UNIVERSITY OF CALIFORNIA
RIVERSIDE

Metabolic Profiling of Primary and Secondary Biosynthetic Pathways in Angiosperms: Comparative Metabonomics and Applications of Hyphenated LC-NMR and LC-MS

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Kayla Anne Kaiser

March 2012

Dissertation Committee:
Dr. Cynthia K. Larive, Chairperson
Dr. Dallas Rabenstein
Dr. Julia Bailey-Serres

The Dissertation of Kayla Anne Kaiser is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I have not walked this road alone, many people supported me along the way, both in person and in memory.

My time here at UCR has been absolutely magical. Thanks to the National Science Foundation, I was able to participate in an interdisciplinary graduate training experience (IGERT) which shaped the course of my graduate education from the day I arrived in Riverside. Receiving financial, professional, and moral support from two esteemed principal investigators, Cynthia Larive and Julia Bailey-Serres, is most certainly a once-in-a-lifetime experience and I feel immensely grateful for their effort. I was lucky enough to have mentors that remained actively engaged in my development as a researcher and as a person, who ultimately sent me across the country and around the world to participate in great science, enhancing my graduate training exponentially. Participating IGERT faculty in other departments including Natasha Raikhel, Kathy Borkovich, Linda Walling, Thomas Girke, Thomas Eulgem, Xumei Chen, Shou-wei Ding, Larry Li, Karine Le Roch, Stefano Lonardi, Dimitrios Morikis, Anandasankar Ray, Harley Smith, Patty Springer, Jason Stajich, Zhenbiao Yang, and Hailing Jin have been of more influence than they are aware.

Young faculty such as Wenwan Zhong, Ryan Julian, Katherine Larsen, Greg Beran, Angelina Chang, and Richard Hooley have been encouraging and willing to share their expertise and experience with myself and other students in our department. They have shown me a path to joining them as tenure-track faculty, and I am humbled by their belief that I am capable of rising to such a rank. Chris Bardeen, Jason Cheng, Yinsheng

Wang, Leonard Mueller, Pingyun Feng, Michael Pirrung, Chris Switzer, and Sean Cutler have each influenced my approach to doing science and living life. The balance and tone of our department is set by these dedicated faculty members, who do an excellent job of providing a workplace that is engaging and dynamic as well as world-class. Francisco Zaera, David Bocian, Robert Haddon, Chris Reed, Eric Chronister, and Tom Morton have played a role in my current success and have influenced my current idea of what it means to be a research professor. I would like to especially thank Mike Rettig and Dallas Rabenstein. Teaching for Mike brought me back to the enjoyment of university life and showed me that true commitment to excellence in teaching can appear effortless after many years of practice. Dallas has always offered me perspective beyond what my mind can even comprehend and I consider his contributions to my understanding of NMR and university education to be invaluable.

My fellow IGERT trainees Colleen Knoth, Eddie Cao, Michelle Brown, Charles Jang, James Kim, Christiana Merrywell, Jolene Diedrich, Theresa Dinh, Augusta Jamin, Sean Boyle, Anna Charisi, Andrew Defries, Mindy (Rodriguez) Salus, Missy (Smith) Scranton, Ayesha Baig, Patrick Schacht, Moses Tataw, Shang Wu, Rae Yumul, Aaron Devries, Jessica Diaz, Robert Koble, and Yifan Lii have been an important component of the support web that caught me any time I started flying out of control. These people helped me survive the difficulty of forging complementation at the interface of chemistry, engineering, computational sciences and cell biology and have been the wind in my sails without even knowing it. IGERT support staff Carolina Stickley, Ronly Schlenk, and

Matt Collin helped me balance the additional workload with grace. Glenn Hicks has been a quiet yet steadfast mentor throughout my biological/professional development.

My chemistry classmates Kim Worsley, Tony Ly, Bryce Davis, Matt Linman, James Wright, Robert Dillon, Julia Zhu, Kelly Theel, Sebastian Jezowski, Shinji Kaneko, Jessy Lemieux, Yingdi Liu, Christen Strollo, Olivia Alley, Candace Guerrero, Derek Chen, Ying Liu, Bruce Ford, Mark Hamza, Emily Spencer, Ben Moore, Ni Li, Samantha Schachermeyer, Jingjing Yao, Ana Gamboa, Angie Garcia, Aaron Moehlig, Sepideh Yaghmaei, Robert Carp, Aholibama Escobar, Thomas Owen, Tiffany Satoorian, have been immensely helpful. Together we weathered the storms and reaped the abundant harvest.

My labmates in the Larive research group Bridget Becker and Albert Korir first introduced me to the beauty of NMR and efficient workflows. Kasie Fang and Christiana Merrywell laid the groundwork for my LC-MS and NMR metabolomic experiments, respectively. Stacie Eldridge was a role model and mentor such as I have never had, nor will ever have again. Jennifer Cruz opened my heart, body and mind, as well as my spirit, to new levels of enlightenment. John Limtiaco was my fellow soldier, marching through the Ph.D. process in step with me, making me feel as an equal and sharing the experience with him has been a great treasure. Chris Jones and Daryl Bulloch pushed me to go farther than I thought I was capable of going. In the words of the late John F. Kennedy, “a rising tide lifts all boats.” Without such dedicated and talented labmmates, I would not have achieved as much. Greg Barding is someone who I respect and admire greatly, it has been a real pleasure working with him. Derek Langslay might be an angel,

not only did I enjoy his temperament and scientific contributions, but also he makes a strong pot of coffee almost every day. Consuelo Beecher is our current protégé and I predict she will have an abundant career. Our undergraduate researchers Thao Nguyen, Fernando Campos, Sarah Gutierrez, Layne Higgins, Archie Taylor, Arpana Vaniya, Nelly Membreno, Bailey Dickey, Erik Velez, John Nguyen, Patrick Salvenson, Amy Morris, Avi Ramkishun, Matt Manighalam, and last but not least, Vishwa Shah breathed life into our days and shared with us the experience of training and mentoring. I am so grateful to have worked with all of you. Visiting professor Szabolcs Béni and his wife Orsyla Molnár contributed their energy and expertise to our lab generously, and I am immensely grateful for their support.

A few more words about Cindy: she has the biggest heart of anyone I know. She is absolutely unmatched in her dedication. Working with her has been like climbing Mt. Everest. Just as a mountain never leaves you feeling unsupported, she is rock-solid. By providing formidable challenges, she exhausted and humbled me. She showed me that by refusing to give up on the ascent, breathtaking heights can be reached. That's how Cindy is. She can do the work of 10 grown men. No tall tales, she is actually that amazing.

I also must acknowledge my stunning collaborators in the Center for Plant Cell Biology: Cristina Branco-Price, Angelika Mustroph, Julian M. Peña-Castro, Piyada Juntawong, Takeshi Fukao, Matthew Collin, Erin Brinton, Seung Cho Lee, Reed Sorenson, Rejbana Alam, Maureen Hummel, and Elaine Yeung. They showed me how to cultivate Arabidopsis and maintained good humor while I persistently presented data

of a physicochemical rather than biological nature. Their suggestions and feedback throughout my time at UCR have been instrumental in developing my understanding of plant physiology, molecular biology, and genetics. The attention and encouragement provided by Julia Bailey-Serres was nourishment for my soul, allowing me to take greater risks in my learning process, knowing that she believed in my potential for success.

Finally, a linchpin in my graduate education was my internship experience at the Max Planck Institute for Chemical Ecology in Jena, Thuringia (Germany). NMR and MS group leaders Bernd Schneider and Ales Svatos, respectively, were beyond hospitable and generous in the exchange of ideas and energy. Fellow NMR group members Christian Paetz, Kamel Shaker, Dirk Holscher, Jingjing Fang, Evalgelos Tatsis, Katia Gruner, and Claudia Cervellati must be acknowledged for their contribution to my grasp of natural product studies beyond model and crop systems (Arabidopsis, rice, corn). Phila Raharivelomanana, a visiting professor, freely shared her personal and professional experiences. My friends at the Institute, who welcomed me and integrated me into their program with open arms: Vojta Klusak, Daniel Giddings Vasso, Holger Danner, Andrea Liliana Clavio McCormick, Alexandra Schmidt, Filippo Scialo, Abu Farhan, Guanjun (Cookie) Li, Marko Petek, Filipe Yon, Variluska Fragoso, Hendrik Wunsche, Maja Miloradovic and Arjen van Doorn. Exchanges with these young researchers will undoubtedly form the basis of future collaborations, and I will never forget my 30th birthday.

A few influential artists and musicians who provided me with inspiration and momentum which helped me finish the experiments and writing summarized in this

dissertation. I hereby acknowledge Beyoncé, Lady Gaga, Rhianna and Shakira for bringing girl power to my ears and helping me keep tempo while pipetting. I also drew strength and inspiration from Paul Simon's Graceland album, no matter where in the world I was stationed. The vitality of Anne Hamersky, as well as Al & Eleanor, Paul & Sue, Dan & Jean, Kitt & Steve, Joe & Allison and all their offspring bring honor to the family name, which has populated my reserve of inner strength and pulled me through dark times.

In closing, I have to acknowledge the staffs at UCR who support our research activities. Without their help we could not sustain this enterprise. Administrative professionals Linda Edwards, Susana Aparicio, Tina Enriquez, Colleen Fleming, Kennett Lai, Jaymie Mateo, Nadine Okuns, Barbara Outzen, Vicki Smola and Christina Youhas have set a professional tone and efficient workflow to keep our paychecks and chemicals arriving on schedule. Instrumentation support staffs Dan Borchardt, Yi Meng, Mike Fournier, Craig Graham, Richard Kondrat, and Jeff Lefler have provided guidance and help in abundance in times when I needed it most. Educational support was provided by Rodney Jenks, Allen Nassimian, Kris Pranata, and Kevin Simpson. To Pris Saavedra, Wayne Kaylor, the custodial crew, the grounds crew, Environmental Health and Safety, the Campus Health Center team, the Recreational Center staff, the Graduate Student Association and Associated Students of UCR staff, Graduate Division and the Teaching Assistant Development Program (TADP) and all the other miracle workers that we academics tend to take for granted (and therefore have been forgotten by me at this moment)… I acknowledge you and your contributions to this work.

Copyright Acknowledgements

The text and figures in Chapter 1 are reprinted from *Metabolic Profiling*; Kayla A. Kaiser, Christiana E. Merrywell, Fang Fang, and Cynthia K. Larive in NMR Spectroscopy in Pharmaceutical Analysis; Ulrike Holzgrabe, Iwona Wawer, and Bernd Diehl, Eds.; Elsevier: San Diego, CA, 2008; pp. 233-268.

