ABSTRACT

Title of Thesis: **TIME SERIES METABOLIC PROFILING ANALYSIS OF THE SHORT TERM *Arabidopsis thaliana* RESPONSE TO ELEVATED CO_2 USING GAS CHROMATOGRAPHY MASS SPECTROMETRY.**

Harin Kanani, Master of Science, 2004

Thesis Directed By: **Dr. Maria Klapa - Assistant Professor**
Department of Chemical Engineering

Metabolic profiling has emerged as a high throughput technique for the quantitative analysis of the cellular physiological state at the metabolic level. It allows for the simultaneous relative quantification of hundreds of low molecular weight intra cellular metabolites. In this analysis, the polar metabolic profiles of *A. thaliana* liquid cultures (grown for 12 days, under light and 23°C) throughout 1-day treatment with 1% CO_2, were measured using gas chromatography-mass spectrometry. Despite the advantages of time series analysis, this is the first plant metabolic profiling study of this type reported in the literature. The time series metabolic profiles were analyzed using multivariate statistical techniques. Data analysis revealed repression of photorespiration, repression of nitrogen assimilation and increase in structural carbohydrates. It is for the first time that the latter phenomenon is observed as a result of elevated CO_2 in the plant environment.

TIME SERIES METABOLIC PROFILING ANALYSIS OF THE SHORT TERM *Arabidopsis thaliana* RESPONSE TO ELEVATED CO_2 USING GAS CHROMATOGRAPHY MASS SPECTROMETRY.

By

Harin Haridas Kanani

Thesis 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
Masters of Science
2004

Advisory Committee:
Dr. Maria Klapa (Advisor), Chair
Dr. John Quackenbush
Dr. Nam Sun Wang

Acknowledgements

Quantitative systems biology approach is an effort towards developing a framework for a comprehensive analysis of a biological system using the concepts and methods developed in historically independent scientific streams. Hence any study or publications in this research area is almost always a result of a combined effort of a group of people – this thesis being no exception.

I would first and foremost like to thank my advisor Dr. Maria Klapa and Dr. John Quackenbush for the initial experiment design and providing the initial conceptual motivation for the current project. I would also like to thank Dr. Tara Vantoi, Lara Linford, Linda Moy and Jeremy Matthew at TIGR for the initial work on carrying out the experiment and extraction of the metabolites which was the starting point for my analysis and thesis. I would like to thank Dr. Brian Bagley and Prof. Judd Nelson associated with the Mass Spectrometry facility of University of Maryland – College Park for their helpful suggestions and support. I would also like to thank Dr. Nam Wang for helping me to develop my thinking through his courses and also agreeing to be on my thesis committee. Specifically I would like to thank him for frequent short personal discussions in his office and his feedback and helpful suggestions on my report and presentations. 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.

My individual transition from a process chemical engineer to a quantitative systems biologist would not have been possible without the long discussions I have had with my advisor Dr. Maria Klapa and my colleague Bhaskar Dutta. Specifically I would like to thank Dr. Klapa for her help in -shaping my thinking for the new post-genomic, non-hypothesis driven era of Quantitative Systems Biology, for constantly providing the motivation for the research, for installing the eye for details in my research and finally for challenging us to do our best work and not allowing us to compromise on any aspects of our research. I would like to thank Bhaskar for our many discussions in the lab, his suggestions and comments on my research and for being a constant companion on those long frustrating nights.

Finally 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 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'.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1. Introduction

Metabolic profiling refers to a high throughput measurement technique, which would allow simultaneous relative quantification of few hundreds of intracellular metabolites extracted from a biological sample. It is thus an extremely useful tool to probe the cellular physiology of a biological system at metabolic level. When combined with multivariate statistical methods, in plants, metabolic profiling can be used to distinguish between various ecotypes [Fiehn et. al., 2000] or mutants [Fiehn et. al., 2000, Roessner et. al., 2001a] and to identify environmental effects [Roessner et. al., 2001b]. These studies have shown that perturbation in one of the cellular levels can be correlated to changes at the metabolic level in plants, using metabolic profiling techniques. Such high throughput measurement and data analysis techniques allow much better understanding of response of the plant to a perturbation as compared to the traditional metabolic analysis techniques [Steuer et. al., 2003].

Comprehensive analysis of biological systems requires the integration of all cellular fingerprints: genome sequence, maps of gene and protein expression, metabolic output, and *in vivo* enzymatic activity [Klapa et. al., 2003]. Hence metabolic profiling can provides a comprehensive framework for measuring the metabolic fingerprints in such integrated analysis thus playing an important role in plants quantitative systems biology studies.

1.1 Motivation

Plant metabolism has been studied using traditional metabolic analysis techniques. However such techniques are limited to the measurement of particular group of metabolites. Thus such an analysis requires a a-priori hypothesis of which metabolites are likely to change in a particular study. The ability of metabolic profiling to quantify major metabolite categories (sugars, amino acids, lipids, alcohols, organic acids) allows an analysis with a much limited *a-priori* hypothesis.

Besides the obvious interest in studying the effect of the change in the ambient CO_2 concentration on the physiology of the plants in light of the global warming issue, the modification of the CO_2 levels in the environment of the plant cultures aimed at satisfying an additional need: to initially focus the holistic analysis of plant physiology in the central carbon metabolism and amino acid biosynthesis networks. Also there exists extensive information about their function both at the metabolic and genomic level. 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. *A. thaliana* was chosen for the current study, as *A. thaliana* has been used a model system for many studies in plant, and its full genome has been sequenced.

1.2 Major objectives and specific aims

In this context, the major objective of the thesis is to obtain the phenotypic fingerprint of each plant at the metabolic level using high throughput metabolic profiling to understand the short term response of plants to elevated CO_2 levels.

The specific aims to be pursued are following:

1. Conduct experiment with elevated CO_2 perturbation:

   A. thaliana plants (Columbia Strain) were grown for 12 days, in liquid media under constant light condition, for 12 days at 23 ℃ temperature. On the 13th day the control set was connected to a cylinder containing air at ambient concentration and the perturbed system was connected to air containing 1% CO_2. In both cases 10% of the air used was C^{13} labeled. Plants were harvested at 0.5 hr, 1 hr, 1.5 hr, 2 hr, 3 hr, 6 hr, 12 hr and 23 hr during the course of the 13th day. The plants were stored at -80 ℃ post harvesting and later the metabolites were extracted using methanol extraction protocol [Rosenner et. al., 2000]. (This part of the experiment was conducted by Dr. Maria Klapa, Dr. Tara van Toi, Lara Linford, Jeremy Matthew, Linda Moy and Dr. John Quackenbush at The Institute of Genomic Research, Rockville, MD. The extracted plant samples obtained from their work were used for the metabolic analysis discussed in this text)

2. Establish a Gas Chromatography-Mass Spectrometry protocol which will allow for relative quantification of the extracted metabolites.

3. Develop a systematic methodology for the quantitative analysis of time series metabolic profiling data which considers:
    - Data filtering / bias elimination
    - Data normalization
    - Multivariate statistical analysis
4. Discuss the obtained results in the context of the known *A. thaliana* physiology.
    - In the context of the acquired experience from the current analysis, modifications were suggested for future experiments

1.3 Thesis Description

Chapter 1: Provides an introduction and motivation for the current analysis along with the specific aims being followed for the current analysis.

Chapter 2: It provides a brief review of the CO_2 metabolism in plants. Past experiments of long term effects of elevated CO_2 on the plant physiology have also been reviewed.

Chapter 3: High throughput methods and instrumental platforms currently available for probing plant metabolism have been reviewed. Out of these methods, the protocol used for current analysis - metabolic profiling using GC-MS is described in detail.

Chapter 4: Discusses the role of multivariate statistics in interpreting data obtained from metabolic profiling by reviewing the experiments performed using metabolic profiling. A description of current statistical methods that can be used for time series metabolic data analysis is provided.

Chapter 5: Provides description of the current experiment, along with details about the metabolic profiling protocol established. The data filtration and normalization procedure developed for time series metabolic data is described and the results obtained from the analysis are presented.

Chapter 6: The results obtained using the multivariate statistical analysis are discussed in the context of the known effects of elevated CO_2 on plant physiology. Based on the current analysis, the advantages of using high throughput metabolic analysis and time series data are discussed by comparing the results obtained from current analysis with past experiments.

Chapter 7: Based on the current analysis the possible future work, which can further improve the understanding of effect of CO_2 on plant physiology is discussed.

Chapter 2. Plant response to elevated CO_2

2.1 Stoichiometry of CO_2 metabolism 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 due to elevated CO_2 levels in the environment, CO_2 metabolism in the plant has been extensively studied.

CO_2 fixation in plants is directly relates to photosynthesis. Photosynthesis is a complex process which uses light energy to transport and convert the CO_2 in the growth environment of the plant into organic compound needed for their growth [Buchanan et. al., 2001]. Depending on the mechanism by which CO_2 is fixed in plants, they are classified into three major categories known as C3, C4 and CAM plants [Buchanan et. al., 2001].

- ***C3 Plants:*** Most plants, including commercial crops like rice, wheat, cotton, belong to the C3 plant category. In C3 plants, the first stable compound produced from CO_2 fixation is 3-phosphoglycerate, a three carbon atom metabolite. The detailed C3 carbon fixation pathway is shown in Figure 2.1 [Buchanan et. al., 2001]
- ***C4 Plants:*** In the plants of this category (e.g. maize, sugarcane tropical grasses) malate (or aspartate) is the first metabolite produced as a result of

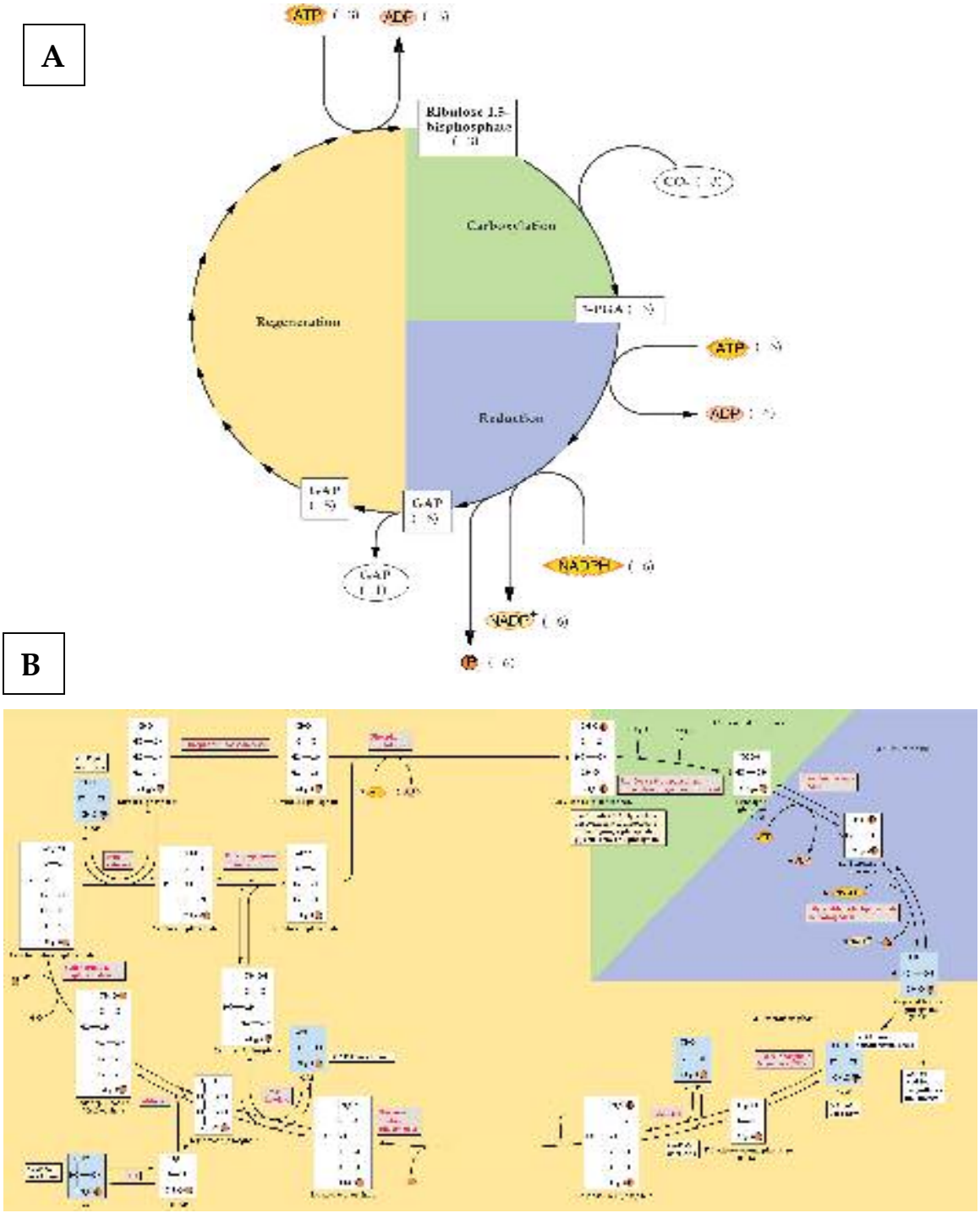


Figure 2.1: CO2 Fixation in C3 Plants (A) Three Stages of Carbon Fixation and important intermediates. The number in bracket indicates the stoichiometric coefficients of the overall reaction (B) Detailed Calvin cycle with stable intermediate metabolites. Copied From Buchanan et. al., 2001.

carbon fixation. Both these metabolites contain four carbon atoms each. These plants contain a two stage carbon fixation process as shown in Figure 2.2 (A) [Buchanan et. al., 2001].

- ***CAM Plants:*** CAM plants are usually encountered in extremely arid environments and usually belong to crassulacean species. Succulent plants such as cacti and pineapple are the characteristic examples of such plant. The first metabolite produced as a result of CO_2 fixation is malate (aspartate) like in C4 plants, but the two carboxylation processes are separated temporally rather than physically as shown in Figure 2.2(B).

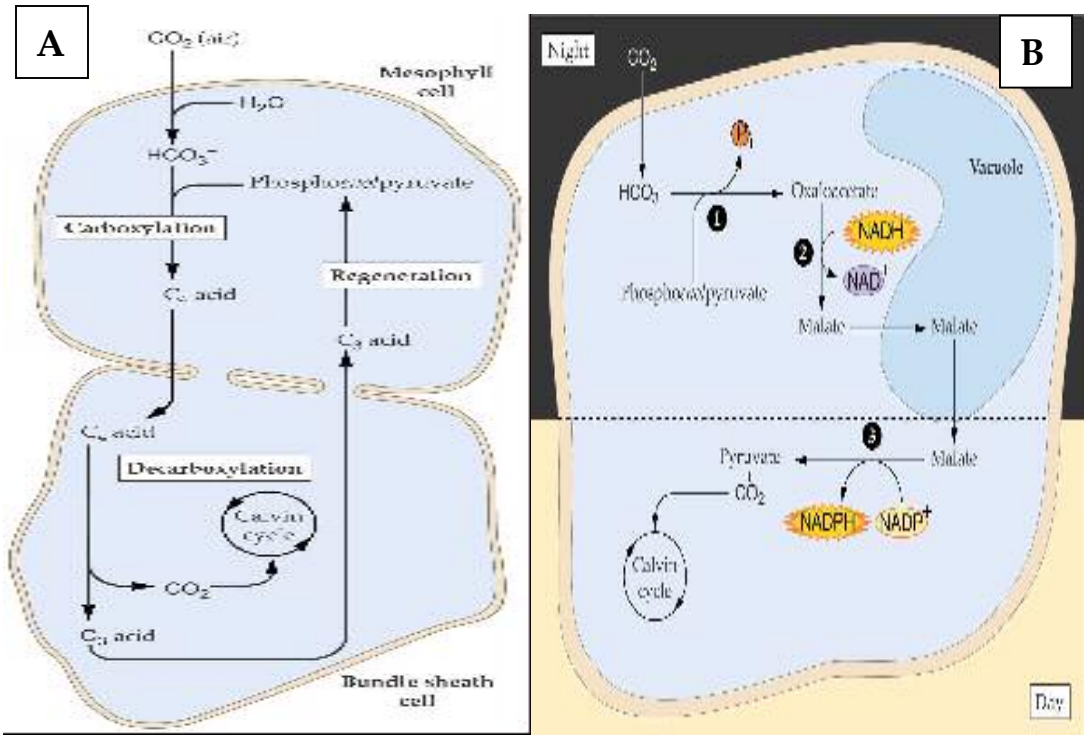


Figure 2.2 Two stage CO2 Fixation (A) C4 Plants (B) CAM Plant. Copied from Buchanan et. al., 2001.

A. thaliana plants used in the current study use C3 photosynthesis like most of the other plants. In C3 plants CO_2 fixation is part of the Calvin cycle and takes place in bundle sheath cells. As shown in Figure 2.1 (A) the Calvin cycle comprises three phases:

(a) ***Carboxylation:*** In carboxylation, CO_2 and water react with ribulose 1,5 bisphosphate (RuBP) to produce 3-phosphoglycerate (3-PGA). The enzyme that catalyzes the CO_2 fixation is ribulose bisphosphate carboxylase/ oxygenase (RuBisCO).

(b) ***Reduction:*** In the reduction phase, 3-PGA is converted to glyceraldehyde-3-phosphate (GAP) while 6 molecules of ATP and NADPH are consumed.

(c) ***Regeneration:*** In the regeneration phase, from the six molecules of GAP, five are used to regenerate three molecules of RuBP and complete the Calvin cycle through a series of reactions shown in Figure 2.1(B) with the simultaneous consumption of three molecules of ATP. The sixth molecule of GAP is subsequently used for anabolic needs of the plant.

In summary the stoichiometry of CO2 fixation in the C3 plants is as follows:

$3\ CO_2 + 3\ H2O + 6\ NADPH + 9\ ATP \rightarrow GAP + 9\ ADP + 6\ NADPH + 6\ Pi$

2.2 Effect of elevated CO_2 on C3 Plant Physiology:

Plant biomass typically consists of carbon, nitrogen, and ionic salts, which have been fixed by the plant over their lifetime [Buchanan et. al., 2001]. These primary nutrients are distributed in plants in different forms, each with specific roles in plant cellular processes. One of the main effects of elevated CO_2 in plants' growth environment is the increase in their growth rate despite significant limitation of other resources (like nitrogen) and environmental stresses [Idso et. al., 2001]. Increased growth rate translates into increased plant biomass for the same life period [Idso et. al., 2001; Paul, 2001; Grondzinski et. al., 1996], including the edible biomass [Idso et. al., 2001] which represents the grains of plant which are consumed by humans. The observed increase in plant biomass at high CO_2 fixation is not uniform for all its constituents. In the following paragraphs this difference will be explained.

Carbohydrates:

In plants, carbohydrates are used for energy storage, biosynthesis, and for various structural roles [Buchanan et. al., 2001]. Carbohydrates are typically produced in the form of starch, sucrose & polysaccharides, starting from the GAP produced in the Calvin cycle. Sucrose which is produced in the leaf cells as a result of photosynthesis, is then transported to other parts of the plants that need carbohydrates. Starch —a polymer of the glucose molecule— is used for

carbohydrate storage in the leaf whenever the sucrose production from photosynthesis exceeds the capacity of the leaf to export it to other parts of plants [Buchanan et. al., 2001]. Polysaccharides, which are polymers made by the combination of two or more sugars, are the primary constituent of the cell wall. During growth, cell wall uses a large part of the plant biomass thus being a major drain on carbon supply [Buchanan et. al., 2001]. Some common constituents of cell walls are cellulose (principal scaffolding component of plant cell wall), cross linking glycans, like xyloglucan and glucuronoarabinoxylans (also known as hemicelluloses, used for cross linking cellulose microfibrils) and pectin (which perform many specific functions like determining wall porosity, pH modulation, cell adhesion, etc.)[Buchanan et. al., 2001].

It has been observed that during the first few days of plant growth under conditions of elevated CO_2 (typically 2 to 3 fold increase with respect to the ambient CO_2 level) the rate of photosynthesis increases and leads to accumulation of non-structural carbohydrates like sucrose and starch in the leaf [Hui et. al., 2001; Paul et. al., 2001]. On average the soluble sugars and starch content increase by 52% and 160%, respectively [Paul et. al., 2001]. This increase results in the feedback inhibition of RubisCO activity, allowing the plant to transfer resources that are being used for photosynthetic activity to other cellular processes [Paul et. al., 2001]. Additionally, enzymes involved in carbohydrate

synthesis, like hexokinase, are also known to regulate photosynthesis [Paul et. al., 2001], however the exact mechanism is not yet fully understood.

***Lipids**:*

Lipids — defined as plant metabolites which are soluble in non-aqueous solvents like chloroform - mainly include fatty acid derived compounds [Buchanan et. al., 2001]. It is speculated that ~200 different fatty acid derivatives can be present in the plants [Buchanan et. al., 2001]. Lipids are produced from Acetyl-CoA, formed from the breakdown of the carbohydrates produced from photosynthesis. Characteristic examples of lipids are the glycerolipids (structural component in the cell membrane), triacylglycerols (storage compounds), and waxes (storage and plant protection compounds). Other lipids have more specific plant functions like plant defense, signaling, electron transport and photoprotection. Fatty acid composition is measured in few photosynthetic related studies. It has been observed that elevated CO_2 did alter the composition of glycerolipids in *Chlorella kessleri* (green algae) [Sato et. al., 2003] and increased cell division in potatoes [Chen et. al., 2001].

Proteins:

Proteins are macromolecules produced from amino acids and constitute the catalysts and regulators of almost all major processes in plants. Total protein content measured using enzymatic assays reduced relatively, in wheat canopies

grown under elevated CO_2 conditions in the presence of $NO3^-$ & $NH4^+$. This decrease however was due to a relatively larger increase in other constituents of biomass like carbohydrates. In absolute term there was an increase in total protein content of the plant. The absolute increase in protein content in wheat canopies grown with $NH4^+$ nutrients was larger than in those grown in $NO3^-$, there by indicating that the increase of protein content is dependent on the source of nitrogen used for plant.

Other Constituents:

Apart from the three most abundant constituents of plant biomass, biomass also comprises minerals (ionic nutrients) and primary/secondary metabolite pools, which play an important role in specific plant functions. Some of the most abundant secondary metabolites are lignins, which play an important structural role in vascular & woody plants[Buchanan et. al., 2001]. The composition of plant biomass in each of these constituents is affected differently by the CO_2 increase in the plant environment which has been reviewed by Idso et. al. [2001], and Grodzinski et. al.[1996]

2.3 Effect of elevated CO_2 on nitrogen assimilation:

In plants most of the nitrogen uptake occurs in roots and then nitrogen is transported to other parts of the plants in the form of nitrate ions. Plants can use

three nitrogen substrates, NO_3^-, NH_4^+ and atmospheric nitrogen as shown in Figure 2.3(A).

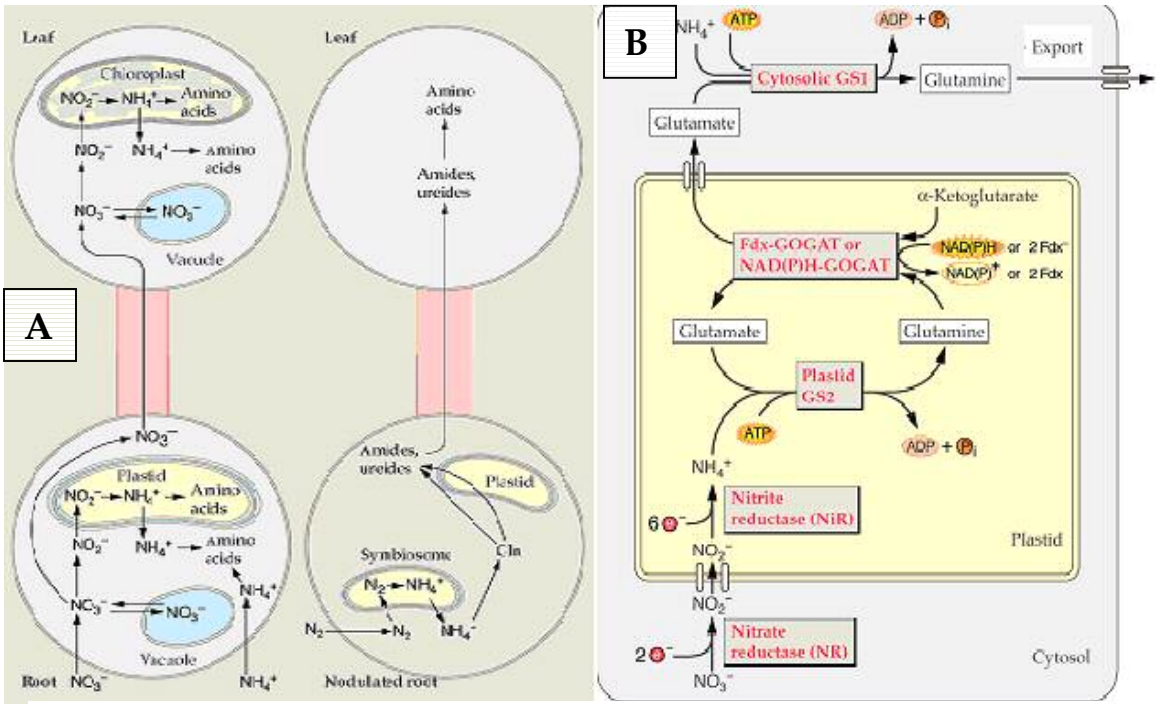


Figure 2.3 (A) Stoichiometry of the primary nitrogen assimilation Mechanisms (B) Enzymes involved in nitrogen assimilation (Buchanan et. al., 2001).

N_2: The assimilation of atmospheric nitrogen in plants takes place only in the presence of symbiotic bacteria.

NO_3^-: NO_3^- present in the soil is converted to NH_4^+ through the formation of nitrite ion. The conversion from NO_3^- to NH_4^+ ion takes place in the plastid in the roots, and in the chloroplasts in leaves as shown in Figure 2.3(A).

NH_4^+: NH_4^+ is assimilated into the N-transport amino acids: glutamate, glutamine, aspartate and asparagine [Buchanan et. al., 2001]. From these nitrogen stores, mainly aspartate and glutamate donate nitrogen for cellular reactions requiring

nitrogen. Carbon and nitrogen availability dictates whether plants choose glutamine or asparagine for nitrogen storage. In dark, adapted plants, due to the lower rate of carbon fixation, carbon availability is low. Under these circumstances, asparagine levels increase dramatically with a simultaneous decrease in glutamine levels, because asparagine stores more nitrogen per carbon atom as compared to glutamine.

The effect of elevated CO_2 on the plant nitrogen content holds extreme significance. While most of the studies which were carried out in controlled (closed) lab or greenhouse environment indicated reduction in the nitrogen content, the experiments being carried out in the field showed the opposite [Idso et. al., 2001]. This "contradiction" can be easily explained if someone takes into consideration that under elevated CO_2 conditions plants redistribute biomass to roots and legumes responsible for symbiotic nitrogen fixation [Idso et. al., one more]. In the field experiments, involving the long term effect of elevated CO_2 (few days to months), plants increase their roots which leads to increase in the nitrogen uptake and compensates for the nitrogen stress that is induced in response to elevated CO_2.

Even though it was argued that a decrease in nitrogen content observed in the lab experiments could be merely a result of dilution effect of increasing biomass, the decrease measured in plant/leaf nitrogen content was much more as

compared to what could be accounted due to increase in biomass. More detailed analysis by researchers [Smart et. al., 1998; Bloom et. al., 2002] indicates that there is a reduction in primary nitrogen assimilation in response to elevated CO2 experiment.

To understand the reduction in plant nitrogen content, Smart et. al. (1998) measured the nitrogen balance in response to elevated CO_2 (1000 ppm) in wheat grown under controlled environment using solution cultured techniques. The plants were also grown under two different NO_3^- (100 and 1000 mM) concentrations in the root zones. They observed:

- Increase in total plant biomass
- Increase in the plant biomass allocation to roots but did not observe an increase in nitrogen uptake per unit area of the root
- A slight increase in NO_3^- uptake in plants
- Reduction in biomass organic nitrogen content which was much more than what could be accounted by the increase in the plant biomass.

Based on the observation of a decrease in plant organic nitrogen content indicating lower nitrogen assimilation, in spite of observing a slight increase in NO_3^- uptake by root, they concluded, that elevated CO_2 conditions interfere NO_3^- assimilation into organic nitrogen.

To further understand the mechanism by which NO_3^- assimilation is hindered by presence of elevated CO_2, Bloom et. al. (2002) carried out a second experiment in which they grew wheat canopies under elevated CO_2 (700 ppm) and ambient conditions with two different nitrogen source, NO_3^- or NH_4^+ . This time they measured:

- O_2 and CO_2 uptake rates in the plants throughout the experiment
- Total nitrogen and protein content of the plants, two weeks after the beginning of the experiment
- The nitrite (NO_2^-) absorption of extracted wheat chloroplast under 0, 0.3, 1 or 3 mM HCO_3^- under *in-vitro* conditions

Observations on NO_3^- to NH_4^+ conversion:

The change in rate of conversion of NO_3^- to NO_2^- and further to NH_4^+ was monitored by measuring the change in O_2 liberated in the process. The analysis of the result obtained indicated that in presence of elevated CO_2 in plant leaves, the transformation of NO_3^- to NH_4^+ (in presence of light) was reduced in short time periods (few hours) as well as long time periods (few days).

The inhibition of NO_3^- fixation in presence of elevated CO_2, could be due to:

- Limitation of NADH (provided by the common malate shuttle in chloroplast and cytoplasm) required for conversion of NO_3^- to NO_2^- in presence of higher photosynthetic rate.

- Limitation of e^- (provided by ferrodoxin) required for the conversion of NO_2^- to NO_3^- which also competes with ferrodoxin requirement during photosynthesis

Since the NO_3^- can be easily stored in plants for a longer time as compared to CO_2, plants utilize the NADH (reducing power) for photosynthesis in preference for NO_3^- assimilation. The combined effect of NADH and ferrodoxin limitation is that under the elevated CO_2 condition the NO_3^- conversion to NH_4^+ is inhibited in presence of higher carbon fixation rate.

Observations on HNO_2 transport:

As part of the primary non-symbiotic nitrogen assimilation process, as shown in Figure 2.3(A,B), HNO_2 is transported from the cytoplasm to the chloroplast in leaves. The study of the effect on NO_2^- transport across chloroplast membrane for different HCO_3^- ion concentration, indicated a decrease in NO_2^- transport with higher HCO_3^- ion concentration inside the chloroplast. This is another way by which the primary nitrogen assimilation using NO_3^- as a substrate is hindered.

Observations on Plant Nitrogen Content:

These observations were confirmed by comparing plant growth rates and protein content in plants grown in NO_3^- and NH_4^+ as the nitrogen source. The comparison between the two showed:

- Plants grown in NH_4^+ and NO_3^- under normal CO_2 levels showed the same biomass content
- In presence of elevated CO_2, the NO_3^- grown plants showed 44% increase in biomass and 24% increase in leaf area as compared to 78% and 49% increase respectively for NH_4^+ grown plants.
- The absolute protein content (after accounting for increase in biomass content) in the shoot protein increased 73% and 32% under NH_4^+ and NO_3^- respectively.

These measurements support the possible hypothesis for the mechanism by which primary nitrogen assimilation is reduced in plants in response to elevated CO2.

2.4 Elevated CO2 effect on Photorespiration:

The carbon fixation in Calvin cycle takes place in presence of RuBisco enzyme. As indicated by the name, RuBisCO gene fixes CO_2 and also catalyzes oxygen fixation. Both CO_2 and O_2 have the same binding site in RuBisCO and hence are competitive substrates. In the oxygenation process, one molecule of RuBP gets converted to one molecule of 3-PGA and one molecule of Phosphoglycolate containing two carbon atoms. Phosphoglycolate then undergoes a series of conversions, which are collectively called as "Photorespiration" [Dey et.al., 1996]. In presence of light, as the temperatures increase, the ratio of carboxylation to

oxygenation shifts in favor of oxygenation due to change in the relative solubility's of CO_2 and O_2, thus reducing carbon fixation efficiency. In order to avoid this condition, in more temperate zone plants use C-4 mechanism using a special leaf anatomy which allows separation of RuBisCo from oxygen. Previous studies measuring the flux control coefficient have confirmed this mechanism [Dey et. al., 1996].

Due to the competition between CO_2 and O_2, at elevated CO_2 levels the photorespiration pathway gets inhibited [Buchanan et. al., 2001]. Hence a photorespiration mutant plant, deficient in secondary nitrogen assimilation associated with photorespiration, does not survive at normal CO2 levels. However when they are grown at 1% CO2 level, due to inhibition of photorespiration, they show a normal growth cycle. Thus 1% CO2 inhibits photorespiration [Buchanan et. al., 2001]

2.5 Role of the experimental design in uncovering the effect of elevated CO_2 on plant physiology:

The review of the literature related to the effect of elevated CO_2 on plant physiology suggests that the experimental design, setup and conditions play an important role in the conclusions that can be derived from the experiment. For the results to be comparable all experimental parameters need to be taken into consideration including mainly, the plant, the time duration of the experiment,

sampling frequency, growth media, light and humidity conditions, the composition of nutrients. In addition the analytical method chosen decides how many and which effects can be measured from the experiment.

2.5.1 Plant Species:

As discussed in the earlier section, plants use three different mechanisms to fix CO_2. The choice of the model plant system should be made in a way that it represents the general class of plants in which the results need to be applied. The choice of plant should be based on:

- Since most of the plants including major important crops are C3 plants, studies using C3 plants can be more useful.
- The plant should also have a short growth cycle since that can significantly affect the project cost and time.
- Maximum information about the biochemical pathways and gene structure should be available.

Arabidopsis thaliana, which is a C-3 plant, has been used extensively to study the interaction between photosynthesis and photorespiration. Recent sequencing of the full A. thaliana genome allows a much better understanding of the plant response at gene expression level. Mutants in *A. thaliana* having defective photorespiration, nitrogen assimilation, stomatal density, insensitivity to light and many more have already been identified and are easily available

commercially, allowing their use in experiment design. Also its short growth cycle makes it an ideal model system to study the effect of elevated CO_2 on C3 plant physiology.

2.5.2 Time Duration:

Most of the studies available in literature refer to duration of the CO_2 "treatment" studies from over few days to up to 10 years [Idso et. al.,2001]. The duration of "treatment" before any measurements are made is important because plants with long life cycles can acclimatize to the elevated CO_2 level with no effect observed in plant physiology. Such a study may be important from the point of view of understanding the effect of elevated CO_2 in the global environment for commercial products, but may not offer a better understanding of the interaction of carbon fixation with other cellular processes.

2.5.3 CO2 level:

The choice of CO_2 level could affect some conclusions drawn from the experiments. In most previous metabolic studies, plants have been treated with 700 & 1000 ppm CO_2 levels. For example, when plants are treated in presence of 700 ppm CO_2 levels, the total secondary plant metabolites show an increase as compared to ambient level in high nutrient environment, but show a decrease in low nutrient environment. However, at 1000 ppm level, the plants show an increase in total secondary metabolites independent of the nutrient level. Thus,

the amount of CO_2 used in the elevated CO_2 atmosphere may also affect the conclusion derived from an experiment.

Apart from the amount of CO_2 used for the experiment, the way the perturbation is applied on the plant, affects the conclusions of the experiment. A gradual increase in the CO_2 levels allows the plant to acclimatize, and hence show a different effect as compared to the one in which sudden changes are made in the CO_2 levels [Hui et. al., 2001].Thus the type of perturbation would depend on the objective of the study.

2.5.4 Nutrient Condition:

Contradictory results have been obtained due to difference in the amount of nutrients, and type of nutrients used for the analysis. As discussed in earlier part of this report, different results about the plant nitrogen content were observed depending on amount of nutrient available to plant [Idso et. al., 2001] and even depending on the type of nitrogen source that was used (NH_4^+ or NO_3^-) [Bloom et. al., 2002]. Hence any results obtained should be discussed keeping in mind the effect of nutrients on the plant response to elevated CO_2 discussed.

2.5.5 Analytical Method:

Most of the experiments related to response of elevated CO_2 focused on measuring carbohydrate accumulation, effect on nitrogen assimilation, secondary carbon metabolites, protein synthesis, and biomass redistribution. For

the analysis of the plant sample typically specific chemical, enzymatic, HPLC & GC based methods suitable for measuring an individual or a class of metabolites was used. A more focused study, which investigated the mechanism of regulation of nitrogen assimilation by elevated CO_2, used very specific instruments to measure CO_2 consumption and oxygen evolution rate [Bloom et. al., 2001]. Labeled carbon and nitrogen substrates were also used to understand the biomass redistribution in plants. In all cases the analytical methods were chosen based upon the experiment hypothesis, which predicted a change in a particular class of the compound.

The current review of literature presented shows 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. Current analytical methods do not allow us to measure the change in all these metabolites simultaneously. Hence a high throughput method which can simultaneously measure changes in metabolites belonging to different classes will allow:

1. Understanding the effect of elevated CO_2 on the metabolism of various plant sub systems.
2. The extent of response of different plant sub systems in response to elevated CO_2.

3. Understanding the interaction between different sub systems of the plants

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 current methods which are focused more in a particular class of compounds. Also since the high throughput method would strive to measure all the metabolites, no prior hypothesis of which class of compounds are expected to change in response to a particular experiment is needed. This allows us to do a less constrained hypothesis based analysis Analytical methods that can be used for such an analysis are discussed in the next chapter.

Chapter 3. Metabolic Profiling of Plants

Metabolic Profiling refers to the high throughput methodology that allows for the simultaneous detection and quantification of low molecular weight metabolites (belonging to different functional categories) that are derived from cellular breakdown of a biological sample. Metabolic profiling analysis of plants comprises each of the following steps:

- **Extraction:** In the extraction process, the cell wall and cell membrane are broken and the intra cellular metabolites are extracted in solvents. The presence of methanol and the heat treatment at 70°C ensures that the proteins are deactivated and does not affect the metabolites.
- **Detection and Quantification:** The small molecular weight metabolites extracted from the plant sample are detected and quantified using various analytical instruments.

In past various extraction methods have been used for analyzing specific group of metabolites which has been reviewed by Katona et. al. (1999). Recently Katona et. al. (1999) and Roessner et. al. (2000) used methanol extraction to extract polar metabolites from apricots and potato tuber samples respectively. Later Fiehn et. al. (2000) extended the protocol to simultaneously extract both the polar and non-polar metabolites. For the current analysis discussed in this report, the methanol

extraction protocol [Roessner et. al., 2000] for extraction of polar metabolites was used. The same is also available in Appendix I for ready reference.

3.1 Analytical Techniques for Metabolic Profiling:

The analytical technique which allows for detection and quantification of mixtures of low molecular weight organic and organo-metalic compounds can be used for measurement of the metabolic profiles. These methods are following:

1. Gas Chromatography –Mass Spectrometry (GC-MS)
2. Liquid Chromatography – Mass Spectrometry (LC-MS)
3. Nuclear Magnetic Resonance (NMR)

The concepts, advantages and limitations of these methods are discussed below.

GC-MS: GC-MS system consists of two parts. In the gas chromatography part the compounds in the sample mixture are separated using chromatography in the gas phase. After separation the compounds enter a mass spectrometer where they are identified and quantified.

LC-MS: LC-MS platform uses an approach very similar to that of GC-MS discussed above except the chromatographic separation takes place in the liquid phase instead of gas phase. The mass spectrometer used can be identical to GC-MS.

NMR: Unlike the mass spectrometry approach, NMR uses the magnetic properties of certain isotopes for metabolic profiling.

Typically in metabolic profiling using GC-MS, in order to make the metabolites volatile they need to be derivatized which is not needed in LC-MS as the separation takes place in liquid phase. Also some of the metabolites like di-phosphate derivatives of sugars cannot be vaporized even after derivatization, hence can not be quantified using GC-MS however can be quantified using LC-MS. In spite of their advantages, due to higher investment and instrumental cost associated with LC-MS, it has not been as widely used as GC-MS systems. So LC-MS use has been mainly limited to analyze selected metabolites [Kopka et. al., 2004] in plants, however it has potential for metabolite profiling in combination with GC-MS technique.

The advantage of using NMR over GC-MS technique is that NMR can also measure the metabolite activity in vivo, and due to its non destructive nature, the same plant can be used for measuring metabolic state at different times [Ratcliff et. al., 2001]. NMR technique can also be used to identify structure of the metabolite unknown metabolites, but its biggest disadvantage is that only the most abundant metabolites (about 50) can be quantified simultaneously using plants [Kopka et. al., 2004]. In the current analysis the most widely used technique for plant metabolic profiling is GC-MS, and details of the same are discussed in the next section.

