mayr U, Griffiths JR, Xu Q: Proteomic and metabolomic analysis of vascular smooth muscle cells: role of PKCdelta. Circ Res 2004, 94(10):e87-96.

Moore B, Zhou L, Rolland F, Hall Q, Cheng WH, Liu YX, Hwang I, Jones T, Sheen J: Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. Science 2003, 300(5617):332-6.

Muller J, Aeschbacher RA, Sprenger N, Boller T, Wiemken A. Disaccharide-mediated regulation of sucrose:fructan-6-fructosyltransferase, a key enzyme of fructan synthesis in barley leaves. Plant Physiol 2000, 123(1):265-74.

Munne-Bosch S, Alegre L. Drought-induced changes in the redox state of alpha-tocopherol, ascorbate, and the diterpene carnosic acid in chloroplasts of Labiatae species differing in carnosic acid contents. Plant Physiol 2003, 131(4):1816-25.

Nelson DE, Koukoumanos M, Bohnert HJ. Myo-inositol-dependent sodium uptake in ice plant. Plant Physiol 1999, 119(1):165-72.

Nelson DL, Cox MM: Lehninger Principles of Biochemistry. New York: Worth Publishers; 2002.

Netting AG. pH, abscisic acid and the integration of metabolism in plants under stressed and non-stressed conditions: cellular responses to stress and their implication for plant water relations. J Exp Bot 2000, 51(343):147-58.

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

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

Nuccio ML, Rhodes D, McNeil SD, Hanson AD. Metabolic engineering of plants for osmotic stress resistance. Curr Opin Plant Biol 1999, 2(2):128-34. .

Oksman-Caldentey KM, Inze D: Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci 2004, 9:433-440.

Paul MJ, Foyer CH: Sink regulation of photosynthesis. J Exp Bot 2001, 52:1383-1400.

Pavlidis P, Noble WS: Analysis of strain and regional variation in gene expression in mouse brain. Genome Biology 2:research 2001, 0042.1-0042.15.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP: Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Nat Acad Sci USA 1994, 91:5022-6.

Poole CF: Recent Advances in the Silylation of Organic Compounds for Gas Chromatography. In: Blau, K., King, G. (Ed.s), Handbook of Derivatives for Chromatography, Heydon & Son Inc, Philadelphia, USA, pp 152-200.

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

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T: The path forward for biofuels and biomaterials. Science 2006, 311:484-489.

Ratcliffe GR, Hill YC: Probing Plant Metabolism with NMR, Annual Reviews in Plant Physiology & Plant Molecular Biology, 2001, Vol. 52, 499-526.

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

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

Rontein D, Basset G, Hanson AD Metabolic engineering of osmoprotectant accumulation in plants. Metab Eng 2002, 4(1):49-56. .

Saeed AI, 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, Quackenbush J: TM4: a free, open-source system for microarray data management and analysis. Biotechniques 2003, 34(2):374-8

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

Sakamoto A, Murata N. The use of bacterial choline oxidase, a glycinebetaine-synthesizing enzyme, to create stress-resistant transgenic plants. Plant Physiol 2001, 125(1):180-8.

Sanchez-Aguayo I, Rodriguez-Galan JM, Garcia R, Torreblanca J, Pardo JM. Salt stress enhances xylem development and expression of S-adenosyl-L-methionine synthase in lignifying tissues of tomato plants. Planta 2004, 220(2):278-85.

Sannazzaro AI, Echeverria M, Alberto EO, Ruiz OA, Menendez AB. Modulation of polyamine balance in Lotus glaber by salinity and arbuscular mycorrhiza. Plant Physiol Biochem 2007, 45(1):39-46.

Schad M, Mungur R, Fiehn O, Kehr J. Metabolic profiling of laser microdissected vascular bundles of Arabidopsis thaliana. Plant Methods 2005,18;1(1):2.

Scheible WR, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelasch D, Thimm O, Udvardi MK, Stitt M: Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. Plant Physiol 2004, 136:2483-2499.

Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995, 270:467-70.

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K: Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 2002, 31:279-292.

Siedow J, Day DA: “Respiration and Photorespiration” in Biochemistry & Molecular Biology of Plants, ed. Buchanan B., Gruissem W., and Jones R., American Society of Plant Physiologists, Rockville, Maryland 2001.

Slater S, Mitsky TA, Houmiel KL, Hao M, Reiser SE, Taylor NB, Tran M, Valentin HE, Rodriguez DJ, Stone DA, Padgette SR, Kishore G, Gruys KJ: Metabolic engineering of Arabidopsis and Brassica for poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer production. Nat Biotechnol 1999, 17:1011-1016.

Smart D, Ritchie K, Bloom A, Bugbee B: Nitrogen balance for wheat canopies (Triticum aestivum cv. Veery 10) grown under elevated and ambient CO2 concentrations. Plant Cell Environ 1998, 21:753-63.