Text and figures in Chapter 2 are reprints from "A comparison of metabolite extraction strategies for ^{1}H NMR-based metabolic profiling using mature leaf tissue from the model plant *Arabidopsis thaliana"* Kayla A. Kaiser, Gregory A. Barding, Jr. and Cynthia K. Larive. Magnetic Resonance in Chemistry. (2009) Volume 47 Issue S1, Pages S147-S156.

Chapter 2 also contains supporting material from "Use of ^{1}H Nuclear Magnetic Resonance To Measure Intracellular Metabolite Levels during Growth and Asexual Sporulation in Neurospora crassa" James D. Kim, Kayla Kaiser, Cynthia K. Larive, and Katherine A. Borkovich. Eukaryotic Cell. (2011) Volume 10 Number 6, Pages 820-831.

Text and figures in Chapter 3 are reprinted with permission from "Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*" Cristina Branco-Price, Kayla A. Kaiser, Charles J. H. Jang, Cynthia K. Larive, and Julia Bailey-Serres. The Plant Journal (2008) Volume 56, Issue 5, Pages 743-755.

To all those who have gone beyond the edge of the map,

who push forward until they are exhausted and can no longer carry on,

to those who awake each day craving new experiences,

unsatisfied with what was “good enough” for yesterday,

to those who dedicate themselves to training future scientists;

together we will conquer what remains unknown and yet to be explored.

This dissertation is dedicated to my husband and our progeny,

along with our academic family, may we grow ever closer to truth.

ABSTRACT OF THE DISSERTATION

Metabolic Profiling of Primary and Secondary Biosynthetic Pathways in Angiosperms: Comparative Metabonomics and Applications of Hyphenated LC-NMR and LC-MS

by

Kayla Anne Kaiser

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2012
Cynthia K. Larive, Chairperson

The goal of this dissertation was to advance plant metabonomics through optimization of biological experimental design, sampling and sample preparation, data acquisition and pre-processing, and multivariable data analysis. The analytical platform most employed for comparative metabonomics was nuclear magnetic resonance (NMR). Liquid-chromatography (LC) coupled to NMR and mass spectrometry (MS) extended metabolic profile coverage from primary into secondary metabolic pathways. Comparative profiling of tissue extracts by LC-MS and complementary analyses by NMR were performed to establish metabolite identity and quantify responses to low-oxygen stress using the model Angiosperm, *Arabidopsis thaliana*.

Angiosperms, or flowering plants, constitute the most diverse and numerous group of land plants, and the most commercially important, producing an estimated 200,000 small molecule metabolites through conserved primary metabolism and divergent secondary pathways. No single metabolic profiling experiment has been able to quantify metabolite abundance over the relevant dynamic range (spanning seven orders of magnitude) and across the necessary structural diversity (chemical space coverage

exceeds all known synthetic molecules). Challenges in plant metabolic profiling also arise through inherent limitations in methods (availability of diverse solid-phases, solvents, pure standards), instrument performance (reproducibility, robustness, speed, sensitivity), and data management (pre-processing, modeling, data basing and sharing).

The field of plant metabolic profiling has grown steadily over the past two decades, although it is still considered an emerging area of research. Comparison of methods for sample preparation provided a foundation for one-dimensional NMR-based comparative metabonomics. Signal overlap was alleviated by liquid-liquid extraction (LLE), solid-phase extraction (SPE), and liquid chromatography (LC). A combined approach involving ultraviolet-visible absorption spectroscopy (UV-Vis), NMR, MS, and tandem MS (MS/MS) was used for dereplication.

Results are presented illustrating the ability of one-dimensional NMR combined with multivariable data analysis (MVDA) to quantify responses of a model plant to a biological problem of fundamental and practical relevance: cellular respiration under oxygen-limited conditions. The suitability of LC-NMR and LC-MS for dereplication is highlighted. Profiling of secondary metabolites by UPLC-MS combined with MVDA was undertaken for comparison with NMR-based comparative metabonomics. Efforts toward quantitative targeted analyses, hurdles and solutions, are highlighted.

SUPPLEMENTARY MATERIALS:

Appendix A: NMR-based Primary Metabolite Database

Appendix B: NMR & MS-based Secondary Metabolite Database

TABLE OF CONTENTS

Chapter Four: Micrometabolomics of hypoxic seedling roots and shoots of Arabidopsis

Chapter Five: LC-NMR and LC-MS Structural Elucidation of Arabidopsis Secondary Metabolites

LIST OF FIGURES

FIGURE PAGE

LIST OF TABLES

TABLE PAGE

Chapter One

Introduction to Metabolic Profiling

Contains excerpts from *Metabolic Profiling*; Kayla A. Kaiser, Christiana E. Merrywell, Fang Fang, and Cynthia K. Larive in NMR Spectroscopy in Pharmaceutical Analysis; Ulrike Holzgrabe, Iwona Wawer, and Bernd Diehl, Eds.; Elsevier: San Diego, CA, 2008; pp. 233-268. Spectra presented in Figure 1.1 are unpublished data (A) and from (B) "Use of ^{1}H NMR to measure intracellular metabolite levels during growth and asexual sporulation in *Neurospora crassa*", James D. Kim, Kayla A. Kaiser, Cynthia K. Larive, and Katherine Borkovich, Eukaryotic Cell, Volume 10, Number 6, Pages 820-831, 2011, (C) "A comparison of metabolite extraction strategies for ^{1}H NMR-based metabolic profiling using mature leaf tissue from the model plant *Arabidopsis thaliana",* Kayla A. Kaiser, Gregory A. Barding, Jr. and Cynthia K. Larive, Magnetic Resonance in Chemistry, Volume 47, Issue S1, Pages S147-S156, 2009, and (D) "Differential Metabolic Regulation Governed by the Rice *SUB1A* Gene during Submergence Stress and Identification of Alanylglycine by ^{1}H NMR Spectroscopy", Gregory A. Barding, Takeshi Fukao, Szabolcs Béni, Julia Bailey-Serres, and Cynthia K.Larive, Journal of Proteome Research, Electronic publication available Nov 11 2011.

The research presented in this dissertation addresses current issues in the analysis of plant metabolism. The post-genomic era is densely populated with databases reporting the results of "-omic" studies, aiming to measure the complement of a particular type of biomolecule within a cell, tissue, organ or organism. Accordingly, metabolomics is the attempt to identify and quantify all endogenous small molecule metabolites in an organism or biofluid sample. In this work, metabolomic approaches as well as more targeted metabolic profiling were applied to study an "-omic" model system, the flowering plant (Angiosperm) *Arabidopsis thaliana*. The fundamental question: Analytically, can we acquire representative metabolite data that is biologically relevant and informative? Vascular plants have evolved over the past 160 million years and contain genetic programming that can be tapped with the latest advances in molecular biology. These include high-throughput transcriptome analyses of temporally and

spatially regulated genes that are nearly 10 years ahead of metabolomics in terms of standardization and publicly-available resources. We aimed to provide information regarding cellular processes through the identification and quantitation of small molecules (< 800 Da) present within a whole plant or organ. In this manner, our goal was to perform comparative studies to reveal dynamic changes in the metabolome in response to environmental perturbation. These methods are applicable to studies of plants with different genetic makeup (i.e. ecotypes or mutants) as well as different plants or other organisms.

The instrumentation at our disposal for small molecule analysis is high-resolution nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy has been shown to illuminate the identity and quantity of chemical nuclei present in a homogeneous liquid sample, such as an extract of either fresh or dried plant material.[1] An alternative approach, particularly useful when it becomes impractical to study mixtures at the working concentrations necessitated by NMR, is to separate components of the complex mixture by chromatography with detection by mass spectrometry (MS), a more sensitive analytical technique. Although MS can inform structural hypotheses, gas-phase ions are distinctly different than solvated endogenous metabolites extracted from intracellular compartments of plant cells, and molecular identity is not always readily established. MS also suffers from difficulties in establishing quantity, since instrument drift and fouling are common over the course of a large-scale metabolomics experiment. In studies of targeted metabolites, such problems are overcome by synthesis of a labeled internal standard, which becomes impractical when the goal of the experiment is to quantify as

many endogenous molecules as possible in a given sample. Approaches that grow the organism under study on isotopically labeled media, or in the presence of a labeled biosynthetic precursor to one specific pathway are gathering momentum but as with other "omic" studies systems biology experiments can be as expensive as $1000 per sample. One approach employed dual labeling using Trp$^{2}H_5$ and Trp$^{13}C_{11}{}^{15}N_2$ in WT and a T-DNA activation-tagged line affecting a MYB-domain transcription factor responsible for regulating anthocyanin biosynthesis.[2] Excised leaves of 8 wk old Arabidopsis were immersed in a solution containing the isotope for 24 h in the light to ascertain relative quantitation of a number of metabolites in secondary pathways, yielding an average relative standard deviation (RSD) of 25% (4 – 45%).

The goal of this dissertation is to consider both analytical best practices in method development prior to conducting a large metabolomic experiment, then to dig deeply enough into the data to contribute to the knowledge of the biological community, which was accomplished through the following objectives:

Objective 1: Establish by comparative metabolomics a protocol for tissue preparation, with high reproducibility and robustness, yielding stable solutions for analysis by ^{1}H NMR (Chapter 2).

Objective 2: Demonstrate using a severe biological perturbation (hypoxia to anoxia) that transcriptomic, translatomic and metabolomic data can be integrated for identification of reconfigured biological processes for survival of cellular energy crisis (Chapter 3).

Objective 3: Observe unique responses of aerial (shoots) and subterranean (roots) organs to ephemeral stress (hypoxia) by alleviating spectral crowding via pH adjustment and observing conditions for qNMR to improve experimental outcomes (Chapter 4).

Objective 4: Identify compounds for which standards are unavailable by chromatographic resolution of individual molecules and structural elucidation by ^{1}H NMR and MS/MS (Chapter 5).

Objective 5: Compare secondary metabolite profiles following severe perturbation (anoxia) using UPLC-MS and XCMS software for data post-processing (Chapter 6).

The first part of this introductory chapter presents an overview of metabolic profiling, discusses challenges, and presents the state-of-the-art in this emerging field. This is followed by a discussion of the motivation for working in plants and studying the problem of low oxygen (hypoxia/anoxia) stress and its metabolic consequences. Predominantly NMR was used in this work (Chapters 2-5), therefore the majority of the discussion in this chapter will focus on issues related to NMR-based metabolic profiling. Minor consideration will be given to MS since it was employed for comparison (Chapters 5-6).