3.2 GC-MS protocol for metabolic profiling of plants:

GC-MS has been used for measuring concentrations of various classes of metabolites in plant derived samples, including sugars, amino acids, fatty acids etc., for a long time. For each class of compound, a separate method was being used. Katona et. al. (1999) created a protocol which could simultaneously measure sugars, alcohols, amino acids in apricots. Roessner et. al. (2000) conducted rigorous systematic analysis for optimizing various parameters of the protocol by estimating and quantifying the sources of variations in the protocol using potato tuber samples. Fiehn et. al. (2000) extended the protocol to also include non-polar metabolites, measuring concentration of metabolites belonging to fatty acids and fatty acid alcohols. Using this protocol they could detect 326 metabolites (214 polar, 112 non-polar) in *A. thaliana* leaf extract. For current analysis protocol developed by Roessner et. al. (2000) to measure polar metabolites was used.

3.2.1 Derivatization of Metabolites:

As per the protocol, the low molecular weight polar metabolites obtained from methanol extraction are used for GC-MS analysis. In order to use these metabolites for GC-MS analysis, the metabolites need to be derivatized. As per the derivatization protocol developed by Roessner et. al. (2000), metabolites containing a ketone or aldehyde groups are transformed into methoxime group.

Similarly active hydrogen atoms (-H) present in the metabolites (in –NH2, -COOH, -OH groups) are replaced by trimethylsilyl (TMS) groups. The detailed derivatization protocol is available in Appendix I.

3.2.2 Separation of Metabolites:

In metabolic profiling using GC-MS, separation of derivatized metabolites takes place using a gas chromatographic process which is schematically represented in Figure 3.1. The process of separation of metabolites is as follows:

- The liquid sample containing derivatized metabolite is injected onto a heated injector where the sample is vaporized
- The derivatized metabolites in gas phase enter a glass capillary column coated with a thin layer of stationary phase, where the metabolites are chromatographically separated.

The separation of the compound is achieved by using the property that the amount of time a compound takes to travel through the column depends on its structure, charge and molecular weight when the chromatographic conditions are held constant. This characteristic time (under given chromatographic conditions) of each derivatized metabolite is called its retention time.

3.2.3 Generating a Mass Spectrum of Metabolite:

The derivatized metabolite separated using gas chromatography enters the mass spectrometer through a heated transfer line. In the current analysis ion trap mass

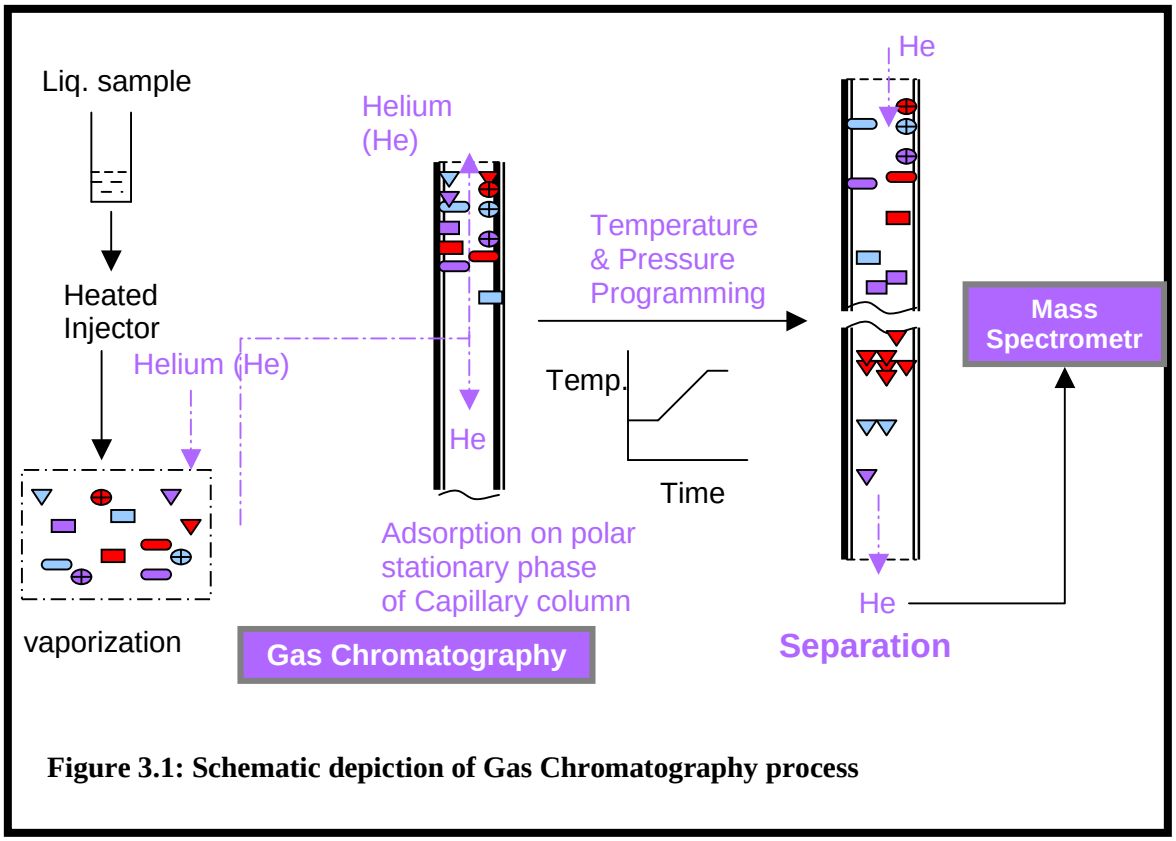


Figure 3.1: Schematic depiction of Gas Chromatography process

spectrometer was used and the working of the same has been shown in Figure 3.2. The mass spectrum of the metabolite is generated through following process:

- The derivatized metabolite enters the mass spectrometer where they are bombarded with electrons (in case of electron ionization) or small gaseous molecules like methane (in case of chemical ionization).
- The electron bombardment breaks down the molecule into smaller fragments, and also ionizes them (due to loss of hydrogen ion).
- The ions are now subjected to an electromagnetic field, and depending on their mass/charge ratio, they follow a certain trajectory in the ion trap of the mass spectrometer.

- The mass spectrometer detects in each scan (a particular time instant when measurement is made) the intensity of ions for a specific range (typically 50-600 m/z) and generates a spectrum (intensity vs. m/z plot) for the particular scan.

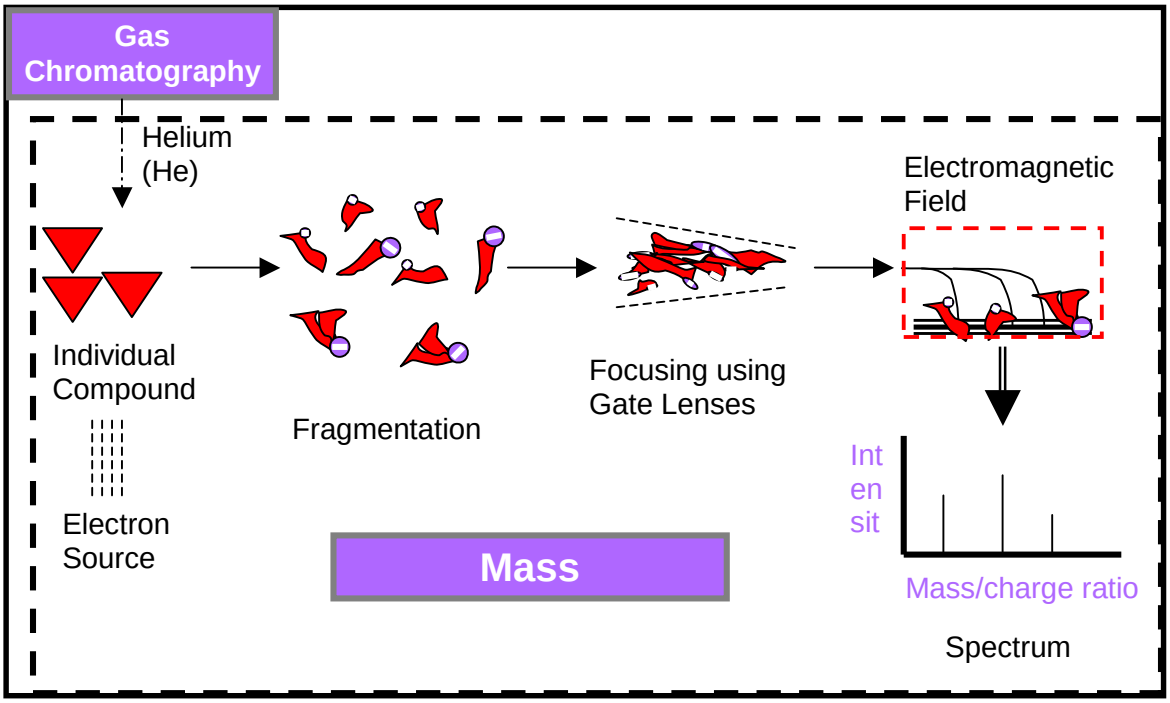


Figure 3.2: Schematic view of Mass Spectrometry

For a specific metabolite, depending on its structure, at a given ionization intensity (measured in electron volts), the fragmentation pattern always remains the same, and is characteristic for the particular molecular structure. The intensity recorded depends upon number of ions present, which in turn, depends upon the number of molecules entering the mass spectrometer. Hence the intensity recorded allows quantification of the compound entering the mass spectrometer. The retention time of the compound and its mass spectrum,

together, allow us to identify the compound. Thus, GC-MS technique allows simultaneous measurement and quantification of compounds. The details about the identification and quantification of the metabolites specifically for metabolic profiling of plants are discussed in the later part of this text.

3.3 Metabolite identification and quantification using GC-MS:

As discussed in the earlier section, the mass spectrometer measures the intensity ions having m/z value in 50-600 range, for each scan which generates a spectrum, as shown in Figure 3.3(A) which is the plot of intensity vs. m/z values recorded in a scan at time 21.91 min. The intensities recorded for each m/z ion change at each scan (i.e. with time) as the compounds entering the mass spectrometer keeps changing. Thus when the spectrum recorded at each scan is combined together it generates three dimensional data (intensity, m/z and time) as shown in Figure 3.3(B). Now by combining the intensity recorded at each m/z value, i.e. combining the intensity recorded for all the ions, at a particular time point, a 2-D plot of total intensity recorded vs. time is generated as shown in Figure 3.3(C). Such a plot represents a two dimensional projection of the 3-D data recorded. This intensity vs. time data is integrated in order to obtain the chromatogram of the sample and to calculate Total Ion intensity peak (TIC peak) area as shown in Figure 3.3 (D). Thus, the TIC peak area represents the total intensity of all the

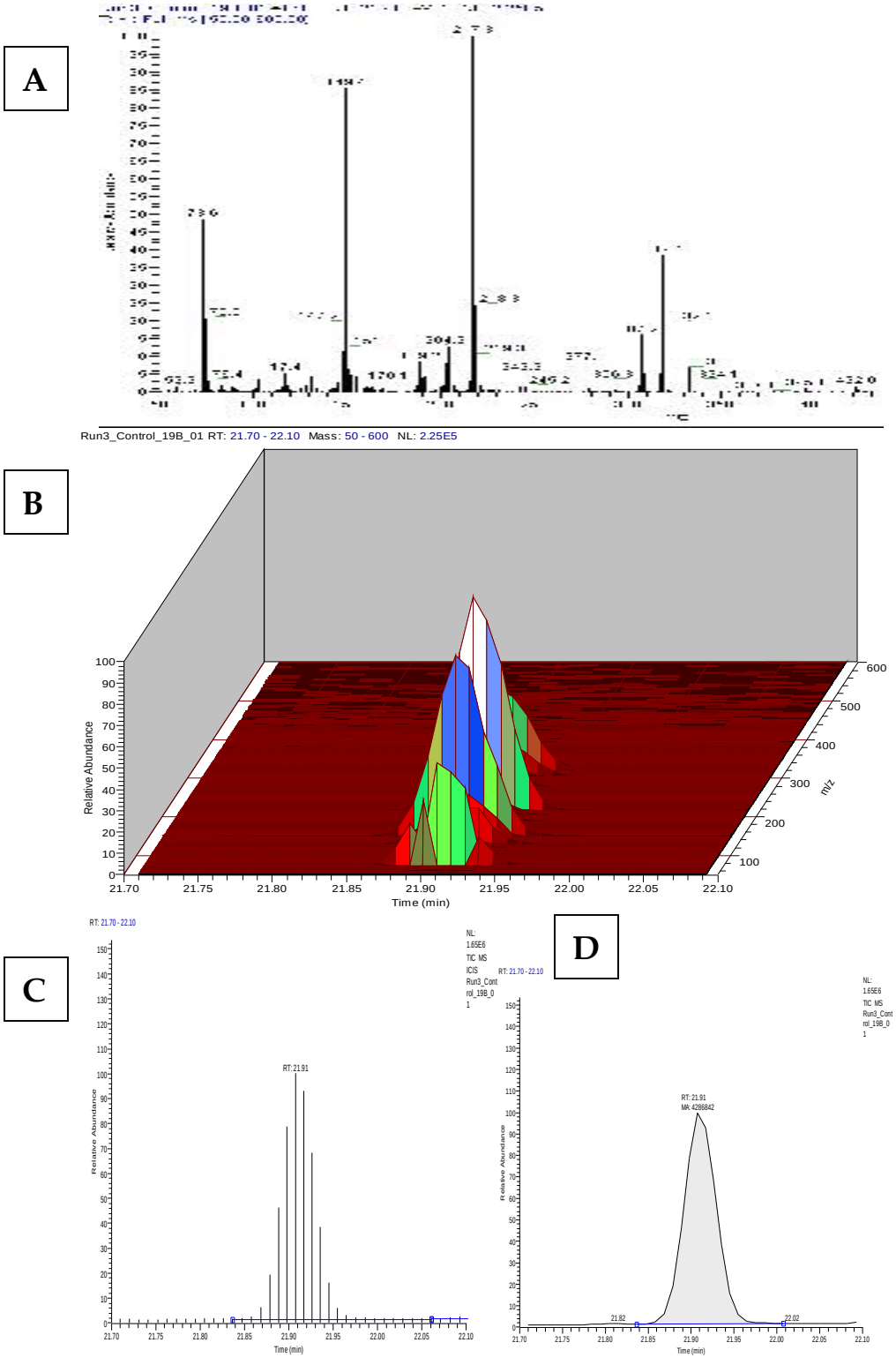


Figure 3.3: (A) Total Intensity Chromatogram (B) 3-D Intensity Map (C) Discrete Intensity measurement (D) Integrated Intensity Peak Area

ions generated, across all scans, and is thus a measure of the concentration of the derivatized metabolite in the injected sample and hence used for quantification of the metabolite. The time at which the highest intensity is recorded, which corresponds to the time at which most of the metabolite elutes, is called the retention time for the metabolite, which remains constant under similar chromatographic conditions.

Identification of an individual metabolite:

Typically, for most GC-MS applications the mass spectrum of a compound is sufficient for its identification. However for metabolic profiling of plant samples, many metabolites are isomers and show very similar mass spectra. Hence, the combination of retention time of a metabolite (under a given chromatographic conditions) and its mass spectrum which is a unique combination for each metabolite is used for identification of the metabolite. For example, the TIC peak shown in Figure 3.3(D) represents TMS derivatized ribitol which has retention time around 21.9 min, and has a mass spectrum which is shown in Figure 3.3(A) matches the standard mass spectra for ribitol TMS derivative available in the commercial NIST Mass spectral library and the Max Planck library on the web (Fiehn et. al., 2000a). Two other compounds, xylitol and arabinose also show a mass spectrum similar to that of ribitol, due to similarities in their structure. However the combination of retention time and spectrum will be unique for

ribitol and will remain the same in all the plant samples as long as the GC-MS conditions are held constant.

Identification of co-elution:

In the case of co-elution of two or more metabolites, the TIC peak at a particular time accounts for the TIC of all co-eluted metabolites. The presence of co-elution can be identified through the variation of the mass spectrum within a given peak. Since the fragmentation pattern of the mass spectrum is unique for a particular metabolite, and it should remain the same throughout all the scans of the peak, any variation in the same is an indication of co-elution of two or more metabolites within the peak. The differences between the fragmentation patterns allows also for the separate identification and quantification of the metabolites.

This can also be seen from the 3-D plot of intensity, m/z and time shown in Figure 3.4. In the case of the 3-D plot for the ribitol peak, all the individual ion fragment intensities have their peak at the same time which corresponds to the retention time of the peak in the chromatogram (21.91 min). However, in case of the co-eluting peak, there is a slight offset in the highest intensity recorded, for some individual ions, which indicates co-elution.

Peak de-convolution of co-eluting metabolites:

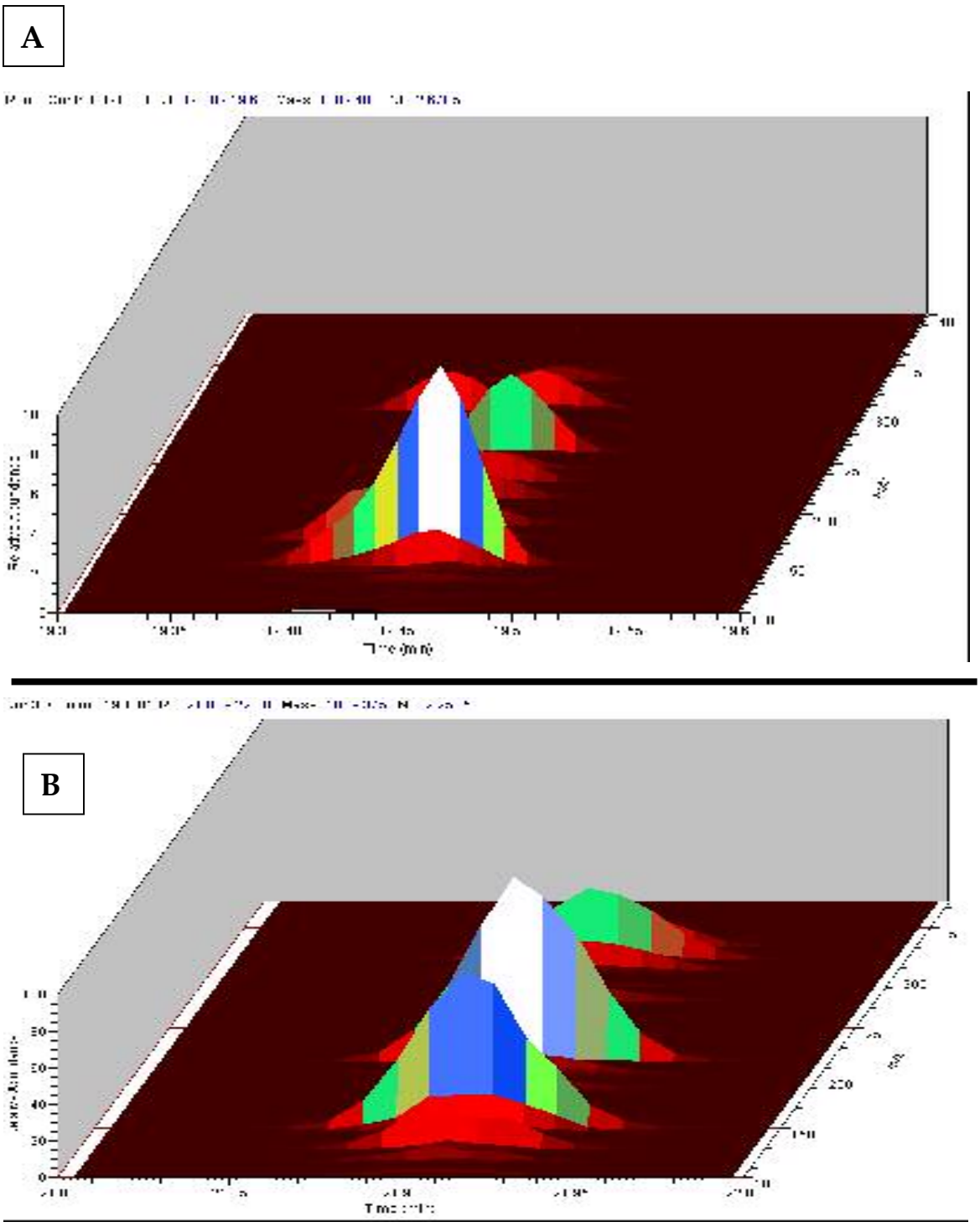


Figure 3.4: 3-D View of (A) a co-eluting aspartate & asparagine peak (B) ribitol peak

Peak de-convolution is an important process in metabolic profiling analysis because 50-70% of the measured metabolites in a plant sample are expected to co-elute. It can be performed either manually of automatically through the use of an appropriate software. Even though the manual approach is more laborious, it is however more reliable than the automatic one. The advantage of using the automatic peak de-convolution approach is that it would reduce the time for detecting all the metabolites considerably, and it can also identify co-elution in which one of the metabolite has very low concentration, which is difficult to identify using visual inspection. The disadvantage of using this approach is that it has not been developed for biological samples, and hence no guidelines exist for various criteria, required by the algorithm, to distinguish between variations in intensity due to presence of a metabolite than that of noise. Hence depending on choice of these parameters for the same chromatogram, the number of metabolites detected could vary from 200 to 600.

Metabolite Quantification :

As discussed in the previous section, TIC is proportional to the amount of the derivatized metabolite run through GC. Just as for a given chromatographic conditions, the retention time of a metabolite remains constant, for a given mass spectrometric condition the mass spectrum also remains constant. This means that the ratio of different intensities recorded in a mass spectrum of the

metabolite remains constant for all the plant samples. This ratio depends upon the fragmentation pattern of the metabolite in the mass spectrometer, on electron bombardment. Thus, to quantify the co-eluting metabolites, marker ions are used in place of total ion intensity. Since there are many co-eluting peaks in chromatogram for the plant sample, for the sake of uniformity, marker ions are used for all the metabolites. Figure 3.5 (A) shows the peak areas for total ion intensity and for each individual marker ion (B)-(E). Such marker ion intensity plots are not the projection of the 3-D plot now (which is the total intensity peak area) shown in Figure 3.4 (B), but instead, a slice of the 3-D plot, at each individual m/z value.

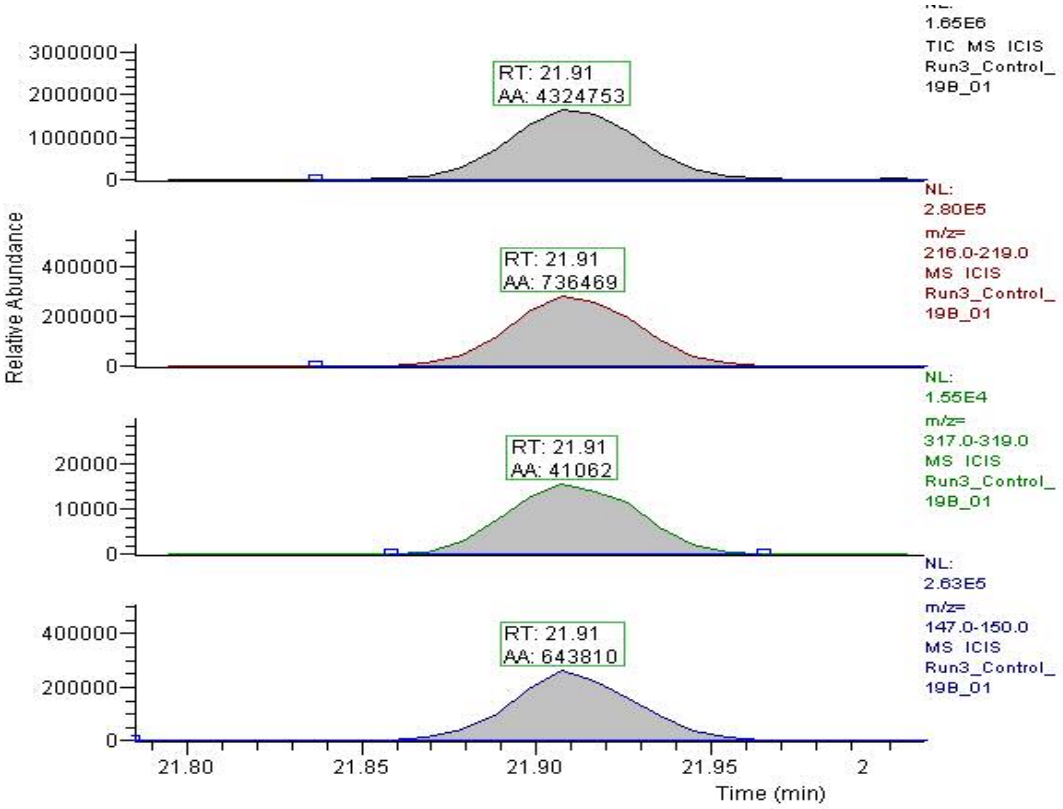


Figure 3.5 Ribitol (a) Total Ion intensity plot (b) 217 m/z Intensity plot (c) 317 m/z intensity plot (c) 147 m/z intensity plot. The ratio of their peak area remains constant for ribitol in all plant samples

3.4 Internal standard Normalization:

The peak area of characteristic m/z used to quantify a particular metabolite is proportional to the amount of this metabolite run through the GC, however, the aim of metabolic profiling is to measure the metabolite concentration in the plant. Since the plant sample goes through many stages of processing before entering the mass spectrometer and variations between different samples exists, the quantification of the metabolite would be affected. In order to facilitate data normalization and to remove the effect of these variations, an internal standard is used for GC-MS Analysis.

The internal standard is chosen based on following conditions:

- The internal standard used should not be produced by the plant
- The internal standard should be representative of the metabolites being measured – so that it undergoes the same variations as other metabolites in the plant sample

The following are possible sources of variation which are accounted using internal standard:

- variations caused during extraction
- variation in the extent of derivatization
- variation in the quantity of sample being injected into the GC-MS.
- variation in GC-MS sensitivity/ ionization efficiency.

The principle behind internal standard normalization is that, since a known quantity of internal standard is added to each plant sample, in absence of any experimental error, internal standard should have the same peak area for all the plant samples. Thus any variation caused in the internal standard peak area also represents the variation the other metabolites in plant sample have undergone. Thus by normalizing the characteristic m/z peak area for each metabolite with the internal standard peak area we obtain relative peak area for each metabolite which is representative of the relative concentration of the metabolites in the plant which can be used to compare the concentration of the metabolites in different plant samples.

Since metabolic profiling quantifies few hundred metabolite samples, in order to compare them across all samples in a non-biased high throughput manner, multivariate statistical methods are required, which are discussed in the next chapter.

Chapter 4.

Multivariate statistical techniques for metabolic data analysis

Metabolic profiling using GC-MS is a powerful tool to probe plant metabolism. GC-MS has been used for a long time to measure individual metabolites in biological samples. Recently it was also used to simultaneously measure metabolites belonging to different functional groups in potato tubers [Roessner et. al., 2000] and in apricots [Katona et. al., 1999]. These developments allowed quantification of a much larger and diverse group of metabolites than was done before, generating a lot more metabolic data as compared to past studies.

Even though more information about the metabolic state of the plant could be extracted in a high throughput manner using these protocols, the data analysis methods used were restricted to variation in individual comparison of samples between different groups. This restricted the use of metabolic profiling to get a biochemical insight of the changes in the physiology of the plant. The use of multivariate data analysis techniques to analyze the metabolic data [Fiehn et. al., 2000] allowed mapping of overall response of the plant to a genetic or environmental change, at the same time identifying metabolites showing differential activity in the two systems in a non-biased manner.

In this chapter we discuss the applications for which the metabolic profiling data can be used and the multivariate statistical techniques required to achieve the objective of the particular analysis.

4.1 Role of multivariate statistical techniques:

There are currently two primary aims of metabolic data analysis using multivariate statistics:

1. To identify the overall change in metabolic state, by clustering various metabolic profiles
2. To identify the metabolites which show a significant difference between the two sets of metabolic profiles.

Currently hierarchical clustering technique (HCL) and principal component analysis (PCA) has been used for the first analysis, where as student t-test has been used to identify differentiated metabolites. The advantages of using multivariate statistical techniques are as follows:

a. Multivariate statistical techniques allow reproducible, non biased analysis of large sets of metabolic data. Hence data analysis does not become a bottle-neck for most studies
b. Multivariate statistical techniques can allow a visual comparison of overall effect of differences based on large number of variables.

In order to achieve this, most multivariate statistical technique uses a measure of distance, which represents the difference between two different metabolic states, based on the measured values of each metabolite. There are many alternate ways of calculating the distance and hence appropriate distance should be chosen based on the problem at hand. In this section we review two commonly used distance measures.

Euclidean Distance:

The geometrical difference (D_{ij}) between any two points in a three dimensional plane defined by X1, X2 & X3 axes is given as

$$D_{ij} = \{(X1_i - X1_j)^2 + (X2_i - X2_j)^2 + (X3_i - X3_j)^2\}^{(1\backslash 2)}$$ **Equation 4.1**

The concept of geometrical distance is extended to Euclidean Distance (ED) in order to compare & quantify the difference between two states (i,j) using N variables and is defined as:

$$ED_{ij} = \{(X1_i - X1_j)^2 + (X2_i - X2_j)^2 + (X3_i - X3_j)^2 + \ldots\ldots + (XN_i - XN_j)^2\}^{(1\backslash 2)}$$

Equation 4.2

Thus Euclidean distance calculates (or is a measure of) the absolute differences for all the variables being measured.

Pearson Correlation Distance:

The Euclidean distance measures the extent of the difference between the metabolites, and uses the absolute values of the variables being measured.

Pearson co-relation measures the similarity and differences in trends irrespective of the absolute values of the variable. The Pearson correlation distance (PCD) between two variables (u, v) across different samples (m), or the correlation between two different samples defined by m different variables is given as:

$PCD(u,v) = covariance(u,v) / \sigma_u * \sigma_v$ **Equation 4.3**

Where

$$cov(u,v) = \frac{\sum_{i=1}^{m}(u_i - \bar{u})(v_i - \bar{v})}{(m-1)}$$

Equation 4.4

Pearson correlation distance value close to 1 would indicate a high degree of correlation between the two variables, where as close to -1 would indicate the opposite correlation between the two. A value close to 0 would indicate that the two samples / variables are not related. Thus Pearson correlation distance is a good measure to identify similarity in the trend or pattern without considering the absolute values. These distance measures can be used by various clustering and significant analysis techniques which are described in the next section.

4.2 Multivariate statistical techniques for metabolic profiling analysis:

From the various multivariate statistical techniques available, as discussed before PCA [Fiehn et. al., 2000], HCL [Roessner et. al. 2001a] and t-test [Fiehn et. al., 2000] have been used in previous metabolic studies. Apart from these methods

other methods like K-Means analysis (and associated FOM analysis), and Significant Analysis of Microarray, which have been used in gene expression analysis, could be used for metabolic profiling analysis. These methods are discussed below:

Hierarchical Clustering Technique (HCL Analysis)

HCL analysis is used for identifying and representing proximity of samples to each other from a group of plant samples (representing metabolic states). In order to perform the clustering, the algorithm calculates distances between all the plant sample using the distance measure chosen by the user. As shown in Figure 4.1 samples which have the lowest distance are linked together to form a cluster. The distance from the cluster of the other samples is now calculated and based on that new linkages are made. Thus HCL analysis can be used to identify presence of different metabolic states amongst plant samples using the hypothesis that distance between metabolic profiles of the plants representing the same metabolic state should be less as compared to distance between metabolic profiles of plants representing different metabolic states. Such an analysis would indicate two separate clusters at the highest clustering level, which would contain plants containing the control set and the perturbed set which has undergone a different environmental condition or which has a different genetic makeup.

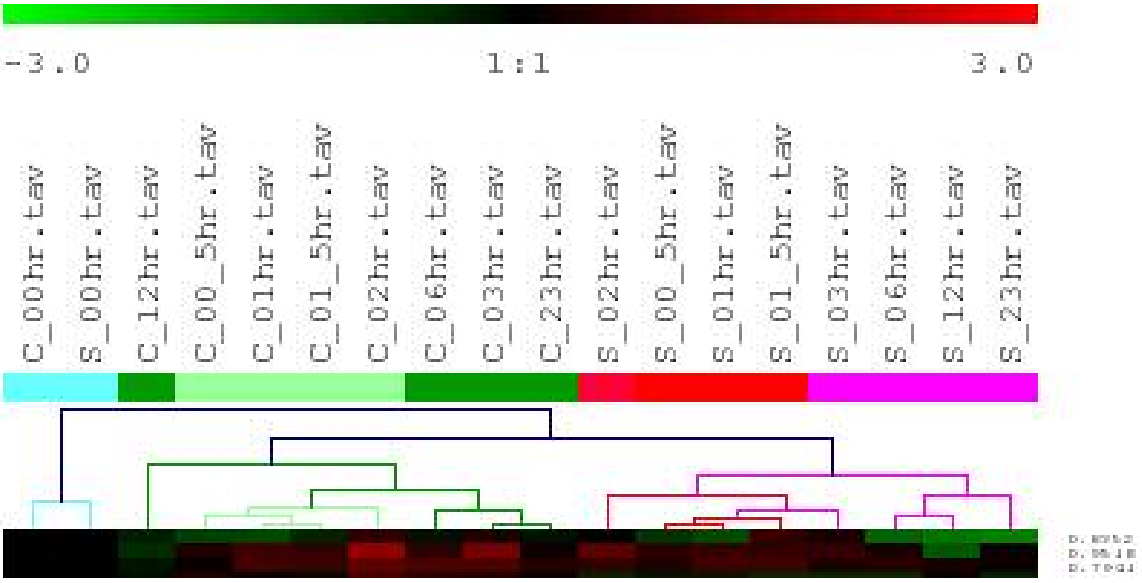


Figure 4.1 Hierarchical Clustering Technique

Principal Component Analysis (PCA):

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. Figure 4.2 indicates a sample 3-D plot using Principal component analysis. Each dot on the plot

represents a plant sample. Plant samples which have almost the same metabolic state cluster close to each other. Hence like HCL analysis, if an environmental or genetic change alters the metabolic state of a plant significantly, the same can be identified from PCA in which they would form a separate cluster.

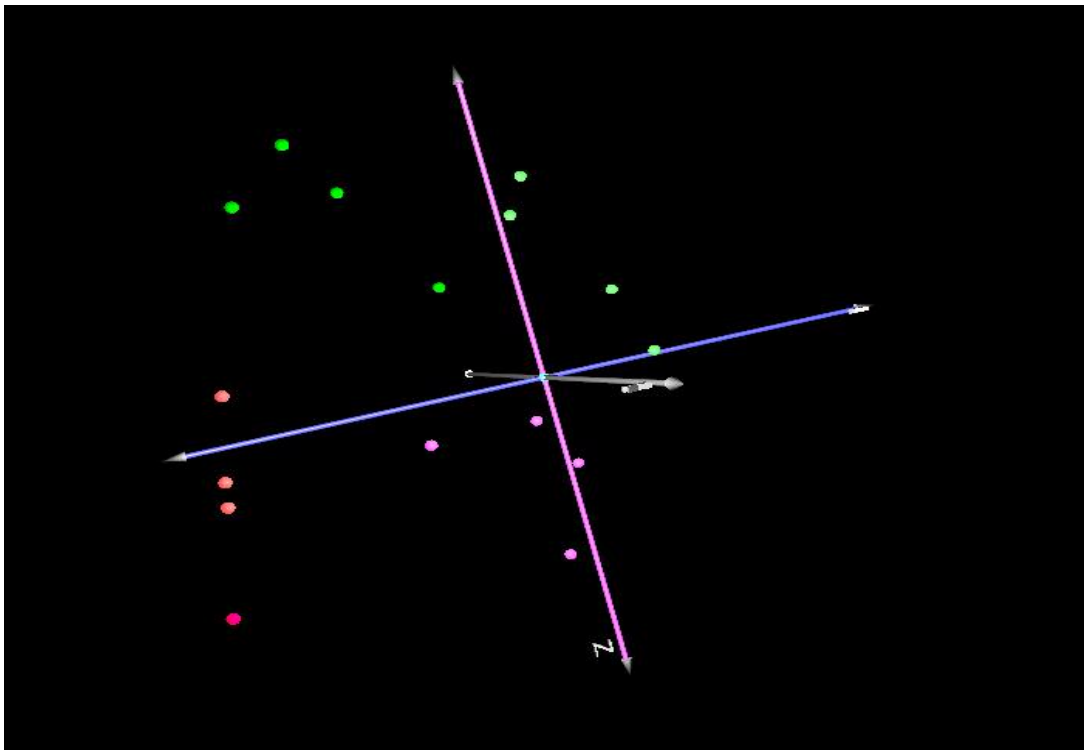


Figure 4.2 Principal component analysis: Projection of different plant samples into three dimensional space mapped by the first three principal components of their metabolic data.

t-test analysis:

In statistics, for any problem which requires to determine if two sets of data belong to the same group or two different groups, t-Tests are used. The decision is achieved by using the mean value of both the groups, the standard deviation (spread) for both the groups and the p-value (which represents the acceptable probability limit) for the analysis. As discussed in the previous analysis most of

the metabolic profiling analysis required comparison between two sets of plants. In order to decide if the metabolite shows different activity between the two sets, the average and standard deviation values for the metabolite relative peak area recorded in both the sets is compared using the p-value between 0.01 – 0.05. Using the t-test, it can than be determined if the values of the metabolites in two sets belong to the same group or a different group. If the t-test indicates that the two sets of values do not represent the same state, than the metabolite is considered to be differentially expressed between the two systems.

Even though t-tests can be used effectively for identifying metabolites which show differential expression, the analysis depends upon the p-value used, which is determined by the user. However t-tests can not be used for time series data in which the aim is to compare metabolite concentration at the same time point in two different groups, and to determine based on this paired comparison if the metabolite shows an increased or decreased activity.

Significant Analysis of Microarray:

Significant analysis of microarray (SAM) is a statistical technique which was developed for analysis of gene expression data obtained from a microarray experiment [Tusher et. al., 2001]. SAM analysis can be performed using four different options, however two options which are important for metabolic profiling data analysis are (a)Two class unpaired SAM (b) Two class paired SAM.

Like t-test, two class paired SAM can be used to identify genes which show differential activity between two different groups. For such an analysis SAM makes an a-priori hypothesis that some of the variables (gene/metabolites) will have significantly different mean expression levels between different sets of samples [Saeed et. al., 2003].

Two class unpaired SAM can thus be used in place of t-test to identify metabolites which show difference between two sets of data, with an added advantage that SAM analysis allows us to determine the limit of the significant change dynamically and also calculates the False detection rate for the given significant change choice. Two class paired SAM, however allows one to one pairing between samples in two groups and identifies gene (or metabolite) which is over/under expressed over the entire pairing. This feature of SAM thus allows identification of metabolites showing significant difference in their time series metabolic profiles.

K-Means Clustering:

K-Means Clustering (KMC) is a clustering technique similar to hierarchical clustering method, and even though has not been used for metabolic profiling data analysis before, it has been used to cluster genes into different clusters [Saeed et. al., 2003]. The proximity of response across samples can be calculated using any distance measure like Euclidean or Pearson correlation distance.

Unlike HCL analysis in which the number of clusters are determined during the analysis, KMC allows distribution of the metabolites into a pre-defined number of groups (Saeed et. al., 2003). By using clustering with Euclidean distance, thus it is possible to cluster all the metabolites which show similar magnitude change close to each other.

For analysis in which the number of groups are known a-priori, KMC analysis can be used directly using the known number of groups. However when such information is not available before conducting KMC analysis, Figure Of Merit (FOM) analysis is performed. FOM analysis provides a measure for the effectiveness of the clustering technique. An example of a FOM graph generated to analyze the efficiency of KMC analysis is shown in Figure 4.3 which provides for different number of clusters, the efficiency of the clustering technique to separate the metabolites. The FOM values are calculated, by removing one of the samples (experiments) out from the analysis and compare it with the original analysis. The hypothesis behind such an analysis is that the most efficient clustering pattern should not be dependent on a single experiment and should not show a significant difference in clustering just by removing one of the experiments. The FOM value is the measure of overall change in the clustering pattern by removing one experiment at a time, for the given number of clusters. Hence lower the FOM value, the higher will be the efficiency of the clustering

technique, as it indicates the clustering pattern is not dependent on a single experiment.