Smith CJ: Carbohydrate Chemistry. In Plant Biochemistry and Molecular Biology, Edited by Lea PJ, Leegood RC: West Sussex, England: John Wiley and Sons; 1993: 73:113.

Sommerville C, Dangl J: Genomics. Plant biology in 2010. Science 2000, 290:2077-2078.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Ansers K, Eisen MB, Brown PO, Botstein D, Futcher B: Comprehensive identification of cell-cycle regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Bio of Cell 1998, 9:3273-97.

Stein SE: An integrated method for spectrum extraction and compound identification from gas chromatography / mass spectrometry data. J Am Soc Of Mass Spectrometry 1999, 10: 770-81.

Stephenopoulous G, Aristidou AA, Nielsen J: Metabolic Engineering: Principals and Methodologies. San Diego: Academic Press; 1998.

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

Stitt M: Rising CO2 levels and their potential significance for carbon flow in photosynthetic cells. Plant Cell Enviro 1991, 14:741-762.

Sumner LW, Mendes P, Dixon RA: Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 2003, 62(6):817-36.

Tabuchi T, Kawaguchi Y, Azuma T, Nanmori T, Yasuda T. Similar regulation patterns of choline monooxygenase, phosphoethanolamine N-methyltransferase and S-adenosyl-L-methionine synthetase in leaves of the halophyte Atriplex nummularia L. Plant Cell Physiol 2005, 46(3):505-13.

Tadege M, Dupuis I, Kuhlemeier C: Ethanolic fermentation: new functions for an old pathway. Trends Plant Sci 1999, 4:320-325.

Taiz L, and Zeiger E: Plant Physiology. Sinauer Associates, Inc., MA, 2002.

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K. Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J 2002, 29(4):417-26.

Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu JK, Shinozaki K: Comparative genomics in salt tolerance between Arabidopsis and arabidopsis-related halophyte salt cress using Arabidopsis microarray. Plant Physiol 2004, 135(3):1697-709.

Tang W, Newton RJ, Li C, Charles TM. Enhanced stress tolerance in transgenic pine expressing the pepper CaPF1 gene is associated with the polyamine biosynthesis. Plant Cell Rep 2007, 26(1):115-24.

Taylor J, King RD, Altmann T, Fiehn O: Application of metabolomics to plant genotype discrimination using statistics and machine learning. Bioinformatics Suppl 2002, 2: S241-8.

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

Thrower JS, Blalock R 3rd, Klinman JP: Steady-state kinetics of substrate binding and iron release in tomato ACC oxidase. Biochemistry 2001, 40:9717-9724.

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K: Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J 2005, 42:218-235.

Trinchant JC, Boscari A, Spennato G, Van de Sype G, Le Rudulier D. Proline betaine accumulation and metabolism in alfalfa plants under sodium chloride stress. Exploring its compartmentalization in nodules. Plant Physiol 2004, 135(3):1583-94.

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB: Missing value estimation methods for DNA microarrays. Bioinformatics 2001, 17:520-525.

Tusher GV, Tibshirani R, Chu G: Significance analysis of microarray applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001, 98:5116-21.

van Loon LC, Geraats BP, Linthorst HJ. Ethylene as a modulator of disease resistance in plants. Trends Plant Sci 2006, 11(4):184-91.

Verala C, Agosin E, Baez M, Klapa M, Stephenopoulos G: Metabolic flux redistribution in Corynebacterium glutamicum in response to osmotic stress. Appl Microbiol Biotchnol 2003, 60:547-55.

Vernon DM, Bohnert HJ. Increased Expression of a myo-Inositol Methyl Transferase in Mesembryanthemum crystallinum Is Part of a Stress Response Distinct from Crassulacean Acid Metabolism Induction. Plant Physiol 1992, 99(4):1695-1698.

Wang KL, Li H, Ecker JR: Ethylene biosynthesis and signaling networks. Plant Cell 2002, 14:S131-S151.

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O: Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci USA 2004, 101:7809-14.

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.

Wingler A, Fritzius T, Wiemken A, Boller T, Aeschbacher RA: Trehalose induces the ADP-glucose pyrophosphorylase gene, ApL3, and starch synthesis in Arabidopsis. Plant Physiol 2000, 124(1):105-14.

Wingler A. The function of trehalose biosynthesis in plants. Phytochemistry 2002, 60(5):437-40.

Xiang W, Windl O, Wünsch G, Dugas W, Kohlmann A, Dierkes N, Westner IM, Kretzschmar HA: Identification of Differentially Expressed Genes in Scrapie-Infected Mouse Brains by Using Global Gene Expression Technology. J Virol 2004, 78(20): 11051-60.

Zhong GV, Burns JK: Profiling ethylene-regulated gene expression in *Arabidopsis thaliana* by microarray analysis. Plant Mol Biol 2003, 53:117-131.