1.1 Metabolic Profiling: Methods & Approaches

Metabolic profiling lies at the intersection of analytical chemistry and biological investigations. It has proven to be a useful approach for annotating genes of unknown function,[3, 4] unraveling the effects of metabolic engineering to facilitate network discovery,[5, 6] to arrive at a better understanding of plant physiology,[7, 8] for natural product discovery,[9, 10] chemical ecology,[11, 12] and taxonomic classification of plants.[13]

In this dissertation, existing metabolomic methodologies were adapted to specifically address questions of interest in survival of low oxygen (hypoxia/anoxia) stress using the model plant *Arabidopsis thaliana*. Living systems aim to maintain homeostasis during periods of stress. Applying a widely targeted approach, we hoped to observe significant metabolic changes in pathways that had not been previously reported.[14] This chapter will serve to articulate challenges in metabolic profiling, describing in detail analytical challenges, instrumental considerations, and post-acquisition processing considerations, critical to consider if one wants to acquire a meaningful dataset. Next, this chapter will articulate the ways and means by which "straw is spun into gold" whereby cleaned matrices of data are subjected to classification and interpretation for biological relevance. Finally, this chapter will provide a foundation and motivation for the study of plants under low oxygen conditions.

1.1.1 Definition of Metabolic Profiling

Metabolic profiling involves the identification and quantification of specific endogenous metabolites. A single metabolic profile can be thought of as a snapshot of

the metabolic state of an organism at a given moment. Comparison of metabolite profiles can reveal distinguishing features between distinct genotypes and differences between stressed and control plants. Metabolic profiling often focuses on single classes of compounds (such as polar lipids, isoprenoids, carbohydrates, amino acids) or members of a particular metabolic pathway (glycolysis, trichloroacetic acid (TCA), branched-chain amino acid biosynthesis, urea cycle). Because of its quantitative and selective nature, metabolic profiling is most useful for hypothesis testing. For example, a metabolically engineered petunia was found to produce more phenylpropanoids, compounds that give rise to floral color and fragrance.[15] These traits lend an advantage for plant fitness (i.e. transfer of genes to subsequent generations through production of seed) because they enhance plant-pollinator interactions.

A related field is metabolomics, sometimes called metabonomics.[16] Although some authors attempt to distinguish between these terms, they are used interchangeably in this dissertation to describe the identification and relative quantification of all endogenous metabolites present in an organism or biofluid sample. Metabolomic investigations are particularly useful when the effect of a genetic transformation,[5, 17] disease state,[18] biotic or abiotic environmental stimulus,[18, 19] or administration of a small molecule[20, 21] is unknown or affects multiple biochemical pathways. The goal of metabolomic investigations is typically a general understanding of the organism's response to a perturbation. Therefore these experiments typically do not focus on a specific pathway or metabolite class. Metabolomic data is sometimes merely qualitative, in these cases it is termed metabolic fingerprinting, and the dataset is used for the

unsupervised classification of samples or the generation of predictive models to describe the difference between several states of an organism.[22] Metabolic fingerprints of several organisms are presented in Figure 1.1.

Metabolomics and metabolic profiling are still undergoing standardization in methodology and data handling. Recommendations have been made for minimum criteria to accept measurements[23] and reporting of associated metadata.[24] The current state of the art will be summarized here as it pertains to the works included in this dissertation. Methods and approaches will be explored with respect to experimental outcomes.

1.1.2 Goals of Metabolic Profiling Experiments

Quantitative experiments can be carried out in two distinct modes: absolute and relative. An experiment aimed at absolute quantitation is considerably more challenging, owing to the fact that for most techniques, analytes with equal concentrations give different detector responses. This is not the case for NMR, where the signal is directly proportional to the number of nuclei giving rise to a resonance, assuming the data has been acquired with appropriate care. This makes NMR an excellent tool for quantitative comparisons of metabolite levels.

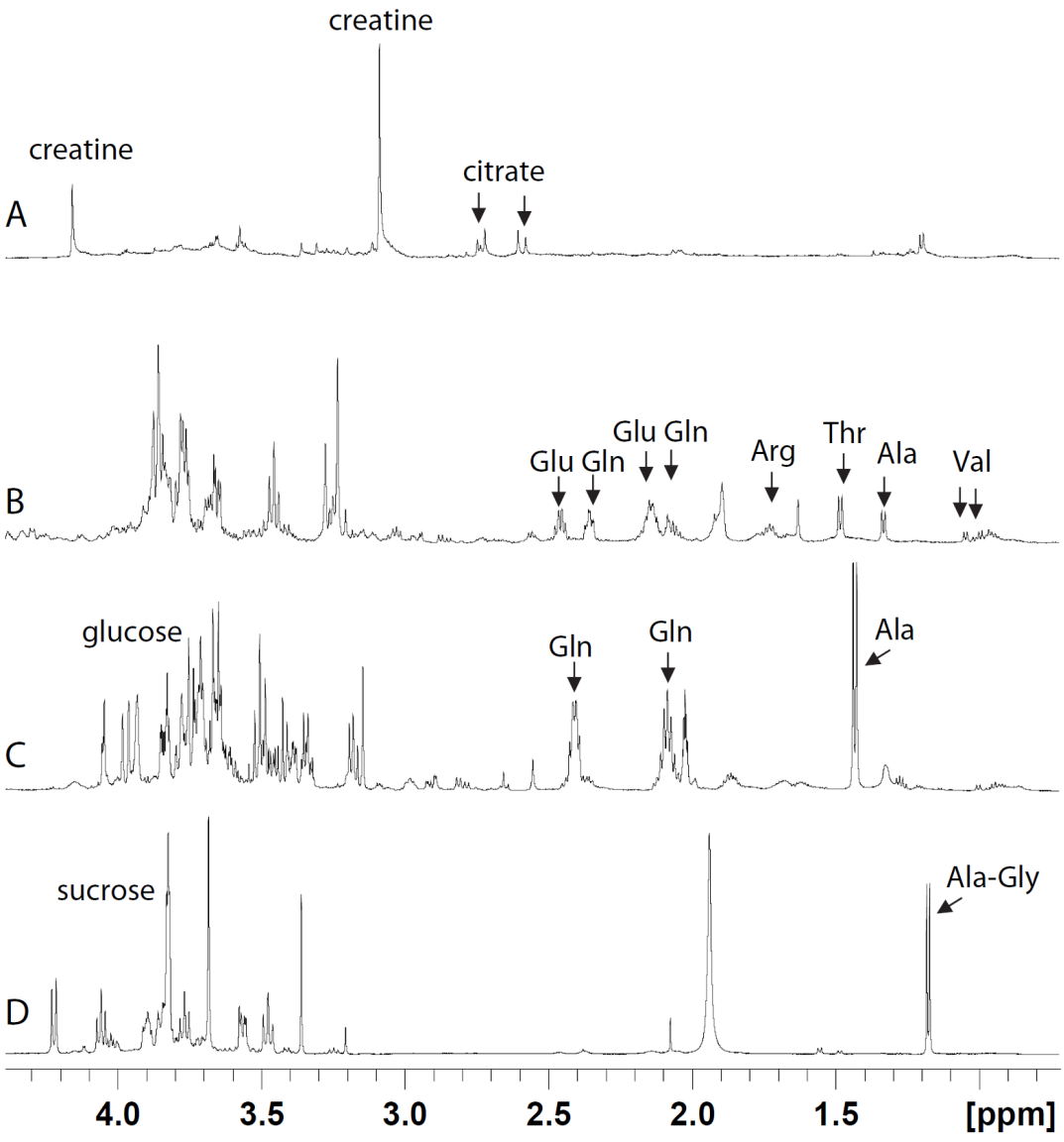


Figure 1.1. Metabolic profiling by 1D ^{1}H NMR spectroscopy records snapshots of molecular phenotype allowing identification and relative quantification of resonances generated by abundant metabolites in human urine (A) and tissue extracts of *Neurospora crassa* (B), *Arabidopsis thaliana* (C), and *Oryza sativa* (D).

More commonly, the objective of metabolic profiling experiments is the discovery of one, or more, measured variables whose values can be used to distinguish

between two populations, classically deemed "case" and "control." This experimental design fits the binary class discrimination type, which has been employed in mainstream biology and medicine for many years.[25] In this type of analysis absolute quantitation is not required and the signals in each case can be reported relative to the signals in each control either as a difference (case – control) or as a ratio (case/control) as long as the two datasets are acquired under identical conditions.[26] The approach of reporting metabolite fold changes has also been used in conjunction with liquid chromatography with electrospray ionization coupled to a quadrupole time-of-flight mass spectrometer (LC/ESI-QTOF-MS) in the characterization of secondary metabolites in seed extracts.[27]

1.1.3 Metabolic Profiling: Past and Present

An engaging history of the word metabolism is afforded by Bing[28] elsewhere, some of which will be presented here. The father of the study of metabolism is generally recognized as the Italian physician and physiologist Sanctorius. His studies were published in 1712, which describe an experiment performed on himself. He sat on a chair mounted to a large balance after eating a meal, waited several hours, then determined that his weight became gradually lighter. He conducted many more experiments, but his study of metabolism was limited to "balance experiments." These were based solely on mass change, but his work lead to more thought about the constant flux of living organisms.

A famous paper by Schwann, a German physiologist, was the first to propose that all organisms are composed of cells. It was published in 1839, using the word

metabolische to describe the fluctuating nature of the molecules constituting a cell. A German chemist, Liebig, was the first to use the word metabolism in 1842. The publication of a textbook by Foster in 1878 solidified the use of the word metabolism as the study of chemical change in cells.

The specific study of amino acid fluctuations was published in 1913 using a variety of titrations for quantitation.[29] The Nobel Prize in Medicine was awarded in 1922 to Meyerhof for identifying the relationship between the consumption of oxygen and the metabolism of lactic acid in muscle tissue. In 1925, Meyerhof continued his studies of metabolism, constructing pathways for glycolysis. The Nobel Prize in Medicine in 1953 was awarded to Krebs for his elucidation of the TCA cycle. For a comprehensive summary of Krebs' contributions to the field of metabolism, the reader is directed elsewhere.[30]

Metabolic profiling was carried out as early as 1969 by MS[31] and issues relating to analysis of biofluids by NMR[32] were reported in 1978 leading to studies of intact cell metabolism[33] by 1984. Work by Vandergreef, Tanaka, and Nicholson in the 1980s using soft-ionization MS, gas chromatography (GC) mass spectrometry, and NMR, respectively, may have been the birth of metabolomics. A metabolic control theory was published in 1985, wherein Derr defined the metabolome as the metabolic composition of a cell. Several areas of study ending in "omics" were born out of the human genome project, conducted between the years 1989 and 2003. These include genomics, proteomics, transcriptomics, and metabolomics.[34]

A useful discussion of the merits and drawbacks to the primary analytical platforms for metabolic profiling (NMR and MS) is provided by Robertson,[16] who states in conclusion

> "most groups who are seriously pursuing metabonomics [sic] are currently or soon will be using both platforms."