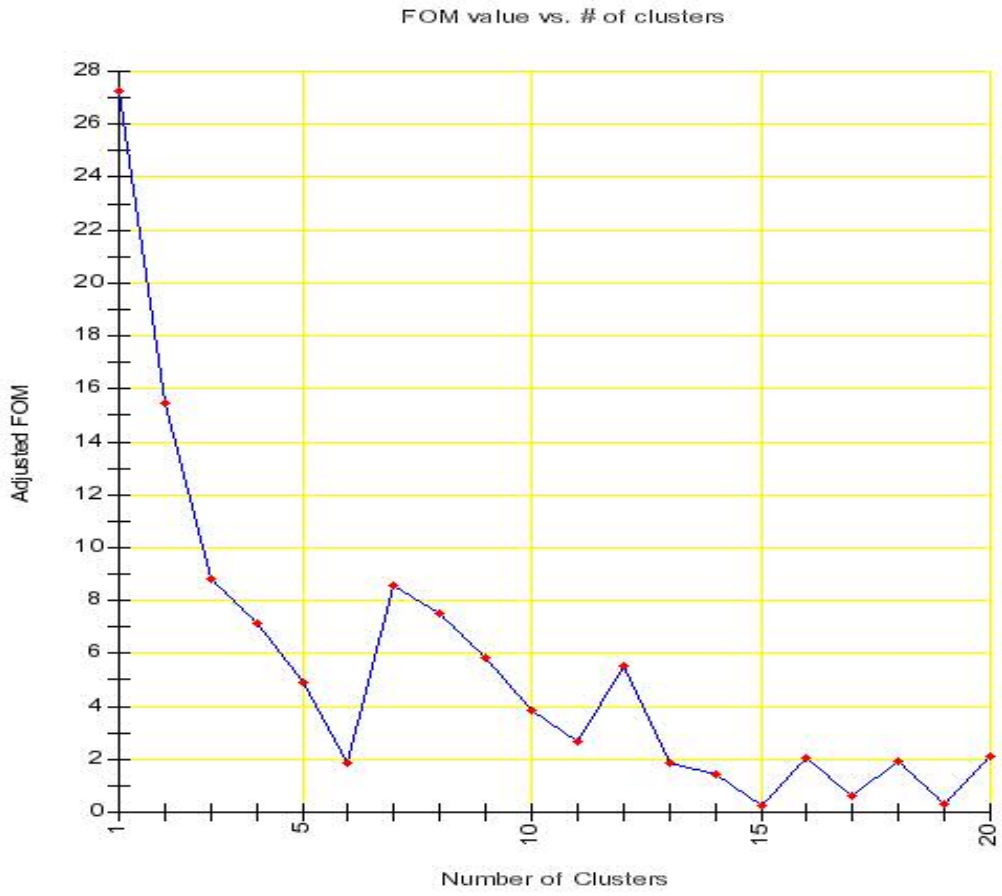


Figure 4.3 FOM Analysis for metabolic data set: FOM curve indicates that the most optimum distribution of the variables is in six groups or fifteen groups where the FOM curve shows local minima.

As can be seen from Figure 4.3, beyond six clusters the FOM value is above for cluster 7 to cluster 13. The next minimum is achieved at 15 clusters. Hence this FOM analysis can be used to conclude that for the given data set can be distributed into 6 or 15 groups with maximum clustering efficiency.

Typically using KMC analysis in conjunction with the Euclidean distance allows distribution of time series profiles of variables in clusters in such a way that the variables which show an increasing trend are clustered together as they all would have similar Euclidean distance, similarly variables which show a decreasing trend with time, or variables which show oscillatory profiles with time can be clustered together. Thus such an analysis can be used for analysis of time series data.

Some of the multivariate techniques described above have been used in plants for different biological studies. These past studies are discussed in the next section

4.3 Current applications of multivariate statistics to metabolic profiling analysis:

Experiment Clustering:

The ability of metabolic profiling to identify different metabolic states is the most commonly used application of metabolic profiling. Untill now this application has been used to differentiate plants having different ecotypes [Fiehn et. al., 2000], genetic mutations or different environmental conditions; also, it can differentiate between different parts of plant. Fiehn et. al. (2000) first used the principal component analysis to differentiate between *A. thaliana* ecotypes (Col-2 and C-24) and their mutants (dgd1, sdd1-1). The ecotypes were known to have

around 100 allelic genes, and the mutants were single gene mutations, with dgd1 deficient in the signaling pathway for photosynthesis which shows a obvious phenotype and sdd1-1 is a mild mutant showing a slight increase in stomatal density. After growing 28-45 plants of each type for two weeks, they were harvested simultaneously and their metabolic profile was obtained by analyzing both the polar and non polar phase. The principal component analysis of the metabolic profiling data indicated the presence of four clusters, with the two ecotypes clusters completely separated from each other and the mutant plant samples formed a separate cluster close to their respective ecotype. The prominent *dgd1* mutant showed a much clearer separation from its parent ecotype where as the mild *sdd1-1* mutant showed lesser difference from its parent ecotype. This showed a strong relationship of different ecotypes and phenotypes with their metabolic state, and showed the ability of metabolic profiling in combination with multivariate statistics to identify this change in the metabolic state in a high throughput, systematic, non based way.

The analysis conducted by Fiehn et. al. (2001) was further confirmed by metabolic profiling of potato tubers for belonging to different transgenic line and grown in different environmental conditions [Roessner et. al., 2001b]. Metabolic profiling of potato tubers obtained from wild type and five other mutant or transgenic lines were analyzed using principal component analysis and

hierarchical clustering. Using just the first two components of the principal component analysis which accounted for 70% of the total variation, all the six clusters could be separated. Once again those mutants or transgenic plants which were defective in genes involved in closely related pathways were clustered close to each other where as the transgenic lines which had modified genes belonging to different pathways showed a much clearer separation in their clusters. Similar clustering pattern was also observed from hierarchical clustering analysis where the plants of the same genotypes clustered together and than with those plants containing mutation in closely related pathways. In another analysis tubers obtained from wild type and three transgenic potato lines were grown under different glucose concentrations. The subsequent metabolic profiles obtained when clustered using first two components of principal component analysis showed that tubers grown in different glucose concentration formed a separate cluster from the normal potato tubers [Roessner et. al., 2001a]. The hierarchical clustering pattern also changed between the transgenic plants in presence of glucose. These experiments and their subsequent non biased, systematic analysis using multivariate statistical methods (PCA and HCL) showed a method by which the relationship between the plant metabolic state and its genetic, phenotypic or environmental conditions can be identified.

Identification of differentiated metabolite:

Fiehn et. al. (2001), used the t-test method to identify metabolites showing significant difference due to presence of mutation. Using t-test with a p-value < 0.01 found 153 out of 326 quantified metabolites to show a significant difference between the dgd1 mutant and wild type (Col-2 ecotype) *A. thaliana* plants metabolic profiles. However in case of sdd1-1 mutant only 53 metabolites indicated a significant difference.

The review demonstrates that multivariate statistical analysis have helped find a relationship between different cellular levels, which was not possible using the standard comparison methods which consider changes in an individual variable at a time.

4.4 TIGR - Multi Experiment Viewer – A new tool for metabolic data analysis:

TIGR - Multi Experiment Viewer (MeV) is a tool developed for multivariate statistical analysis of data generated using high throughput measurement techniques [Saeed et. al., 2002]. This java based tool developed by The Institute of Genomic Research (TIGR) primarily for gene expression data analysis, allows a common platform for conducting different multivariate analysis. MeV allows data analysis of gene expression data using 17 different statistical techniques including the multivariate techniques PCA, HCL, KMC, SAM, t-test described above. It also allows use of 10 different distance measures including the

Euclidean and Pearson correlation distance described above. The option of color coding a cluster of experiment also allows a visual comparison between different clustering / data analysis techniques. Thus MeV provides an extensive platform for metabolic profiling data analysis which has not been used in the past.

4.5 Metabolic Profiling for identifying correlation in metabolites

Since metabolic profiling is a technique which can simultaneously measure metabolites belonging to different parts of a biochemical network of a plant, it has been considered as one of the promising technologies for identification of unknown metabolic pathways [Weckwerth et. al., 2002]. The current approach suggested for finding unknown pathways is through measuring correlation between metabolites in certain samples. The hypothesis behind such an analysis is that metabolites which show a correlation in their variations between different plant samples, should be closely related through biochemical pathways. Even though an effort to find such correlation in the relative concentration of metabolites of central carbon metabolism of potato tubers grown under different conditions [Carrari et. al., 2003] indicated that while fructose-6-phosphate and glucose-6-phosphate did show a very strong correlation in their relative concentration across all samples and lysine and methionine did show a mild correlation in their relative concentrations, citrate - iso-citrate, malate-citrate, succinate-fumarate which are known to be related in the TCA cycle did not show

such a correlation. Similarly an analysis of correlation in dgd1 mutants of A. thaliana indicated presence of very strong correlation for leucine, iso-leucine and serine with threonine (strong) as well as valine (weak). These results indicate that even though the presence of correlation between metabolites does indicate a possibility of the metabolites being closely related through a biochemical pathway, the absence of correlation need not necessarily indicate that the metabolites are not closely related.

As a first attempt to identify the correlation between metabolites Kose et. al. (2001) created a correlation matrix called "clique-metabolite" matrix, which measured the correlation between all the metabolites detected and created a network which connected metabolites based on their correlation with each other. In order to account for observation that metabolites directly related to each other did not always show a strong correlation, a modified approach to measure correlation which takes into account "dynamic fluctuations" in the experimental setup which could disturb such correlation was proposed by Stuer et. al. (2003). In the modeling approach first a non-linear model is created for the system using the known biochemical pathway and which models the presence of the "dynamic fluctuation" and then proposed an approximated linear model which is simpler computationally. Using this approximated model, they obtained correlations in glycolysis pathway using metabolic data which obtained

correlations very close to the non-linear model. However when they attempted to reverse engineer the network using this linear model, they could not obtain conclusive results in absence of time-series data. Arkin et. al. (1997) conducted a in-vitro time series analysis for "Correlation Matrix Construction" (CMC) using enzymes and metabolites of the glycolysis pathway in a continuously stirred tank reactor at steady state. Using the time series data for variation in 14 metabolites, they could reconstruct the known relationship between the metabolites.

Thus in absence of time series metabolic data, most efforts until now [Kose et. al.,2001], [Carrari et. al.,2003], [Stuer et. al., 2003] have focused on measuring correlation in deviation of metabolite concentration among plant samples grown together as part of one set. The current analysis has not yet identified any strong correlation between metabolites which are not known to be bio-chemically linked. Moreover, the simplistic correlation approach to reconstruct a network has been unsuccessful to reconstruct the known biochemical network. The success of such a reconstruction in the simplified in-vitro system using time series data indicates the value of time series data, where the correlation is being measured not only at a single metabolic state but at a multiple closely related metabolic states (represented by each time point) in an effort to better understand and interpret the correlation between different metabolites.

Thus from the review of the current literature and discussion it is clear that even though metabolic profiling can be used currently to identify changes in the overall metabolic state of the plant and for identifying which metabolites change the most due to change in the biological state. Even though such an analysis can be useful to identify genotype or phenotype of a plant, it can not be directly used to understand complex biochemical relationship or for an integrated analysis aimed at understanding interactions between different cellular levels. Time series metabolic profiling studies which would allow comparison between multiple metabolic states of the same plants represented by different time points would allow a much better understanding of the biochemical interactions as previously observed by other researchers. However methodology to identify change in the metabolic state, to identify metabolites showing differential expression or to measure correlations between metabolites has not been developed for the time series metabolic profiling data. Also before conducting a data analysis, a normalization strategy which would allow better comparison between different time profiles would also be needed. In the following chapters we present results obtained from data analysis of time series metabolic profiling and their significance in context of previously known biological information.

Chapter 5. Results

5.1 Experimental Setup:

For the current experiment, *A. thaliana* plants were grown for 12 days in a liquid culture in 500 ml shake flasks, under constant light condition at 23°C temperature. On the thirteenth day, the plants were supplied ambient air (79% N_2, 21% O_2, 0.03% CO_2 concentration) for control system and air of elevated CO_2 level (78% N_2, 21% O_2, 1% CO_2) for the perturbed system. In both experiments, 10% of the CO_2 present was ^{13}C labeled. For each experiment 20 shake flasks were used. In each flask, 200 ml of liquid media was added. The liquid media was prepared using Gamborg medium (Sigma,USA) and sucrose. The sucrose concentration in the media was 20 gm/lit or 58.5 mM. The pH of the media was adjusted to 5.7. For the current experiment seeds of *A. thaliana* Columbia strain were used. Approximately 2 mg of seeds were used per flask which is equivalent to approximately 100 seeds / flask. The seeds were stored overnight in a refrigerator at 4°C temperature after which they were inoculated in a bio-safety cabinet.

The flasks were closed with a stopper containing two glass tubes, one of which was long and immersed in the liquid media, whereas the other shorter tube was connected to vapor phase of the flask as shown in Figure 5.1(B). On the thirteenth day, (after growing *A. thaliana* for the first 12 days at atmospheric condition) the

longer tube of each flask was connected to a cylinder (containing the desired air composition) through a manifold as shown in Figure 5.1(A).

A.  **B.** 

Figure 5.1: A. Picture of the experimental setup in the growth chamber.
B. Picture of a shake-flask in this setup.

Before connecting the tubes to cylinder, plants from 3 shake flasks in the control experiment and 4 shake flasks in the perturbed experiment were harvested. On the thirteenth day, plants from two shake flasks were harvested at each of the time points 0.5 hr, 1 hr, 1.5 hr, 2 hr, 3 hr, 6 hr, 12 hr and 23 hr. The harvested plants were cleaned with distilled water, dried, weighed and then were frozen in liquid nitrogen to stop their metabolism. Later, they were stored in -80 °C freezers.

5.2 Metabolic Profiling using Gas Chromatography – Mass Spectrometry:

Gas Chromatography-Mass Spectrometry (GC-MS) was used for identification and quantification of metabolites in the plant sample for this study. The

instrument used was Thermo Finnigan make GCQ-Polaris ion trap GC-MS. The general protocol developed by Roessner et. al.(2000) was followed, however modifications were made in order to account for specific conditions and difference in available equipments of the experiment.

5.2.1 Plant Grinding:

Plant sample stored at -80 °C was transferred to a pre-chilled mortar containing liquid nitrogen for grinding. It was ground to a fine powder under the presence of liquid nitrogen. In the previous studies involving metabolic profiling, a specific tissue of the plant was typically used like potato tubers [Roessner et. al., 2000; Roessner et. al.,2001a], Arabidopsis thaliana leaves [Fiehn et. al., 2000] and apricot fruit [Katona et. al., 1999]. Since the aim of the current experiment is to understand the overall metabolic response of the whole plant to elevated CO_2 condition, the whole plant was used for grinding.

5.2.2 Metabolite Extraction:

1 gm of freshly ground plant was transferred in a 50 ml conical tube. 500 µl of 2 mg/ml ribitol (Sigma-Aldrich) solution in water was added as an internal standard so that the ribitol concentration in the sample is 1 mg/gm of fresh plant. After adding 28 ml of methanol (Merck, USA), the plant was homogenized using a tissue homogenizer. The resulting solution was transferred to four 15 ml.

conical tubes and kept in water bath at 70°C for 15 minutes. Later, an equal volume of autoclaved de-mineralized water was added to each 15 ml tube (approximately 7 ml each). Each 15 ml tube solution was further divided — equally — into two conical tubes, such that the entire plant sample is distributed equally in 8 conical tubes which were labeled A-H and served as duplicates to allow multiple analysis of the same plant sample. The solution was then stirred vigorously and later centrifuged at 2200g for 5 minutes at 23 °C. Subsequently, the samples were dried in a SpeedVac (make, model) at room temperature until all the methanol and water was dried. 1.4 ml of methanol was used for 100 μg of potato tubers as per the protocol. Since *A. thaliana* roots are known to be difficult to extract, 28 ml of methanol was used for 1 gm of plant sample, instead of 14 ml suggested by the protocol. The dried samples in the 15 ml conical tube were stored at -20°C freezer for further analysis.

(This part of the experiment was conducted by Dr. Maria Klapa, Dr. Tara van Toi, Lara Linford, Jeremy Matthew, Linda Moy and Dr. John Quackenbush at The Institute of Genomic Research, Rockville, MD. The extracted plant samples obtained from their work were used for the metabolic analysis discussed in this text)

5.2.3 Metabolite Derivatization Protocol:

To derivatize the plant sample, the dried methanol extracted plant samples were transferred from 15 ml conical tubes to 2 ml glass vial. 100 µl of 20 mg/ml methoxyamine hydrochloride (Sigma-Aldrich) solution in pyridine (HPLC Grade, Aldrich) was added to each sample and kept for 90 minutes at 30 °C. 100 µl of derivatizing agent n-methyl (n-trimethylsilyl) trifluoro acetamide (MSTFA) (Regis Tech, NC,USA) was added to the solution using a glass syringe and allowed to react for 30 minutes at 37 °C. 30-40 µl of the derivatized sample was than transferred to a high recovery 1.5 ml autosampler vial. As mentioned in previous section, metabolites extracted from 1 gm of plant sample were divided into eight 15 ml conical tubes, hence each conical tube contained metabolites extracted from approximately 125 µg of plant.

The derivatization protocol (described in Appendix I) was optimized for metabolites extracted from 100 µg of potato tubers; hence for metabolites derived from 125 µg plant sample, the reagent volume was increased by 25% — from 80 µl to 100µl, to account for the additional metabolites. Also in the present case, the solution containing external retention time standards were not used, instead, the internal standard — ribitol was used as retention time standards.

Every day, before starting the plant sample analysis, calibration curves were prepared using samples with different concentration of ribitol for operations in linear range of the machine. In preparing the calibration samples, 2 mg of ribitol was dissolved in 1 ml of pyridine to obtain a solution of 2 µg/µl. Solutions containing 0.2 µg/µl, 0.02 µg/µl and 0.002 µg/µl of ribitol were prepared using successive dilution, starting with 2 µg/µl solution with pyridine. Methoxyamine hydrochloride solution (40 mg/ml) in pyridine was also prepared. In order to derivatize the calibration sample 50 µL of each ribitol solution (containing 100 µg, 10 µg, 1 µg and 0.1 µg of ribitol) was mixed with 50 µl of 40 mg/ml (containing 2 mg of methoxyamine hydrochloride) and allowed to react for 90 minutes. The rest of the protocol was followed exactly as the plant sample protocol described above.

5.2.4 GC-MS Conditions:

To identify and quantify the metabolites, ion trap GC-MS (GCQ, Polaris, Thermo-Finnigan make) was used. It was also equipped with an autosampler (Thermo-Finnigan). Here too, except for few modifications, the protocol used was similar to the one described in Roessner et. al.(2001a).

Injector Condition:

Using the autosampler, 1 µl of derivatized sample was injected into a heated injector at 230°C. In order to reduce sample carry over and to increase the life of the autosampler syringe, the syringe was flushed three times with 10 µl of GCResolve Hexane (make) before injection of sample, and 6 µl of pyridine after injection of sample. The pyridine wash increased syringe life by removing residual sample solution, which on drying, formed a hard, thin layer on the glass surface of the syringe. Before injection, the syringe was flushed with 4 µl of the plant sample to reduce the variation in the quantity of sample injected which could have been caused by presence of air bubbles or solvent carry over. Three pull ups were also performed in the auto sampler vial, before withdrawing 1 µl of the sample. These measures, along with optimization of the right injection depth, allowed us to decrease the standard deviation between multiple injections in the ribitol (internal standard) peak area from 12-15% to 2-6%.

Teflon coated high temperature septum was used in order to avoid contact of pyridine with plastic: thus reducing contamination of the sample with septum material. A glass liner containing glass wool was used in the injector for preventing deposition of non-volatile residue on the column surface, which in turn, increased the life of the chromatography column. The plant samples tend to dirty the liner at faster rates compared to other organic samples, and hence the

liner should be changed more frequently to prevent residue built up on the glass wool of the liner. The split ratio used was 75:1. This was optimized to ensure operation in linear range of the machine. More details about choosing the right split ratio are given in the section dealing with quantification of metabolite concentration.

GC-MS Conditions:

The separation of the metabolites using Gas Chromatography was achieved using a ZB-50 (Phenomenex, CA, USA) 30 m long, 0.25 mm diameter with a 0.25 μm thick glass capillary column, equivalent to SPB-50 column used as per the Roessner et. al.(2000) protocol. Oven temperature is held at 70°C for five minutes, after which is heated to 320°C at the rate of 5 °C/min, and finally held constant at 320°C for 1 min. For calibration samples, the same heating rate was used, but the highest temperature was 185°C instead of 320 °C. The pressure was controlled automatically to ensure a constant gas velocity of 40 m/sec. The transfer line was maintained at 250°C and the ion volume was maintained at 200°C. The ionization lens was maintained at 1400 eV. The data was recorded starting from 5 minutes after injection, in a full scan mode, i.e. at all time. The intensity of all the ions in the range of 50-600 m/z is recorded, as compared to the SIM mode used by Roessner et. al.(2000), which measures intensity of only specific ions for a particular time. The scan mode allows better unbiased detection of metabolites in

the plant sample as compared to SIM mode in which you have to choose *a-priori* the ions that are to be monitored; this may miss out on detection of metabolites which were not expected to be present. The mass spectrometer was auto-tuned using calibration gas recommended by the manufacturer.

5.2.5 Identification of metabolites:

The GC-MS analysis of the plant samples, using the protocol described in the earlier section, results in a chromatogram with multiple peaks as shown in Figure 5.2 (A). Each peak represents concentration of a single metabolite or two-three co-eluting metabolites as shown in Figure 5.2 (B). The metabolites can be identified using their retention time and mass spectrum, which is a unique combination for all metabolites as discussed in Chapter 3. For example, using the protocol above, TMS derivative of ribitol had retention time of 21.91 min. as can be seen from Figure 5.2 (B) and had characteristic mass spectrum as shown in Figure 5.2 (C). The spectrum and the retention time for the TMS derivative were the same for all the samples containing ribitol and hence can be used for identifying ribitol peak in the plant chromatogram.

The list of retention times for the derivatized metabolites in plants and their mass spectrum has been published before and can be downloaded from the web. However due to variation between two instruments, there are also variation in

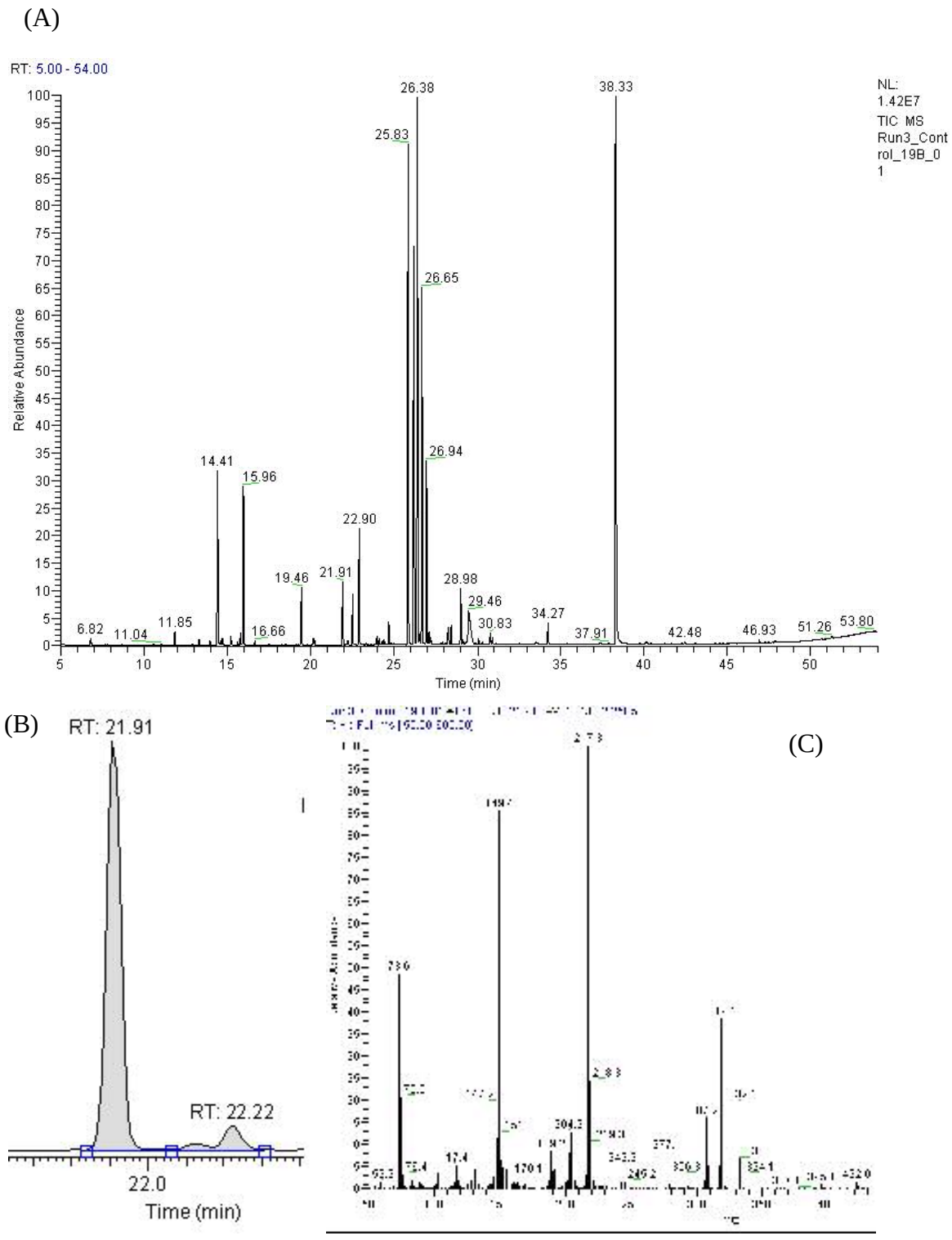


Figure 5.2: (a) Complete Chromatogram of the plant sample (b) Each peak of chromatogram represents a particular compound (Ribitol – RT 21.91) or a group of co-eluting compounds (Xylitol and Arabinose RT: 22.22) (c) Mass spectrum recorded at a specific retention time (21.91 min).

the retention times of the metabolites. Different mass spectrometers can also show variation in mass spectrum obtained due to differences in sensitivities and ionization condition. These differences are even more prominent when comparing mass spectra (obtained from quadruple mass spectrometer) and ion trap mass spectrometer. To take care of these variations, a library of around 30 known metabolites containing their retention time and mass spectrum was created using standard compounds, using the same GC-MS conditions as the plant samples. On comparison with the TMS derivatives library, in most cases, the retention times were observed to be around 0.5- 2 min more in our instrument. That was due to the differences in the chromatographic conditions. Moreover, the variation in retention time was not constant for all the metabolites. The comparison of the retention time and the variation for all the metabolites identified is available in Appendix II. The mass spectrum of the standard metabolites was also compared to the TMS derivatives library and to the NIST Mass Spec library.

After creating the library of standard compounds, they were identified in the plant samples by looking for a peak that had the mass spectrum of the standard compound, and approximately the same retention time recorded for that standard in the library. After identifying peaks for the standard metabolites in plant sample chromatogram, more metabolites were identified in the plant

sample using the retention time and mass spectrums of the TMS derivatives in plant that were published (Fiehn et. al., 2000) and commercially available NIST mass spectral library. The retention time difference between standards library and the published library was used to estimate the possible retention times of other metabolites whose standards were unavailable. Due to possibility of errors present in the library, only those compounds which matched both the retention time and mass spectrum from one or more sources were treated as a positive match, and others were treated as unknown metabolites.

Using the method described above 212 different derivatized metabolites were detected in the plant sample, out of which, 70 had a known structure. These metabolites were either represented in the chromatogram by an individual peak or in some cases, two or three co-eluting metabolites were presented together by a single peak. Such a co-elution occurs due to very close retention times of two or more metabolites. By conducting manual peak de-convolution for the chromatogram as described in Chapter 3 of the report, retention time and marker ion combination for a particular metabolite were obtained. Since approximately 60-70% of the metabolites detected were present in the form of co-eluting peaks, for the purpose of uniformity, all the peaks were represented by using their marker ion instead of the total ion intensity peak. The 212 metabolites detected

using this method (known as well as unknown structure), their retention time and their marker ions are listed in Appendix III.

5.2.6 Quantification of Metabolites:

Relative quantification of the metabolites was carried out using marker ion peak area as described in Chapter 3 using ribitol as an internal standard. The marker ions used for identifying the metabolites, were also used for quantification of the metabolite. In order to ensure accurate relative quantification of metabolites, the instrument should have operations in its linear range. The linear range of machine depends upon the sensitivity of the instrument, hence varies from instrument to instrument (and sometimes even for the same instrument). In order to identify this range, calibration curves were run at different split ratios which changes the amount of metabolites entering the mass spectrometer for the same injected quantity. The linear range of operation for our instrument, with the injection/GC-MS conditions described above was obtained at 75:1 split ratio. In order to ensure that this range was valid for all the samples, the equipment was tested everyday using ribitol calibration samples. The calibration curves obtained are shown in Figure 5.3 along with calibration curves used to identify the linear range. In order to confirm that characteristic ion peak area can be used instead of total ion intensity peak area we plotted the ratio of ribitol marker ion (217 m/z)

peak area to total intensity peak area over the entire operational range, which is shown in Table 5.1 and Figure 5.4.

As can be seen from the table and the plot, the ratio does decrease from 0.17 to 0.12 over the entire concentration range spanning three orders of magnitude. Since in metabolic profiling we rarely expect a single metabolite to have a concentration variation up to three orders of magnitude, the error caused by this deviation would be very small.

In order to confirm this, in plant samples total ion intensity as well as marker ion intensity peak areas were calculated for non-coeluting compounds and their normalized profile was compared. (the normalization procedure is discussed in later part of this report). Figure 5.5 shows this comparison of normalized total intensity peak area and marker ion peak area for two metabolites succinate and glycine. As can be seen from the figure both the total ion current and the marker ions give almost the same profile and hence marker ion can be used for quantification.

The advantages of using marker ion intensity for quantification are as follows:

- it allows quantification of co-eluting metabolites,
- it allows quantification with better signal to noise ratio

Typically the noise observed in GC-MS is due to constant leaks into the mass spectrometer which mostly remain constant throughout the sample run for a

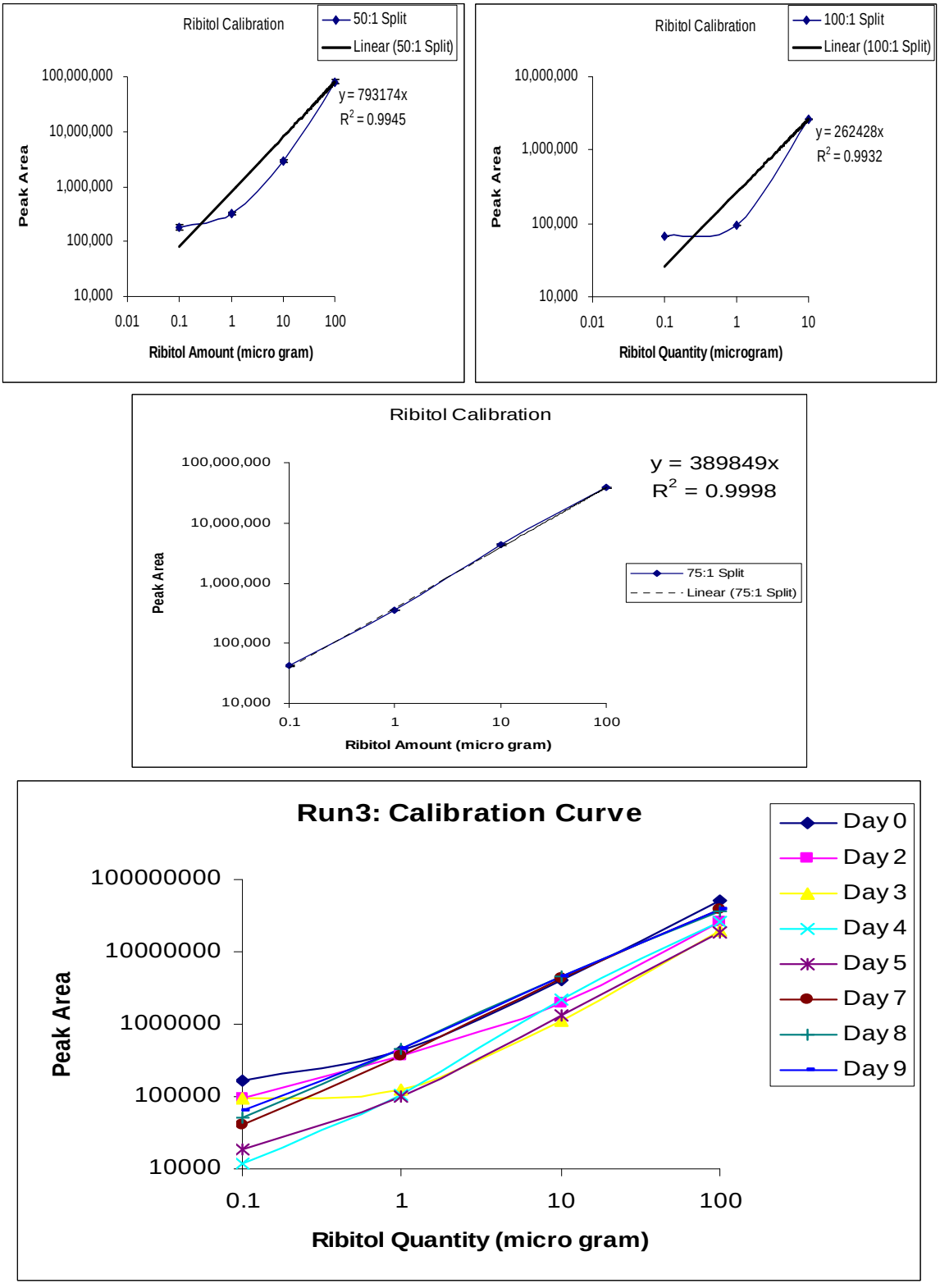


Figure 5.3: Identifying and ensuring the linear range using ribitol samples

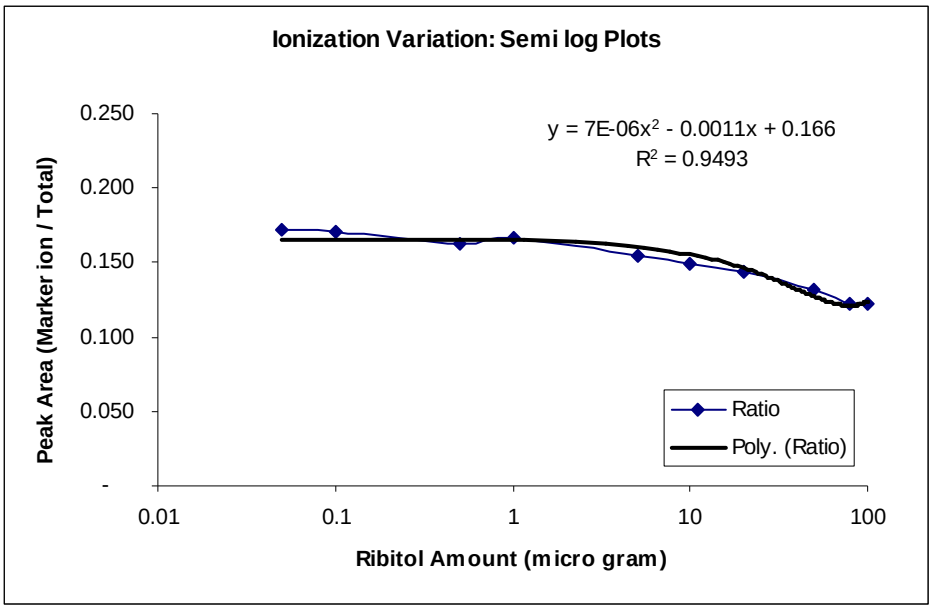


Figure 5.4 Variation in ratio of ribitol marker ion intensity peak area to total intensity peak area

Ribitol	Peak Area Ratio		
Microgm	Injection 1	Injection 2	Average
100	0.122	0.124	0.123
80	0.125	0.120	0.122
50	0.131	0.131	0.131
20	0.141	0.146	0.144
10	0.148	0.150	0.149
5	0.155	0.154	0.155
1	0.169	0.164	0.166
0.5	0.163	0.163	0.163
0.1	0.170	0.172	0.171
0.05	0.172		0.172
		average ---->	0.150

Table 5.1: Ratio of ribitol marker ion intensity peak area to total intensity peak area, for different ribitol quantities.

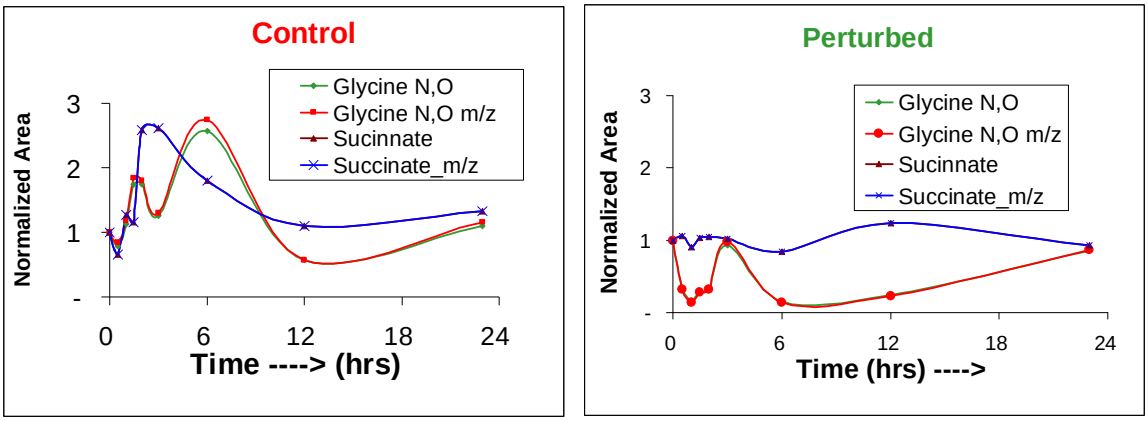


Figure 5.5 Comparison of normalized profile for TIC and marker ion (specific m/z) for glycine and succinate.

given GC-MS system. Hence the noise recorded is typically higher for certain m/z values which correspond to the ions generated by these impurities. So the noise does contribute to total ion current peak area reducing the signal to noise ratio. However, by choosing a marker ion whose m/z value has lower noise levels, the metabolite can be quantified without the effect of the constant noise factors. This can be seen more clearly from Figure 5.6. As we can see from figure 5.6(a), using the total ion intensity chromatogram only two metabolites can be quantified with sufficient signal to noise ratio in the chromatogram between 31 to 32.5 min retention time. Even though the chromatogram does indicate the presence of more compounds, the peaks are very small and have irregular peak areas, so they cannot be quantified. However the 3-Dimension intensity map of the same region Figure 5.6(b), shows that the noise intensity is high only for certain m/z values. As shown in Figure 5.6(c) by choosing different marker ions which do not have high noise levels, not only the presence of co-elution of two

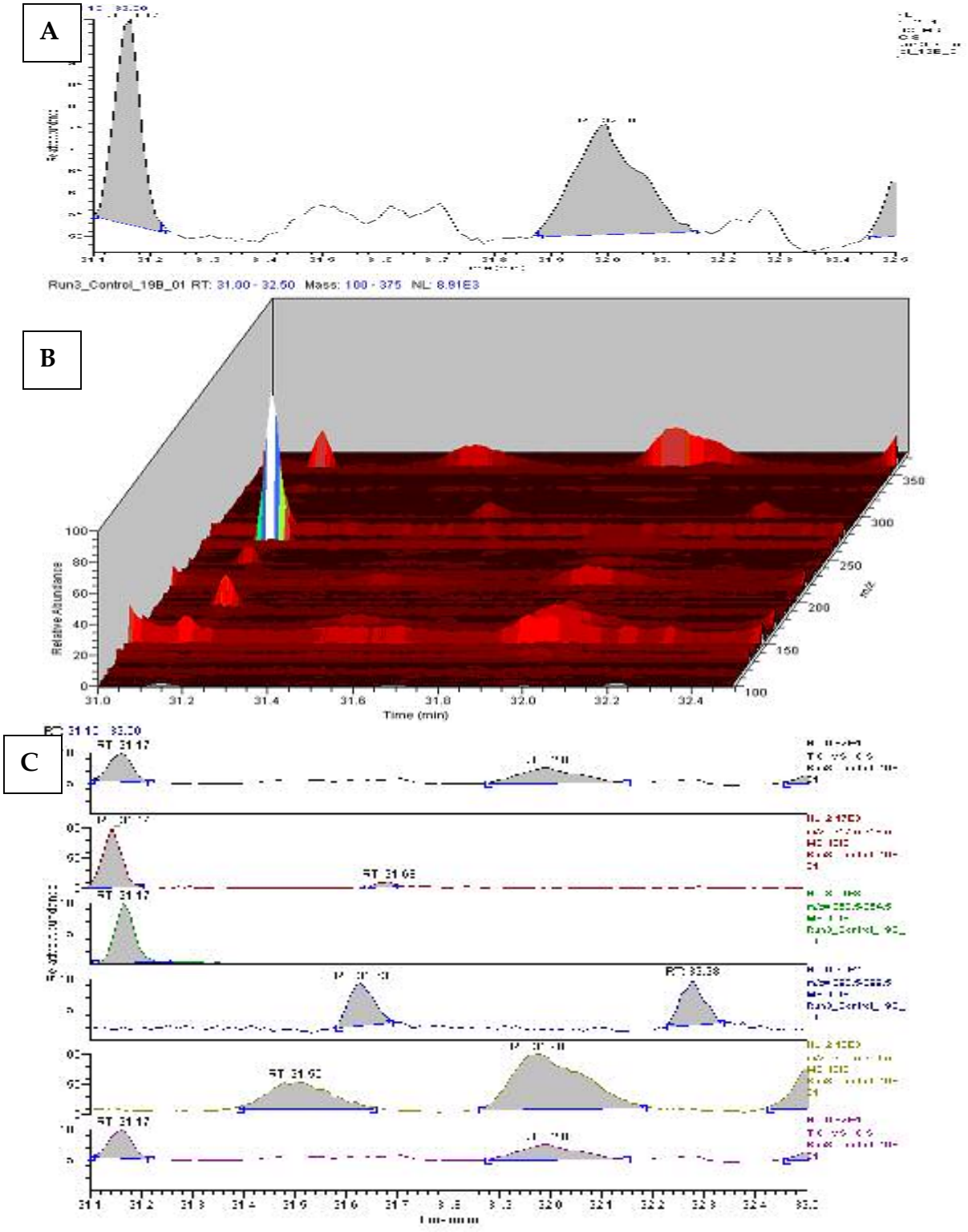


Figure 5.6: Better Signal to noise ratio using marker ion. (A) Total Ion Intensity Chromatogram (B) 3-D Intensity Map (C) Chromatogram for specific marker ions

peaks (TIC peak having retention time 31.17 min) be detected but also other peaks present at retention time 31.50, 31.63, 31.68, 31.98 and 32.28 can be detected which were part of the noise before. Due to these advantages associated with using marker ions, they were used for quantification of all the metabolites detected.

Since 10% of the CO_2 used was labeled, there was a possibility of redistribution of isotopomers of an ion as the labeled CO_2 is fixed in the plant. In order to avoid errors arising out of labeling effects, instead of a specific m/z, a range of m/z which included all possible isotopomers of the marker ions was used when possible. For e.g. in order to quantify the peak area of a Glutamate marker ion having 246 m/z, a range of 246-249 m/z was used for quantification. Whenever the metabolite concentrations were low, and the marker ion represented <5% of the total intensity, more than one marker ions of the same metabolite was used for quantification so as to obtain stronger signal and better separation from neighboring peaks.

5.3 Data normalization and filtering:

The objective of the experiment is to identify the effect of elevated CO_2 in *A. thaliana* growth atmosphere. The methodology being used at the experiment design level is of having a control system which undergoes the same conditions as the perturbed system, except for a certain period of time in its growth

(depending on what time point the plant sample represents) for which it experienced the presence of elevated CO_2 in its atmosphere. However in real environment, it is difficult to control all the parameters which could affect the final measurement of the metabolite concentration. One of the aims of the normalization and filtering process is to remove or minimize the biases created by experimental, instrumental deviations or error, allowing us to obtain conclusions independent of these errors.

The other aim of the normalization and filtering process specific to the metabolic profiling process is to scale the data in such a way so as to allow:

(a) better visualization or comparison between experiments / metabolite profiles

(b) use of multivariate statistics for data analysis, ensuring that a particular metabolite/ group of metabolites create a bias in the analysis.

This is particularly important in metabolic profiling, because the variation in concentration of different metabolites in plant samples can be up to three orders of magnitude. Due to these differences in some of the analysis, variations in high concentration metabolites may get additional weightage, which may not be desirable. Thus scaling the data to one level allows us to have almost equal weightage to variation in each metabolite concentration.

In order to achieve these aims, there were six stages of data filtering and normalization, starting from the raw peak area obtained from quantification:

1. Normalization using internal standard – gives relative peak areas,
2. Filtering specific metabolic profiles containing gross errors,
3. Filtering injections with gross errors – injection outlier analysis,
4. Filtering plant samples which show abnormal metabolism – biological outlier analysis,
5. Normalizing average relative peak area of each time point w. r. t. time zero average relative peak area for that time point – gives normalized peak area and,
6. Logarithm w. r. t. base 2 of ratio of normalized peak area of perturbed system at time t over control system at the same time point.

The identification and quantification strategy discussed in the previous sections gives a vector containing two hundred peak areas of marker ions from each injection. Since the control set contained 19 plant samples and each sample was injected three times, a 57*200 matrix containing peak areas of marker ions (having values ranging from ~1,000 to ~1,000,000) is obtained. Similarly from the perturbed system, which had 20 plant samples, a 60*200 matrix is obtained. This is the starting raw data used for normalization process which is described below.

5.3.1 Normalization w. r. t. Internal Standard:

Internal standard is a compound of known concentration added externally to the sample in order to facilitate quantification of peak in GC-MS analysis. As discussed in Chapter 3 of this report, internal standard has been used regularly: in many GC-MS analysis applications in general and metabolic profiling in particular. In this normalization, we divide each metabolite marker ion peak area w. r. t. the ribitol marker ion peak area in the same injection. As discussed in detail in earlier chapter, this process allows removal of variations generated due to errors in methanol extraction procedure, drying and GC-MS instrumental variations. In absence of these variations, the ribitol peak area would have been constant for all plant samples, for all injections. But as can be seen from Figure 5.7 (a) and (b), there is variation in ribitol marker ion peak area between injections and between different plant samples, representing presence of the errors discussed. The average standard deviation between the injections is 3%, with minimum deviation being 1% and maximum deviation being 9%. Since in most cases, all the three injections of the same plant samples were run simultaneously, using the same derivatized plant sample, this deviation

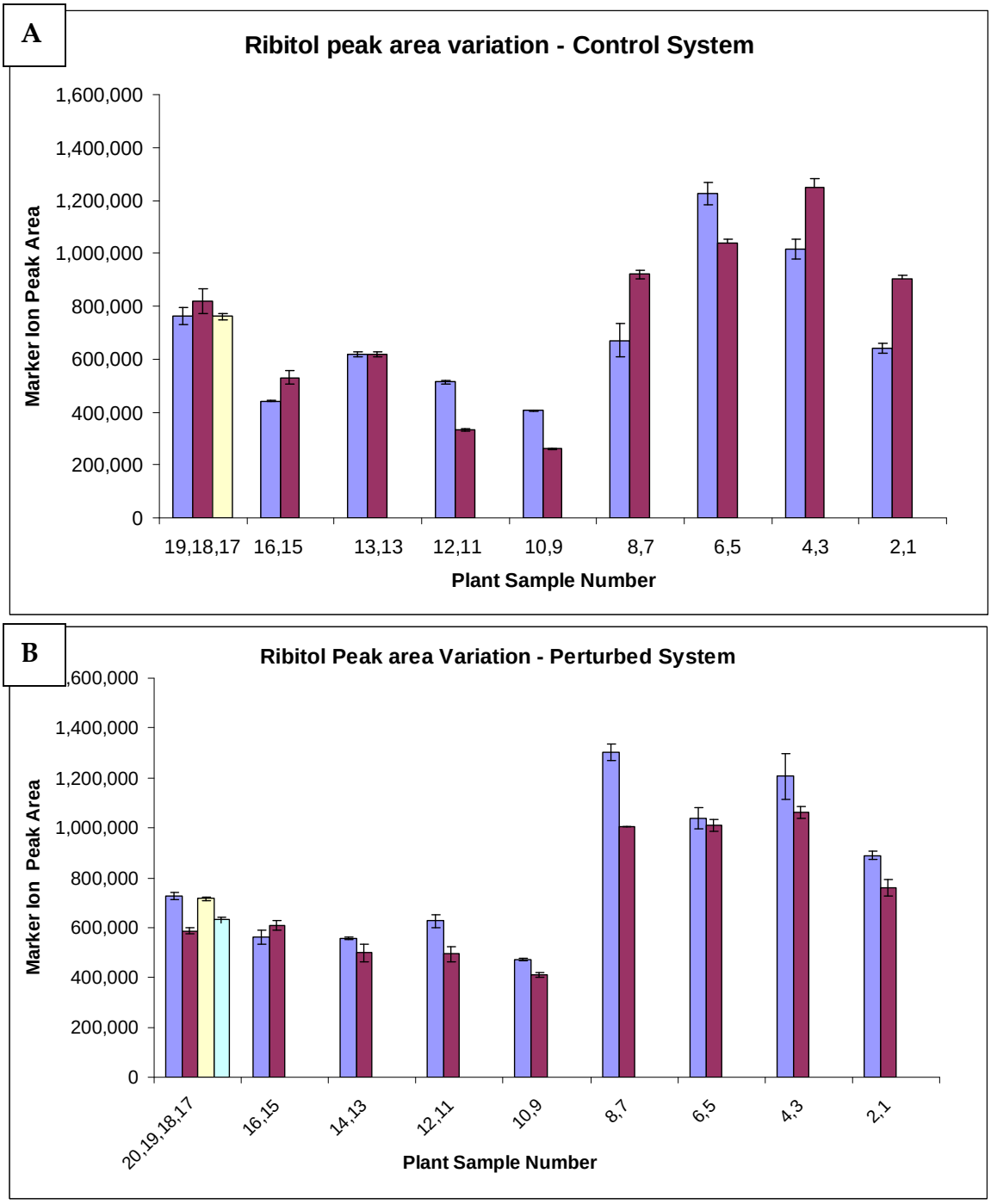


Figure 5.7 Ribitol marker ion peak area for different plant samples in (A) Control plant samples (B) Perturbed Plant samples.

represents the deviation caused by various errors related to sample injection or variation in the split ratio of the GC-MS. The total variation between the ribitol marker ion peak area is due to variations caused in methanol extraction, drying, derivatization, injection errors and instrumental variation. The standard deviation in ribitol caused by all these factors (measured as standard deviation of ribitol peak area for all injections) is 35%. Thus, in absence of normalization, using internal standard could have resulted in a deviation of about 35%, which would have been unaccounted for, consequently creating a bias in the analysis.

The average ribitol marker ion peak area for all injections is ~750,000, hence on normalizing the raw peak area matrix w. r. t. ribitol area, relative peak area matrix of the same size was obtained, in which the relative peak area values varied from ~5 to ~0.001.

5.3.2 Filtering specific metabolic profiles:

The metabolic profiles which could contain gross errors or those which could create error in our analysis were identified and removed from further analysis at this stage of data normalization. In order to identify metabolites having gross error, a matrix indicating overall consistency of the metabolic profile was produced, which is given in Appendix IV. The metabolites to be removed were selected based on following criterion:

1. From the total presence score of the metabolite in the matrix, metabolites which were present in less than 89% time points (i.e. absent in more than two time points out of total 18 time points) were removed from the analysis.
2. Using the total standard deviation of the metabolite in the matrix, metabolites which showed very large total standard deviation were also removed. These metabolites typically had very low relative area which was the cause of the problem.
3. For metabolites which exhibit dual derivatization forms, profile of one of the derivatization form was removed. For e.g. fructose is present in two derivatization form: fructose methoxime-1 5 TMS or fructose methoxime-2, 5 TMS. These are two derivatization forms which are different in positions of the methoxime group that was generated during derivatization from the same fructose compound. Even though these dual derivatization forms have been reported by other researchers in their library, there is no reference of how they are used in the data analysis. Since each derivatized molecule represents one non-derivatized molecule, the peak areas of the two derivatization forms should be added up in the order to obtain the total metabolic profile. However due to co-elution of metabolites it is not always possible to add the total peak areas. Also it is

observed from derivatization chemistry that the ratio in which these are distributed should remain constant. This is also seen from comparing their final normalized profiles, shown in Figure 5.8. Since these metabolites give the same biological information for our analysis, one of the peaks of the two derivatization form was removed.

4. Similarly some of the amino acids also show two derivatization states, for example glycine N, O TMS and glycine N,N,O TMS. The second form contains an additional TMS group in the amine group, i.e. both the H atoms of $-NH_2$ group are replaced by TMS group. This additional derivatization was detected only in some of the plant samples, and typically also showed other problems like high injection standard deviation (present only in some of the time points), low percentage presence, and high min/max ratio. This indicated that the second derivatization form did not truly represent the original metabolite present in the plant sample. Since enough information was not available to add to the peak areas of the two derivatization forms, they were removed from further analysis.
5. Some of the peaks showed a very high deviation between the minimum and maximum recorded concentration (ratio ~1000) in the control system.

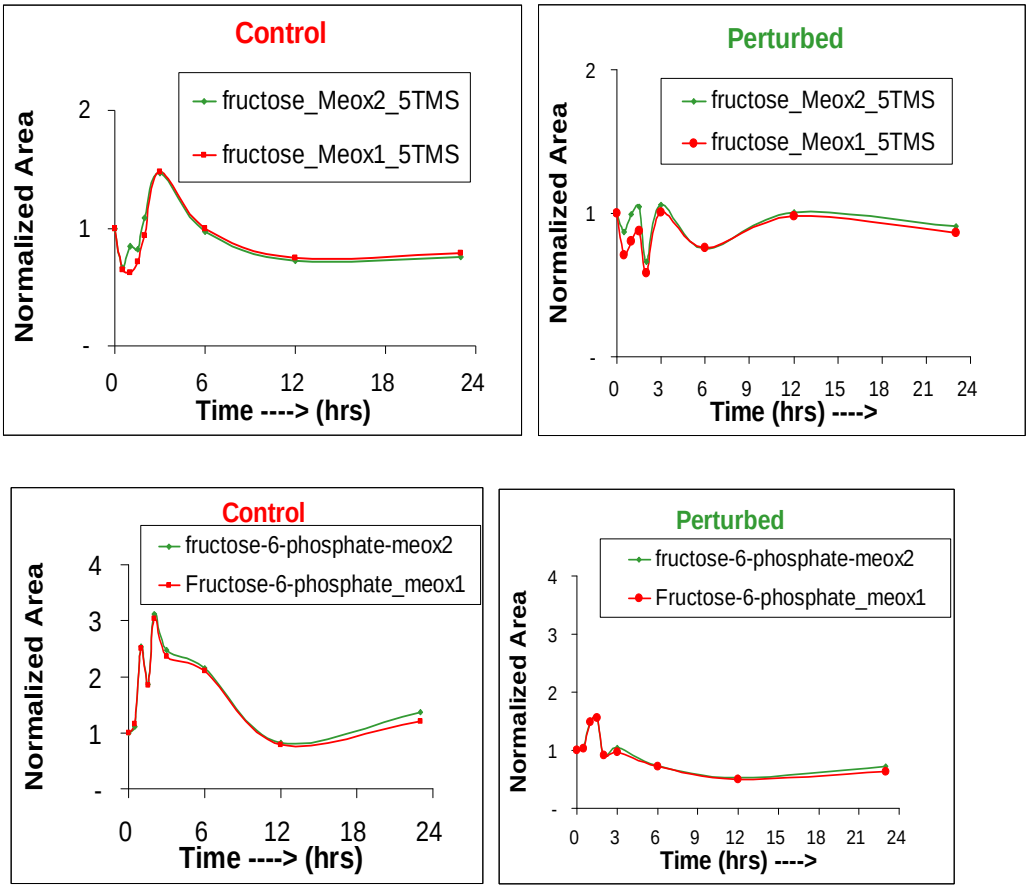


Figure 5.8 Normalized profiles for multiple derivatization forms of sugars and sugar derivatives.

These peaks were mostly unknown peaks. Since we did not expect the metabolism to change so much in absence of perturbation, the profile of these peaks was removed from further analysis.

6. Sucrose peak was removed as sucrose was present in the plant growth media, and since the plants couldn't be completely washed to remove sucrose on the plant surface, it was difficult to differentiate between sucrose produced in the plant and sucrose impurity due to media so the same was removed.

Based on this criteria around 50 metabolite profiles were removed (shown in Appendix V) from analysis, and the relative data matrix was reduced to 57*150 for control and 60* 150 for Perturbed system.

5.3.3 Injection Outlier Analysis:

The aim of injection outlier analysis is to detect those injections which contain gross errors due to abnormal instrumental variations. In order to detect these errors, the Multi Experiment Viewer (MeV, Version 2.1) of TIGR TM-4 package [Saeed et. al., 2003] was used. Specifically the Hierarchical clustering Algorithm (HCL) was used with Euclidean distances. The CY 3 intensity column of the standard TAV file was replaced by raw peak area of ribitol marker ion for that injection, and CY 5 intensity column was replaced by peak area of marker ion of an individual metabolite. The gene label column contents were replaced by metabolite names. This was done so that the CY5 over CY3 ratio represents the relative area of the metabolite. 57 TAV files, each representing an individual injection, were prepared for the control system and 60 TAV files were prepared for the perturbed system. The injections were now clustered and the clustering pattern obtained for control and perturbed system is shown in Figure 5.9. From the clustering pattern of the control injections we can see that in most cases, injections of the same plant samples cluster together, or cluster along with its biological replicate which were harvested at the same time. Only injections

19B_01, 1G_02 and 08G_01 cluster separately. Looking closely we can see that even though in the clustering pattern 08G_01 seems to be clustering separately, it is only one level away in the pattern from the cluster containing two other injections. Hence 19B_01 and 1G_02 injections were identified as those having some problems and were considered outliers and removed from further analysis. Using similar analysis in the perturbed system injections 9D_02, 4D_01, 2D_01 and 1D_01 were identified as outliers and removed from further analysis.

After removing the injection outliers, the relative areas for three injections for a plant sample (two in case of plant samples in which outlier injection was removed) were averaged, giving average relative area for each metabolite of the plant sample. This generated a 19*150 matrix for the control system and 20*150 matrix for perturbed system.

5.3.4 Biological Outlier Analysis:

The above analysis indicates that the three injections of the same plant sample cluster together under ideal conditions, since they represent the metabolite concentration for the same derivatized plant sample. Similarly, we would expect that the plants belonging to two flasks which were harvested at the same time point, should have almost the same metabolite concentration since they represent the same metabolic state of the plant. So, under perfect experimental conditions, the biological duplicates (representing flasks removed at the same time point)

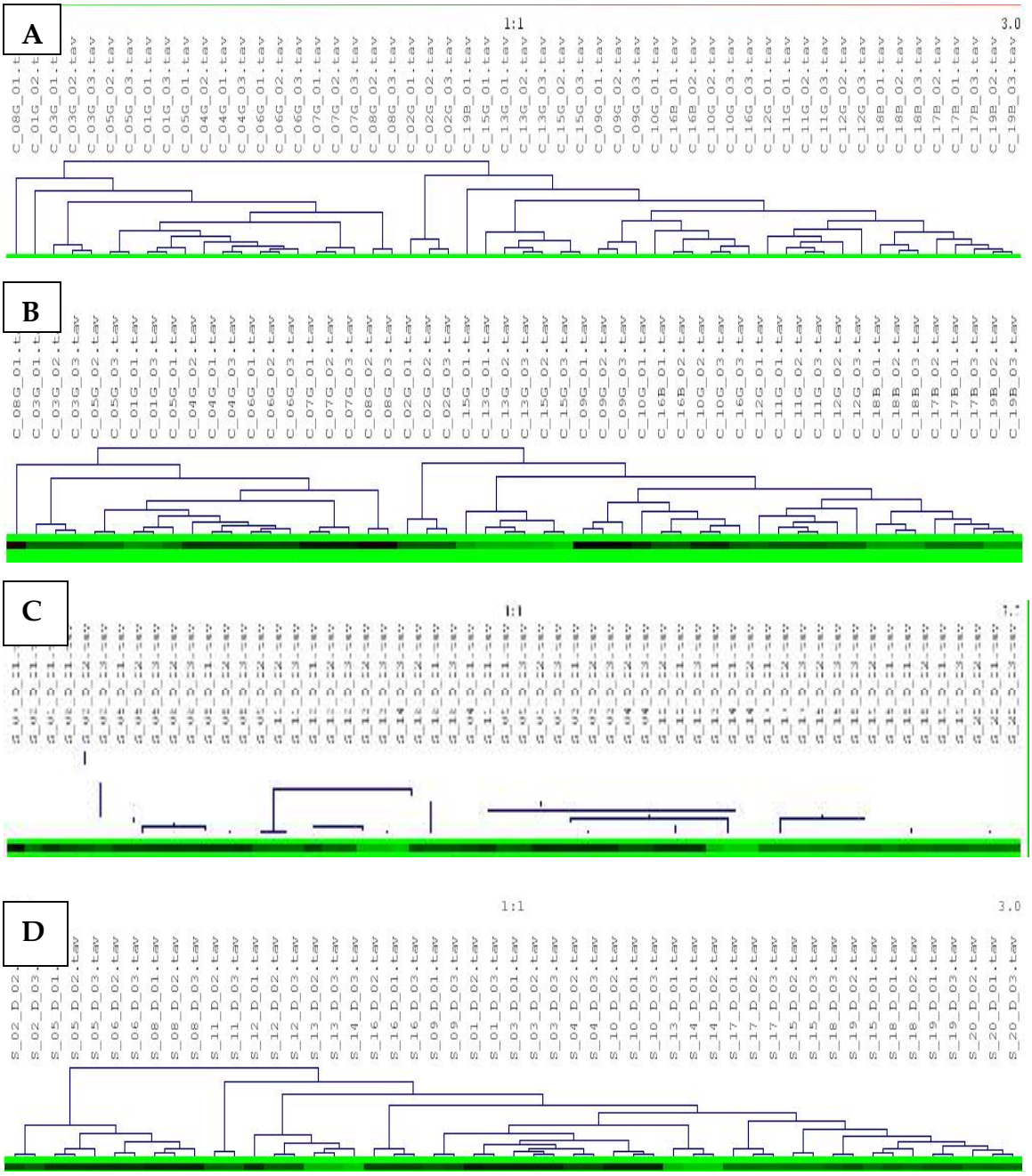


Figure 5.9 Injection hierarchical clustering – (A) control samples before outlier removal (B) control sample after outlier removal (C) Perturbed samples before outlier removal (D) Perturbed samples after outlier removal

should cluster together in HCL analysis of the averaged relative area for the plant sample. This was partially evident from the injection clustering analysis where the injections of the biological duplicates clustered close to each other.

Based on this hypothesis we conducted biological outlier analysis. 19 and 20 TAV files were prepared for 19 and 20 plant samples in the control and perturbed system respectively using averaged relative areas. To prepare the TAV files, the CY3 column was replaced with a value 1 in all the files and CY 5 was replaced with the average relative area of the metabolite. The clustering pattern obtained from hierarchical clustering analysis is shown in Figure 5.10 We can see that even though the biological duplicates cluster together or very close for most time points, in some cases like plant samples 1, 2 in control and 15, 16 in perturbed cluster far away from each other. These are possible indications of the presence of an outlier, however since only two biological duplicates were available it was not possible to determine which of the two samples is an outlier. We did have more than two plant samples in case of time zero however the three bioduplicates at time zero in control cluster together (17-19) and in perturbed 18-20 cluster immediately with a cluster containing 17, the fourth bioduplicate. Hence none of the biological outliers were removed in the current analysis. However if three flasks were harvested at each time point, it may have been possible to remove the biological outlier which may cause bias in the analysis.

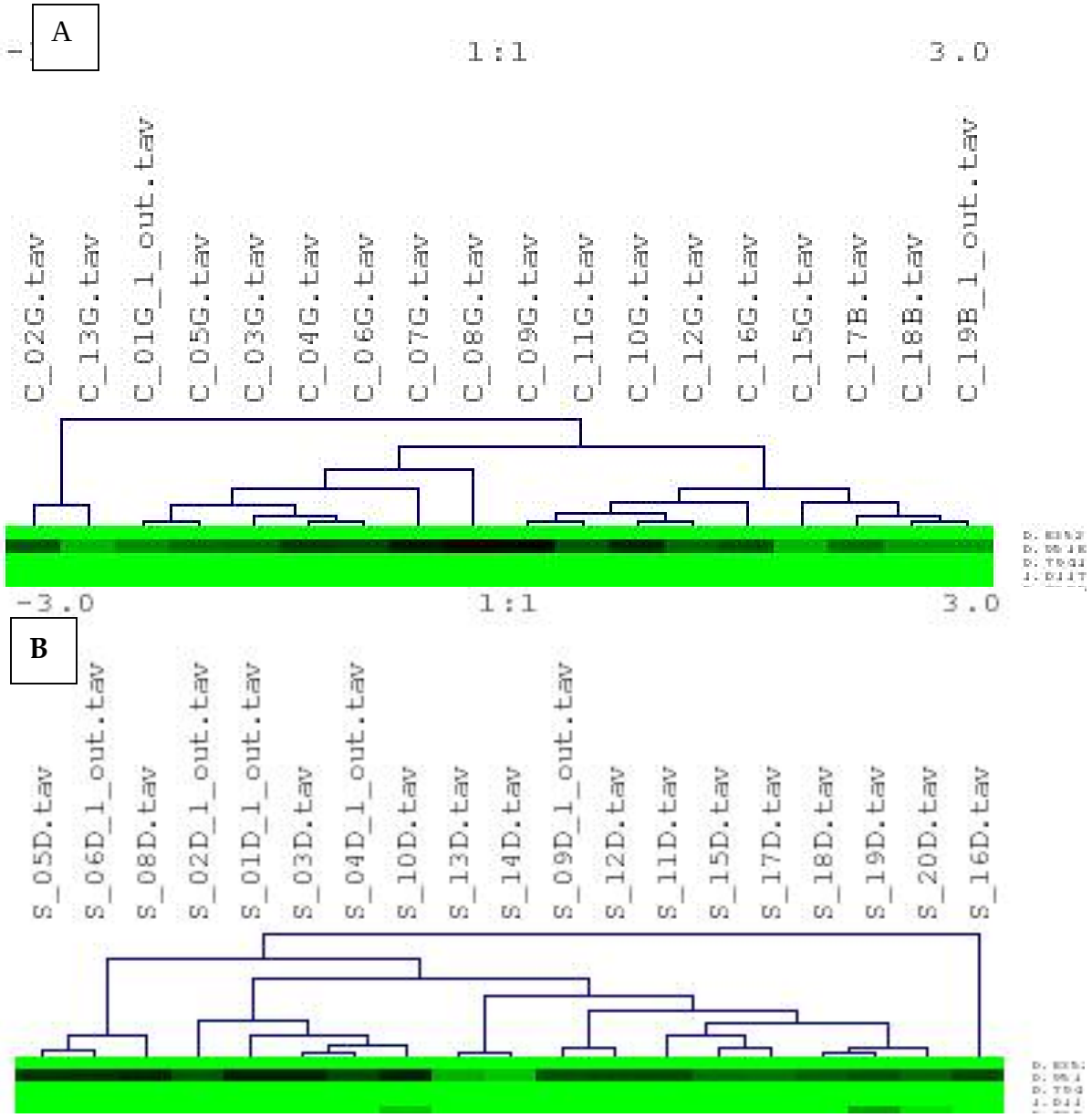


Figure 5.10 Biological Outlier Analysis (A) Control Samples (B) Perturbed Samples.

At the end of this analysis, the relative areas of a metabolite representing a time point (present in all the plant samples), were averaged, creating 9*150 matrix containing average relative areas of each metabolite at the time point.

To estimate the usefulness of the outlier removal process and the variability in measurements (due to experimental conditions), a summary of the injection, bio diversity and total standard deviation at each time point in control and perturbed system was prepared after outlier removal (given in Table 5.2).The average injection, biological and total standard deviation for each individual metabolite is given in Appendix VI. The standard deviation analysis indicates that the injection standard deviation for the current method is ~ 10% —smaller as compared to standard deviation between plant samples representing biological states — which is ~ 30%. Thus indicating biological variation is larger than instrumental variation in metabolic profiling technique used.

5.3.5 Normalization w. r. t Time Zero:

The aim of the current data analysis is to compare the metabolic profile of each metabolite in the control and the perturbed system to understand the effect of the elevated CO2 on the plant metabolism. The other aim of the analysis is to compare the variations in different metabolites within the system. In the current experiments, the plants of the perturbed and the control system were grown under identical conditions for the first 12 days.

Before connecting the system to the manifold, three plants in the control system and four plants in the perturbed system were harvested at time zero on the thirteenth day. Since these plants were grown for the first 12 days under same

			Avg. Deviation at time pt		
	Plants	**Time**	**Injection**	**Biological**	**Total**
Control	17-19	-	6%	27%	26%
Control	15-16	23.0	7%	39%	34%
Control	13	12.0	5%	*	*
Control	11-12	6.0	9%	36%	32%
Control	9-10	3.0	7%	33%	29%
Control	7-8	2.0	13%	51%	47%
Control	5-6	1.5	6%	23%	21%
Control	3-4	1.0	9%	29%	28%
Control	1-2	0.5	8%	38%	33%
Control	**Average**		**8%**	**34%**	**31%**
Perturbed	17-20	-	5%	28%	27%
Perturbed	15-16	23.0	8%	35%	31%
Perturbed	13-14	12.0	8%	15%	17%
Perturbed	11-12	6.0	19%	35%	39%
Perturbed	9-10	3.0	4%	30%	25%
Perturbed	7-8	2.0	8%	39%	32%
Perturbed	5-6	1.5	10%	26%	27%
Perturbed	3-4	1.0	8%	20%	19%
Perturbed	1-2	0.5	5%	24%	22%
Perturbed	**Average**		**8%**	**28%**	**27%**
Average			**8%**	**31%**	**29%**

* only one sample was available for this time point

Table 5.2 Standard Deviation Analysis

conditions, under ideal conditions, these plants represent the same metabolic state – which is the metabolic state of the plant grown as per the protocol described earlier in this report for 12 days. However since the control and perturbed experiment was performed on different days, due to various factors related to plant growth environment which were not controlled or couldn't be controlled, the plants show variation in the metabolic state in the control and the

perturbed system at time zero. Since we want to compare the effect of elevated CO2 during the thirteenth day, in order to remove the effect of variation in the first 12 days, the average relative peak area for each metabolite at each time point is divided by the average relative peak area of the metabolite at time zero of the same set. In the current analysis this ratio is called normalized peak area for the metabolite. By carrying out this normalization, the starting point of each metabolic profile becomes 1, which is a common reference point for comparison as shown in Figure 5.11. Also normalization w.r.t. time zero scales the data from ~0.001 – ~5 (in case of relative peak area) to ~0.1 – ~10, as in the short term response, for most metabolites the concentration variation is not more/less than 10 times the initial concentration. Thus scaling all the profiles to the same order of magnitude, allows better comparison between different metabolic profiles and also removes any bias for changes in high concentration metabolites during the statistical analysis. The normalization generates two 9*150 matrix having data of the order of 0.1 to 10.

5.3.6 Ratio Perturbed over control:

After normalizing the data w.r.t. time zero, we obtain normalized metabolic profile for the control and the perturbed system. In order to facilitate comparison between the perturbed and the control system, instead of comparing variations in two different profiles, they can be combined into a single profile representing

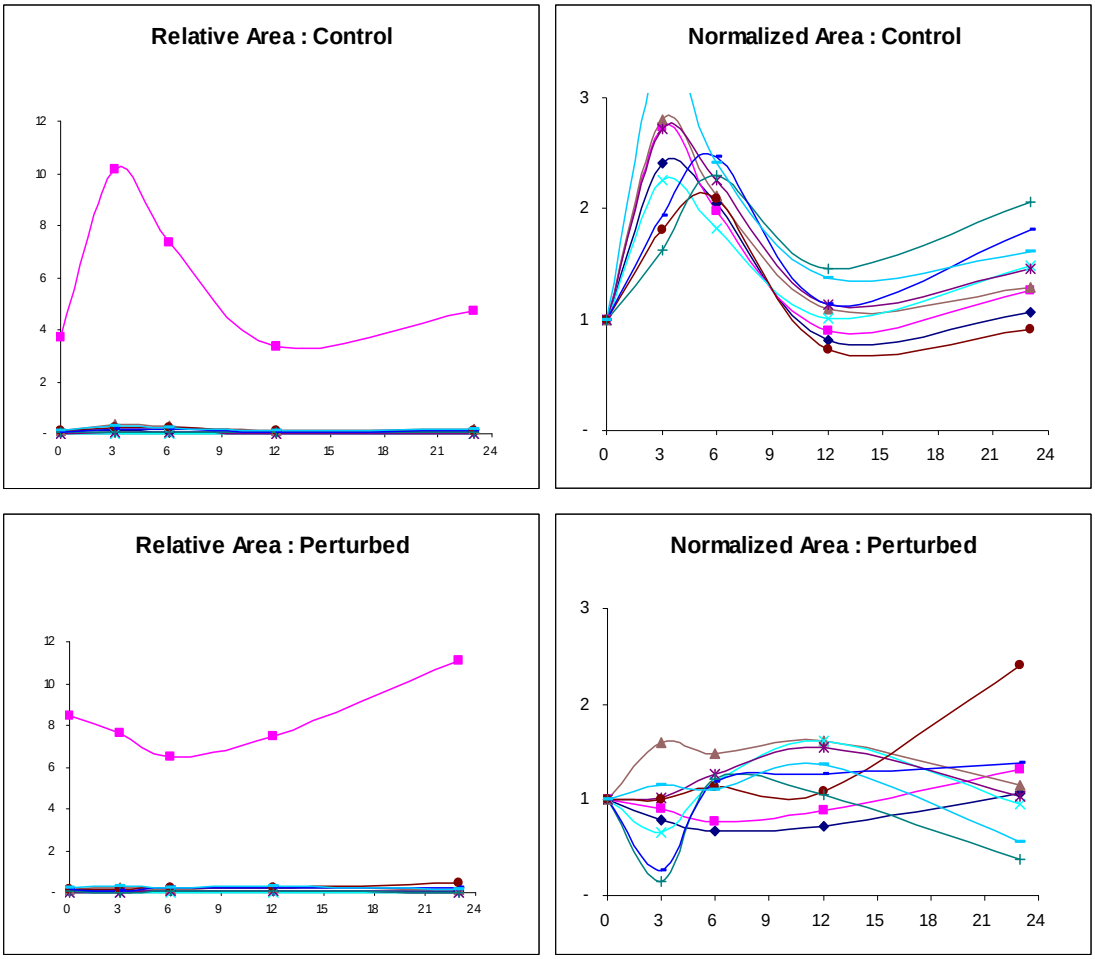


Figure 5.11: Normalization w.r.t. time zero.

the differential response of the control and perturbed system. This can be done by taking the ratio of normalized perturbed area to normalized control area, at the same time point. As seen from Figure 5.12, the ratio allows us to see the effect of elevated CO2 by directly looking at one graph (or data point) instead of two. This also allows us to obtain those metabolites which show a similarity in their response to elevated CO2. Thus by taking the ratio, we combine the two 9*150 matrix into a single 9*150 matrix containing the ratio.

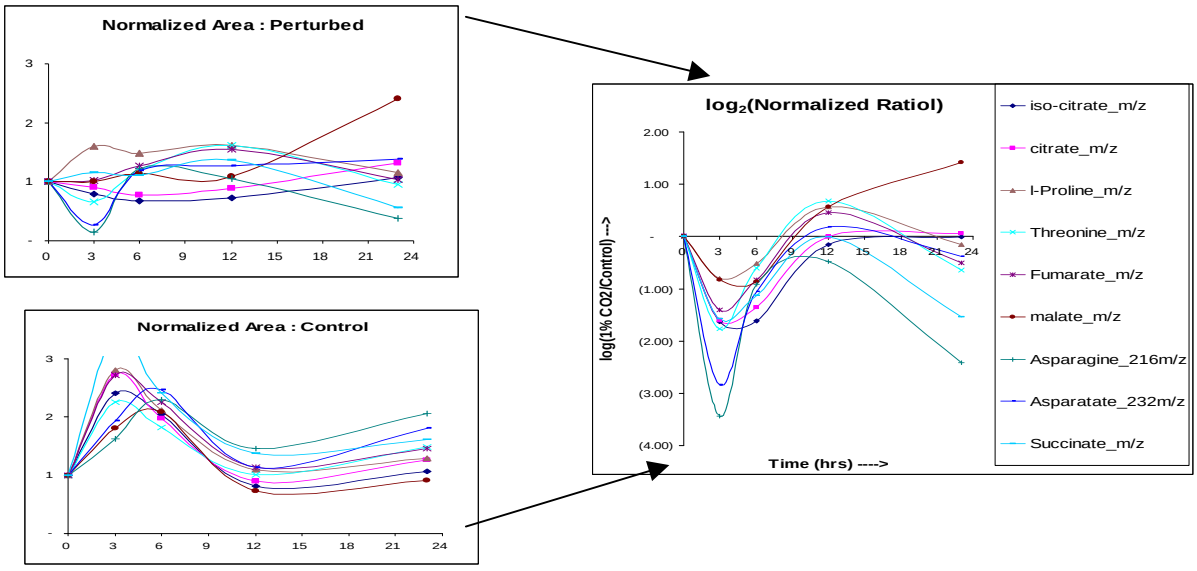


Figure 5.12 Ratio of Normalized Area

5.4 Data analysis using multivariate techniques:

As shown in Figure 5.13, as the result of the data normalization process described above, 150 metabolic profiles are obtained and the aim is to identify those metabolites which show differential expression in the perturbed system. This can facilitate understanding the effect of elevated CO_2 on plant physiology. The other aim of the data analysis is to identify co-relation in different metabolic response to the perturbed system.

To achieve goals of the data analysis in a non-biased, high throughput way, multivariate statistical techniques can be used. Since these techniques have been extensively used in high throughput gene expression analysis, software tools developed for gene-expression analysis were used for the current analysis. This also created a common methodology for the analysis of gene expression and

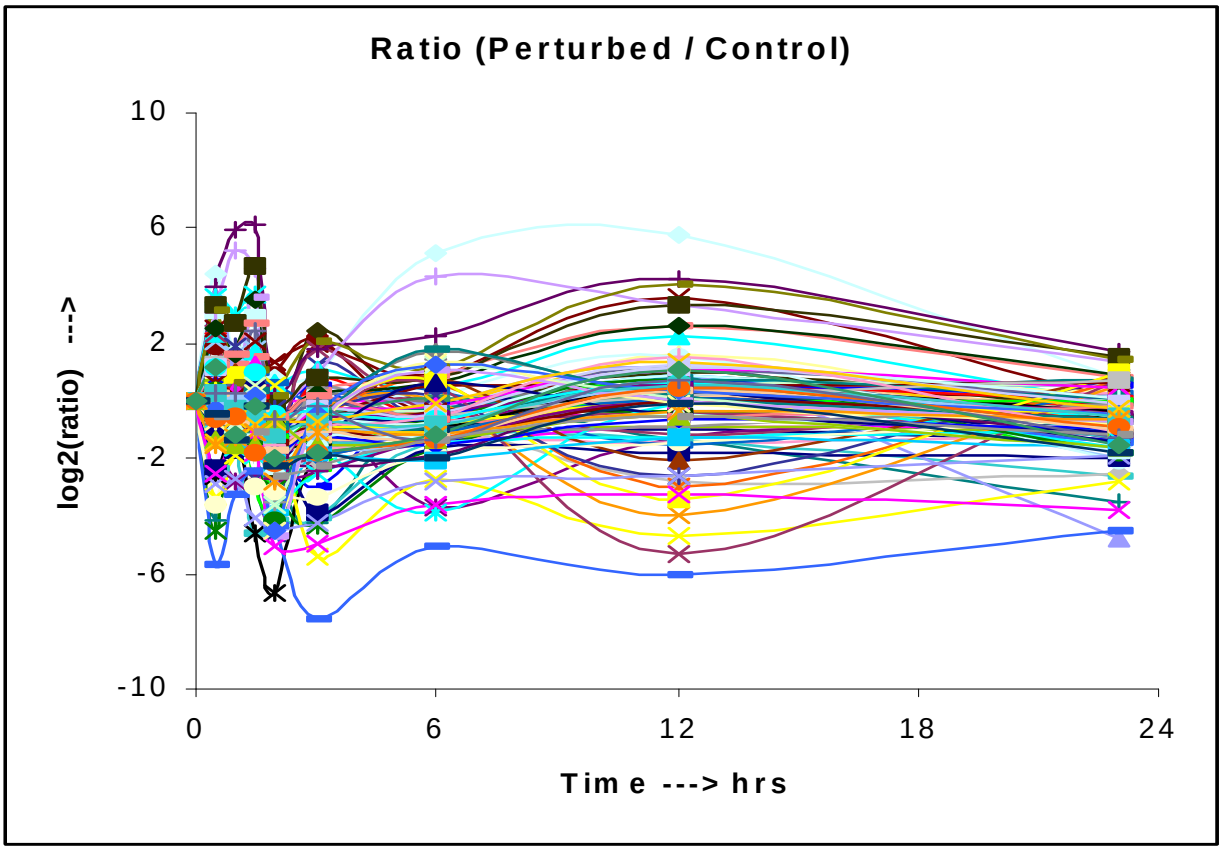


Figure 5.13 Graph of log-ratio vs. time for 150 metabolites obtained after normalization.

metabolic expression analysis which will be important for achieving the larger goal of integrated data analysis of the experiment. Hence Multi Experiment Viewer (MeV) of TIGR – TM4 software was used for the current analysis as it allows visualization and analysis of the data using different multivariate statistics algorithm.