A review of recent literature and our direct experience shows this statement to be true.[35, 36] Although some purveyors of metabolic profiling champion either NMR or MS as superior, and encourage exclusive use of one or the other, these scientists seem to be in the minority. Therefore, consideration of both analytical platforms will be given in this dissertation.

The journal *Metabolomics*, considered the gold standard for reporting in our field, outlined the metabolomics discovery process (pipeline) in 2005 consisting of the following stages: experimental design, instrumentation optimization, data gathering, data analysis/modeling, and finally model cross-validation.[37] In our experience, the robust statistical treatment of data, projection of that data onto biochemical pathways, and generation of biological conclusions are the ultimate aims and often the most challenging aspects of a metabolic profiling experiment. Each of these steps and their importance in determining the quality/integrity of measurements in plant metabolomics will be considered in turn. Alternative workflows are used by others.[37-39] One inherent challenge in metabolomics is that it is cross-disciplinary; successful studies require innovation in handling and analysis of complex mixtures, biochemical sample preparation, and interpretation of results by chemometrics and databasing.[40]

Quality is more difficult to quantify in metabolomics experiments due to a difference of opinion in what constitutes a "metabolite" versus a "signal" in a large dataset, and whether "signal" quantity is equivalent with experimental quality. A current view is that quality metabolomic data is acquired in a form which maximizes the dynamic range of the instrumentation (each has inherent limitations) in such a way that provides an accurate reflection, when measured at the bulk level, of what is happening at the cellular level when a biological event (treatment) occurs. Challenges to achieving this desired outcome are presented in the following section. This review focuses primarily on NMR spectroscopy with minor consideration of LC-MS; these techniques were used in subsequent chapters for analytical and biological studies.

1.2 Challenges in Plant Metabolic Profiling

Challenges in plant metabolic profiling arise from biological, analytical, and technical sources. In this section, what is known about the dynamic range of metabolite concentrations is discussed to lead to the understanding that it is not currently analytically possible to carry out metabolomics in a single experiment. Considerations in sampling and number of replicates in experimental design are covered in detail. Quantitative analysis poses a greater number of constraints on the experimental design, and various approaches are outlined here as they relate to the original work conducted for this dissertation. Alternative approaches for intact tissue measurements and preparation of tissue extracts are discussed briefly, more consideration is given to the latter as the subject of Chapter 2.

1.2.1 Dynamic Range of Metabolite Measurements

Although metabolic pathways are largely conserved across organisms, it is difficult to find published values for average *in vivo* metabolite concentrations even for common model biological systems such as *Escherichia coli* (bacterium), *Saccharomyces cerevisiae* (fungus), *Caenorhabditis elegans* (invertebrate), *Drosophila melanogaster* (insect), and *Rattus norvegicus* (vertebrate). The concentration data reported for 63 metabolites in *Bacillus subtilis,*[41] obtained using MS, indicates that the range of metabolite abundance spans five orders of magnitude, but the majority of metabolites measured (~83%) span only two orders of magnitude. It has been proposed that plant metabolites span seven orders of magnitude,[24] from plant hormones present in trace amounts to C and N transport compounds such as sucrose and asparagine, respectively, that are found at much higher levels.

Quantitative ^{1}H NMR has demonstrated an achievable dynamic range of greater than 300:1, which potentially allows a large number of metabolites to be sampled quantitatively by this technique in a single experiment.[42] This is demonstrated by Moing et al.,[43] who report concentrations of 17 metabolites in extracts of *Arabidopsis thaliana* rosette leaves (60 d) over a wide range of concentrations (0.02 – 25.21 mg / g DW). Quantitation by NMR was shown to be linear and correlated well with other standard methods, including enzymatic assay and high-pressure liquid chromatography (HPLC) determinations for selected metabolites. In a study comparing two stationary phases to characterize penicillin fermentation broth extracts using LC-MS, the authors report

absolute concentrations of 19 metabolites (in mg/L) via external calibration and standard addition,[44] where concentrations spanned a similarly wide range (0.2-700 mg/L).

LC-MS based metabolite profiling revealed a limit of detection between 1 and 10 μg/L.[44] In this technique, the analyte response is dependant on the ionization yield of the sample the ESI source and ion saturation of the electron multiplier detector. Detector saturation is independent of the analyte, whereas ionization in the ESI source depends on the analyte structure and concentration as well as the sample matrix. About half of the nearly 200 metabolites studied in antibiotic fermentation broths by Preinerstorfer et al.[44] showed a linear range over four orders of magnitude.

1.2.2 Obtaining a Valid, Representative Sample

Care needs to be taken in deciding what constitutes the sample. In plant studies, a sample may consist of one leaf, many leaves from one plant, a single whole plant, or a homogenous pool of leaves from many plants. Growth media will have a metabolic consequence, as well as access to water, temperature, light availability, light cycles or humidity. Natural diurnal and seasonal variations in metabolic profiles have been observed across organisms of all kingdoms: bacteria, fungi, plants,[45] and animals.[46] These factors can be somewhat controlled by careful experimental design.[47] For example in plant studies, planting and harvesting at regular time intervals can average out variations in the growth environment if the resulting tissues are pooled.[48]

1.2.3 Replicates and Variability

The term replicate is often used in the literature without indication of what type of replicate was conducted in the experiment. A machine replicate means that a single identical sample was analyzed multiple times by the same instrument. For NMR experiments, there is little variability (on the order of 1% using a high-purity standard[49]) between machine replicates and for this reason machine replicates are not typically measured. In an interlaboratory comparison of identical urine samples using NMR, errors arose from factors such as efficiency of water suppression and it was discovered that different acquisition parameters were used.[50] When identical acquisition parameters were used, the differences fell to 1%.

For LC-MS experiments precision was determined by repeated measurement of standard mixtures at three concentration levels. Precision was below 10% RSD for a majority of metabolites and greater than 10% RSD for between 5-30% of metabolites, depending on the solid phase and pH of the mobile phase used for the chromatographic separation. These parameters were determined using a training set of 36 metabolites.[44]

An analytical replicate is a single homogenous sample subjected to the complete analytical protocol, including the sample preparation procedure.[51] In particular, the efficiency and reproducibility of extraction must be determined for each set of metabolic profiling experiments when evaluating samples that require extraction steps.[52] It is difficult to estimate the magnitude of variation caused by sample preparation because it depends on the protocol, the sample under investigation, and the care taken by the experimenter who prepares the sample for analysis. Variation between analytical

replicates has been minimized through the use of automated sample handling.[49] An advantage of NMR analysis of biological fluid samples is that minimal sample preparation is required, limiting the error contributed by sample manipulations.

For LC-MS analysis sample preparation tends to be more involved to protect the integrity of the instrumentation and provide reliable results. A commonly used technique for sample cleanup and enrichment of target molecules is solid-phase extraction (SPE). For ten commonly used post-harvest fungicides, which were detected in the matrix of plant material via SPE enrichment prior to separation and detection by LC-MS, the intra- and inter-day RSD ranged between 2-23% (n = 6).[53] In another study, enrichment of glucosinolates was achieved using SPE in extracts of broccoli seeds[54] prior to structural characterization by direct-infusion ESI-MS and distortionless enhancement by polarization transfer (DEPT) NMR experiments.[55]

Biological replicates are samples derived from separate organisms in which an attempt has been made to control all possible sources of biological variability such as genotype, age, diet, growth environment and even the phase of an organism's diurnal cycle at the time of sampling. To evaluate the robustness of the metabolic profiling experimental design, biological replicates can be split into analytical replicates and assayed several times to constitute machine replicates (Figure 1.2). It is essential that the variation between analytical replicates be less than between biological replicates.[56] For example, Fiehn et al.[57] estimated that the analytical variance was ~8% using GC/MS for metabolic profiling of plant extracts whereas the biological variance was estimated at 26-

56%. As with analytical variability, the biological variability depends on many factors and should be established for each experiment.

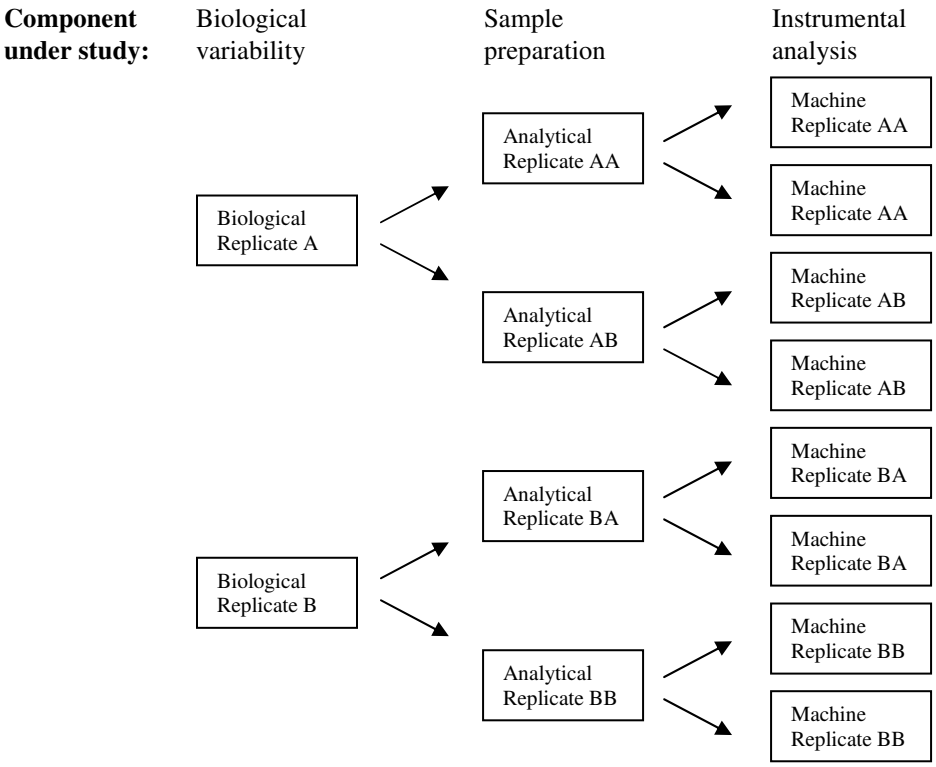


Figure 1.2 Strategy for evaluating the robustness of metabolic profiling methods through the use of replicates.