In order to use the MeV software, data files in the TAV format are needed. Two sets of TAV files were generated for the current analysis. For analysis requiring separate metabolic profiles of control and perturbed system, TAV files were

generated by replacing the CY3 column values with relative area of the metabolite at time zero, and CY5 column values with relative area of the metabolite at time 't': so that the ratio of CY5/CY3 represents the normalized area of the metabolite at time t. To carry out data analysis using ratio of normalized peak area, TAV file for a particular time t was generated by replacing the CY3 intensity column with control Normalized peak area and Cy5 intensity column with perturbed normalized peak area, so that the ratio CY5/CY3 would indicate the ratio of the metabolites.

5.4.1 Experiment Clustering with Principal Component Analysis:

In order to understand the effect of elevated CO_2 on the overall metabolic state, principal component analysis and HCL algorithm was used. Using the TAV files for normalized peak area for control and perturbed, Principal Component Analysis was carried out using Pearson Co-relation distance. The projection of experiments using Pearson co-relation distance, the first three principal components accounted for around 65% of the data and separated the perturbed plant samples from the control as shown in Figure 5.14 (A). Also the plant samples representing the initial response of plants in 0.5-2hrs were separated from the samples representing longer response of plant (3-23 hr) in both the control and the perturbed system. The point representing time zero of the perturbed and the control system are located at the center of the plot.

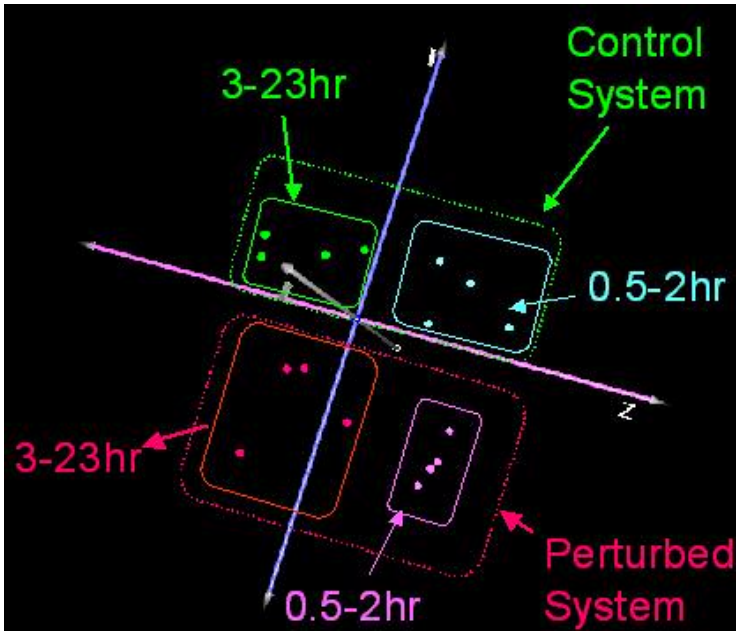
Hierarchical clustering using Pearson co-relation distance gave a similar result as Principal component analysis which is shown in figure 5.14(B). In both the clustering patterns the control experiments formed a separate cluster from the control system. Also within the control and perturbed normalized area, the initial time points 0.5-2 hr formed a separate cluster as compared to the longer response.

Hence from both the perturbed and the control systems we could see that the plants had a different initial metabolic state as compared to the longer response. Since this differential response was observed in the initial time points for both the controlled and the perturbed system, we decided to analyze the longer time points separately (from the samples representing the shorter time points) for further analysis.

5.4.2 Identifying Significant Metabolites:

For identifying metabolites that show differential expression in response to elevated CO_2 in a non-biased, reproducible, high throughput manner, multivariate statistical techniques were used. At present, no known example of identifying metabolites, which show a differential response to environmental perturbation using time series metabolic data, is available in the literature. Since one of the aim of the current research is to carry out an integrated genomic and metabolomic analysis of the system, the established methods and tools of gene

A



B

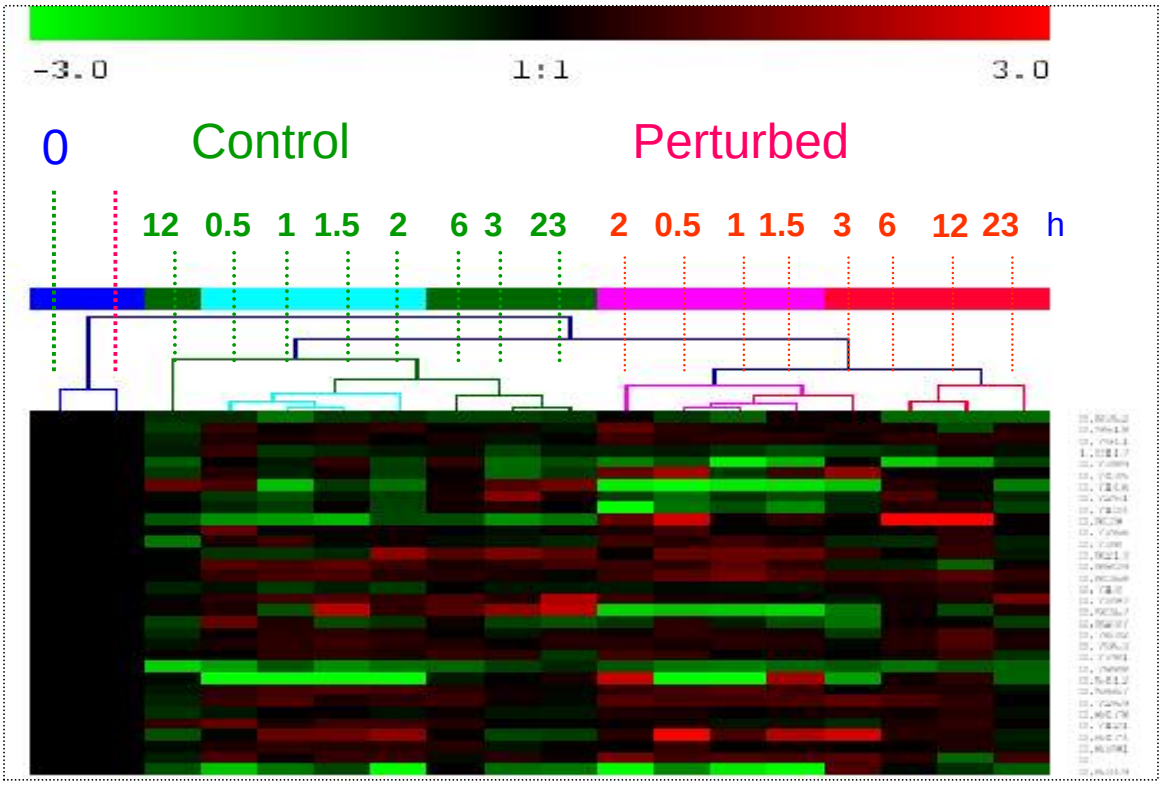


Figure 5.14 Clustering of experiments using using pearson Co-relation using (A) Principal component analysis (B) Hierarchical clustering analysis

expression analysis for identifying differentially expressed metabolites were used for the metabolic expression analysis.

Significant Analysis of Microarray (SAM), allows identification of genes which show differential expression; the same was used to identify differentially expressed metabolite. In order to identify metabolites which show significant differential expression when comparing normalized metabolite of each time point of the perturbed set to the same time point in the control set, Two Class - paired SAM option was used. TAV files of the perturbed and control representing the same time point were paired with each other. The standard default for number of permutations and SO percentile (100 and 5% respectively) were used. The imputed matrix was calculated using the default 10 nearest neighbors of K – nearest neighbor imputers. A delta value of 0.91 with Median number of false significant genes 0.657 (the smallest number above zero false significant gene) was used. The SAM graph obtained from this analysis is shown in Figure 5.15(A). The SAM analysis identified 37 negative significant metabolites, i.e. metabolites which show reduction in the perturbed system as compared to control system (enclosed by green ellipse). Figure 5.15(B) indicates a typical SAM graph from gene expression analysis. Comparing these two graphs we can see that, the SAM curve for metabolic expression does not pass through the origin and shows a negative intercept as compared to SAM curve for gene fr

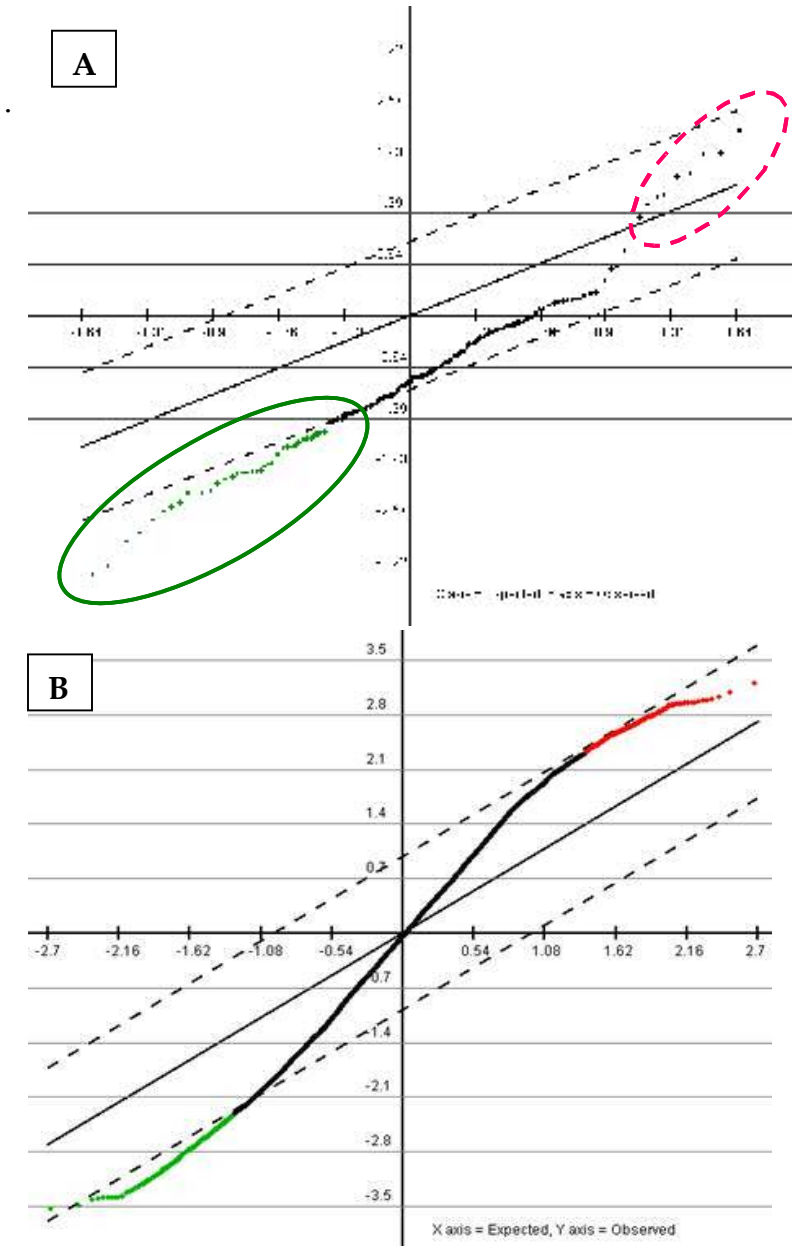


Figure 5.15 Two class paired SAM Analysis (A) metabolic profile (B) Gene expression

expression. Due to this effect no positively significant metabolites were identified from the SAM analysis. However the curve shows a marked change in the slope for the last nine metabolites in the first quadrant indicating positive differential

expression of the metabolites. Hence these metabolites were marked as differential expressed (indicated by the red ellipse).

In order to ascertain these results, K-Means Clustering (with Euclidean distance) of the ratio of normalized area of the metabolites was used. In order to identify the number of clusters that should be used for the K-Means analysis FOM analysis was used and the graph obtained from the same is shown in Figure 5.16 (A). Based on the FOM analysis, K-Means analysis was performed for 9 clusters.

The K-Means Clustering obtained using this analysis is shown in Figure 5.16(B). The comparison between SAM and K-Means clustering indicated that the nine positively significant metabolites found from SAM, form a separate cluster (cluster with profiles shown in red color) using K-Means Clustering. The same nine metabolites also form a separate cluster in PCA of the metabolites using Euclidean distance as shown in Figure 5.17. Thus combining all the three analysis we can conclude that these nine metabolites show differential expression, and the list of these metabolites is available in Table 5.3. In case of 37 negatively significant metabolites found from SAM analysis, these metabolites are distributed in three K-Means Cluster 1,3 and 8. However the K-Means clusters also include certain other metabolites showing reduction in their normalized area, but these are not part of the 37 negatively significant metabolites found from SAM analysis. The comparison of the metabolites can be seen from Table 5.3.

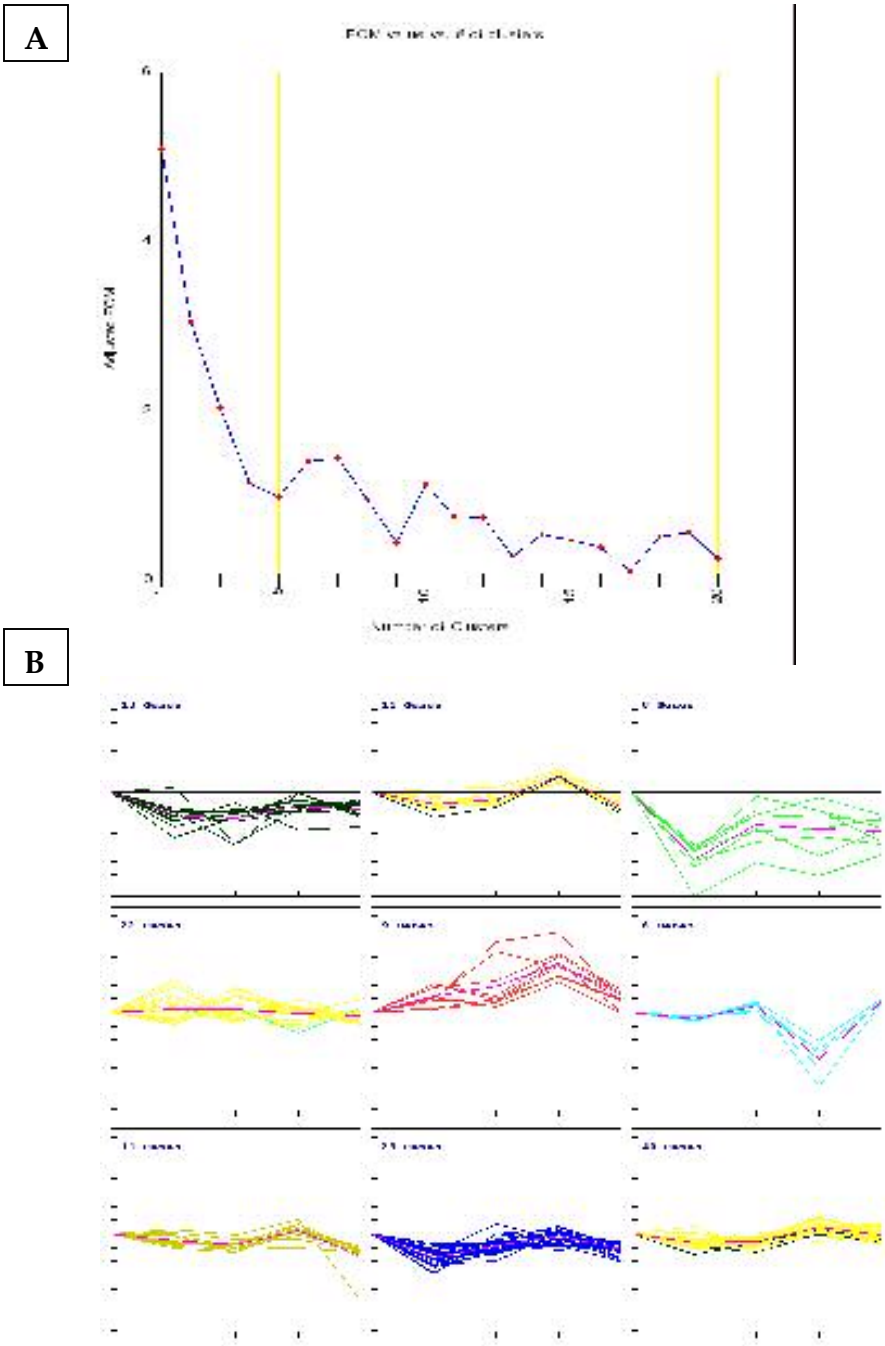


Figure 5.16 (A) FOM Analysis using Euclidean distance (B) K-Means clustering using Euclidean distance

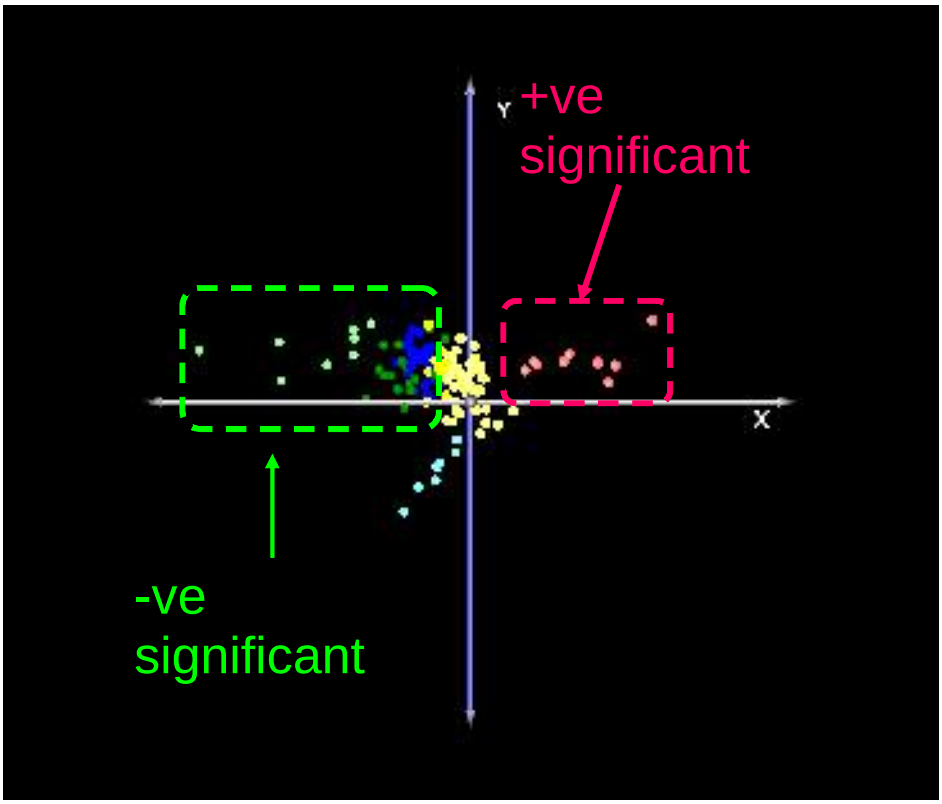


Figure 5.17 Principal Component Analysis for metabolites with

SAM		K - Means Clusters			
-ve Significant	+ve	CL-1	CL-3	CL-8	CL - 5
Asparagine	Glycerol	Fructose- 6 - P	Asparagine	Citrate	Glycerol
Glutamate	Arabinose	Glucose - 6 - P	Glutamate	iso-Citrate	Arabinose
Ornithine	Xylitol	Serine	Ornithine	Glycerate	Xylitol
Citrate	6 unknowns	9 Unknowns	Ethanolamine	Inositol – P	6 unknowns
iso-Citrate			4 Unknowns	Phenylaline	
Lactate		*Glycine N, O*		Glucarate	
Fructose - 6 – P				Lactate	
Glucose - 6 – P				10 unknowns	
Inositol – P					
Glycerate				*4-Aminobutyrate*	
Phenylaline				*Asparatate*	
Serine				*β - Alanine*	
Glucarate				*Succinate*	
Ethanol Amine				*8 Unknowns*	
23 Unknowns					

Table 5.3 Comparison of +ve and -ve significant metabolites from SAM with K-Means Clusters

5.4.3 Identifying co-relation between Metabolites:

In order to identify metabolites showing similar response to elevated CO2 in the atmosphere, K-Means clustering with Pearson Co-relation distance was used. The Clustering Pattern obtained using 9 clusters is shown in Figure 5.18. The color used for a particular metabolite is the same as the one used in Euclidean clustering. We can see that using Pearson correlation we now have a different clustering pattern due to regrouping of the metabolites, such that most of the metabolites showing similar profile, cluster together. The list of metabolites found in each cluster is given in Table 5.4

5.4.4 Labeling analysis:

In control and the perturbed systems, as discussed in section 5.1, 10% of the CO_2 used was C^{13} labeled. The purpose of using the labeled CO_2 was to measure the difference in rates at which the labeling in the control and perturbed system would change for different metabolites, with time. In order to understand effect of the change in labeling, the redistribution of the isotopomers was studied for each individual isotopomers, using the change in the peak area. For e.g. Figure 5.19(a) represents the Glutamate peak in the chromatograph and its typical mass spectrum is shown in Figure 5.19(b).

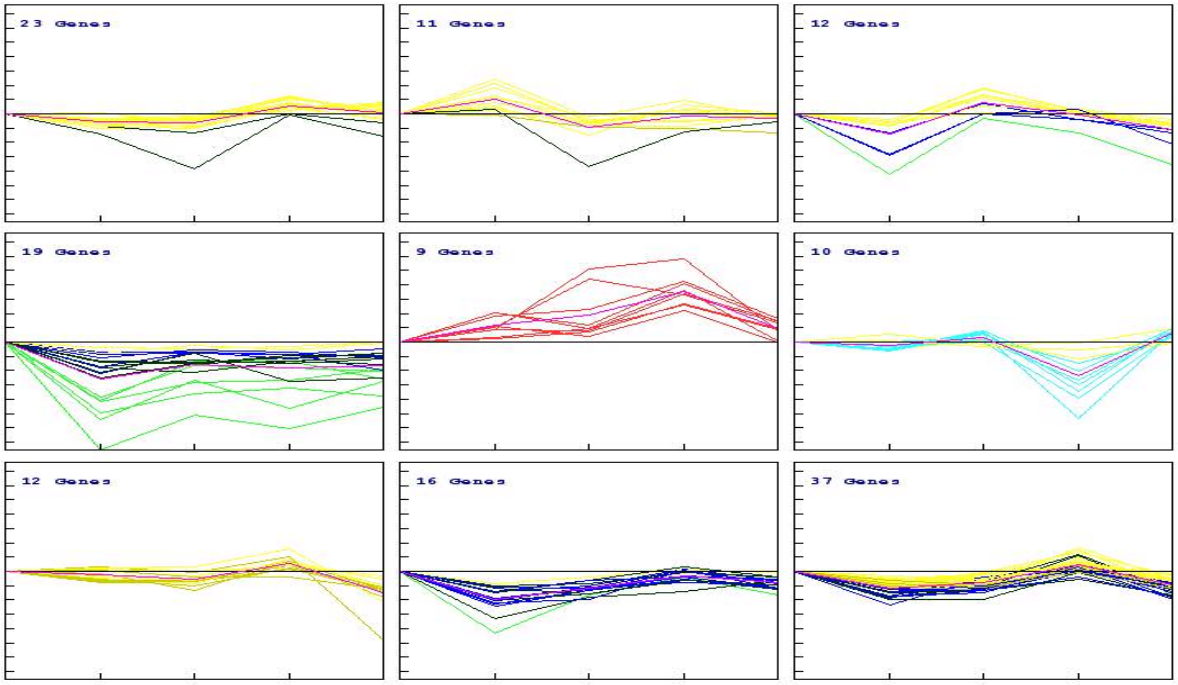


Figure 5.18 K-Means Clustering – Pearson co-relation distance to identify metabolites having similar response to elevated CO2 perturbation.

Cluster 1	**Cluster 2**	**Cluster 5**	**Cluster 9**
3-phosphoglycerate	3-hydroxyglutarate	arabinose	2-methylmalate
Aconitate	Glycine	glycerol	4-aminobutyrate
Fructose	Lactate	xylitol	Aconitate
Galactose	Oxalate	6 unknowns	Aspartate
Gluconate	Threonate	**Cluster 6**	citrate
Glucose	6 unknowns	10 unknowns	Cytosine
Inositol	**Cluster 3**	**Cluster 7**	Fumarate
proline	beta-alanine	sorbitol	Glutamine
Mannose	Ethanolamine	tyrosine	Glycerate
Shikimate	10 unknowns	10 unknowns	Homocystine
Sorbitol-6-P	**Cluster 4**	**Cluster 8**	inositol-6-P
11 unknowns	Asparagine	ascorbate	iso-citrate
	glucose-6-P	fructose-6-P	malate
	ornithine	glutamate	Phenylaline
	Serine	succinate	phosphoric acid
	15 unknowns	10 unknowns	Glucarate
			Threonine
			21 unknowns

Table 5.4 Clusters of metabolites obtained using K-Means Clustering

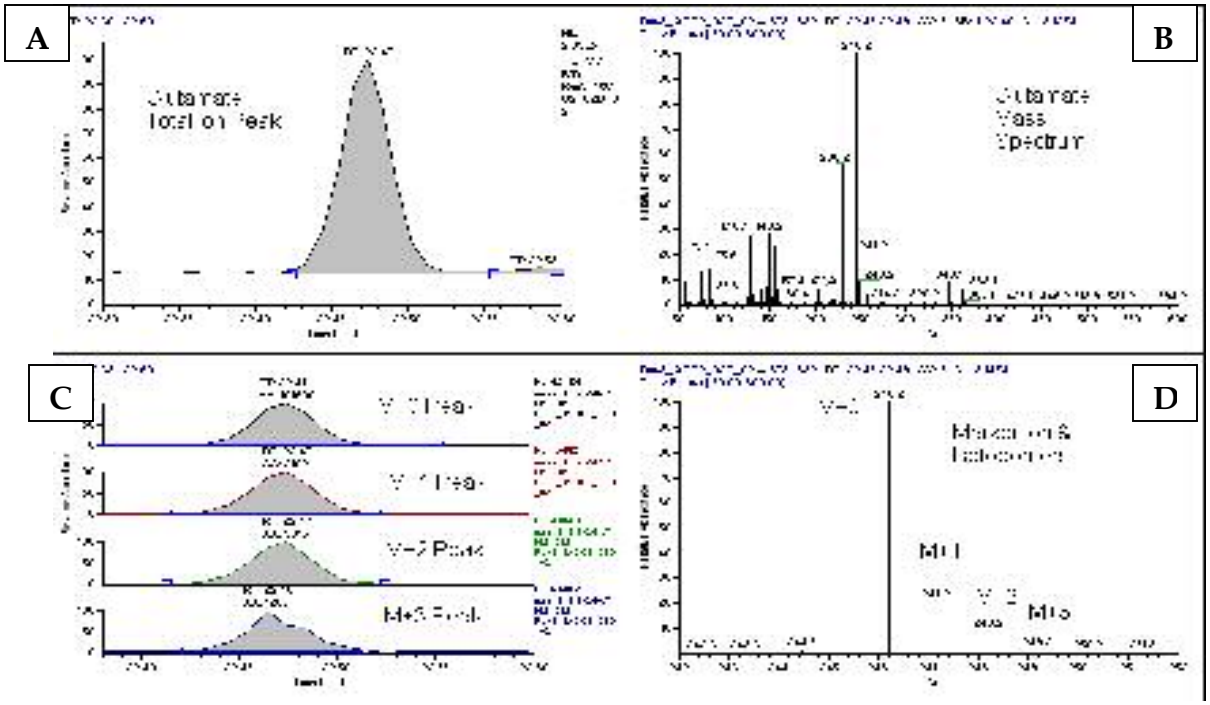


Figure 5.19 (A) Glutamate TIC Peak (B) Mass Spectrum (C) Glutamate marker ion individual peak area (D) Isotopomers of the marker ion

We can also quantify the relative abundance of each isotopomer using its peak area as shown in Figure 5.19(c). The ratio of peak area that sentence fragment represents the relative abundance of the isotopomers in the plant sample. Since the time zero plants were harvested before connecting the plants to cylinder containing labeled CO_2, the distribution of isotopomers in these plant samples indicates natural abundance of the isotopomers in plants. Any deviation from this distribution in other plant samples would be the result of labeled CO_2 used in the experiment. In order to identify this deviation, we calculated the fraction of a particular isotopomer peak area over total area of all the isotopomers combined, which is shown in Figure 5.20(a). In order to better compare the

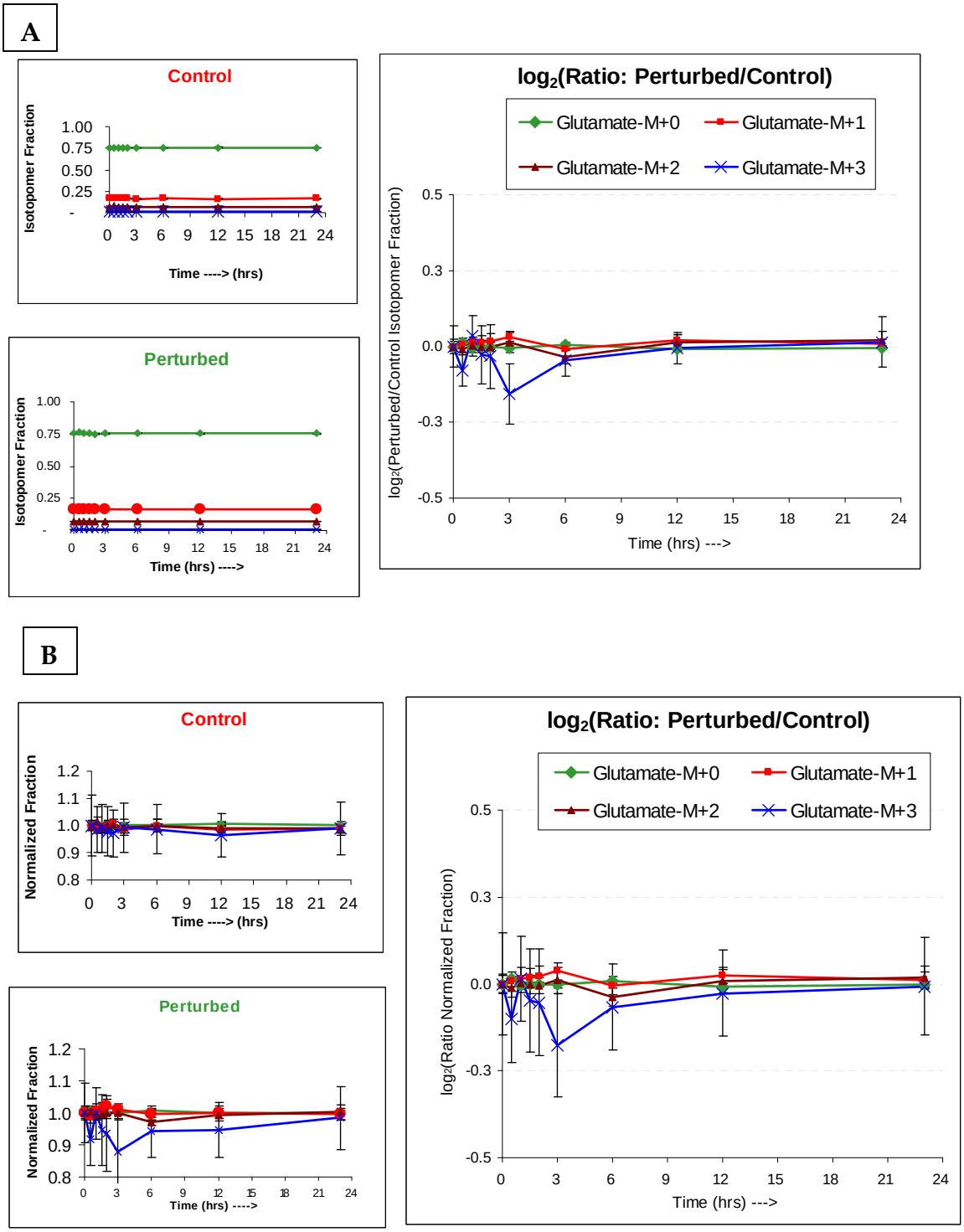


Figure 5.20 (A) Individual isotopomers fraction in control, perturbed system and their ratio (B Time zero Normalized isotopomers fraction in control, perturbed system and their ratio.

profiles the fraction at each time point was divided by the fraction measured at time zero which represents the natural abundance. This normalized mass fractions are shown in Figure 5.20(b). The ratio of the normalized fraction at each time point in perturbed system, to the normalized fraction at the same time point in the control system, was taken as shown in Figure 5.20(b). By viewing the ratio, one can compare the rate at which the metabolite in plant gets labeled. If the fixation of labeled CO_2 was faster in the perturbed system as compared to control, that would increase the Ratio normalized fraction [M+1] and decrease the ratio of normalized fraction for [M+0] isotopomer. However since in the current experiment, only 10% of the CO_2 used was labeled, we do not expect an increase of more than 10% in the isotopomer area fraction. However we can see from Figure 5.19 that the instrument variability itself is of the order of 10% and hence it would be difficult to separate the variation in the measured isotopomer fraction, due to use of labeled CO_2 from that of variation due to instrumental variation confusing sentence. Only [M+3] isotopomer of glutamate shows some deviation in normalized ratio, however this is more likely to be the result of noise and other instrumental variation which significantly affects [M+3] peak area as it is the smallest peak with lowest signal to noise ratio, as can be seen from Figure 5.20(c). Similar such comparison of isotopomers for more metabolites is given in Appendix VII of the report.

Chapter 6. Discussion of Results

6.1 Metabolic profiling Protocol:

The experimental protocol used in the described analysis allowed for the identification and quantification of ~212 polar metabolites in the plant samples. This number is comparable to the 214 polar metabolites that were detected in *A. thaliana* leaves using similar protocol [Fiehn et. al., 2000a]. The current analysis was performed at a split ratio of 75:1, i.e. only 1/75 of the 1 µl sample injected was actually used for analysis as compared to split ratio of 25:1 in previous studies [Roessner et. al., 2000, Fiehn et. al., 2000] which used quadrupole mass spectrometer. This is an indication that the ion trap mass spectrometer provides a more sensitive analysis platform for metabolic profiling. However the ion-trap MS has the disadvantages of low reproducibility at high metabolite concentration due to the saturation and space-charge effects [Kitson et. al., 1996]. These effects were not anticipated in the used ion-trap MS (Thermo Finnigan, Inc), because it has been designed in such a way that the ionization source is separated from the ion-trap. In any case, the split ratio was chosen in such a way that the calibration curves of ribitol (Figure 5.3) were linear over a wide range of concentrations, thereby ensuring that the equipment was functioning at the linear detection range for the measured metabolites. The use of marker ions for the metabolite quantification, allows for co-eluting metabolites to be quantified;

this would not have been possible using TIC peaks. It also enables the quantification of metabolites in very low concentration in the plant sample, by allowing better signal to noise ratio. From current analysis, a library of metabolites along with their retention times and marker ions has been developed. Since previous libraries do not contain the marker ion for each metabolite that gives the best separation of co-eluting metabolites, such a library will allow for faster analysis of future metabolic profiles that are obtained from the metabolic profiling protocol established in the context of the presented work.

Data Filtering and Normalization:

For the comparison between metabolic profiling data of a control and a perturbed set of samples to provide any meaningful results, the variation between the samples because of the applied perturbation should be higher than the variation due to biases in the experimental setup and protocol and biodiversity. As it has been shown in Chapter 5 the variability due to the experimental setup and protocol is estimated from the variation between the various (3) injections of each samples, while the variability due to biodiversity is estimated from the variation between the duplicate samples at each time point. In previous metabolic profiling study the average instrumental and biological variabilities were estimated ~8% and ~35% ,respectively [Fiehn et. al., 2000a]. Similar values were obtained in the current analysis (see Table 5.1).

6.2 Discussion of the metabolic profiling data in the context of *A. thaliana* physiology:

As discussed in Chapter 4, The typical experiment design in most of the previously reported plant metabolomics studies was involved the growth of two sets of plant samples - having either different genetic backgrounds [Roessner et. al., 2001a, Roessner et. al., 2001b, Fiehn et. al., 2000a] or being subject to different environmental conditions [Roessner et. al., 2001a] - for a specific period of time. PCA and HCL was then subsequently used to show the difference in the metabolic profiles of the two plant sets. In the current analysis, for the first time time-series data over the growth of two plant sets in different environmental conditions were obtained and their metabolic profile was measured. PCA & HCL analysis of the two sets of data (control and perturbed) shown in Figure 5.14 clearly indicates that the presence of elevated CO_2 (even for a very short exposure) alters plant metabolism significantly, as plants grown under elevated CO_2 condition form a separate cluster from the control system. This indicates that even if plants are not grown for their entire growth cycle under different conditions, they exhibit a differential metabolic response, which can be measured by metabolic profiling technique. The clear separation of the plant samples in the perturbed and control system, from the beginning, also indicates that the change in the metabolic state of the plant, in response to the perturbation, is bigger as

compared to the changes within a system, of the normal growth cycle of the plant.

6.2.1 Identification of Differentially expressed Metabolites:

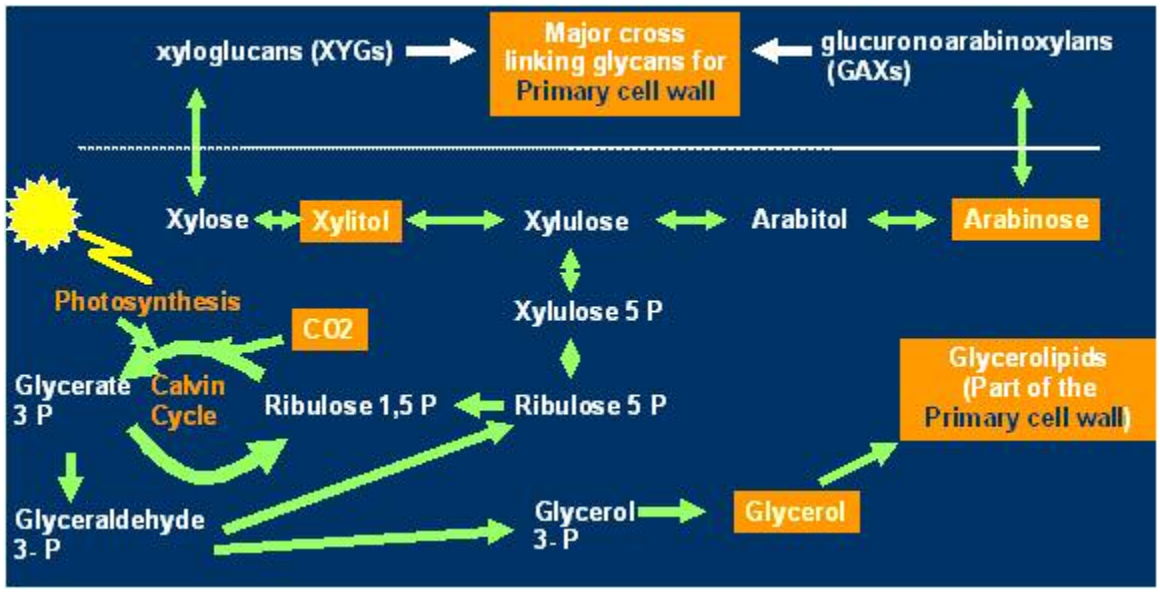
The metabolic analysis techniques, used in previous long term metabolic studies, of elevated CO_2, were designed to measure for the change in the concentration of specific metabolites because of the applied treatment. The high throughput nature of metabolic profiling allows for the simultaneous measurement of many metabolites in different functional groups. Previous studies using metabolic profiles of plants, used classical t-test to identify metabolites; this showed a differentiated expression between two sets of plants. This analysis is akin to comparing two "averaged metabolic snapshots" of plants and identifying differentiated metabolites. In the current analysis, however, the aim is to compare multiple metabolic states, represented by plants harvested at different time points of a control system, with those of a perturbed system, at the same time points. The t-test approach cannot be used for such a problem. Instead, two class paired Significant Analysis of Microarray (SAM) was used. Using this analysis, negatively and positively differentiated metabolites can be identified. However, comparison of the SAM graph obtained from metabolic profiling results, to that of gene expression analysis, as can be seen from Figure 5.15, indicates that the metabolic profiling analysis shows a negative intercept on the

Y axis (representing Observed variation), which is not shown by the gene expression data. Since the SAM analysis is designed for microarray gene expression analysis, in which all the samples are adjusted to have the same total mRNA concentration, the increase in gene expression of certain genes has to be compensated by an almost equivalent decrease in the expression of other genes. The negative intercept in the metabolic analysis can possibly be explained as follows:

1. In the current study, only the concentration of the polar metabolites was measured. An increase in the polar to non-polar metabolite ratio in the plant biomass due to the applied perturbation means that the total polar metabolite concentration per gram of plant is decreased. Measuring only the polar metabolites, it is expected that most of them are under produced in the perturbed compared to the control system. Since no non-polar metabolites were measured such decrease is not "compensated" in some by the equivalent increase in the non-polar metabolite concentration and therefore the Y-intercept is negative.
2. After performing normalization of data w. r. t. time zero, the data for all the metabolites gets scaled around 1 and the information about the relative concentration of the metabolite is lost. Due to this, a 10 % change in a fructose concentration receives the same weightage in the analysis as

50% change in xylitol even though xylitol has relative area almost 100 times less as compared to fructose. Now consider a scenario in which a small increase (for e.g. about 10%) in the metabolite with high relative concentration (e.g. fructose, glucose or sucrose) is achieved through large reductions (for e.g. about 75%) in many metabolites with smaller concentration, the total normalized area of the large molecules being produced increases only by 10% but the normalized area of the smaller concentration metabolite would change by 75%. Under such a condition, the total normalized area would indicate a decrease, though in absolute terms there is no real decrease of mass in such an analysis.