The question of sample size has been addressed rigorously by Dobbin and Simon,[58, 59] who conclude that there is no "rule of thumb" for how many replicates are necessary. Instead, the authors propose the use of formal calculations based on several parameters: the goals of the experiment, whether pooling is employed in the sampling procedure, and the relative magnitudes of different sources of variability. Although their calculations are applied to data from microarray experiments, the data from microarrays

and metabolic fingerprints is comparable in their degree of complexity and high dimensionality. In plant studies, samples from several plants are often combined to average inherent biological variability.[60, 61] In one such study, it was found that up to 30% of the major metabolites varied significantly in tomato fruit from plants grown under identical conditions.[62]

When devising the analysis sequence, samples should be randomly analyzed to avoid any unrelated effects due to systematic instrument changes. In addition, measures should be taken to control analytical performance deterioration. The use of an internal standard helps to compensate for such effects, but metabolomic studies by their nature limit the effectiveness of this approach due to the great diversity of analytes measured. An alternative to an internal standard is to use a quality control (QC), which is a pool of all samples reflecting a global picture of all analytes present in the samples.[6, 63, 64] This QC sample should be placed in the sequence outlined in Figure 1.3 adapted from van der Greef et al.[6] QC samples should be injected each 3-5 hours.[64] The QC sample can be spiked with an internal standard (STD).[63]

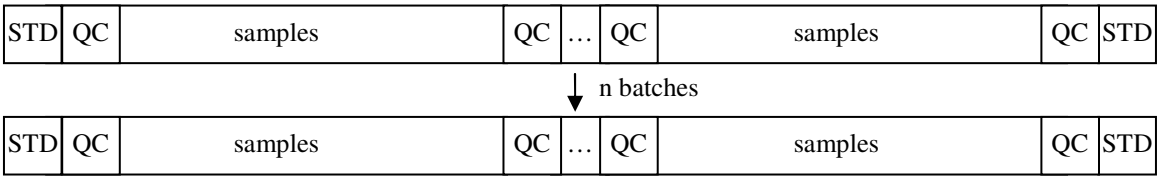


Figure 1.3 Standardized analysis design for large-scale metabolomics studies. Calibration standards (STD) in a mixture allow quantification of specific metabolites. Biological calibration samples (QC) correct for intrabatch temporal trends and interbatch differences. Samples are randomized with respect to run order and batch.

Multivariate data analysis (post-acquisition) should reveal a tight cluster among QC samples to establish absence of systematic bias. Also included in the analytical sequence is a mixture of standards (STD) in the absence of matrix for the purposes of ensuring instrument integrity. Anecdotal evidence suggests the first few injections on an LC-MS system do not provide reproducible results. It is recommended that the first batch of samples not be considered, because subsequent injections of the biological matrix are required to attain reproducibility.[63]

1.2.4 Standards for Assessment of Recovery and Quantitation

Internal or external standards are used in NMR metabolic profiling experiments to provide a chemical shift reference and for purposes of absolute quantitation. Care is required in measurements where the goal is quantitation, especially as the resonances of most NMR reference standards are due to isolated nuclei with long relaxation times. In theory, any compound can be used as an NMR standard providing it is chemically stable, does not interact with sample components and produces a resonance that does not overlap with those of the sample. Because of the complexity of spectra obtained for biological samples, the later requirement practically limits the choice of reference compounds. For ^{1}H NMR experiments sodium trimethlysilyl propionate-d_4 (TSP-d_4) or sodium 2,2-dimethyl-2-silapentane-5-sulfonate-d_6 (DSS-d_6) are common standards since they produce a single well-resolved resonance at a frequency well-removed from the signals produced by most organic compounds.[65-68] Tryptophan, phenylalanine and fumarate

were spiked into flower nectar samples for absolute quantitation via 1D NMR using 10 mg + 0.01 μg of internal standard and the method of standard addition.[69, 70]

When TSP is to be used as a concentration reference, care needs to be taken to ensure that atmospheric moisture is not absorbed by the TSP when its mass is initially recorded. One strategy to avoid such accuracy errors is to use commercially available TSP in D_2O stored in small volume ampules that can be utilized immediately after opening.[71] TSP has also been shown to adsorb to glass surfaces over time, resulting in significant losses in its effective solution concentration.[72]

Adequate coverage of chemical space using a mixture of L-tryptophan, L-phenylalanine, p-coumaric acid, caffeic acid, sinapic acid, benzoic acid, quercetin, kaempferol, rutin, trans-resveratrol, naringenin, chlorogenic acid, trans-cinnamic acid, isorhamnetin, ferulic acid and tomatine was useful to monitor the quality of the chromatogram and reproducibility of retention times throughout the runs and to aid in metabolite identification in strawberry and tomato.[73] Absolute quantitation by LC-MS is typically achieved using stable isotope-labeled internal standards.[74]

1.2.5 Intact Tissue Measurements versus Tissue Extracts

Metabolic profiling studies of specific tissues are often performed using tissue extracts rather than whole tissues because extracts provide much of the same information as whole tissues, but yield much better resolved NMR spectra than can be obtained by high-resolution magic angle spinning (HR-MAS). By spinning the sample at the magic angle (54.7^o), spectral broadening due to dipolar interactions and susceptibility

differences within the sample are minimized. So-called magnetic resonance spectroscopy (MRS) has been useful in plant[75] and mammalian[76] biology for intact tissue studies, although it was not employed in this work.

There are many methods available for extracting metabolites from tissues and preparing them for NMR analysis, but perchloric acid extractions are among the most widely utilized.[77, 78] Although most water soluble metabolites are recovered by this extraction method, hydrophobic metabolites such as lipids are lost. In cases where these hydrophobic metabolites are of interest, an organic solvent such as chloroform can be used to extract the tissue pellet remaining after perchloric acid extraction. Combining the data from these two extraction methods has been shown to represent a more complete picture of an organism's metabolic profile.[77] One important disadvantage of perchloric acid extraction is that this strong oxidizing acid can react with some metabolites, leading to altered and potentially misleading metabolic profiles. As an alternative, tissues may be extracted using an aqueous buffer or a mixed solvent such as water/acetonitrile. As it is very difficult to establish quantitative recoveries, most metabolic profiling studies that use tissue extracts rely on relative quantitation in comparing sample groups. An optimized sample preparation protocol for NMR-based metabolomics using *Arabidopsis thaliana*[79] is presented in more detail in Chapter 2.

Intact-tissue MS (Svatoš et al.)[80] is an emerging alternative to solvent extraction followed by LC-MS. Arabidopsis leaves, petals and sepals have been examined using this method. In the liquid-extraction for surface analysis (LESA) method, solvent containing a MALDI matrix and internal standard is spotted onto tissue, which is struck

with a laser to generate ions. Ions are detected by time of flight (TOF) MS. Ions are injected into the mass analyzer with similar kinetic energies and separated in a drift tube with lighter singly charged ions (low mass-to-charge ratio, m/z) arriving early and heavier singly charged ions arriving later (Figure 1.4), encountering a series of transducers prior to image collection by a charged-couple device (CCD).

Images in time are analogous to extracted ion chromatograms (EIC) in LC-MS analysis. False colors can be applied for visualization. Mass resolution on the order of 200-300 ppm and mass accuracy within 0.5 Da has been achieved with this approach. A synthetic organic fluorescent dye (Cy-3) is used to confirm quality of spotting, in tissues that are photosynthetic the autofluorescence of chlorophyll is captured to confirm tissue is not damaged in sample preparation. The laser can be rastered across the sample to achieve surface coverage and can operate either in focused (high-resolution) or unfocused mode (total imaging).

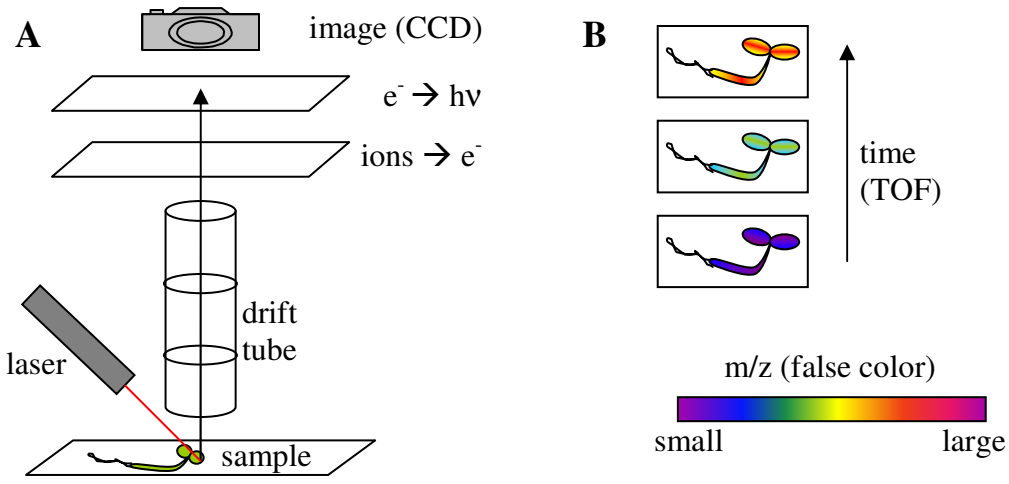


Figure 1.4 Mass spectrometry imaging schemes used in plant metabolic profiling. (A) MALDI is coupled to TOF, images are collected by a CCD. (B) Each time point recorded is an extracted ion image. False colors are applied for visualization. Composite images of biological interest are presented.