Using the SAM analysis, 37 negatively significant and 9 positively significant metabolites were identified. In order to confirm this result, we also performed a K-Means analysis, and Principal component analysis for the metabolites. The 9 positively significant metabolites, clustered together in K-Means analysis (Cluster 5, with red colored profiles in Figure 5.16), and also formed a separate cluster in principal component analysis, as seen from Figure 5.17. From the 9 positively significant metabolites obtained using all the three analysis, (from Table 5.3) only three of them have known structures. The three metabolites are Xylitol, Arabinose and Glycerol, and their profile is shown in Figure 6. 1. As can be seen from Figure 6.1, all the three metabolites are used in production of



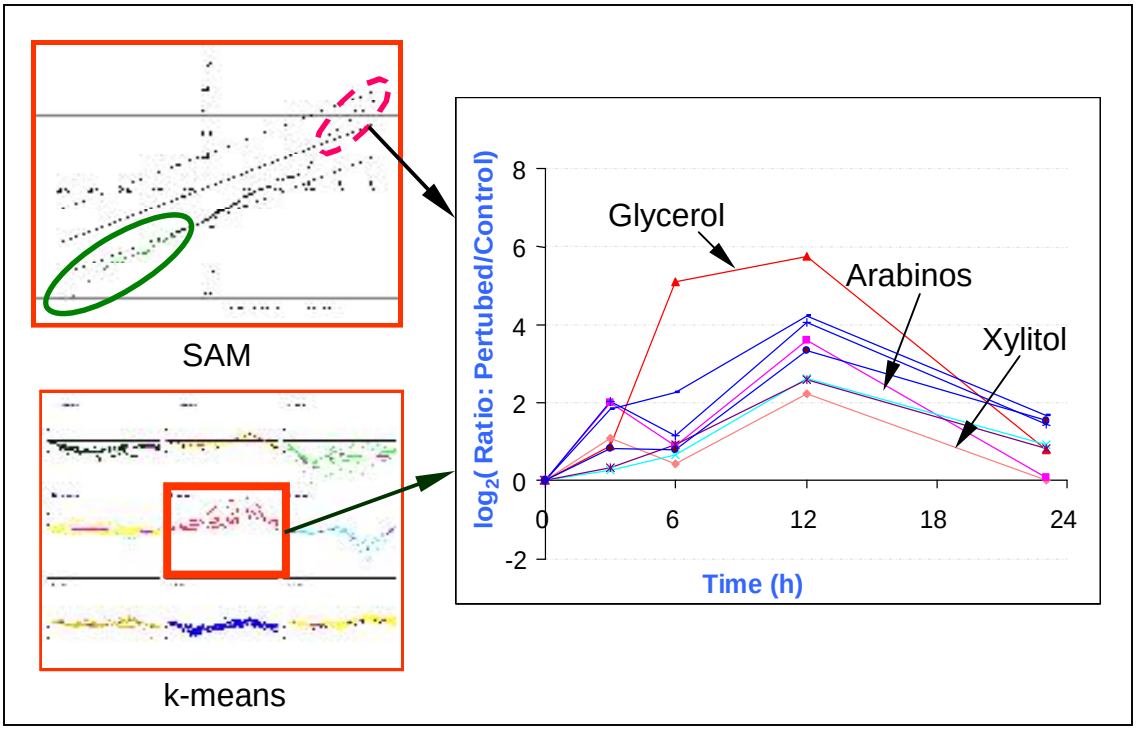


Figure 6.1: Positive Significant Metabolites: Constituents of primary cell wall

primary cell wall. Xylitol and arabinose are used for the production of xyloglucan and glucuroarabinoxylans, which are structural carbohydrates used for cross linking in the primary cell wall. Specifically, the xyloglucan is known to

increase during cell wall expansion, facilitating the loosening of the primary cell wall [Dey et. al., 1997]. Apart from structural carbohydrates, the plant cell wall also contains glycerolipids in the cell membrane. Since the current metabolic analysis was performed only for polar metabolites, and not for non-polar metabolites, the increase in the lipid could not be measured; however glycerol (which is the polar part of the glycerolipids used in the cell membrane), showed an increase in the perturbed system, with the same order of magnitude as the one shown by xylitol and arabinose, thus, indirectly indicating an increase in the plant lipid content, which has not been observed before.

The long term exposure of elevated CO_2 has shown accumulation of carbohydrates in plants, as the first major response of plant to elevated CO_2. However this increase has been observed in non structural energy storage carbohydrates like sucrose and starch. Since sucrose was also used in the growth liquid media, which could have contaminated the plant sample, sucrose was excluded from the current analysis. Glucose and fructose the other two major constituents of the hexoses pool and constituents of starch, even though unidentified as positively significant, did, however, as shown in Figure 6.4.2, show more increase in the perturbed system after an initial decrease at the end of 23 hours. However comparison of Figure 6.1 and Figure 6.2 indicates a much more dramatic change in the increase in structural carbohydrates, (which form

the primary cell wall) as compared to the hexoses pool (non-structural carbohydrates).

This has not been reported in the literature before, and is similar to the increase in the Glycerolipids, or the lipid content in the plant. The increase in structural carbohydrates and glycerol lipids — constituents of the primary cell wall —

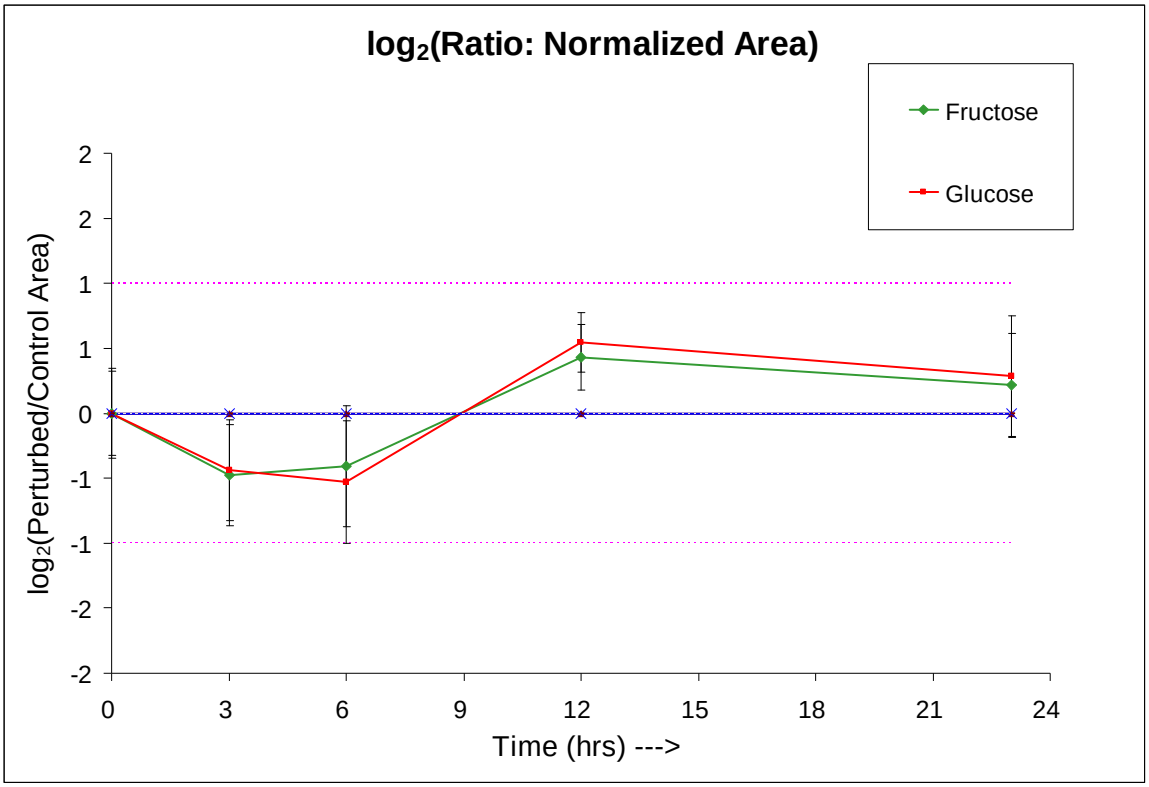


Figure 6.2 Response of major non-structural carbohydrates

could be the immediate plant response to elevated CO_2. This may not be a long term effect and hence may have been undetected in the previous study. The other possible explanation can be, that since the previous metabolic analysis were designed specifically to measure effect of constituents like starch, glucose, sucrose, etc., changes in metabolites, specific to cell wall, would not have been

measured or detected. This also indicates one of the advantages of using high throughput metabolic analysis technique, as compared to focused metabolic analysis techniques, since the former is comparatively sensitive and can detect results not directly related to initial hypothesis whereas the latter, usually, only proves or disproves a hypothesis. The observed increase in the lipid content and carbohydrates, also supports the possible explanation of the negative intercept in the SAM graph, as discussed before.

A similar analysis of the 37 negatively significant metabolites, identified from SAM, shows that these metabolites are distributed in three K-Means cluster as shown in Figure 6.3. Apart from containing the 37 negatively significant metabolites, the K-Means cluster also contain 12 more metabolites which show a decrease in concentration. These metabolites could also be identified from the SAM analysis by slightly reducing the delta value. Even though the negatively significant metabolites cannot be separated from the non-significant metabolites completely, in the PCA clustering of metabolites using Euclidean distance; the eight most negatively significant metabolites in cluster 8, separate out *clearly* from the rest of the metabolites. As can be seen from Figure 6.3, the most negatively significant metabolites, contain two of the four nitrogen storage and transport metabolites of the plant.

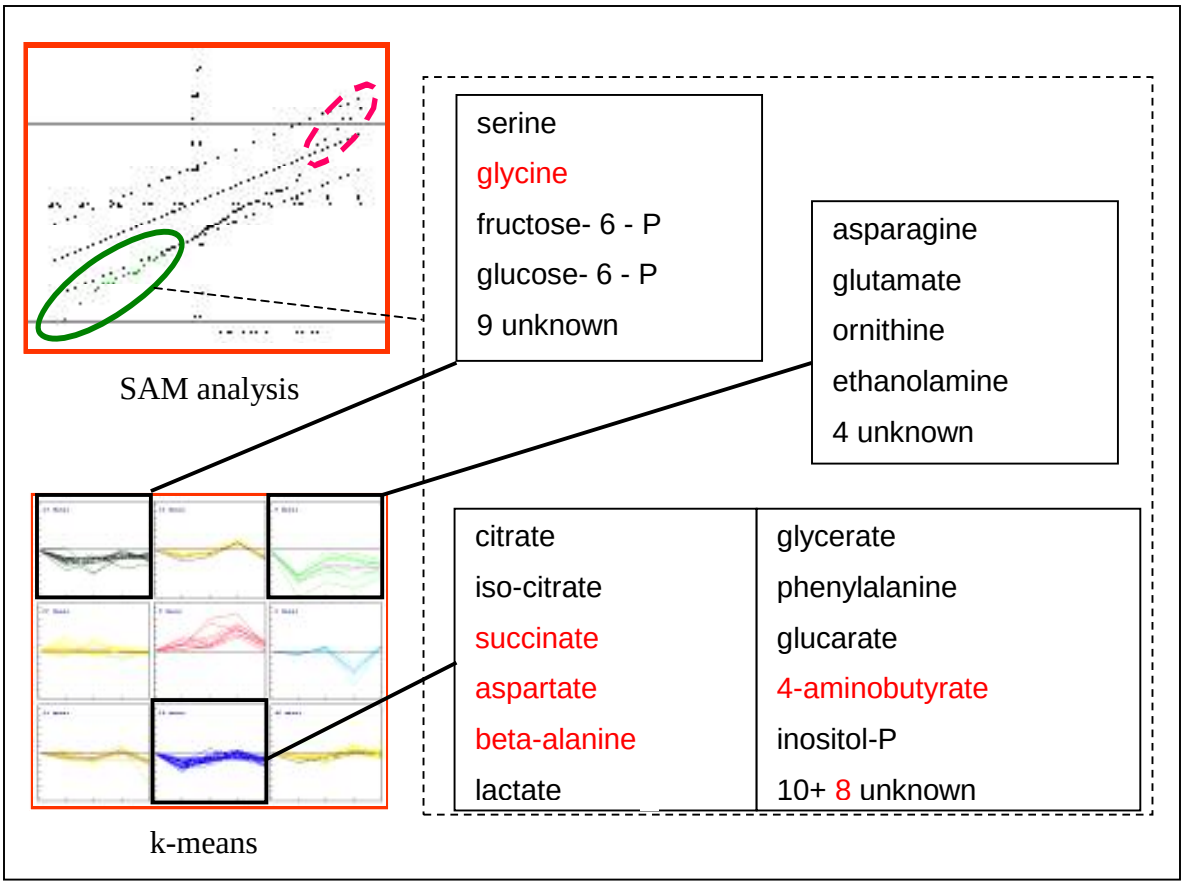


Figure 6.3 Comparison of negatively significant metabolites obtained from K-Means and SAM. (The metabolite names in red indicate metabolites part of KMC analysis but not found from SAM).

From the other two nitrogen storage metabolites, aspartate (even though not identified from SAM analysis) clusters along with other negatively significant metabolites in K-Means clustering. Glutamine also shows a small decrease in concentration, as shown in Figure 6.4 This decrease, in metabolites, is consistent with the observation in the long term by other researchers in decrease in organic nitrogen content (as discussed in Chapter 2). Thus the elevated CO_2 immediately down-regulates the nitrogen assimilation, as most of the metabolic stores show a significant decrease in nitrogen content during the first three hours itself. As discussed before, in presence of light, the plant uses glutamine as the principal

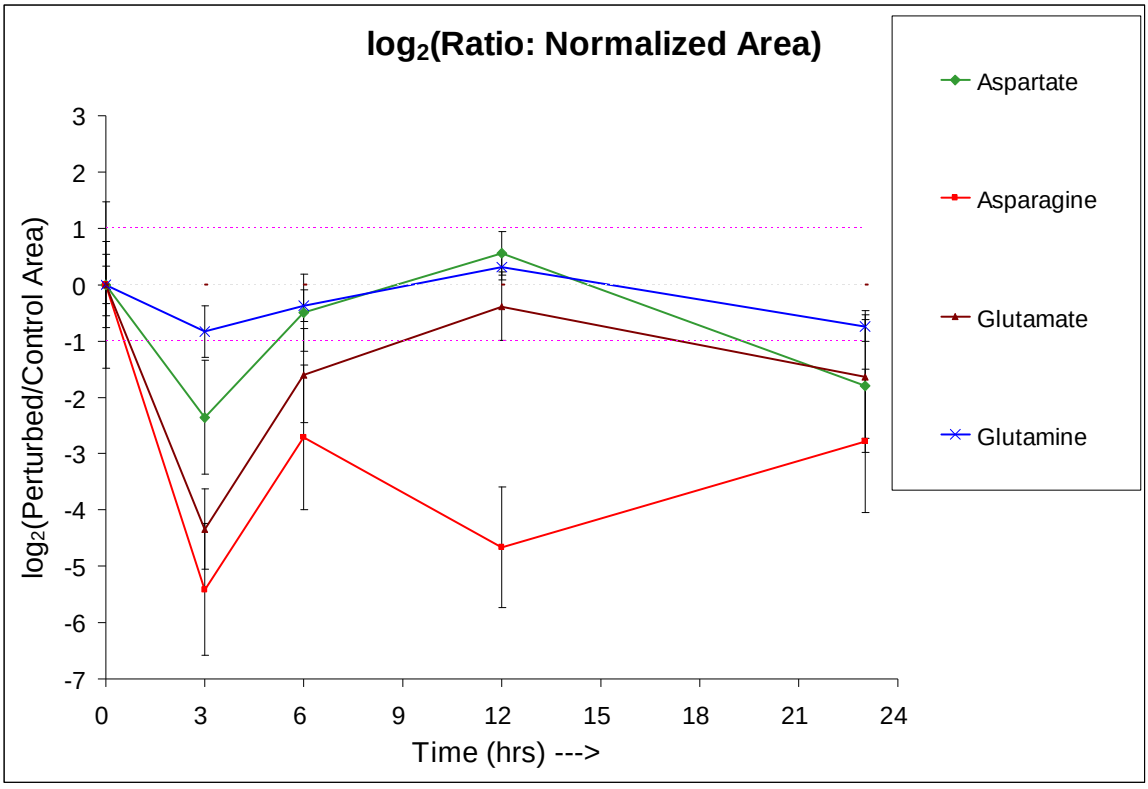


Figure 6.4 Response of plant nitrogen stores

nitrogen storage metabolite. Since the plants in the current experiment were grown under constant light conditions, glutamine may have been the principal nitrogen store for the plant. Glutamine also had a higher relative concentration as compared to aspartate and asparagine in the plant samples. Since glutamine does not show a large decrease in concentration, this supports the explanation for reduced nitrogen content as a result of transfer of reduction resources to photosynthesis in response to elevates CO_2, since the nitrogen is only available for a short term from the plant nitrogen stores. From the other metabolites detected, three negative, significant metabolites — glycine, serine and glycerate — belong to photorespiration pathway shown in Figure 6.5.

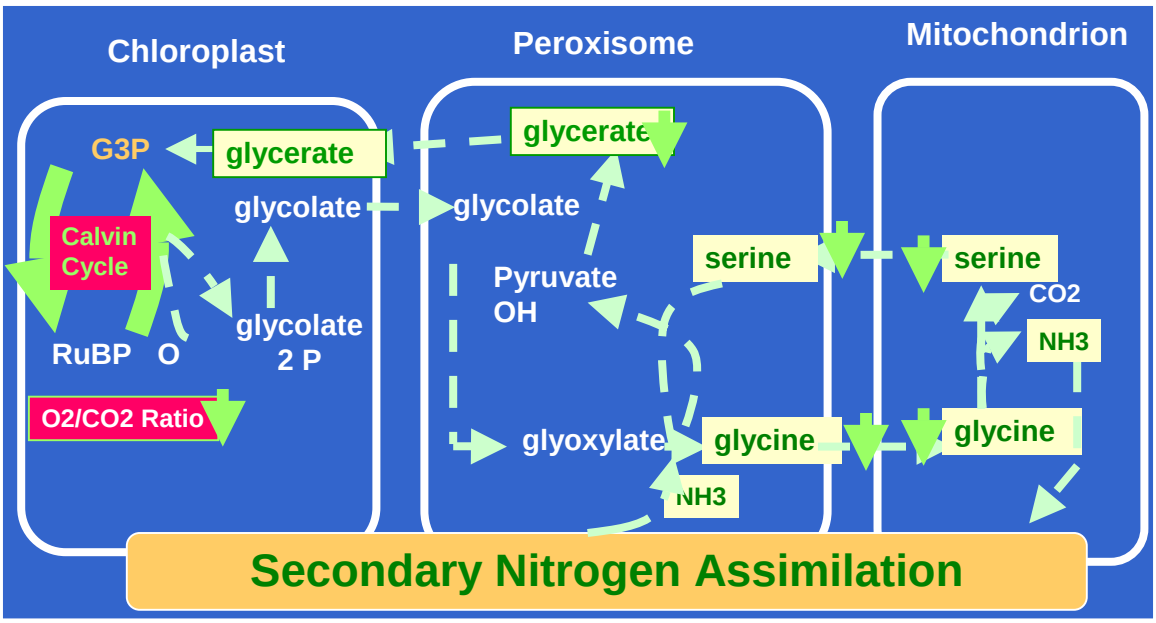


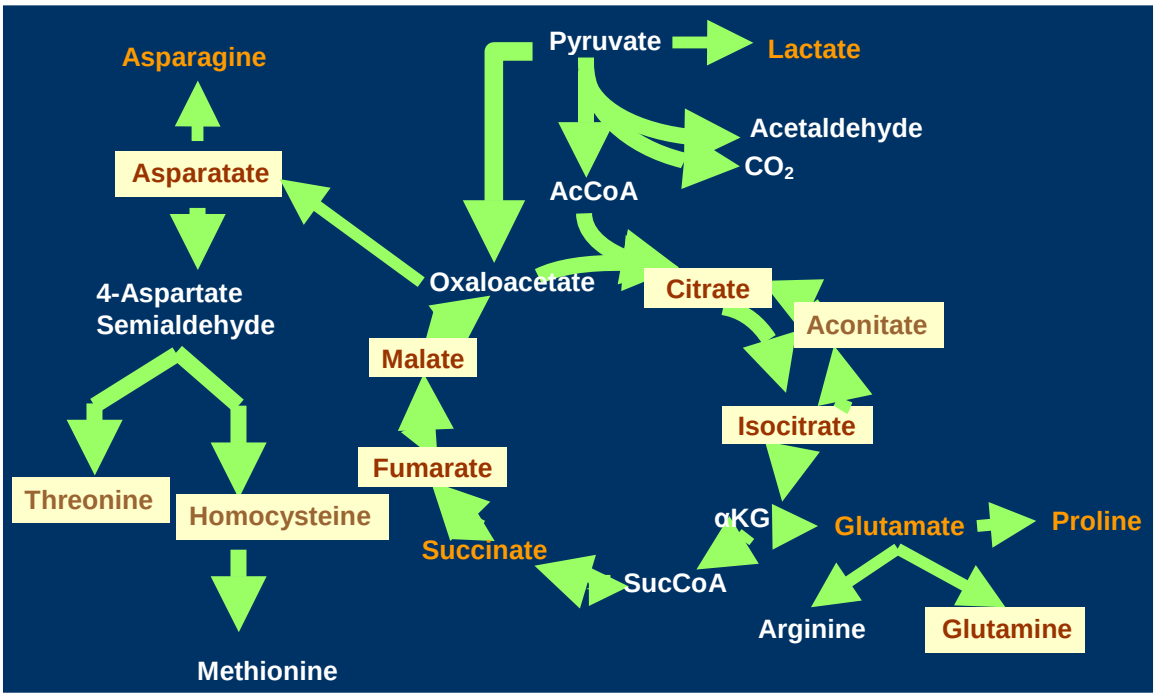
Figure 6.5 All observable metabolites of photorespiration pathway in plants exhibit a reduced

As discussed in Chapter 2, presence of 1% CO_2 in the growth environment is known to reduce the photorespiration pathway because of the competition for RuBisCo between CO_2 and O_2. Since the accumulation of glycine and serine have known to be associated with increased photorespiration, the decrease in all the three observable metabolites of photorespiration, support the previous studies that photorespiration is reduced in presence of elevated CO_2. The other metabolites showing a decrease were possibly related metabolically to the nitrogen storage metabolites discussed above (beta alanine, 4-aminobutyrate and ornithine). Apart from that, the hexose phosphate pool (Fructose -6- phosphate, Glucose-6-Phosphate and inositol phosphate), which is the starting point for production of larger molecular weight carbohydrates, part of the TCA cycle

metabolites (citrate, iso-citrate and succinate) show a decrease in concentration in the perturbed system along with lactic acid, phenylalanine, glucaric acid and glyceric acid. Based on the current information available, the exact reason for their decrease could not be ascertained.

6.2.2 Identifying correlation of metabolic data:

By performing k-Means clustering, using Pearson correlation distance, we obtained clusters of metabolites which showed a correlation in their pattern of response, irrespective of their absolute concentration. As shown in Figure 5.17, the 9 clusters obtained from the analysis, were different from the 9 clusters obtained from Euclidean analysis (which clustered metabolites based on their absolute value of change rather than the pattern of change). The response of known metabolites belonging to Cluster 9 and Cluster 1 — the two largest clusters in the analysis — are shown in Figure 6.6 and Figure 6.7. The response of positively significant metabolites, which still clustered together, has been already shown in Figure 6.1. All the known metabolites of Cluster 1 belong to TCA cycle or are closely related to TCA cycle as shown in Figure 6.6. From the 14 observable metabolites related to TCA cycle, when Pearson correlation is used, nine of the TCA cycle metabolites cluster together, indicating presence of strong correlation in their metabolic response. The metabolites that did not cluster with the same group were succinate, glutamate, proline, asparagine and lactate, while,



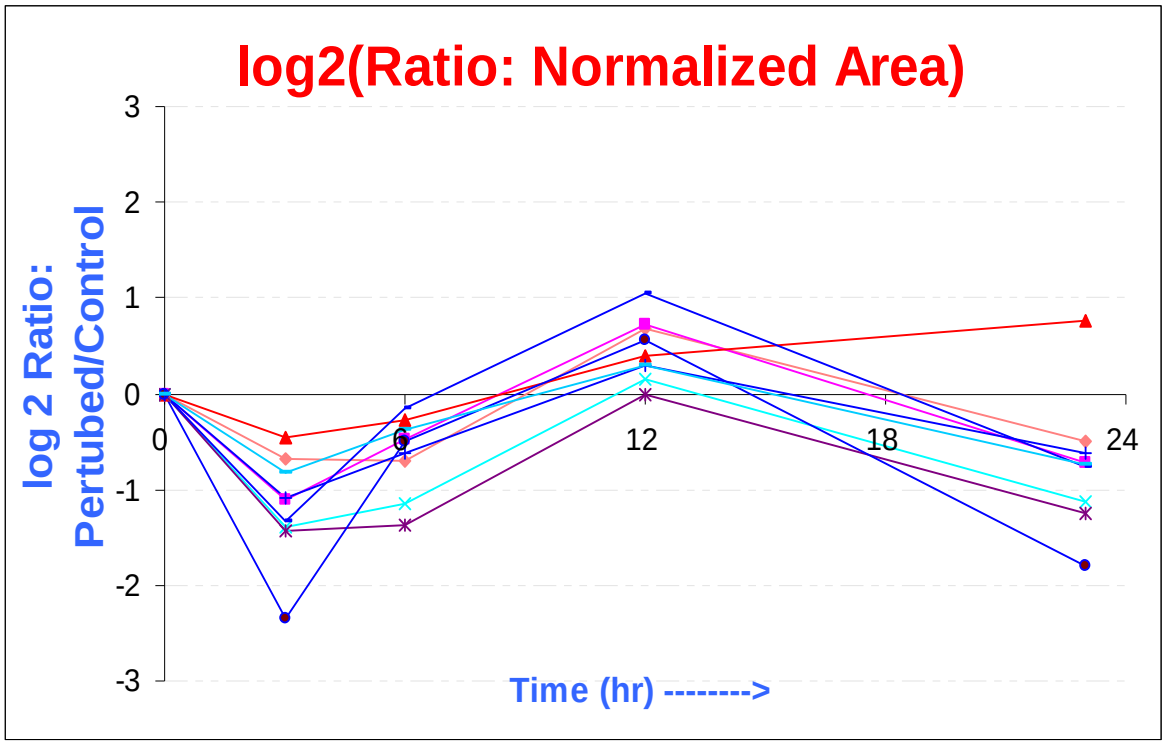
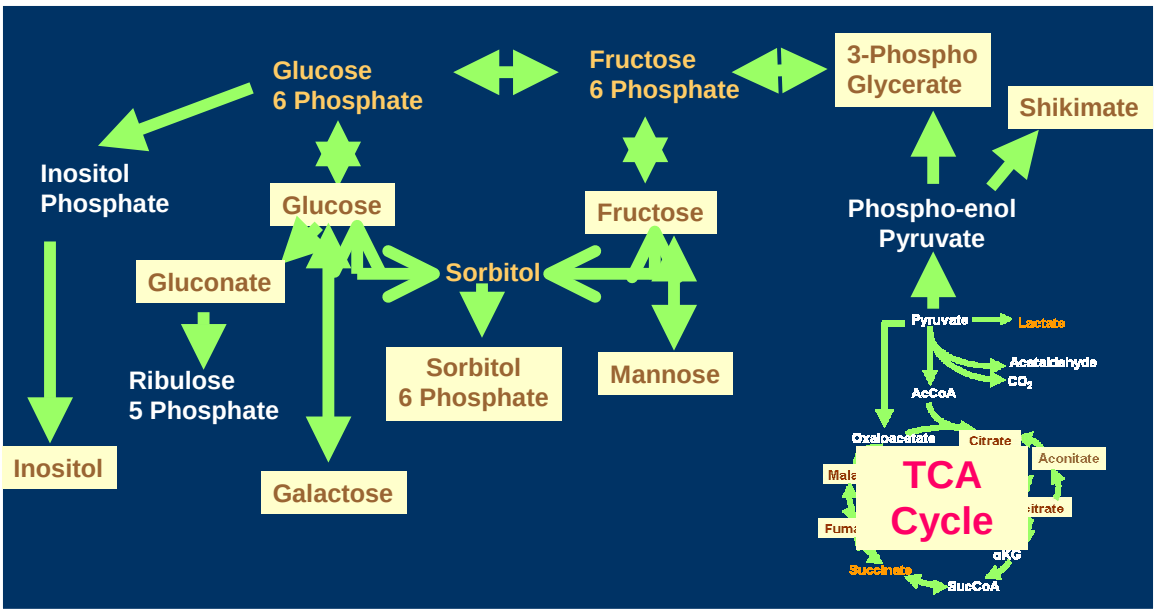


Figure 6.6 K-Means cluster showing correlation in TCA cycle and related pathway metabolites.



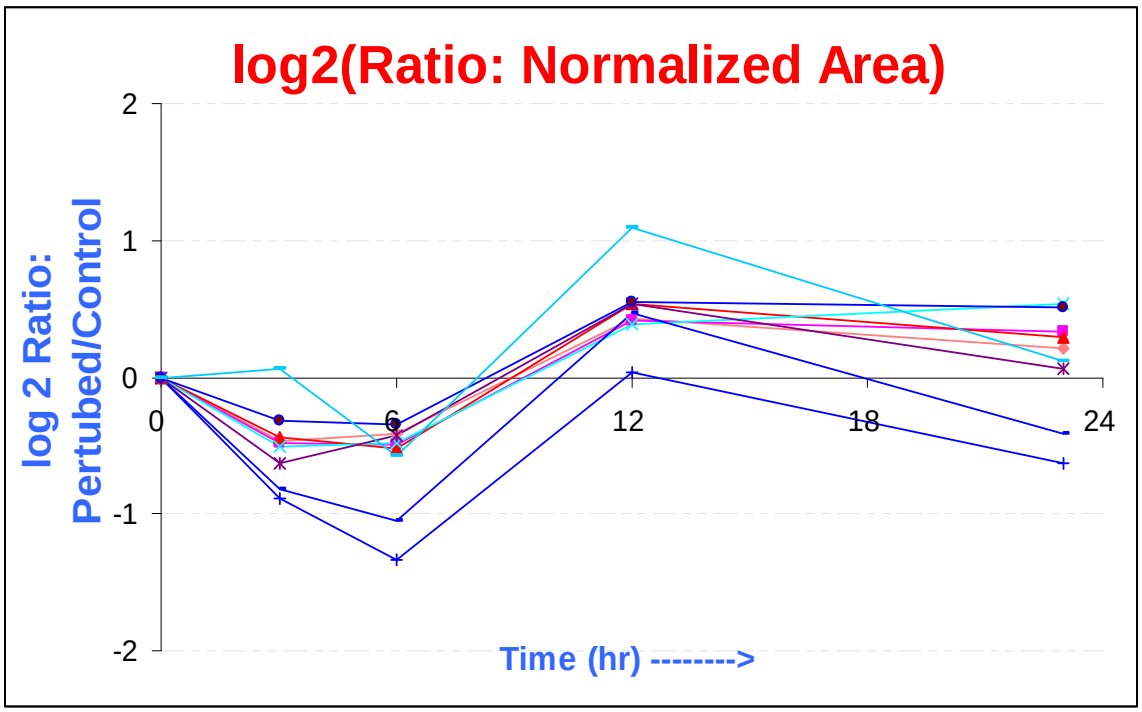


Figure 6.7 K-Means cluster showing correlation in metabolites related to sugar synthesis

some other metabolites not directly related to TCA cycle (like glucaric acid inositol phosphate, phosphoric acid) clustered along with the TCA cycle metabolites.

Similarly as shown in Figure 6.7, metabolites belonging to, or related to, glycolysis pathway, cluster together in Cluster 1. As in previous case, some of the metabolites belonging to the pathway like glucose-6-phosphate, fructose-6-phosphate, inositol-6-phosphate, sorbitol, do not cluster in the same group, and some other metabolites like proline, though not directly related, clusters together with the metabolites belonging to glycolysis group. As shown in Figure 6.1, in the cluster containing positively significant metabolites, (from the three known metabolites, xylitol and arabinose which belong to Pentose phosphate pathway) show a stronger correlation as compared to glycerol, even though all three of them are part of the primary cell wall. The current analysis thus indicates, that clustering techniques using Pearson co-relation is a powerful tool to identify correlation between metabolic profiles. However, the strong correlation, or absence of correlation, in metabolic data cannot be directly used to conclude a positive metabolic relationship between the metabolites. The main problem in such an analysis is that metabolic profiling technique measures intracellular metabolite concentration, and not metabolic fluxes. Biochemically related chemical transformations may be directly related by their metabolic fluxes,

however the change in flux may or may not create a proportional change in metabolite concentration, hence the correlation may not be observed in the concentrations measured - for e.g. glucose can be converted to glucose-6-phosphate using hexokinase enzyme, and hence glucose and glucose-6-phosphate are expected to show a correlation in their responses, especially when the reaction is closer to equilibrium conditions. When the equilibrium condition is satisfied, concentration of metabolically active glucose and glucose-6-phosphate are in a constant ratio, and hence should show a high degree of correlation in their response. However, this is not observed in the current analysis – which is due to the fact that the variation in concentration of glucose and glucose-6-phosphate is a net result of changes in metabolic flux of all the chemical reactions in which they participate. Since these metabolites take part in multiple reactions – the net change in their concentration is not highly correlated. On the other hand, metabolites like xylitol, arabinose and glycerol which are all used in the primary cell wall, would show a proportionate increase in their response, which represents a similar response in cell wall production. In such cases, metabolites related by a common function may show a similar response. The metabolites which are linked both by common functions and through metabolism (glucose and fructose, xylitol & arabinose) show the strongest co-relation in their response. Thus the dual role, of some of the metabolites, as a

structural element as well as a metabolic substrate, needs to be considered while interpreting correlations in metabolic time profiles (or deviation within different samples at the same time).

Certain metabolites like fructose-6-phosphate, 3-phosphoglycerate, may not be part of any macromolecules. For such metabolites, the concentration measured using metabolic profiling technique, represents the metabolically active substrate. Also, if the metabolic reactions are at equilibrium condition, the substrate and product would show a constant ratio (which is the equilibrium constant for the reaction). Now if we assume a quasi-steady rate for such metabolites, i.e. if we assume that over a period, the metabolite may change their concentration but reach equilibrium at a much faster rate, and hence assumed to be under equilibrium conditions all the time; we would obtain a perfect co-relation between the response of the two profiles, as their concentration always are related by the equilibrium constant. In such a case, a correlation value very close to +1 would be obtained between the two profiles. A small deviation from this assumption would still, give a correlation coefficient close to +1. Even though analysis with such assumptions have been used in a bacterial system, in plants, it is very difficult, experimentally, to create and monitor an experimental setup which will ensure such a quasi-steady state, especially when using a destructive measurement technique like metabolic profiling using GC-MS (as

compared to NMR). Considering this, metabolites which may be closely related, may not show correlation in their metabolic response if the quasi steady state condition is not achieved during the experimental conditions. In the current analysis, most of the TCA cycle metabolites (except succinate), which mostly are present in metabolically active form, show a strong correlation in their response, usually, indicating a presence of near quasi steady state condition for the TCA cycle reactions.

Thus Pearson correlation distances, allow identification of metabolite clusters which show a similar time response to a perturbation, however, the clustering pattern cannot always identify metabolites belonging to a pathway. This is due to limitations of the experimental design and metabolic profiling.

6.3 Significance of Metabolic Profiling for Plant Physiological Studies:

Most of the previous studies concerning the response of the plants to elevated CO_2 levels involved the quantification of a small number of metabolites either of a particular functional group of interest in plant studies or because previous analyses or information about the metabolic pathway structure had indicated that these metabolites might show a significant change due to the applied perturbation. One of the major disadvantages of such a "hypothesis-driven" approach is that, while the change in a metabolite of interest can be accurately quantified, the change in other metabolites that might have been simultaneously

changing is not measured and no conclusions about their correlation can be derived. In this context, the high-throughput metabolic profiling analysis can provide the advantages of a "data-driven" approach, which allow for the simultaneous measurement of the changes in a large number of metabolites in the plant biomass.

Finally, for the data-driven, systems biology approach in plant research, metabolic profiling can prove to be a very important high throughput tool for measuring the metabolic state of the plant. The current analysis of the time-series metabolic data, using different multivariate statistical techniques (MEV, SAM) and methods that are developed for gene expression analysis, shows a common platform can be used to perform data analysis for an integrated study.

Metabolic Profiling analysis provides one extensive cellular fingerprint, which could be used in conjunction with information from the other cellular levels in the context of systems biology research. However, while comparing the genomic or proteomic data, with metabolic profiling data, one needs to consider that changes in metabolite concentration do not directly translate in changes in metabolic fluxes, the latter being directly comparable to gene and protein expression. Change in a reaction flux does not always mean similar change in the concentration of its substrate(s) or product(s) and vice versa.

Chapter 7: Future Work

Metabolic profiling provides a high throughput method which allows for an extensive phenotypic fingerprint of the plants to be obtained. From the current analysis many areas, which need to be improved, were identified, in order to make metabolic profiling a more powerful tool, which can be easily established.

7.1 Metabolic Profiling Protocol:

In the current metabolic profiling analysis, protocol using GC-MS has been established, and the parameters required for optimum operation of the instrument have been identified. Based on the analysis following parameters were identified which needs future work:

- ***Derivatization Protocol:*** Current derivatization protocol gives rise to multiple derivatization forms of the metabolites. A modified protocol which resolves this issue would resolve the limitation of the current protocol.
- ***Internal Standard:*** In the current analysis a single compound was used as internal standard. Even though this can account for most experimental errors, it may not account for variations in derivatization which may be dependent on the function of the metabolite.
- ***Peak Identification:*** Manual peak de-convolution approach was used in the current analysis. In the past, a software tool like AMDIS [Fiehn, 2001a]

had been used by researchers for conducting peak de-convolution. However, no systematic method or approach is available to select the parameters for AMDIS. Study of systematic use of AMDIS would allow a much better de-convolution greatly improving the current protocol.

7.2 Metabolic Profiling Data Filtering and Normalization:

With the increasing number of metabolites that can now be identified with metabolic profiling techniques, a systematic approach for data filtering is also required. Few screening techniques have been suggested in the current analysis however a more uniform and commonly used process should be developed.

7.3 Data analysis using multivariate statistics:

K-Means and SAM analysis were used to identify metabolites which show a differentiated expression. However, the SAM graph obtained was different from the standard SAM graph; thus the positive and negatively significant metabolites could not be identified by specifying a particular delta value. This could be result of either absence of lipid metabolites, or the difference between the normalization of the metabolic and genetic analysis. Also, using the lipid data, a more complete study needs to be conducted to identify the cause and correct the same. Alternately, another systematic method needs to be developed for identifying differentiated metabolites using time profiles of the different metabolites.

The current analytical technique, showed the use of Pearson correlation distance in the metabolic analysis, to identify correlated metabolites and to map the plant metabolic profile using K-Means and Principle Component Analysis respectively. Even though the technique did cluster metabolites that show similar profiles, some of the metabolites showing similar profiles (Glucose-6-phosphate and Fructose-6-phosphate) were clustered separately. Hence a method which allows a better, more consistent representation of Pearson correlation between metabolic profiles is needed.

7.4 Design of more elaborate experiments:

The current analysis has shown the amount of information that can be obtained by systematically designing a perturbation in a previously well studied system. Related individual perturbation of the same system, would allow a much better understanding of the perturbed system by comparing the similarities and differences in the response of the system to various perturbations.

For such future experiments, based on current analysis following modifications are suggested:

Labeling:

10% of the CO_2 used was labeled C^{13}. However, the current analysis indicated that due to instrumental variability, the changes in the labeling of the metabolites, because of partially labeled substrate, could not be detected. Hence

for future experiments, either a much larger portion of the substrate should be C-13 labeled (50% to 100%) or alternatively labeled substrate should not be used at all, as it would interfere with the marker ion quantification.

Selection of Time points:

Since this analysis was the first analysis involved in immediate response of the elevated CO_2, the time points chosen for the analysis were distributed so as to get the immediate response in first three hours, and the longer 1 day response. The analysis indicated that the longer time points were much more useful for the analysis as compared to shorter time points, where both the control and perturbed system showed large deviations. Also many metabolites showed a change in their response during 12 to 23 hour periods. Hence the time points for harvesting should be distributed uniformly, throughout the time period, and should specifically have a time point between 12 to 24 hours.

Nutrient Source:

Additionally, the results in the literature indicate that by using NH4+ ions as the primary nitrogen source, a much larger increase in plant biomass is allowed by removing the effect of nitrogen stress. Therefore, an experiment may be conducted using NH_4^+ as the source, in order to understand the plant response to elevated CO_2, in absence of nitrogen stress.

For metabolic profiling, the current analysis was conducted for polar metabolites; the lipid metabolites were excluded from the analysis. However there were indications that the lipid phase may be showing a change in response to the perturbation, hence the same should be included in the future analysis.

7.5 Future applications for metabolic profiling:

In order to extend the role of metabolic profiling technique to larger research areas in plant genetics and biochemistry, more experiments with different experimental goals need to be performed, some of which are listed below:

Metabolic Profiling for screening mutants:

Even though it has been demonstrated that a single mutation changes the plant metabolism (and hence this can be used as a screening technique to identify mutant plants), the same has not been shown till now. For identifying possible challenges in these issues and finding their solutions, experiments should be designed to screen for mutants.

Metabolic profiling for plant hormones study:

The role of all plant hormones which control various plant physiological properties at metabolic level is still not well understood. Due to its extensive ability to profile metabolites, metabolic profiling can be used effectively to understand the role of hormones in plant.

Appendices

Appendix I. Protocols

Methanol Extraction Protocol:

1. Purpose

This protocol describes the extraction of metabolites from Arabidopsis thaliana plant sample.

2. Instruments

2.1 Homogenizer
2.2 Water Bath
2.3 Conical Tubes, 15 and 50 ml
2.4 Tube Stand
2.5 Pipettor
2.6 Pipette tube 10 mL
2.7 Pipette Tips, 1 mL
2.8 Vortexor
2.9 Balance accurate upto 1 mg
2.10 Dry ice box
2.11 Timer
2.12 Permenant Marker

3. MATERIALS

3.1 Methanol
3.2 Ethanol
3.3 De-ionized water
3.4 Ribitol Solution
3.5 Dry ice

4. REAGENT PREPARATION

4.1 Ribitol Solution (2 mg/ml)

4.1.1 Measure Ribitol in a measuring plate. The weight of the ribitol should be 1 mg * (No. of plant samples)* (Weight of each plant sample) * 2. For a set of forty samples, one gram each, measure 1*40*1*2 = 80 mg

4.1.2 Transfer the Ribitol measured in a 50 ml tube (or a reagent bottle if the quantity of ribitol is above 100 mg)

4.1.3 Add de-ionized water to the tube using 25 ml pipette so as to produce a solution having ribitol concentration of 2 mg/ml. So for 80 mg Ribitol add 160 ml water.

4.1.4 Ensure that you have enough Ribitol solution for the whole batch of samples.

4.1.5 Store the solution at 4 deg. C temperature when not in use.

5. PROCEDURE

5.1 Plant Grinding:

5.1.1 Before starting Grinding ensure that the water bath to be used for extraction is set at 70 deg C.

5.1.2 Follow the procedure given in Sop# M001 for Grinding.

5.1.3 Take a labeled 50 ml tube with conical bottom. Mark the conical tube for 2.5 ml mark (which is typically close to the junction of cylindrical and conical bottom section).

5.1.4 Transfer the powdery ground plant to the conical tube, up to 2.5 ml mark which approximately corresponds to 1 gm of plant.

5.1.5 Keep the tube in Dry ice box and transfer it close to the Fume hood along with other supplies required for extraction.

5.2 Methanol Addition

5.2.1 Take a tube tray for holding 50 ml conical tubes.

5.2.2 Take empty 50 ml conical tubes (the number of tubes should be equal to number of samples being processed at a time) and place them on the stand.

5.2.3 Measure 28 ml of methanol and add it in each tube.

(Note: The Rosenner et. al. protocol used 1.4 ml for 0.1 mg potato tubers. Since Arabidopsis thaliana has more dry biomass and it's root are more difficult to extract we use double the quantity)

5.2.4 Measure 500 μL of 2 mg/ml Ribitol solution using a 1 mL Pipette tip and add it to 28 mL methanol in each tube.

5.2.5 Close the lid, and shake the tube to ensure proper mixing.

5.2.6 Thus ensure that you have 50 ml tubes containing 28 mL methanol and 0.5 mL Ribitol (2 mg/mL concentration in water) before starting homogenization. (This gives a concentration of 1 mg of Ribitol for a gram of plant, if some other concentration is desired adjust volume accordingly).

5.3 Homogenization

5.3.1 Fill a 100 ml measuring cylinder with de-ionized water and insert the homogenizer tip into the cylinder. Start the homogenizer and run it for few seconds. Clean the homogenizer with a paper napkin.

5.3.2 Spray Ethanol on the homogenizer and wipe it again with a paper napkin.

5.3.3 Take out one 50 ml conical tube containing 1 gm (2.5 ml) of ground plant material from the dry ice box.

5.3.4 Transfer the 28.5 ml methanol and ribitol solution to the 50 ml conical tube containing the plant.

5.3.5 Homogenize the methanol, plant and ribitol solution mixture, for 2 to 5 minutes depending on the type of sample. The homogenizer may become hot so in between after every one minute, give it a break for a few seconds. Ensure that it becomes a homogeneous mixture with no large solid plant pieces left.

5.3.6 Stir the homogenized mixture using a vortexor, and later with hand to ensure that no solid particles settle down, and that they are uniformly distributed in the entire tube.

5.3.7 Divide the 28-30 ml volume homogenized solution in four 15 ml conical bottom tubes (labeled Plant Number followed by A, B, C, D) equally such that each tube has 7-7.5 ml each of the homogenized material.

5.3.8 Repeat the above procedure starting from 5.3.1 (i.e. cleaning the homogenizer) for one more plant.

5.4 Extraction in Water Bath

5.4.1 Put the eight 15 ml conical tubes (A, B, C, D tubes for Plant 1 and Plant 2) on a stand and put the stand along with the tubes in the Water bath.
5.4.2 Ensure that the tube lids are not tightly closed, before you put them in the bath, to allow gas expansion and avoid tube bursts on heating due to pressure.
5.4.3 Set the timer for 15 mins and start the timer.
5.4.4 Check the temperature in the thermometer as the digital temperature indicator may have a small error.
5.4.5 After 15 minutes take the tubes out of the water bath.
5.4.6 You will observe precipitated solid matter at the bottom part of the tube. This are methylated metabolites.

5.5 Water Treatment:

5.5.1 Set two 50 ml conical tubes prefilled with de-ionized water inside the fume hood.
5.5.2 Attatch a 15 ml pipette tube to the pipettor.
5.5.3 Open caps for two 15 ml conical tubes containing the plant – methanol mixture (which was just removed from the water bath).
5.5.4 Add 7 ml of de-ionized water to each tube in order to neutralize the methylated compound. (Total volume required 7*8 = 56 ml). Close the lids tightly.
5.5.5 Shake each tube individually on the vortexer in order to ensure very high degree of mixing. In between also mix the solution in the tube with hand rotating it by 360 degrees about a horizontal axis.
5.5.6 In the end ensure that the solution is more transluscent as compared to after the methanol treatment and only fine solid particles remain uniformly distributed in the solution.
5.5.7 Repeat the above procedure for all the eight tubes.

5.6 Sample Division:

5.6.1 Take four empty 15 ml tubes per sample. Label them Plant #, E, F, G, H on the Cap and the tube. Place the tubes on a stand.

5.6.2 Shake once more the tube containing solution in order to make sure there is no setteling of the solid residue in the tube.
5.6.3 Transfer 7 ml sample from tube A (containing approximately 14 ml liquid) to tube E.
5.6.4 Similarly from tube B transfer 7 ml to F, from C transfer 7 ml to G, from D transfer 7 ml to H.

5.7 Centrifugation:

5.7.1 Place the 15 ml conical tubes containing 7 ml of sample each in the centrifuge. Loosen slightly the caps of the tube, in order to avoide tube bursts.
5.7.2 Change the units of the centrifuge to "cfm", the default units are in rpm.
5.7.3 Set the speed to 2000 cfm, time to 5 mins and temperature to 25 deg C
5.7.4 Start the centrifuge.
5.7.5 At the end of the centrifuge operation take the tubes out, place them on the stand without shaking it and close the lid of the centrifuge.

5.8 Drying:
5.8.1 Ensure that the vacuum pump of the SpeedVac is running. Open the top lid and place all the tubes that need to be dried in the speedVac unit, ensuring that the tubes are placed in a way so that they balance the centrifuge.
5.8.2 Close the lid and apply vacuum. Also start the rotation of the SpeedVac.
5.8.3 Wait till you here a "clicking" sound indicating that the vacuum seal is in effect. In case you don't hear the same after 15 mins, it's an indication that the vacuum pump is not working properly.
5.8.4 Dry the samples for 6-16 hours, till the sample is dry, and not sticky.
5.8.5 Store the samples in 4/-20 deg C freezer in upright position.
5.8.6 This samples are now ready for derivatization for metabolic profiling.

Derivatization Protocol: Derivatization

Lipid phase

The remaining lipid phase is treated as follows: Take out 100 µL for LC/MS analyses and refrigerate. To the remaining ~ 700 µL lipophilic phase, add 900 µL $CHCl_3$. Add 1 mL MeOH containing 3% v/v H_2SO_4. Transmethylate lipids and free fatty acids for 4 h at 100°C. Take care that your glass vial is sealed with a teflonized seal, and not with rubber. Otherwise, you will lose a lot of your solution and will find many rubber additives in your sample.

Extract your solution two times by adding 4 mL H_2O, vortexing, centrifuging at 4000 rpm, and discarding the water phase. Dry the remaining chloroform phase over anhydrous Na_2SO_4 and transfer the supernatant into a new glass vial. Concentrate to about 80 µL. Add 10 µL pyridine and 10 µL MSTFA to the remaining 70 µL portion, silylate for 30 min at 37°C, and inject 2 µL into the GC/MS with a split ratio of 25:1.

Polar phase

Add 50 µL of methoxyamine hydrochloride (20 mg/mL pyridine) to the dried (1 mL) fraction of your polar phase. Incubate for 90 min at 30°C with continuous shaking. Add 80 µL of MSTFA for 30 min at 37°C and wait 120 min at 25°C before injection. Inject 2 µL, split 25:1. The second portion of the dry polar fraction can be used for LC/MS analyses or stored frozen at -80°C.

Appendix II. Retention Time Comparison

UMCP			Max Planck List			RT
Peak No	Name	RT	Mol. Wt.	RT	Type	Difference
1	LACTIC ACID,O,O-TMS	6.70	234	6.447	organic acid	0.25
3	ALANINE,N,O-TMS	7.71	233	7.206	amino acid	0.50
6	GLYCINE,N,O-TMS	8.60	219	8.039	amino acid	0.56
7	PYRUVIC ACID MEOX TMS	9.04	189	8.434	organic acid	0.61
9	VALINE,N,O-TMS	10.80	261	9.869	amino acid	0.93
10	ETHANOLAMINE,N,N,O-TMS	11.02	277	10.120	amine	0.90
8	OXALIC ACID TMS	10.60	234	10.145	organic acid	0.46
13	GLYCEROL 3TMS	11.85	308	10.633	alcohol	1.22
15	LEUCINE, N,O-TMS	12.47	275	11.315	amino acid	1.16
18	ISOLEUCINE,N,O-TMS	13.15	275	11.925	amino acid	1.23
19	GLYCINE,N,N,O-TMS	13.29	291	12.163	amino acid	1.13
20	SERINE,O,O-TMS	13.95	249	12.540	amino acid	1.41
21	PROLINE,N,O-TMS	14.40	259	12.946	amino acid	1.45
22	PHOSPHORIC ACID,O,O,O-TMS	14.40	314	13.062	inorganic acid	1.34
24	GLYCERIC ACID,O,O,O-TMS	14.69	322	13.346	organic acid	1.34

UMCP			Max Planck List			RT
Peak No	Name	RT	Mol. Wt.	RT	Type	Difference
25	SERINE,N,O,O-TMS	15.21	321	13.824	amino acid	1.39
29	FUMARIC ACID TMS	15.78	260	14.063	organic acid	1.72
28	THREONINE,N,O,O-TMS	15.61	335	14.230	amino acid	1.38
30	SUCCINIC ACID 2TMS	15.95	262	14.406	organic acid	1.54
32	B-ALANINE TMS	16.64	305	15.243	amino acid	1.40
35	HOMOSERINE 3TMS	17.47	335	15.986	amino acid	1.48
39	2-METHYL BENZOIC AVID TMS	18.25				
38	2-METHYLMALIC ACID 3TMS	18.47	364	17.005	organic acid	1.47
40	3-HYDROXY GLYTARIC ACID TMS	18.47				
43	4-AMINOBUTYRIC ACID 3TMS	19.40	319	28.156	organic acid	(8.76)
44	MALIC ACID TMS	19.44	350	17.890	organic acid	1.55
46	ASPARAGINE,N,N,N,O-TMS	20.13	420	19.435	amino acid	0.70
47	L-HYDROXYPROLINE,N,O,O-TMS	20.15	347	17.949	amino acid	2.20
49	ASPARTIC ACID,N,O,O-TMS	20.15	349	18.590	amino acid	1.56
50	THREONIC		424	18.740	organic acid	

UMCP			Max Planck List			RT
Peak No	Name	RT	Mol. Wt.	RT	Type	RT Difference
	ACID,O,O,O,O-TMS	20.21				1.47
55	RIBITOL TMS	21.90		20.512	sugar alcohol quantify standard	1.39
57	XYLITOL 5TMS	22.09	512	20.668	sugar alcohol	1.42
58	GLUTAMIC ACID 3TMS	22.22	363	20.710	amino acid	1.49
59	ARABINOSE MEOX1 4TMS	22.54	467	20.735	monosaccharide	1.83
60	CYTOSINE 2TMS	22.89		21.717	pyrimidine	1.17
62	HOMOCYSTEINE,N,N,O-TMS	22.90	351	21.789	amino acid	1.11
61	GLUTAMINE,N,N,N,O-TMS	22.90	434	21.797	amino acid	1.10
68	PHENYLALANINE,N,O-TMS	23.61	309	22.083	amino acid	1.53
75	ORNITHINE,N,N,N',O-TMS	24.65	420	23.497	amino acid	
77	ASPARAGINE,N,N,O-TMS	24.66	348	23.039	amino acid	1.62
83	FRUCTOSE MEOX1 5TMS	25.81	569	24.530	monosaccharide	1.28
84	ACONITIC ACID 3TMS	25.81	390	24.596	organic acid	1.21
85	MANNOSE MEOX TMS	25.82	569	24.599	monosaccharide	1.22
86	SORBTOL TMS	25.91		24.629	sugar alcohol	1.28
90	FRUCTOSE MEOX2 5TMS	26.15	569	24.896	monosaccharide	1.25
89	GALACTOSE MEOX1 TMS	26.15	569	24.899	monosaccharide	1.25

UMCP			Max Planck List			RT
Peak No	Name	RT	Mol. Wt.	RT	Type	Difference
88	SHIKIMIC ACID TMS	26.15	462	24.912	organic acid	1.24
91	GLUCOSE MEOX1 5TMS	26.35	569	25.132	monosaccharide	1.22
93	CITRIC ACID TMS	26.63	480	25.223	organic acid	1.41
94	GLUCOSAMINE MEOX1 TMS	26.64		25.265	monosaccharide	1.28
95	GLUTAMINE,N,N,O-TMS	26.91	362	25.291	amino acid	1.62
96	GLUCOSE MEOX2 5TMS	26.91	569	25.361	monosaccharide	1.65
99	LYSINE,N,N,N',O-TMS	27.16	434	25.545	amino acid	1.62
98	ISOCITRIC ACID TMS	27.05	480	25.614	organic acid	1.44
101	3-PHOSPHOGLYCERATE TMS	27.58	474	26.029	phosphorylated compound	1.55
104	GLUCONIC ACID,O,O,O,O,O,O-TMS	27.87		26.669	organic acid	1.20
114	SACCHARIC ACID TMS	28.77		27.411	organic acid	1.36
116	INOSITOL,O,O,O,O,O,O-TMS	28.95	612	27.701	sugar alcohol	1.22
117	ASCORBIC ACID TMS	28.95		27.732	organic acid	1.25
122	TYROSINE	29.54				
156	SORBITOL-6-PHOSPHATE TMS	34.94		33.793	phosphorylated compound	1.15
158	FRUCTOSE-6-PHOSPHATE	35.03		33.839	phosphorylated	1.19

UMCP			Max Planck List			RT
Peak No	Name	RT	Mol. Wt.	RT	Type	Difference
	MEOX1 TMS				compound	
159	FRUCTOSE-6-PHOSPHATE MEOX2 TMS	35.20		34.000	phosphorylated compound	1.20
160	GLUCOSE-6-PHOSPHATE MEOX1 TMS	35.38		34.193	phosphorylated compound	1.19
161	GLUCOSE-6-PHOSPHATE MEOX2 TMS	35.58		34.377	phosphorylated compound	1.20
163	TRYPTOPHANE, N,N′,O-TMS	35.75	420	34.693	amino acid	1.06
170	SUCROSE TMS	38.34		37.149	disaccharide	1.19

Note: The retention time & spectrum of metabolites marked in bold letters were verified using the standard substance obtained from Sigma.

Appendix III. Metabolite List with Marker Ions & Retention Times

Peak No	Name	RT	Marker ions		
1	LACTIC ACID,O,O-TMS	6.70	117	191	219
2	unknown01	6.81	299	281	
3	ALANINE,N,O-TMS	7.71	116	190	218
4	unknown02	7.79	177		
5	unknown03	8.50	280.5-285		
6	GLYCINE,N,O-TMS	8.60	204		
7	PYRUVIC ACID MEOX TMS	9.04	115	174	189
8	OXALIC ACID TMS	10.60	190	220	
9	VALINE,N,O-TMS	10.80	144-145.5		
10	ETHANOLAMINE,N,N,O-TMS	11.02	106	114	174
11	unknown06	11.23	176	217	
12	unknown07	11.50	216	172	
13	GLYCEROL 3TMS	11.85	202-207		
14	unknown07b	12.12	228	183	
15	LEUCINE, N,O-TMS	12.47	158	232	260
16	unknown8a	12.90	149-152	167	191
17	unknown08	13.11	241-242.5	163-164.5	
18	ISOLEUCINE,N,O-TMS	13.15	100	158-159.5	218-219.5
19	GLYCINE,N,N,O-TMS	13.29	174-176.5	248-250.5	276
20	SERINE,O,O-TMS	13.95	131-135		
21	PROLINE,N,O-TMS	14.40	140-144.5	214-219	
22	PHOSPHORIC ACID,O,O,O-TMS	14.40	283	299-301	387-390
23	unknown9	14.58	262		
24	GLYCERIC ACID,O,O,O-TMS	14.69	292		

Peak No	Name	RT	Marker ions		
25	SERINE,N,O,O-TMS	15.21	188-189	204-206	218-220
26	unknown12	15.25	134	184	285
27	unknown13	15.36	298-301	344-347	
28	THREONINE,N,O,O-TMS	15.61	100-103	218-220	291-293
29	FUMARIC ACID TMS	15.78	217		
30	SUCCINIC ACID 2TMS	15.95	335	173	
31	Unknown14	16.53	203-205	262-264	149
32	B-ALANINE TMS	16.64	174-176.5	248-251	290-293
33	unknown15	16.81	129-131	219-221	103-105
34	unknown15a	17.04	149	180-182	197.5-200
35	HOMOSERINE 3TMS	17.47	128-130	218-220	
36	unknown15b	17.65	174-175		
37	unknown15c	17.79	214-215		
39	2-METHYL BENZOIC AVID TMS	18.25	119-120.5	193-195.5	
38	2-METHYLMALIC ACID 3TMS	18.47	246-249		
40	3-HYDROXY GLYTARIC ACID TMS	18.47	163	190	231
41	unknown16	18.58	221-224.5	298-300.5	
42	unknown18	19.13	232		
43	4-AMINOBUTYRIC ACID 3TMS	19.40	304		
44	MALIC ACID TMS	19.44	232-234		
45	unknown19	19.81	114	290-	

Peak No	Name	RT	Marker ions		
				292	
46	ASPARAGINE,N,N,N,O-TMS	20.13	216		
47	L-HYDROXYPROLINE,N,O,O-TMS	20.15	140	348	
48	ASPARTIC ACID,N,O,O-TMS	20.15	232-234.5		
49	THREONIC ACID,O,O,O,O-TMS	20.21	318-321		
50	unknown22	20.63	217		
51	METHIONINE TMS	20.99	175-176		
52	unknown25	21.13	188		
52b	CYSTEIN TMS	21.30	142-144		
53	unknown27 227	21.58	227-228		
54	unknown28 154	21.58	153-159		
55	RIBITOL TMS	21.90	216-219		
56	Ribitol	21.90	317-322		
57	XYLITOL 5TMS	22.09	307	319	
58	ARABINOSE MEOX1 4TMS	22.22	277	306-309.5	389-390.5
59	GLUTAMIC ACID 3TMS	22.54	245-249		
60	CYTOSINE 2TMS	22.89	98-100	170-171.5	240-241.5
62	HOMOCYSTEINE,N,N,O-TMS	22.90	20-203.5	219-220.5	234-235.5
61	GLUTAMINE,N,N,N,O-TMS	22.90	154-158		
63	unknown31	23.12	292	333	
64	unknown32	23.29	275		
65	unknown32b	23.30	215	240	254
66	unknown33	23.37	182	229	257

Peak No	Name	RT	Marker ions		
67	unknown33b	23.41	320		
68	PHENYLALANINE,N,O-TMS	23.61	218-219.5	192-193.5	
68b	unknown34	23.71	245-246.5		
69	unknown35	23.84	292-294		
70	unknown36	23.95	204	257	347
71	unknown37	24.09	217-218	257-259	292-294
72	unknown37b	24.22	142	185-186	275-277
73	unknown38	24.28	292-294	333-335	
74	unknown39	24.40	257-259	319	347
75	ORNITHINE,N,N,N',O-TMS	24.65	142	174	200
76	unknown40	24.65	171-173		
77	ASPARAGINE,N,N,O-TMS	24.66	116-118	132	159
77b	unknown41	24.79	116-117	159-160.5	215-216
78	unknown41a	24.88	199	289	
79	unknown41b	24.95	232-234		
80	unknown42	25.00	206	348	
80b	unknown43	25.12	181	230	257
81	unknown44	25.24	260		
82	unknown44b	25.46	298-301		
82b	MANNITOL TMS	25.58	205	277	319
83	ACONITIC ACID 3TMS	25.81	209-212	375-376	
84	FRUCTOSE MEOX1 5TMS	25.81	306-309		
85	MANNOSE MEOX TMS	25.82	158-		

Peak No	Name	RT	Marker ions		
			162.5		
86	SORBTOL TMS	25.91	315	357	387
87	ACONITIC ACID TMS	25.91	211	375	385
87b	GALACTITOL TMS	25.96	191-193	330-333	342
88	SHIKIMIC ACID TMS	26.15	239	243	255
89	GALACTOSE MEOX1 TMS	26.15	160	229	319
90	FRUCTOSE MEOX2 5TMS	26.15	306-309		
91	GLUCOSE MEOX1 5TMS	26.35	159-163		
92	Unknown 45a	26.49	204	191	
93	CITRIC ACID TMS	26.63	272.5-276		
94	GLUCOSE MEOX2 5TMS	26.64	205-207.5	318-320.5	
95	GLUTAMINE,N,N,O-TMS	26.91	155-159		
96	GLUCOSAMINE MEOX1 TMS	26.91	203		
97	unknown45	26.92	133	369	
98	ISOCITRIC ACID - TMS	27.05	244.5-247.5		
99	LYSINE,N,N,N',O-TMS	27.16	155-157.5	229-231.5	317-318.5
100	unknown46	27.44	204		
101	3-PHOSPHOGLYCERATE TMS	27.58	227	298-300.5	357
102	unknown48	27.71	189.24		
103	unknown49	27.78	189	273	
104	GLUCONIC ACID,O,O,O,O,O,O-TMS	27.87	333		
105	unknown51	28.02	405-408		
106	unknown52	28.08	333		
107	unknown53	28.10	299	319	
108	unknown54	28.12	204		
109	unknown55	28.19	191	204	217

Peak No	Name	RT	Marker ions		
110	unknown56	28.24	272-275	362	365
110b	unknown56b	28.31	156-158	173	205
111	unknown57	28.39	273-275	362-365	
112	unknown58	28.54	220-223	205	
113	unknown59	28.67	237	312	
114	SACCHARIC ACID TMS	28.77	333	393	
115	Unknown 59b	28.94	269.5-272.5		
116	ASCORBIC ACID TMS	28.95	330-334	358-361.5	
117	INOSITOL,O,O,O,O,O,O-TMS	28.95	305-307		
118	unknown60	29.07	288	312	
119	unknown61	29.09	361		
120	unknown62	29.27	361-364		
121	unknown63	29.43	203-206	318-321	
122	TYROSINE	29.54	192-193	218-221	280-283
123	unknown64	29.70	361-363		
124	unknown65	30.02	361		
125	unknown66	30.14	318-321		
126	unknown67	30.24	174		
127	unknown68	30.33	204-206	319-321	
127b	unknown69	30.52	232-235		
128	unknown70	30.54	298-301	354-359	428-430
129	unknown71	30.77	203	227-	

Peak No	Name	RT	Marker ions		
				230	
130	unknown72	30.80	331		
131	unknown73	30.94	361-364		
132	unknown74	31.12	358-360		
133	unknown75	31.14	264-266		
134	unknown76	31.26	292-293	299.5-301	
135	unknown77	31.33	221-223.5	383	
136	unknown77b	31.68	144	156	174
137	unknown78	31.95	155	186	
138	unknown79	31.98	201	257	
139	unknown80	32.21	269		
140	unknown80b	32.56	116-119	344-346.5	
141	unknown81	32.64	210	228	
142	unknown81b	32.68	268-270		
143	unknown82	32.70	156		
144	unknown83	32.82	268	281	298-301.5
145	unknown83b	33.02	203-205		
146	unknown84	33.23	156	289	
147	unknown85	33.40	228-230		
148	Unknown 86	33.62	299	315	357
149	unknown87	33.72	119-121	290-292	
150	unknown88	33.89	203-206		
151	unknown89	34.04	268	430	
152	unknown90	34.26	117-119	359	

Peak No	Name	RT	Marker ions		
153	unknown91	34.51	180	229	245
154	unknown92	34.78	130	155	
155	unknown93	34.87	243	358	
156	SORBITOL-6-PHOSPHATE TMS	34.94	315-317	387-391	
157	unknown94	34.98	281	355	
158	FRUCTOSE-6-PPHOSPHATE MEOX1 TMS	35.03	315-318	457-459	
159	FRUCTOSE-6-PHOSPHATE MEOX2 TMS	35.20	315-318	457-459	
160	GLUCOSE-6-PHOSPHATE MEOX1 TMS	35.38	315-318	387-390	
161	GLUCOSE-6-PHOSPHATE MEOX2 TMS	35.58	315-319	387-389	
162	INOSITOL DERIVATIVE TMS	35.74	315	318	
163	TRYPTOPHANE,N,N´,O-TMS	35.75	100	202-204.5	291-293
164	unknown102	35.76	277-279	333-335.5	
164b	unknown103	36.43	197	229	315
165	unknown105	36.65	204	321	361
166	unknown106	37.00	299	315	354-357
167	unknown107	37.47	139	169	319
168	unknown108	37.76	273	362	
169	unknown109	38.05	129	204	305-307
170	SUCROSE TMS	38.34	168-171.5		
171	unknown110	38.63	343	415-417	
172	unknown111	38.93	221	281	299
173	unknown119	39.13	186	214-216	415-417
173b	unknown120	39.24	-		
174	unknown121	39.25	-		
175	unknown123	39.59	169	361	

Peak No	Name	RT	Marker ions		
176	unknown124	39.75	169	191	361
176b	unknown125	40.00	169	204	319
177	unknown126	40.16	169	204	231
178	unknown127	40.52	259-261	298-301	
179	unknown128	40.77	281	298-301	355
180	unknown129	41.17	298-301	315-319	461
181	unknown130	41.26	169	194	267
181b	unknown131	41.31	204-205	361-362	149
182	unknown132	41.66	203-206		
183	unknown133	42.29	230	236	245
184	unknown134	42.51	298-300	355-358	428-431
185	unknown135	42.60	298-300	355-358	428-431
186	unknown136	42.93	245	324-327.5	
187	unknown139	43.07	204		
188	unknown140	43.13	422-446		
188b	unknown143	43.28	197	348-350	
189	unknown144	44.26	324	368	410
190	unknown145	44.72	297	408	
191	unknown147	45.01	204	361	
192	unknown148	45.16	361	391	
193	unknown149	45.34	324		
193b	Unknown 150	46.10	299	315	
194	unknown155	46.92	324	361	450
195	unknown157	47.47	361	437	
196	unknown159	47.86	354		
197	unknown160	49.45	484-486.5		

Peak No	Name	RT	Marker ions		
198	unknown161	50.12	193		
199	unknown162	50.22	382		
200	unknown163	51.24	396		

Appendix IV. Presence/Quality Test for Metabolites

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
9	l-Valine	"no peak"		-	0%	0.085	1.0	-	11%	6%
18	iso_leucine	"no peak"		-	0%	0.018	12.5	77%	33%	17%
15	Leucine	"no peak"		-	0%	0.014	16.8	79%	33%	17%
186	unknown136	0.003	1.2	12%	22%	0.003	1.0	-	11%	17%
35	l-homoserine	0.018	1.0	-	11%	0.042	2.8	46%	33%	22%
81	unknown44	0.001	1.1	3%	33%	0.002	1.0	-	11%	22%
25	Serine N,O,O	0.073	23.6	130%	22%	0.587	156	100%	44%	33%
3	Alanine	0.032	2.3	57%	22%	0.093	8.5	58%	44%	33%
174	unknown121	0.190	9.6	121%	33%	0.369	4.7	58%	44%	39%
145	unknown83b	0.002	1.4	14%	44%	0.005	3.8	80%	33%	39%
48	Hydroxyproline	0.001	3.7	61%	44%	0.001	1.4	20%	33%	39%
142	unknown81b	0.005	4.3	61%	56%	0.004	2.4	43%	33%	44%
153	unknown91	0.002	4.5	60%	44%	0.003	2.3	34%	44%	44%
7	Pyruvic_Acid	0.002	2.1	46%	33%	0.002	1.8	25%	56%	44%
167	unknown107	0.021	6.5	109%	33%	0.029	21.7	120%	67%	50%
175	unknown123	0.012	3.9	60%	44%	0.012	2.6	36%	56%	50%
185	unknown135	0.004	1.8	26%	44%	0.003	2.5	30%	56%	50%
47	Hydroxyproline	0.003	3.2	55%	56%	0.002	3.0	51%	44%	50%
146	unknown84			41%	67%			66%	33%	50%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
		0.002	3.6			0.010	3.5			
171	unknown110	0.002	3.1	44%	78%	0.010	12.5	120%	22%	50%
151	unknown89	0.010	13.6	64%	67%	0.007	15.2	88%	56%	61%
136	unknown77b	0.008	9.4	69%	67%	0.009	1.9	25%	56%	61%
155	unknown93	0.003	5.7	46%	78%	0.005	2.7	42%	44%	61%
134	unknown76	0.003	2.9	39%	44%	0.003	4.7	43%	78%	61%
143	unknown82	0.002	1.8	19%	67%	0.002	3.1	41%	56%	61%
52	unknown25	0.001	2.6	39%	67%	0.001	6.7	56%	56%	61%
173	unknown119	0.009	8.3	60%	89%	0.003	1.9	26%	44%	67%
154	unknown92	0.006	2.7	46%	33%	0.005	3.4	44%	100%	67%
36	unknown15b	0.005	16.9	56%	67%	0.005	5.1	54%	67%	67%
135	unknown77	0.002	2.3	34%	78%	0.002	1.9	26%	56%	67%
189	unknown144	0.012	22.8	81%	89%	0.020	3.3	51%	56%	72%
163	Tryptophan	0.003	3.0	43%	56%	0.003	2.4	28%	89%	72%
188	unknown140	0.001	2.8	33%	78%	0.000	3.7	57%	78%	78%
23	unknown9	0.142	3.0	34%	78%	0.221	3.7	34%	89%	83%
32	beta-alanine	0.078	228	100%	100%	0.116	143	92%	78%	89%
	unknown133	0.009	8.9	65%	100%	0.017	7.0	56%	78%	89%
138	unknown79	0.004	2.3	34%	89%	0.005	2.5	36%	89%	89%
169	unknown109	5.031	9362	140%	100%	4.980	3,095	98%	89%	94%
133	unknown75	0.030	26.5	105%	100%	0.017	30.8	92%	89%	94%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
181	unknown130	0.020	3.5	49%	100%	0.016	6.3	55%	89%	94%
180	unknown129	0.016	4.0	52%	100%	0.013	5.8	54%	89%	94%
80	unknown42	0.009	9.4	69%	100%	0.005	4.5	52%	89%	94%
147	unknown85	0.008	12.6	91%	89%	0.006	1.9	23%	100%	94%
164	unknown102	0.007	3.5	43%	89%	0.009	2.7	40%	100%	94%
178	unknown127	0.004	2.4	35%	89%	0.005	3.7	49%	100%	94%
168	unknown108	0.003	3.5	38%	89%	0.004	3.3	44%	100%	94%
149	unknown87	0.003	4.3	56%	89%	0.004	3.6	49%	100%	94%
190	unknown145	0.003	2.8	29%	89%	0.002	3.1	34%	100%	94%
193	unknown149	0.001	2.8	35%	89%	0.002	2.2	32%	100%	94%
22	Phosphoric_Acid	8.206	2.5	30%	100%	6.929	1.6	16%	100%	100%
93	Citrate	6.430	3.6	48%	100%	7.791	1.7	17%	100%	100%
61	Glutamine	5.640	2.5	32%	100%	5.539	2.3	22%	100%	100%
170	Sucrose	3.884	2.8	33%	100%	3.803	2.3	24%	100%	100%
84	fructose_Meox1	3.083	2.4	30%	100%	2.839	1.7	17%	100%	100%
90	fructose_Meox2	2.518	2.2	27%	100%	2.393	1.6	15%	100%	100%
95	Glutamine _N,N.O	2.143	1824	109%	100%	1.238	9,779	176%	100%	100%
91	glucose_meox1	1.800	2.1	26%	100%	1.767	1.7	17%	100%	100%
94	glucose_meox2	1.586	3.5	46%	100%	1.393	2.6	34%	100%	100%
54	unknown28	1.089	58.8	111%	100%	0.956	36.0	68%	100%	100%
55	Ribitol_217			0%	100%			0%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
		1.000	1.0			1.000	1.0			
117	Inositol	0.967	2.4	35%	100%	0.956	1.9	22%	100%	100%
59	Glutamate	0.704	7.8	76%	100%	0.557	17.0	85%	100%	100%
2	unknown01	0.520	4.4	49%	100%	0.571	4.2	33%	100%	100%
56	Ribitol	0.509	1.1	5%	100%	0.509	1.2	5%	100%	100%
96	Glucosamine	0.457	1602	111%	100%	0.234	338	198%	100%	100%
152	unknown90	0.455	7.2	77%	100%	0.451	4.2	43%	100%	100%
111	unknown57	0.429	4.7	56%	100%	0.280	101	58%	100%	100%
109	unknown55	0.339	18.0	84%	100%	0.286	20.9	92%	100%	100%
130	unknown72	0.332	3.7	48%	100%	0.295	2.2	27%	100%	100%
53	unknown27	0.302	75.3	115%	100%	0.259	101	72%	100%	100%
85	Mannose	0.268	2.2	27%	100%	0.260	1.8	20%	100%	100%
70	unknown36	0.267	316.5	114%	100%	0.129	64.9	159%	100%	100%
121	unknown63	0.256	3.9	45%	100%	0.224	4.0	49%	100%	100%
83	Aconitic _acid	0.250	2.8	33%	100%	0.244	2.2	30%	100%	100%
21	l-Proline	0.244	2.8	38%	100%	0.203	3.0	34%	100%	100%
108	unknown54	0.240	11.8	82%	100%	0.204	15.0	88%	100%	100%
44	Malate	0.207	5.0	55%	100%	0.235	2.5	37%	100%	100%
89	Galactose _meox1	0.193	2.2	29%	100%	0.200	1.9	19%	100%	100%
30	Succinate	0.186	3.7	47%	100%	0.283	2.9	26%	100%	100%
92	Hexopyranose	0.168	50.3	93%	100%	0.159	10.1	82%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
77	Asparagine _N,N	0.142	12.5	100%	100%	0.075	12.1	100%	100%	100%
110	unknown56	0.133	5.0	56%	100%	0.090	11.1	50%	100%	100%
131	unknown73	0.132	4.3	55%	100%	0.073	17.8	76%	100%	100%
71	unknown37	0.123	2.8	41%	100%	0.116	5.1	61%	100%	100%
129	unknown71	0.114	4.3	53%	100%	0.129	36.7	130%	100%	100%
99	Lysine	0.114	30.4	88%	100%	0.097	125	107%	100%	100%
88	Shikimic _acid	0.114	2.4	33%	100%	0.104	1.7	18%	100%	100%
49	Asparatate	0.112	188	138%	100%	0.103	346	107%	100%	100%
19	Glycine_ N,N,O	0.108	52.0	92%	100%	0.128	285	121%	100%	100%
26	unknown12	0.107	9.3	78%	100%	0.048	6.9	74%	100%	100%
20	Serine_O,O	0.104	3.4	38%	100%	0.108	6.7	54%	100%	100%
43	4-Amino butyrate	0.090	32.1	81%	100%	0.104	1900	94%	100%	100%
66	Unknown33	0.084	207	114%	100%	0.112	193.8	85%	100%	100%
74	Unknown39	0.078	22.8	95%	100%	0.056	48.4	141%	100%	100%
98	iso-citrate	0.076	3.6	44%	100%	0.099	1.7	18%	100%	100%
24	Glycerate	0.070	5.6	59%	100%	0.102	2.3	27%	100%	100%
73	Unknown38	0.067	4.6	49%	100%	0.074	2.6	36%	100%	100%
194	unknown155	0.065	8.4	72%	100%	0.109	2.6	42%	100%	100%
127	unknown68	0.060	6.7	59%	100%	0.052	5.0	53%	100%	100%
38	2- methyl_benzoate	0.049	3.1	31%	100%	0.036	1.7	16%	100%	100%
160	glucose-6-P			73%	100%			65%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
		0.045	6.6			0.052	6.3			
182	unknown132	0.044	6.2	58%	100%	0.040	4.6	39%	100%	100%
124	unknown65	0.040	52.9	125%	100%	0.074	12.9	79%	100%	100%
13	Glycerol	0.036	4.3	52%	100%	0.177	34.0	150%	100%	100%
58	Arabinose	0.036	19.4	57%	100%	0.050	4.0	46%	100%	100%
1	Lactate	0.036	2.6	33%	100%	0.046	4.1	50%	100%	100%
103	Unknown49	0.036	5.2	56%	100%	0.027	30.1	56%	100%	100%
45	unknown19	0.035	4.6	49%	100%	0.038	7.3	50%	100%	100%
46	Asparagine	0.034	40.4	84%	100%	0.029	78.3	110%	100%	100%
60	Cytosine,	0.032	2.8	33%	100%	0.032	2.4	24%	100%	100%
29	Fumarate	0.032	2.9	36%	100%	0.039	1.7	21%	100%	100%
118	Unknown60	0.031	8.9	76%	100%	0.024	49.0	70%	100%	100%
16	Unknown8a	0.029	5.7	48%	100%	0.045	3.3	41%	100%	100%
187	Unknown139	0.029	3.6	41%	100%	0.027	1.9	20%	100%	100%
162	inositol-P-comp	0.028	4.3	48%	100%	0.023	3.7	45%	100%	100%
122	Tyrosine	0.025	11.2	80%	100%	0.028	8.7	67%	100%	100%
97	unknown45	0.024	22.3	94%	100%	0.015	16.6	124%	100%	100%
72	unknown37b	0.024	5.0	54%	100%	0.012	2.8	37%	100%	100%
86	Sorbitol	0.024	9.1	72%	100%	0.029	5.1	53%	100%	100%
102	unknown48	0.023	5.0	54%	100%	0.017	42.1	62%	100%	100%
105	unknown51	0.023	18.1	89%	100%	0.018	9.5	59%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
10	Ethanolamine	0.022	14.2	74%	100%	0.025	15.0	95%	100%	100%
68	Phenylaline	0.022	3.4	43%	100%	0.025	1.5	15%	100%	100%
6	Glycine_N,O	0.022	4.1	48%	100%	0.062	6.7	70%	100%	100%
27	unknown13	0.021	3.0	43%	100%	0.014	2.6	31%	100%	100%
76	unknown40	0.020	3.6	38%	100%	0.011	3.1	33%	100%	100%
166	unknown106	0.020	17.4	140%	100%	0.012	10.6	95%	100%	100%
176	unknown124	0.019	2.9	37%	100%	0.018	2.1	26%	100%	100%
196	unknown159	0.019	5.2	51%	100%	0.023	3.6	44%	100%	100%
104	gluconic_acid	0.019	5.4	54%	100%	0.024	4.5	57%	100%	100%
114	saccharic_acid	0.019	7.8	68%	100%	0.015	4.3	63%	100%	100%
69	unknown35	0.018	7.5	77%	100%	0.012	3.4	35%	100%	100%
39	3-hydroxy-glutaric_acid	0.018	11.3	94%	100%	0.042	11.8	91%	100%	100%
148	sugar_phospho comp	0.017	4.8	52%	100%	0.017	4.5	54%	100%	100%
125	unknown66	0.017	8.0	80%	100%	0.018	6.1	63%	100%	100%
33	unknown15	0.017	3.6	45%	100%	0.016	3.1	37%	100%	100%
78	unknown41a	0.017	31.3	109%	100%	0.024	9.9	67%	100%	100%
101	3-Phospho glycerate	0.016	11.9	92%	100%	0.015	4.3	60%	100%	100%
172	unknown111	0.016	16.0	140%	100%	0.010	8.9	94%	100%	100%
17	unknown08	0.016	12.0	90%	100%	0.020	4.4	54%	100%	100%
75	Ornithine	0.016	4.4	63%	100%	0.011	4.7	57%	100%	100%
115	Hexadecanoic			35%	100%			23%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
	acid	0.016	2.6			0.015	1.9			
41	unknown16	0.016	3.6	48%	100%	0.014	6.4	42%	100%	100%
79	unknown41b	0.015	3.6	37%	100%	0.012	2.0	24%	100%	100%
195	unknown157	0.014	6.2	46%	100%	0.012	19.8	47%	100%	100%
51	unknown22	0.014	3.2	40%	100%	0.013	1.6	16%	100%	100%
116	Ascorbic acid	0.014	2.9	36%	100%	0.014	4.0	57%	100%	100%
177	unknown126	0.014	4.4	43%	100%	0.013	3.4	38%	100%	100%
42	unknown18	0.013	8.2	59%	100%	0.029	192.0	96%	100%	100%
8	Oxalic_Acid	0.013	2.2	26%	100%	0.026	5.6	68%	100%	100%
141	unknown81	0.013	5.8	53%	100%	0.012	1.9	23%	100%	100%
65	unknown32b	0.012	3.0	40%	100%	0.011	2.4	36%	100%	100%
62	Homocystine	0.011	3.2	36%	100%	0.011	2.5	29%	100%	100%
107	unknown53	0.011	4.3	50%	100%	0.010	3.1	39%	100%	100%
179	unknown128	0.011	13.6	114%	100%	0.012	23.4	116%	100%	100%
34	unknown15a	0.011	3.5	43%	100%	0.011	2.8	32%	100%	100%
128	unknown70	0.011	6.6	55%	100%	0.013	17.2	123%	100%	100%
158	Fructose-6-Phos_meox1	0.011	3.8	44%	100%	0.011	3.2	37%	100%	100%
132	unknown74	0.010	4.2	34%	100%	0.010	3.8	35%	100%	100%
37	unknown15c	0.010	2.4	32%	100%	0.012	2.0	18%	100%	100%
82	unknown44b	0.010	6.2	50%	100%	0.010	8.7	74%	100%	100%
150	unknown88	0.010	5.6	54%	100%	0.013	2.2	26%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
200	unknown163	0.010	15.3	68%	100%	0.012	2.6	34%	100%	100%
63	unknown31	0.009	3.3	42%	100%	0.010	3.0	34%	100%	100%
50	Threonate	0.009	5.4	49%	100%	0.009	4.0	50%	100%	100%
28	Threonine	0.009	3.3	38%	100%	0.012	2.6	27%	100%	100%
144	unknown83	0.009	5.4	58%	100%	0.010	11.9	77%	100%	100%
4	unknown02	0.008	2.5	29%	100%	0.009	1.7	18%	100%	100%
100	unknown46	0.008	4.9	48%	100%	0.010	3.3	48%	100%	100%
184	unknown134	0.008	17.2	125%	100%	0.005	10.5	88%	100%	100%
165	unknown105	0.008	3.1	42%	100%	0.025	23.8	186%	100%	100%
137	unknown78	0.008	9.9	77%	100%	0.009	2.5	29%	100%	100%
140	unknown80b	0.008	4.7	55%	100%	0.009	2.8	33%	100%	100%
159	fructose-6-P-meox2	0.007	3.8	44%	100%	0.008	2.9	34%	100%	100%
31	Unknown14	0.007	6.3	53%	100%	0.008	2.6	38%	100%	100%
192	unknown148	0.007	3.9	31%	100%	0.005	3.7	33%	100%	100%
161	glucose-6-P-meox2	0.007	4.2	50%	100%	0.008	3.3	41%	100%	100%
64	unknown32	0.007	3.8	53%	100%	0.002	5.1	62%	100%	100%
112	unknown58	0.006	7.2	82%	100%	0.005	36.9	62%	100%	100%
156	sorbitol-6-phosphate	0.006	4.5	44%	100%	0.006	2.6	35%	100%	100%
106	unknown52	0.006	6.5	64%	100%	0.008	3.0	44%	100%	100%
119	unknown61	0.006	14.3	74%	100%	0.008	12.4	90%	100%	100%
123	unknown64			58%	100%			47%	100%	100%