1.2.6 Protein Precipitation and Quenching of Metabolic Activity

A common method for removing proteins from tissue extracts is via precipitation by an organic solvent such as acetonitrile or methanol. In metabolic profiling experiments, it is important that the protein precipitation process does not cause excessive loss of small molecule metabolites. More studies have been conducted in mammalian systems, with some work in worms and fungi. Comprehensive studies in plants are still needed and experiments to evaluate extract stability are recommended for any proposed metabolomic study. Using mammalian serum, loss of metabolites was minimized by optimizing the organic solvent to plasma ratio as well as rate of solvent addition, with the best results obtained when organic solvent is added slowly over several hours to allow the metabolite-plasma protein binding equilibria time to adjust.[81]

Heat treatment (50°C or 90°C for between 2-30 minutes, once or repeatedly) has been proposed[82, 83] as an alternative to precipitation by organic solvent (acetonitrile or methanol) and/or low pH in extracts and biofluids. When comparing a variety of techniques for sample preparation for metabolite profiling of human serum, a Bradford assay revealed that acetonitrile-treated serum samples contained 6% residual protein while methanol and ethanol reduced residual protein to 2 and 4%, respectively. Heat treatment alone was insufficient for complete precipitation of proteins, inadequately quenching metabolic activity, but when combined with methanol, the metabolic profile was reproducible and contained the most "features" of all sample preparation techniques compared to analysis with nanoLC-ESI-MS in positive mode.

1.3 Data Acquisition

Once the sample has been prepared for analysis, consideration must be paid as to how to interrogate it. One significant benefit of NMR is that it is a non-destructive technique, making replicate analyses possible to obtain structural as well as quantitative information. LC-MS-based metabolic profiling, although destructive, yields quantitative analyses and provides structural information about analytes by in-source fragmentation or collision-induced dissociation tandem mass spectrometry (MS/MS). The following section will treat NMR and LC-MS separately to discuss considerations that apply to plant metabolite profiling and comparative metabolomics.

After experimental design parameters have been determined and sample preparation has been optimized to address the biological question under investigation, analytical data collection can begin. As with many biological experiments, the amount of tissue available for analysis is limited either by the nature of the species under investigation (clinical samples are limited, research animals/plants tend to be of small size) and the difficulty in time and space associated with production of biological material (access to facilities for growth/cultivation are limited as well as the technical skill involved in preparing tissue practically limits the amount of tissue that can be raised by one researcher/technician). Therefore optimizing sensitivity on the analytical side becomes of critical importance to record datasets that can withstand statistical evaluation of metabolite abundance and can inform the investigator of changes in metabolite abundance related to the applied biological treatments (i.e. stress).

1.3.1 NMR-specific Considerations

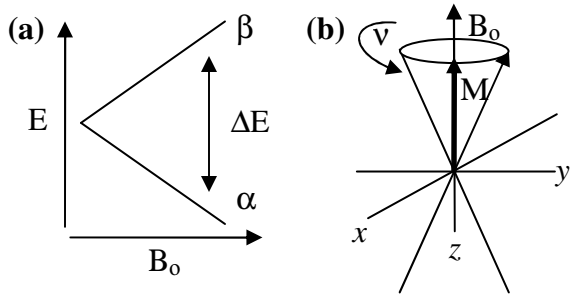
Elegant primers on NMR are available elsewhere,[84-87] for the purposes of this chapter several terms will be defined as pertaining to this work. In the NMR experiment a population of spins is split into quantized energy levels by an external magnetic field (B_o). Spin is a property of atomic nuclei introduced to explain spectroscopic observations in the presence of a magnetic field, generally limited to nuclei with odd numbers of nuclear particles. Many biologically relevant nuclei possess spin, e.g., ^{1}H, ^{13}C, ^{15}N, ^{31}P.

The "spin" or spinning charge in an external magnetic field possesses angular momentum like a spinning bicycle tire. When perturbed, it will spiral around equilibrium rather than swinging through it like a pendulum (oscillator). This process is called precession, and the precessional frequency of a nucleus depends first on its identity (magnetogyric ratio, γ) and its local electronic and magnetic environment. The dispersion of resonant frequencies among nuclei in a sample can be increased by applying a stronger external magnetic field (B_o). The basics of NMR are shown in Figure 1.5, which illustrates an energy level diagram, a representation of a nuclear spin, and the fundamental equations which relate the aforementioned parameters (γ,B_o) to practical considerations of relevance to metabolomics, such as sensitivity.

Figure 1.5a summarizes the Stern-Gerlach experiment, which demonstrated that nuclear spin (½) is quantized (Equation 1.1) into low (α) or high (β) energy states. At equilibrium, the lower energy state (α, with B_o) is more populated. This can be called a "spin excess," which is small at room temperature, hence limiting the sensitivity of the NMR experiment. Even if there is only a small spin excess, with a large number of

nuclei in a sample the precession will become visible. Ensembles of nuclei (Figure 1.5b) can be represented by a bulk magnetization vector (M) which points along the positive z-axis, parallel to B_o. The precessional frequency (ν) of a nucleus also increases with increasing magnetic field strength, responsible for dispersion of resonances at high field (Equation 1.2). The population difference between energy states increases with increasing magnetic field strength and decreases with increasing temperature (Equation 1.3).

The energy gap between states also increases with B_o for a spin ½ nucleus such as hydrogen, allowed precessional directions are either forward (also called +1/2, lower energy, α state), which is more populated than the alternative, reverse (also called -1/2, higher energy, β state) at thermal equilibrium. The populations of these states behave according to Boltzmann's distribution, the small difference in energy between the two spin states that gives rise to nearly equal populations is what limits sensitivity.



$$\Delta E = h\gamma B_o/2\pi \qquad (1.1)$$

$$\nu = \gamma B_o/2\pi \qquad (1.2)$$

$$N_\beta/N_\alpha = \exp(-\Delta E/kT) \qquad (1.3)$$

Figure 1.5 NMR basics (a) energy level diagram for a spin ½ nucleus in magnetic field (B_o) shows quantization (Equation 1.1) of low (α) or high (β) energy states. Ensembles of nuclei (b) can be represented by a bulk magnetization vector (M) which points along the positive z-axis, parallel to B_o. The precessional frequency (ν) of a nucleus also increases with increasing magnetic field strength, responsible for dispersion of resonances at high field (Equation 1.2). At equilibrium, the lower energy state (α, with B_o) is more populated. The population difference between states increases with increasing magnetic field strength and decreases with increasing temperature (Equation 1.3).

Radio Frequency (RF) irradiation (B_1) is applied as a pulse delivered to a coil surrounding the sample containing NMR-active atomic nuclei. When the frequency of incident radiation matches the precessional frequency of the nuclei under investigation, called the Larmor frequency, absorbance of energy (resonance) occurs. Pulses used in this work cause equalization ($\pi/2$, 90^o) or inversion (π, 180^o) of populations of spins. Applying a discontinuous pulse (gradient, G) across the sample can encode the spatial position of a nucleus in an ensemble of spins. Shaped pulses (p) can be used to excite only a specific range of precessional frequencies. "Spin gymnastics" in the form of a series of well-timed and power-optimized pulses can be applied to nuclei such that specific transitions between the energy levels are achieved by pulse sequences. It is important to optimize the parameters used by pulse sequences to maximize the experimental signal-to-noise (S/N) in the experiment.

After the RF pulse, as the population of nuclei reestablishes equilibrium, an alternating current is induced in the coil around the sample due to the coherence generated in the precessing spins. This free induction decay (FID) contains information regarding the electronic/magnetic environment of the nuclei in the sample, which is digitized and recorded in the time-domain as it decays. The time-domain data is converted to the frequency-domain spectrum via Fourier transformation. Nuclei which experience a relatively large amount of electron shielding are more isolated from the effect of the B_o field, and their resonances are found at lower frequencies (upfield) in the NMR spectrum. Deshielded nuclei, which can also be called electron deficient,

experience a downfield shift to higher frequencies in the spectrum. In this way, chemical shift provides indirect structural information.

The resonant frequencies exhibited by a sample can also be informative regarding the position of nuclear spins in molecules and the nature of their connectivity. Just as two stir bars engage each other if positioned proximally, neighboring spins influence the number of fine energy levels (multiplicity) a given spin will exhibit. For two neighboring spin ½ nuclei which are electronically different and connected by a chemical bond, said to be first-order coupled, the resonances of each nucleus will be split into two peaks, centered around the chemical shift representative of the electronic environment each respective nucleus. The spacing between the peaks is the coupling constant, J. If more than two dissimilar spin ½ nuclei are coupled, the multiplicity will be equal to the number of neighboring spins plus one (n + 1) under the conditions of first-order coupling. Two neighboring nuclei which are in similar electronic environments such that the difference in their chemical shifts is on the same order as their coupling constant will exhibit strong coupling, resulting in distorted intensities often containing more peaks than expected by the n + 1 rule.

Because the precession frequencies of nuclei are in MHz (10^6 per second), and in solution small molecules undergo rotation about single bonds about 10^{10} per second, a signal intensity proportional to the number of chemically equivalent nuclei is observed for protons bound to a freely rotating carbon. In cases where rotation is restricted, protons bound to the same carbon may be magnetically inequivalent.

The NMR experiment, as with any absorption spectroscopy, can be quantitative where the system is calibrated, either using a calibration curve or by comparison of integrals for the unknown sample against a standard of known concentration. The intensity of the spectral features, recorded when nuclei have had enough time to fully reestablish equilibrium populations between transient scans, can be useful to assign nuclei to molecules in simple mixtures based on ratios of the integrated areas of resonances.

Taken together, the information from an NMR experiment (chemical shift, multiplicity, integral ratios) can inform us about molecular structure. Nuclei from two different molecules in a complex mixture may generate NMR signals with the same chemical shift (accidental equivalency). Depending on the goal of the experiment, this may or may not be a problem because intensity is additive.[88] Two-dimensional (2D) NMR techniques can be used to further explore molecular connectivity and to disperse overlapping signals through a second orthogonal dimension. Spectral overlap is a complicating factor in any metabolomic experiment, and spectral deconvolution[88-90] is commonly applied.

The rate of spin-spin relaxation and magnetic field inhomogeneity contribute to spectral linewidths, therefore time spent shimming the magnet pays dividends in resolution.[91, 92] For measurements in high field magnets, line widths below 2.5 Hz at half-height for the chemical shift reference TMSP-d_4 spiked into a plant extract indicate that magnetic field homogeneity is acceptable for acquisition of a one-dimensional (1D) proton spectrum, although 0.7 Hz is attainable.

Signal averaging in NMR is commonly used to improve the spectral S/N ratio (S/N) and for purposes of phase cycling to reduce spectral artifacts, but at the expense of sample throughput. Depending on the amount of tissue available and the number of samples in a dataset, reasonable compromises between recycle time (time per scan) and number of scans to achieve adequate S/N for quantitative interrogation must be found on a case-by-case basis. Using a recycle time of 3 seconds for illustration, Table 1.1 provides values for S/N as it relates to number of scans (NS) which can be co-added provided that the sample is stable over the total experiment time.

Table 1.1 Improving S/N (S/N) requires compromise between sensitivity and total experiment time. Experiments lasting more than two hours for ^{1}H NMR metabolomics may be considered too costly in terms of instrument time, where a survey spectrum is more important for fingerprinting and lead generation.

Relative S/N (S/N)	Number of scans (NS)	Total experiment time
1	1	0.05 min
2	4	0.20 min
4	16	0.80 min
8	64	3.20 min
16	256	12.8 min
32	1024	51.8 min
64	4096	6.82 hr
128	16384	13.6 hr
256	65536	2.28 d

1.3.1.1 1D ^{1}H NMR as a Survey or Fingerprinting Technique

For a first-pass inspection of the dataset or when a change in relative levels is assessed as a function of a perturbation, it is sufficient to acquire spectra under steady state conditions.[93] The time between pulses can be less than 5 times the spin-lattice

relaxation time, T_1, (on the order of 0.5-3 seconds for most metabolite resonances).[94] Although resonance intensity will be reduced if recycle times < 5 T_1 are used, signal averaging can more than compensate for the loss in intensity if the total experiment time is kept constant.

A small portion of the ^{1}H NMR spectrum resulting from Arabidopsis root extract is presented in Figure 1.6, integrated over non-targeted 0.02 ppm intervals. The resulting integral intensities and corresponding chemical shifts are inputs for chemometric modeling. Resonances from the same multiplet are often distributed between multiple "bins," contributing to the redundancy and a large number of correlated variables in the resulting data matrix.

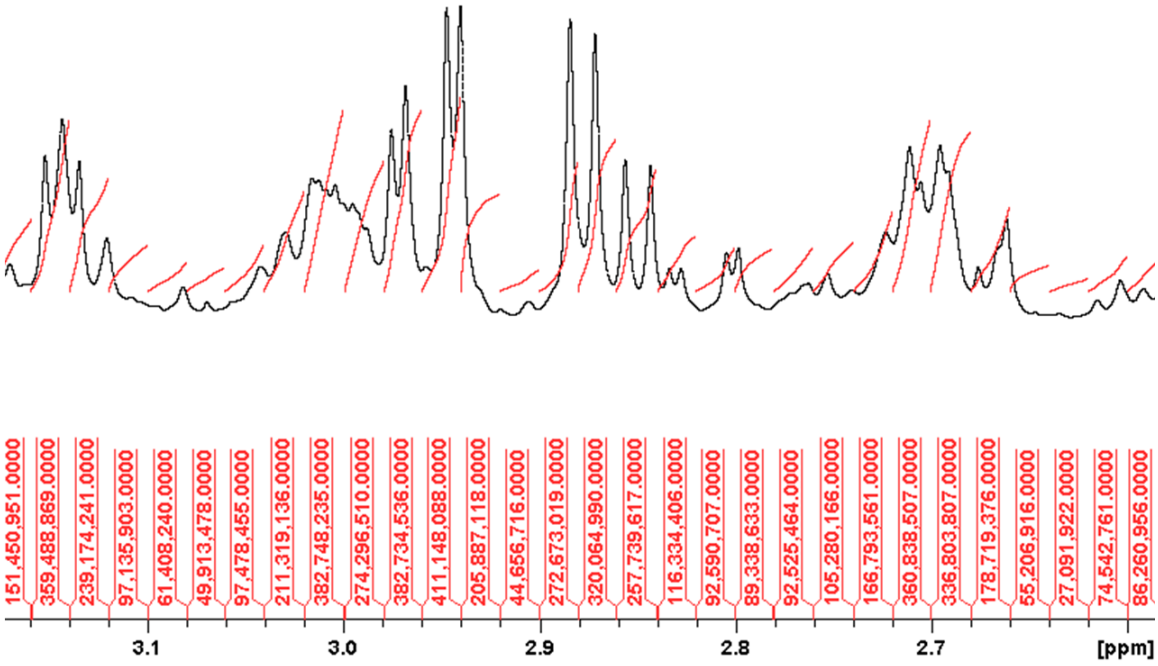


Figure 1.6 NMR spectrum of 7-d-old Arabidopsis seedling root extract divided into "bins" of width 0.02 ppm generally covers the width occupied by a single resonance or spectral feature.

Multivariate statistics can be applied to a dataset of "binned" spectra representing different biological treatments or sample handling conditions to identify spectral regions of high variance. Such variance may be contributed by small deviations in chemical shift due to differences in pH between samples, and can be reduced by buffering or titrating the pH of each sample to a defined value prior to analysis. More useful is when a metabolite, responding differently to the treatment or handling condition under investigation, is identified by comparing fingerprints to each other. By using multivariate statistics combined with a spectral fingerprint provided by ^{1}H NMR spectroscopy, molecular signatures can be recognized with corresponding magnitude and direction of variance with respect to treatment.

1.3.1.2 One-Dimensional ^{1}H NMR for Quantitation

Spectral overlap complicates integration but regions can usually be found for tens of metabolites where overlap is not a significant problem. Requirements for quantitative NMR measurements will be more thoroughly explained in Section 1.3.1.4. Integrated areas of metabolites in stressed or treated tissue extracts are compared with respective controls, to provide relative quantitation.

Another advantage in NMR-based metabolite analysis is that all but the smallest metabolites (succinate, pyruvate, glycine, etc.) give rise to more than one resonance in the spectrum. These can be quantitatively evaluated separately for internal cross-validation of resonance assignments as well as calculated fold changes. For example, γ-aminobutyric acid (GABA) has three distinct methylene resonances at 3.01, 1.89, and

2.28 ppm for the α, β, and γ protons, respectively. To compliment internal cross-validation, it is possible to use homonuclear or heteronuclear 2D NMR experiments to establish connectivities between ^{1}H and/or ^{13}C nuclei in a metabolite, depending on the type of experiment selected. A few are described in the following section.

1.3.1.3 Two-Dimensional NMR

Considerations for 2D experiments are similar to those of 1D proton NMR. Depending on the nature of the experiment, specific consideration should be given to calibration of pulses. 2D NMR increases the effective spectral resolution by spreading information over two dimensions. Because of the greater dispersion of ^{13}C spectra, heteronuclear ^{1}H-^{13}C spectra can provide much greater resolution than a 1D ^{1}H spectrum.[85,95]

An experiment which can be useful for metabolomics specifically is the Homonuclear 2D J-resolved (2D J-Res) spectrum, wherein the chemical shift in the direct dimension is decoupled from its multiplicity in the indirect dimension.[19, 20, 96, 97] By taking a projection at an angle (45^{o}) one can obtain a decoupled proton spectrum. This reduces spectral overlap,[98] and 1D projections of 2D J-Res experiments can yield quantitative information about metabolites in the sample.

Proton-proton correlation experiments such as COrrelation SpectroscopY (COSY) and TOtal Correlation SpectroscopY (TOCSY) are generally useful to assign or confirm molecular assignments to resonances observed in 1D proton spectra. In particular the double-quantum filtered (DQF) - COSY is recommended in cases where severe crowding

obscures crosspeaks near the diagonal. An additional advantage of this pulse sequence is that it can be carried out on samples containing large amounts of water (90% H_2O, 10% D_2O). Presaturation together with the DQF ensure the suppression of the solvent resonance. A DQF-COSY spectrum of *Arabidopsis* seedling extract is presented in Figure 1.7.

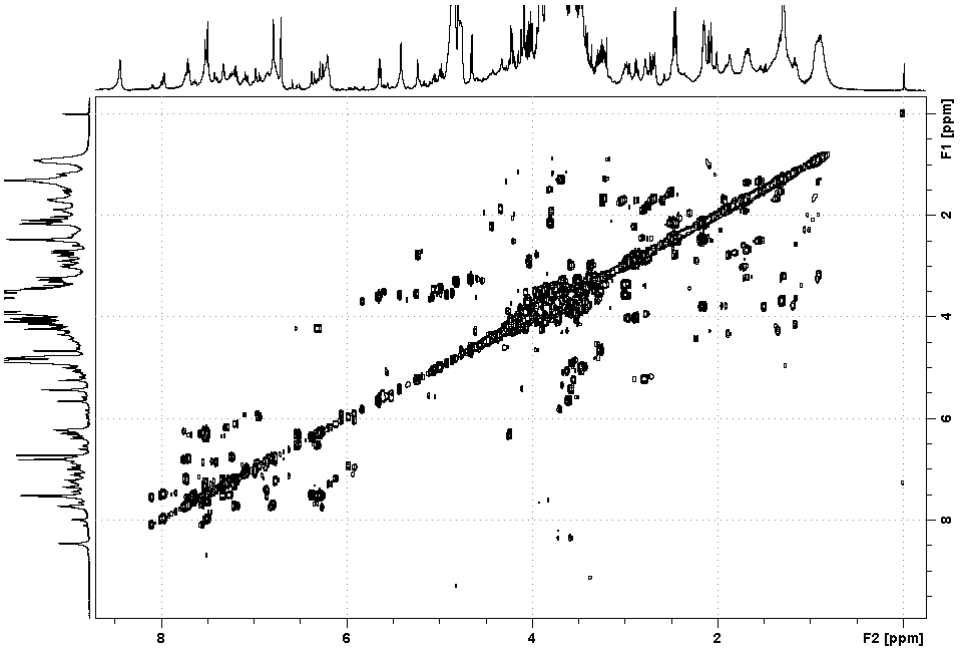


Figure 1.7 DQF-COSY (Bruker cosydfphpr) spectrum resulting from SPE-enrichment of ~1.8 g DW *Arabidopsis thaliana* (Col-0) seedling (7 d old) tissue. Delay (d1) of 1.5 seconds was used between scans; the water resonance was attenuated by presaturation using a pl9 value of 50 dB.

The phase-sensitive TOCSY (Bruker mlevphpr) spectrum is an efficient way to establish correlations within spin systems such as aliphatic side chains and sugar rings. During the TOCSY spin-lock (MLEV-17 in the experiments reported in this dissertation) the spins "see" only B_1 as the effective field; therefore their chemical shift differences become negligible and the spin systems are all higher order, leading to cross peaks between all protons within a spin system. A short spin lock (30 ms) approximates a

COSY spectrum, while longer times result in the total correlation. In Figure 1.8, a mixing time of 80 ms was used.