		Control				Perturbed				Total
Peak No	Filename Sample ID Time	avg	min/ max	Total Std. Dev.	% Pres.	avg	min/ max	Total Std. Dev.	% Prs	% Pres.
		0.005	8.4			0.005	4.4			
157	unknown94	0.005	3.7	49%	100%	0.006	13.9	113%	100%	100%
198	unknown161	0.005	2.5	39%	100%	0.004	2.4	29%	100%	100%
191	unknown147	0.005	4.4	55%	100%	0.005	4.3	39%	100%	100%
57	Xylitol	0.004	6.7	48%	100%	0.006	4.6	55%	100%	100%
113	unknown59	0.004	2.7	32%	100%	0.003	1.4	12%	100%	100%
40	2-methyl malate	0.004	2.9	37%	100%	0.005	2.1	22%	100%	100%
67	unknown33b	0.003	4.2	47%	100%	0.006	8.9	78%	100%	100%
87	actonic_ acid	0.003	6.3	58%	100%	0.009	18.5	113%	100%	100%
120	unknown62	0.003	5.1	56%	100%	0.003	3.8	50%	100%	100%
139	unknown80	0.003	5.8	58%	100%	0.003	4.8	54%	100%	100%
197	unknown160	0.002	5.9	58%	100%	0.004	20.8	80%	100%	100%
12	unknown07	0.002	4.6	48%	100%	0.002	16.3	58%	100%	100%
126	unknown67	0.002	4.9	50%	100%	0.002	5.6	42%	100%	100%
11	unknown06	0.001	12.4	104%	100%	0.001	4.3	47%	100%	100%
14	unknown07b	0.001	2.4	27%	100%	0.001	1.8	17%	100%	100%
199	unknown162	0.001	6.6	58%	100%	0.001	3.4	43%	100%	100%
5	unknown03	0.001	2.0	25%	100%	0.001	1.5	16%	100%	100%

Appendix V. List of Metabolites Removed

Sr. No	Metabolite	Cause
1	Valine	< 89% Present
2	Alanine	< 89% Present
3	Leucine	< 89% Present
4	iso leucine	< 89% Present
5	homoserine	< 89% Present
6	hydroxy proline	< 89% Present
7	pyruvic acid	< 89% Present
8	Tryptophan	< 89% Present
9	unknown15b_m/z	< 89% Present
10	unknown25_m/z	< 89% Present
11	unknown44_m/z	< 89% Present
12	unknown77b	< 89% Present
13	unknown81b	< 89% Present
14	unknown82	< 89% Present
15	unknown83b	< 89% Present
16	unknown84	< 89% Present
17	unknown89	< 89% Present
18	unknown91	< 89% Present
19	unknown107	< 89% Present
20	unknown121	< 89% Present
21	unknown123	< 89% Present
22	unknown 127	< 89% Present
23	unknown135	< 89% Present
24	unknown136	< 89% Present
25	unknown 76	< 89% Present
26	unknown 92	< 89% Present
27	unknown 77	< 89% Present
28	unknown 93	< 89% Present
29	unknown 110	< 89% Present
30	unknown 119	< 89% Present
31	unknown 140	< 89% Present
32	lysine	high std. dev
33	Glucosamine	high std. dev
34	unknown 39	high std. dev.

Sr. No	Metabolite	Cause
35	unknown 105	high std. dev.
36	unknown 155	high std. dev.
37	Unknown 28	min/max high
38	Unknown 27	min/max high
39	unknown 75	min/max high
40	unknown 109	min/max high
41	fructose_Meox1_5TMS	Dual Derivatization
42	fructose-6-phosphate-meox1	Dual Derivatization
43	glucose_meox2_5TMS	Dual Derivatization
44	glucose-6-phosphate-meox2	Dual Derivatization
45	Glutamine N,N,O	Dual Derivatization
46	Glycine_N,N,O	Dual Derivatization
47	Serine_N,O,O	Dual Derivatization
48	Asparagine N,N,N, O	Dual Derivatization
49	sucrose	saturated
50	ribitol 319	additional marker ion

Appendix VI. Standard Deviation Analysis

		Average over all Time Points		
Peak	Metabolite	% Standard Deviation		
No		Inject.	Biological	Total
	Average	11%	32%	28%
55	Ribitol_217	0%	0%	0%
93	citrate	2%	25%	19%
117	Inositol	2%	18%	14%
90	fructose_Meox2_5	2%	20%	15%
13	glycerol	2%	43%	32%
38	2-methyl_benzoic_acid	2%	13%	10%
37	unknown15c	2%	20%	15%
61	Glutamine	2%	21%	16%
71	unknown37	2%	31%	23%
44	malate	3%	25%	19%
91	glucose_meox1_5	3%	22%	17%
73	unknown38	3%	21%	16%
104	gluconic_acid	3%	24%	18%
100	unknown46	3%	35%	26%
63	unknown31	3%	18%	14%
152	unknown90	3%	29%	23%
68	Phenylaline_N,O_	3%	19%	14%
22	Phosphoric_Acid	3%	14%	12%
30	Succinate	3%	33%	26%
70	unknown36	3%	79%	59%
41	unknown16	3%	26%	21%
78	unknown41a	3%	47%	35%
4	unknown02	4%	16%	13%
58	Arabinose_Meox1	4%	49%	37%
140	unknown80b	4%	20%	17%
137	unknown78	4%	24%	19%
98	iso-citrate	4%	26%	20%
121	unknown63	4%	32%	24%

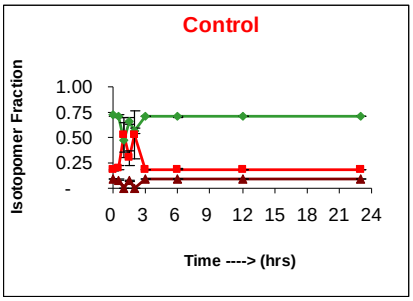
		Average over all Time Points		
Peak	Metabolite	% Standard Deviation		
130	unknown72	4%	19%	16%
159	fructose-6-phosphate-meox2	4%	21%	16%
182	unknown132	4%	29%	22%
24	Glyceric_Acid_O,O,O-	4%	36%	27%
160	glucose-6-phosphate-meox1	4%	26%	20%
2	unknown01	4%	21%	17%
85	mannose	5%	24%	20%
111	unknown57	5%	50%	38%
110	unknown56	5%	45%	33%
89	galactose_meox1	5%	23%	18%
113	unknown59	5%	16%	13%
150	unknown88	5%	33%	26%
33	unknown15	5%	22%	18%
26	unknown12	5%	37%	27%
79	unknown41b	5%	20%	16%
115	hexadecanoic_acid	5%	17%	14%
176	unknown124	5%	27%	21%
162	inositol-phosphate-compound	6%	29%	22%
118	unknown60	6%	48%	37%
57	Xylitol_5_	6%	36%	28%
21	l-Proline	6%	24%	20%
114	saccharic_acid	6%	36%	27%
147	unknown85	6%	22%	18%
60	Cytosine,_2_	6%	23%	18%
50	Threonic_acid_O,O,O,O	7%	28%	24%
112	unknown58	7%	31%	25%
200	unknown163	7%	30%	25%
83	aconitic_acid (1)	7%	27%	22%
139	unknown80	7%	27%	19%
40	2-METHYL_MALIC_ACID	7%	16%	15%
64	unknown32	7%	39%	29%
199	unknown162	7%	32%	25%
187	unknown139	7%	21%	19%
29	Fumarate	7%	17%	16%

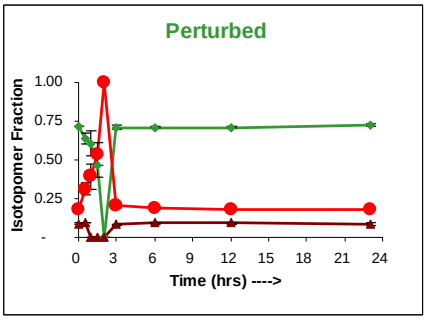
		Average over all Time Points		
		% Standard Deviation		
Peak	Metabolite			
198	unknown161	7%	25%	21%
148	sugar_phospho_comp	8%	33%	30%
65	unknown32b	8%	26%	22%
92	hexopyranose	8%	56%	45%
34	unknown15a	8%	19%	17%
132	unknown74	8%	17%	17%
141	unknown81	8%	31%	25%
189	unknown144	8%	42%	27%
196	unknown159	8%	28%	25%
86	sorbitol	8%	27%	23%
193	unknown149	8%	34%	27%
51	unknown22	8%	24%	20%
88	shikimic_acid	8%	18%	18%
126	unknown67	8%	24%	20%
14	unknown07b	8%	10%	13%
106	unknown52	8%	32%	26%
125	unknown66	9%	40%	34%
190	unknown145	9%	34%	30%
101	3-Phosphoglycerate	9%	25%	23%
20	Serine_O,O_	9%	28%	23%
164	unknown102	9%	20%	19%
69	unknown35	9%	34%	31%
116	ascorbic_acid	9%	21%	21%
156	sorbitol-6-phosphate	9%	23%	20%
178	unknown127	9%	48%	27%
103	unknown49	10%	44%	35%
119	unknown61	10%	49%	40%
	Lactate	10%	22%	21%
17	unknown08	10%	20%	19%
197	unknown160	10%	39%	32%
149	unknown87	11%	27%	26%
192	unknown148	11%	35%	27%
28	Threonine	11%	21%	20%
120	unknown62	11%	26%	26%

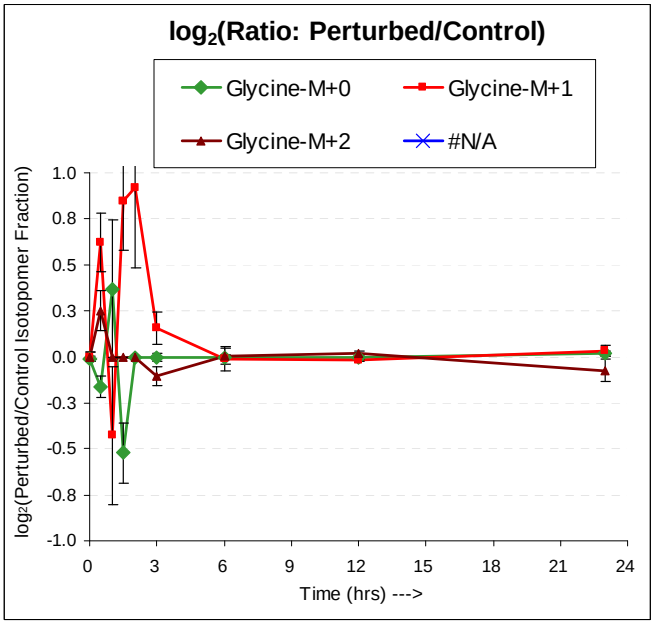
		Average over all Time Points		
Peak	Metabolite	% Standard Deviation		
195	unknown157	11%	33%	28%
177	unknown126	11%	29%	23%
181	unknown130	11%	26%	25%
12	unknown07	12%	25%	23%
127	unknown68	12%	40%	31%
109	unknown55	12%	57%	47%
43	4-Aminobutyric_acid	12%	25%	23%
62	homocystine_2_	13%	27%	26%
16	unknown8a	13%	21%	22%
180	unknown129	14%	32%	30%
191	unknown147	14%	28%	28%
39	3-hydroxy-glutaric_acid	14%	43%	37%
183	unknown133	14%	49%	45%
138	unknown79	14%	31%	30%
32	beta-alanine_3_	14%	40%	34%
108	unknown54	14%	51%	42%
6	Glycine_N,O	15%	51%	43%
8	Oxalic_Acid	15%	36%	32%
107	unknown53	15%	27%	28%
129	unknown71	15%	41%	36%
102	unknown48	15%	48%	43%
124	unknown65	15%	44%	41%
168	unknown108	16%	32%	37%
5	unknown03	18%	16%	22%
80	unknown42	18%	42%	35%
157	unknown94	19%	25%	33%
123	unknown64	19%	26%	31%
59	Glutamate_Tri_	19%	54%	47%
128	unknown70	21%	29%	36%
166	unknown106_phosphoderivative	21%	34%	42%
82	unknown44b	22%	25%	34%
76	unknown40	22%	40%	35%
122	Tyrosine	23%	43%	49%
184	unknown134	23%	28%	37%

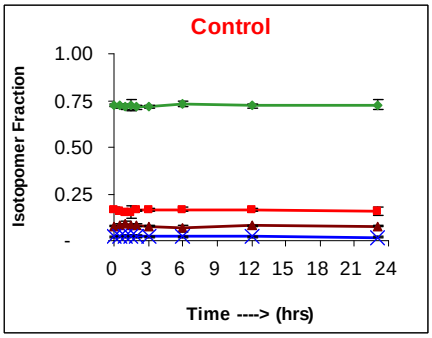
		Average over all Time Points		
Peak	Metabolite	% Standard Deviation		
72	unknown37b	23%	45%	41%
144	unknown83	23%	35%	46%
11	unknown06	23%	28%	34%
179	unknown128	24%	31%	45%
67	unknown33b	24%	51%	43%
105	unknown51	24%	65%	54%
172	unknown111	25%	31%	46%
45	unknown19	25%	53%	45%
10	Ethanolamine	25%	55%	48%
42	unknown18	25%	60%	51%
131	unknown73	26%	47%	47%
49	Asparatate_232m/z	26%	59%	53%
87	actonic_acid(2)	26%	38%	52%
66	unknown33	26%	52%	50%
77	Asparagine_N,N	27%	38%	39%
75	ornithine	28%	31%	35%
97	unknown45	36%	48%	60%
31	Unknown14	37%	51%	60%
27	unknown13	47%	33%	54%

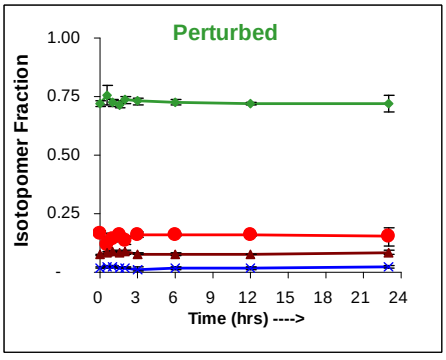
Appendix VII. Control and Perturbed Graphs

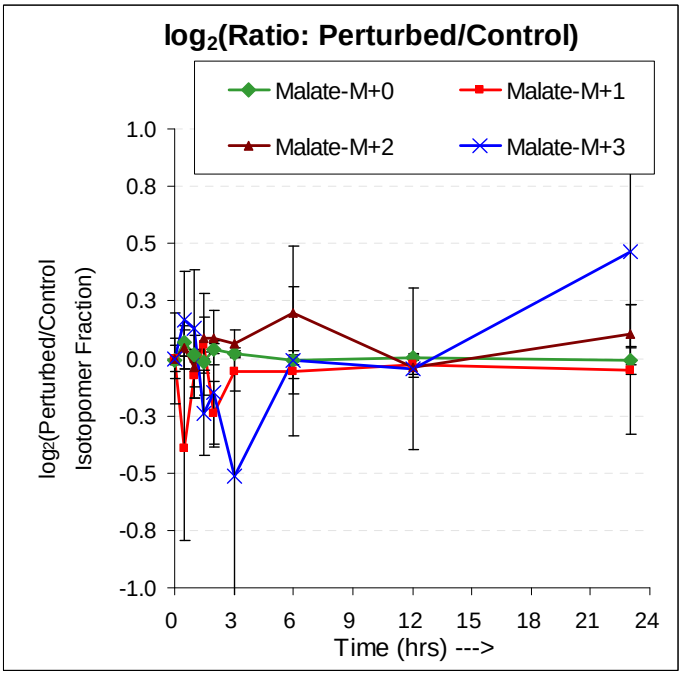


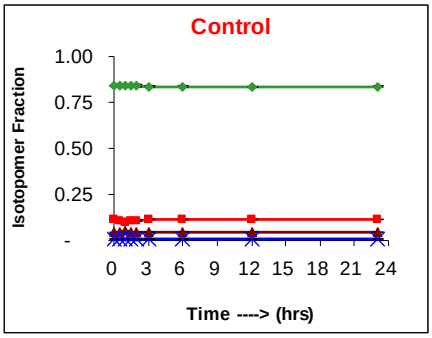


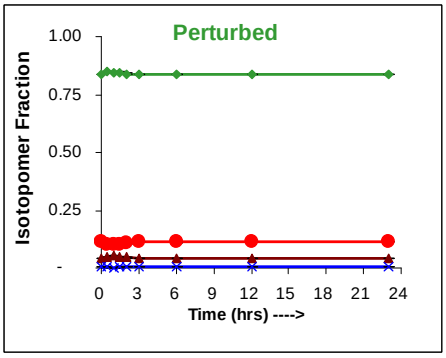


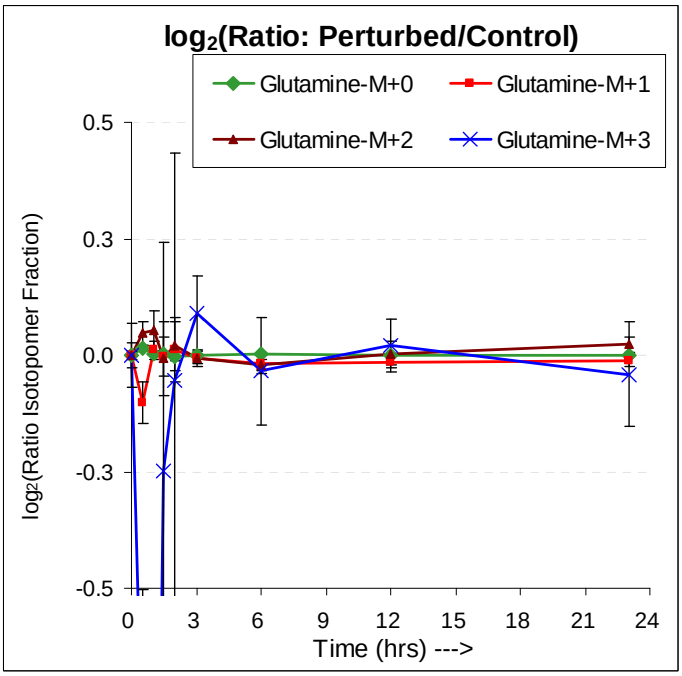


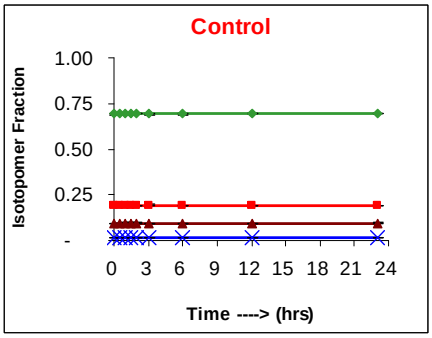


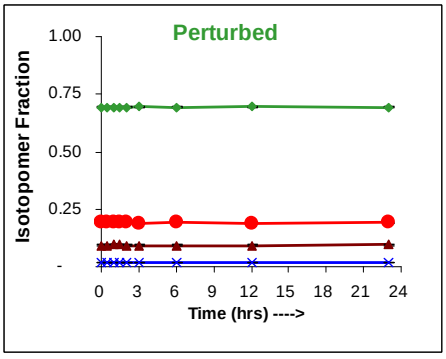


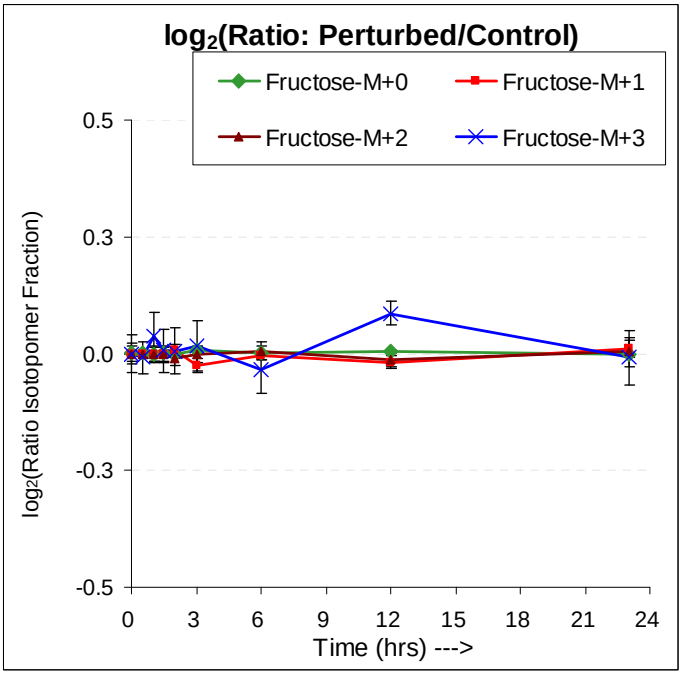

Control
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time ----> (hrs)

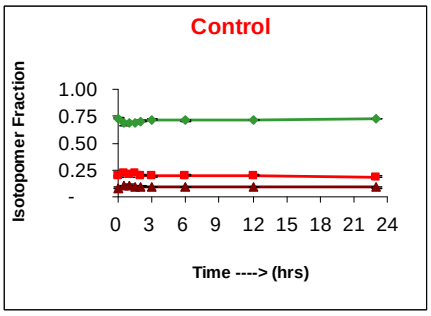

Perturbed
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time (hrs) ---->

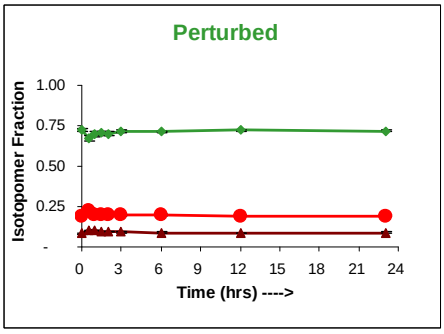

log2(Ratio: Perturbed/Control)
Glutamine-M+0
Glutamine-M+1
Glutamine-M+2
Glutamine-M+3
log2(Ratio Isotopomer Fraction)
0.5
0.3
0.0
-0.3
-0.5
0 3 6 9 12 15 18 21 24
Time (hrs) --->

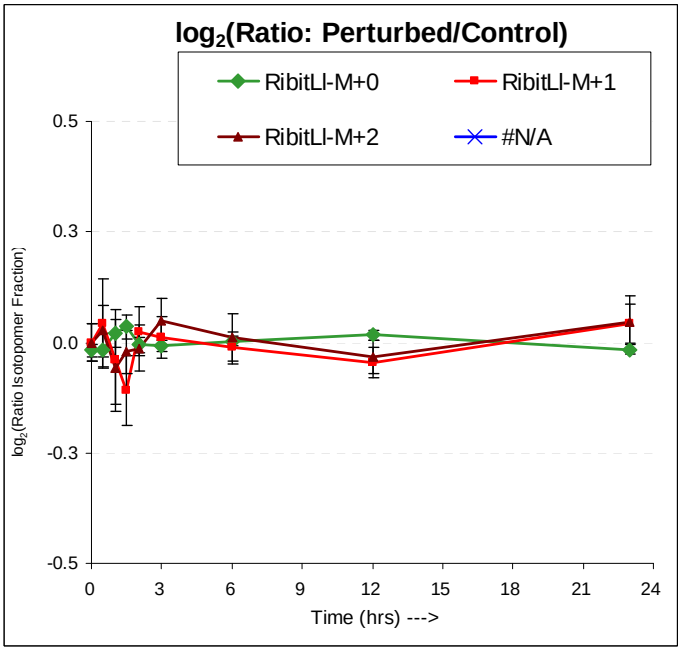

Control
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time ----> (hrs)

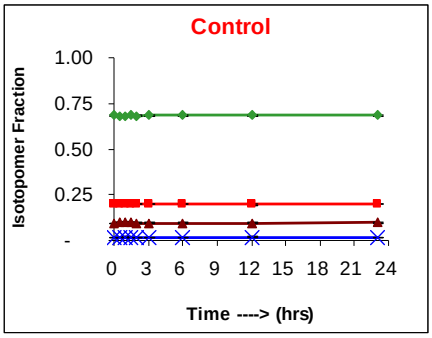

Perturbed
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time (hrs) ---->

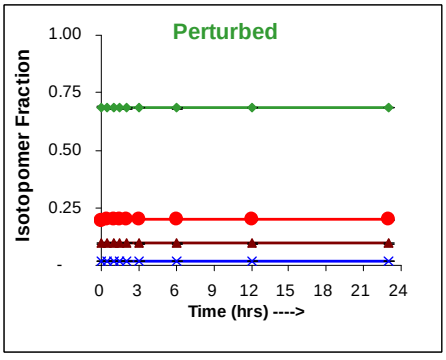

log2(Ratio: Perturbed/Control)
Fructose-M+0
Fructose-M+1
Fructose-M+2
Fructose-M+3
log2(Ratio Isotopomer Fraction)
0.5
0.3
0.0
-0.3
-0.5
0 3 6 9 12 15 18 21 24
Time (hrs) --->

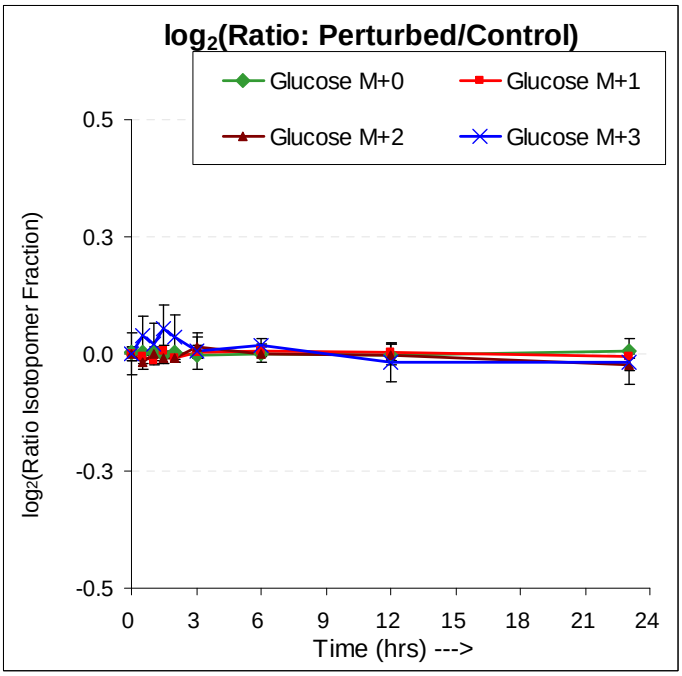

Control
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time ----> (hrs)


Perturbed
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time (hrs) ---->


log2(Ratio: Perturbed/Control)
RibitLl-M+0
RibitLl-M+1
RibitLl-M+2
#N/A
log2(Ratio Isotopomer Fraction)
0.5
0.3
0.0
-0.3
-0.5
0 3 6 9 12 15 18 21 24
Time (hrs) --->


Control
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time ----> (hrs)


Perturbed
Isotopomer Fraction
1.00
0.75
0.50
0.25
-
0 3 6 9 12 15 18 21 24
Time (hrs) ---->


log2(Ratio: Perturbed/Control)
Glucose M+0
Glucose M+1
Glucose M+2
Glucose M+3
log2(Ratio Isotopomer Fraction)
0.5
0.3
0.0
-0.3
-0.5
0 3 6 9 12 15 18 21 24
Time (hrs) --->

References

Arkin A., Shen P., and Ross J., A test case of Correlation metric construction of a reaction pathway from measurements, SCIENCE (1997), Vol. 277, 1275-1279.

Bloom A. J., Smart D., Nguyen D., and Searls P., Nitrogen assimilation and growth of wheat under elevated carbon dioxide., PNAS (2002), Vol. 99, 1730-1735.

Buchanan B., Gruissem W., and Jones R., Biochemistry & Molecular Biology of Plants, American Society of Plant Physiologists, Rockville, Maryland (2001).

Carrari F., Wochinak E. U., Willmitzer L., and Fernie A. R., Engineering central metabolism in crop species, learning the system, Metabolic Engineering (2003), 191-200.

Chen C and Settler T, Response of potato tuber cell division and growth to shade and elevated CO2, Ann Bot (Lond) (2003), Vol. 91, 373-81.

Dey P., and Harborne J., Plant Biochemistry, Acadamic Press Inc., San Diago, California (1997).

Duran A. L., Yang J., Wang L., and Sumner L., Metabolomics spectral formatting, alignment and conversion tools (MSFACTS), Bioinformatics (2003), Vol. 19,2283-2293.

Fiehn A. G., Protocol for Plant Leaf Metabolite Filtering, http://www.mpimp-golm.mpg.de/fiehn/blatt-protokoll-e.html

Fiehn O., Kopka J., Dormann P., Altmann T., Trethway R, and Willmitzer L., Mtabolic Profiling for plant functional genomics, Nature Bioltechnology(2000), Vol. 18, 1157-1161.

Fiehn O., Integrated studies in plant biology using multiparallel techniques, Current opinions in biotechnology (2001), Vol. 12, 82-86.

Grodzinski B., Woodrow L., Leonardos E. D., Dixon M., and Tsujita M., Plant response to short and long term exposure to high carbon dioxide levels in closed environments, Adv. Space Res.(1996), Vol. 18, No. 4/5, 203-211.

Horning M. G., Moss A. M., and Horning E. C., Formation and Gas-Liquid Chromatographic Behavior of Isomeric Steroid Ketone Methoxime Derivatives, Analytical Biochemistry(1968), Vol. 22, 284-294.

Hui D., Sims D., Johnson D., Cheng W. and 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.

Idso S., and Idso K., Effect of atmospheric CO2 enrichment on plant constituents related to animal and humane health, Environmental and Experimental Botany (2001), Vol. 45, 179-199.

Katona Zs. F., Sass P., Molnar-Perl**,** Simultaneous determination of sugar, sugar alcohols, acids and amino acids in apricots by gas chromatography-mass spectrometry, Journal of Chromatography A, 847(1999) 91-102.

Kitson F. G., Larsen B., McEwen C. N., Gas Chromatography and mass spectrometry: A practical guide, Acadamic Press, New York, 1996.

Klapa M. I., Aont J. C. A. and Stephanopolous G., Systematic quantification of complex metabolic flux networks using stable isotope mass spectrometry, European Journal of Biochemistry (2003), Vol. 270, 3525-3542.

Klapa M. I., and Quackenbush J., The quest for the mechanisms of life, Biotechnology and Bioengineering (2003), Vol. 84, 739-742.

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

Kose F., Weckwerth W., Linke T., and Fiehn O., Visualizing plant metabolomic correlation network using clique-metabolite metrices, Bioinformatics (2001), Vol. 17, 1198-1208.

Larios B., Ag⎫era, Cabello P., Maldonado J. M., and Haba P., The rate of CO2 assimilation controls the expression and activity of glutamate synthase through sugar formation in sunflower (*Helianthus annuus L.)* leaves, Journal of Experimental Botany, Vol. 55, 69-75.

Metabolite Mass Spectra Library, http://www.mpimp-golm.mpg.de/mms-library/details-e.html

Paul M., and Foyer C., Sink Regulation of photosynthesis, Journal of experimental botany (2001), Vol. 52, 1283-1400.

Raamsdonk L., Tenusink B., Broadhurst D., Zhang N., Hayes A., Walsh M., Berden J., Brindle K., Kell D., Rowland J., Westerhoff H. V., Dam K. v., and Oliver S. G., A functional genomic strategy that uses metabolome data to reveal the phenotype of silent mutations, Nature Biotechnology (2001), Vol. 19, 45-50.

Ratcliffe G. R. and Hill Y. C., Probing Plant Metabolism with NMR, Annual Reviews in Plant Physiology & Plant Molecular Biology (2001), Vol. 52, 499-526.

Roessner U., Wagner C., Kopka J., Trethwey R. and Willmitzer L., Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, The Plant Journal (2000), 23(1), 131-142.

Roessner U., Leudemann A., Brust D., Fiehn O., Linke T., Willmitzer L., and Fernie A., Metabolic profiling allows comprehensive phenotyping of genetically and environmentally modified plant systems, The Plant Cell (2001), Vol. 13, 11-29.

Roessner U., Willmitzer L., and Fernie A., High Resolution Metabolic Phenotyping of genetically and environmentally diverse Potato Tuber Systems. Identification of Phenocopies, Plant Physiology (2001), Vol. 127, 749-764

Sato T, Tsuzuchi M, and Kawaguchi A., Glycerolipid synthesis in Chlorella kessleri 11 h. II. Effect of the CO2 concentration during growth, Biochim Biophys Acta (2003) Vol.1633, 35-42.

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),Vol. 34, 374-8.

Samoilov M., Arkin A., and Ross J., On the deduction of chemical reaction pathways from measurements of time series of concentrations, Chaos (2001), Vol. 11, 108-114.

Sicher R. C., Effects of elevated carbon dioxide on nitrogen metabolism of barley primary leaves, American Society of Plant Biologists, 2001 Annual meeting abstracts, Abstract # 399, http://abstracts.aspb.org/aspp2001/public/P34/0180.html

Smart D, Ritchie K, Bloom A, and Bugbee B,Nitrogen balance for wheat canopies (Triticum aestivum cv. Veery 10) grown under elevated and ambient CO2 concentrations, Plant Cell Environ (1998),Vol. 21, 753-63.

Stuer R., Kurths J., Fiehn O., and Weckwerth W., Observing and interpreting correlations in metabolomic networks, Bioinformatics (2003), Vol. 19, 1019-1026.

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

Taylor C., Factories of the future? Metabolic engineering in plant cells, The Plant cell (1998), Vol. 10, 641-644.

Taylor J., King R., Altmann T., and Fiehn O., Applications of metabolomics to plant genotype discrimination using statistics and machine learning, Bioinformatics (2002), Vol. 18, S241-S247.

Trenthwey R. N., Gene discovery via metabolic profiling, Current opinions in Biotechnology (2001), Vol. 12, 135-138.

Tusher G. V., Tibshirani R., and Chu Gilbert., Significance analysis of microarray applied to the ionizing radiation response, PNAS (2001), Vol. 98, 5116-5121.

Weckwerth W., and Fiehn O., Can we discover novel pathways using metabolomic analysis?, Current opinions in Biotechnology, 2002, Vol. 13, 156-160.