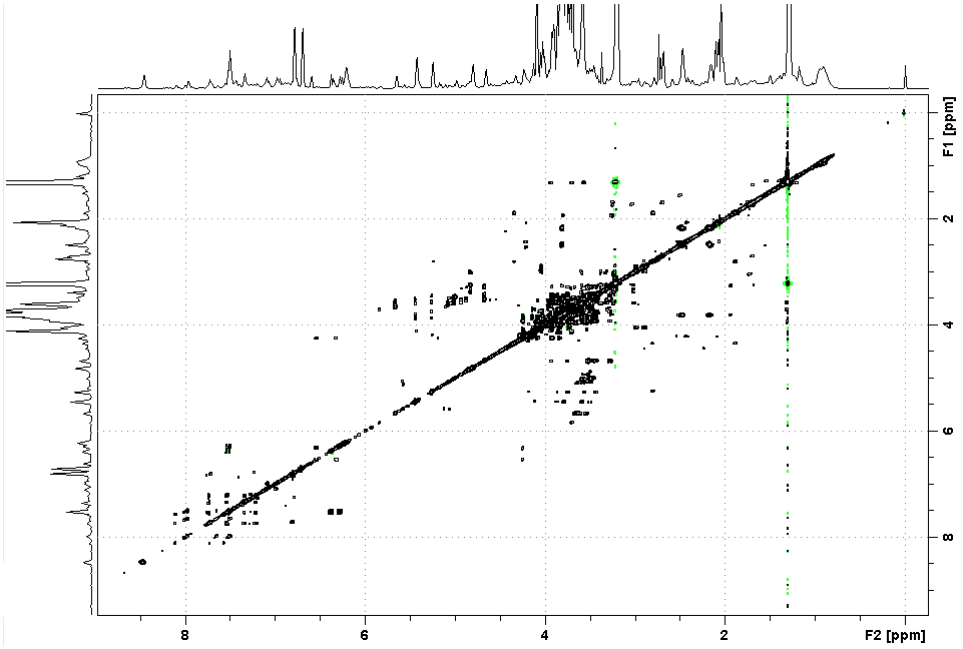


Figure 1.8 TOCSY (Bruker mlevphpr) spectrum resulting from SPE-enrichment of ~1.8 g DW *Arabidopsis thaliana* (Col-0) seedling (7 d old) tissue. A delay (d1) of 2 seconds was used between scans. The water resonance was attenuated by presaturation using a value of pl9 of 50 dB. The mixing time (d9) was 80 ms.

Diffusion-Ordered SpectroscopY (DOSY) is a pseudo 2D technique wherein nuclei are phase encoded by a gradient pulse according to their spatial position within the sample. After an incremented delay, the phase of the magnetization is decoded by application of a second gradient pulse. The magnetization is fully recovered if the nuclear position remains constant with respect to the direction of the applied gradient. Signal intensity will vary depending on the degree of diffusion that has taken place during the delay.[99] The pseudo-2D DOSY spectrum plots chemical shift on the direct

dimension and diffusion coefficient in the indirect dimension. DOSY has been applied to the analysis of plant extracts.[100, 101] It is generally possible to use DOSY to assign resonances to molecules, based on the idea that all protons connected to the same molecule should have identical diffusion coefficients, however in our experience it is not useful for extremely crowded spectra, especially when the small molecules of interest have similar diffusion coefficients (Figure 1.9).

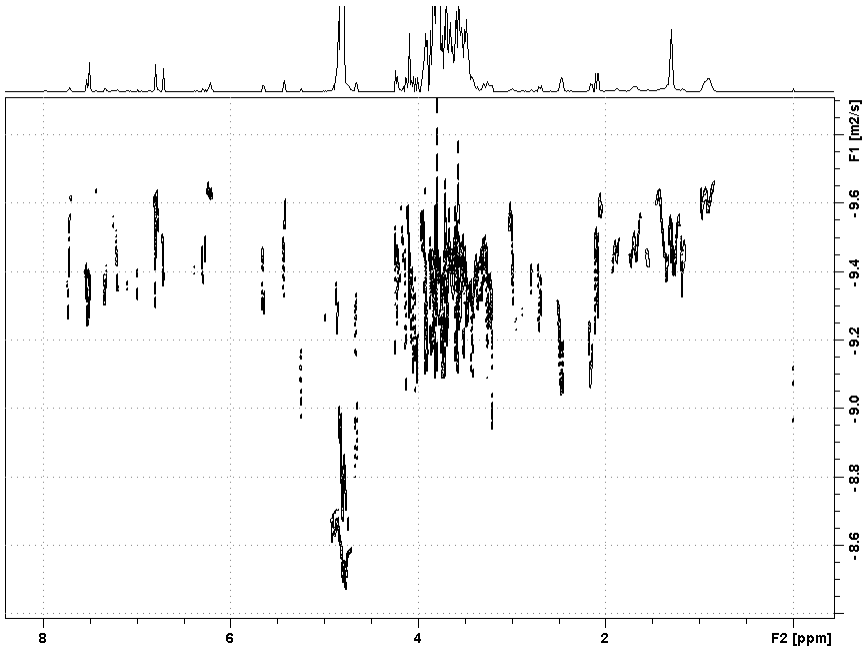


Figure 1.9 DOSY spectrum for SPE-enriched ~1.8 g DW *Arabidopsis thaliana* (Col-0) seedling (7 d old) tissue measured with the Stimulated Echo experiment (Bruker stebpgp1s) using bipolar gradients.

This chapter summarizes considerations for "Methods" in metabolomics and metabolic profiling, which will be applied in Chapters 3-6. Thus far we have seen how NMR can be applied for intact mixture analysis. Taken together, the information

supplied by 1D and 2D NMR experiments cooperatively establish metabolite quantity and identity.

1.3.1.4 Requirements for Quantitative NMR Measurements

Careful selection of NMR data acquisition parameters is required to produce the best possible data for the system being studied and it is essential that all datasets are acquired with the same parameter set if any direct quantitative comparison is to be made.[65, 102, 103] Important parameters to consider in optimizing experiments are the excitation pulse width, receiver gain, acquisition and relaxation times and solvent suppression parameters, which are addressed in greater detail in the following section. Because sample ionic strength can vary significantly among biological samples, it is advisable to tune the probe for each new sample. Similarly, it is common practice to optimize field homogeneity by shimming[91] on each sample to achieve a minimum standard established by the laboratory for quality, often specified by the line width at half height of a prominent singlet resonance, typically that of the internal standard. Equally important is the linewidth at the height of the ^{13}C satellite peaks, indicative of baseline resolution.

If several magnets are available, the magnet with the highest field strength should be employed for metabolic profiling analyses. Bertram et al.[104] recorded the metabolic profile of urine on four NMR spectrometers (250MHz, 400MHz, 500MHz, 800MHz) and reported the expected increase in spectral resolution and S/N with field strength. The greater dispersion of high field magnets can facilitate resolution of resonances that would

be overlapped at lower field strengths. In addition, because NMR sensitivity increases as $B_o^{3/2}$, higher field instruments can reduce the experimental time or increase spectral S/N.

Relative quantitation is often sufficient in metabolic profiling studies, though there are cases in which the absolute concentration of a metabolite is desired. Quantitative analysis requires spectra measured with adequate digital resolution, a repetition time sufficient for relaxation of the signals of interest, and with a high S/N ratio. For example, to be 99% certain that the measured integral falls within $\pm$ 1% of the true value, a S/N ratio of 250 is required.[65] Proper post-acquisition processing of the NMR spectrum can also affect the reliability of quantitative data. The spectrum must be appropriately apodized, properly phased and baseline corrected, and integration limits carefully selected.[105] However, the major challenge in quantitative metabolic profiling is ensuring that the resonances of the metabolite and the quantitation standard are fully relaxed. This can be accomplished by setting the recycling time to at least five times the longest T_1 relaxation time of the resonances of interest. This is not always a practical approach since metabolites are typically present at low concentration and lengthy data acquisitions may be required to obtain adequate S/N ratios. In such cases, a shorter repetition time can be used and a T_1 correction factor applied to the resulting integrals.[65]

1.3.1.5 Water Suppression

Most biologically-derived samples require some method of suppressing the water resonance to obtain high quality ^{1}H NMR spectra. NMR spectrometers have analog-to-digital converters with a limited dynamic range, making it difficult to digitize and detect

metabolite resonances in the presence of an intense solvent signal. Incomplete water suppression can also produce baseline roll and phase distortions that hinder quantitative measurements. The simplest methods of solvent suppression are those that rely on selective saturation of the solvent resonance during the recycle delay. More sophisticated techniques include WET (water suppression through enhanced T_1 effects) which utilizes selective excitation of the solvent followed by dephasing gradients and WATERGATE which uses gradient tailored excitation.[60, 65, 67]

Arabinar et al.[66] compared five water suppression techniques for robustness, repeatability, sensitivity, and practicality in the analysis of urine samples. The pulse sequences evaluated can be separated into two groups, those that leave the intrinsic water signal intact: WET and ES (excitation sculpting using a double-pulsed field gradient echo), and those that incorporate selective saturation of the solvent resonance: NOEPR (NOESY pulse train with presaturation during the recycle delay and mixing time), ESCW (ES with continuous wave on-resonance saturation), and ESWGL (ES with adiabatic frequency modulation). WET was the least effective at suppressing the water resonance and produced a baseline distortion that covered a fairly broad chemical shift range. ES and NOEPR were more effective than WET at suppressing the solvent resonance but produced baseline distortions in the spectral region around the water chemical shift (4.5 – 5.3 ppm). The best results were achieved when excitation sculpting was combined with selective saturation using either continuous wave (CW) or adiabatic frequency modulation. The spectra produced by these methods showed excellent water suppression,

flat baselines, and high reproducibility and were amenable to automated spectral processing.

1.3.2 Liquid Chromatography Experiments

The field of plant metabolic profiling continues to develop, aided by innovative approaches to sample preparation and analysis as well as developments in instrumentation. Figure 1.10 presents a timeline of records available through Web of Knowledge when a topic (TS) search string encompassing metabolism, metabolite, metabolomics, etc. (metabol*) is used together with "plant." Publication records were restricted by the predominant analytical techniques: GC, LC and NMR. Milestones in instrumentation or database/methods infrastructure are labeled.

Growth throughout the post-genome era is clearly observed. From this graph it appears over the last decade a majority of publications utilized LC (grey bars). This is likely due to the diversity of solid phase supports available combined with the ubiquity of the instrumentation and ease of quantitation. Alternative chromatographic approaches utilize GC (black bars), however it is more cumbersome in sample preparation and may suffer from bias because GC relies on a derivatization step prior to analysis, whereas for LC (coupled to ESI-MS) derivatization is not necessary. The reliability of GC retention indices and establishment of databases

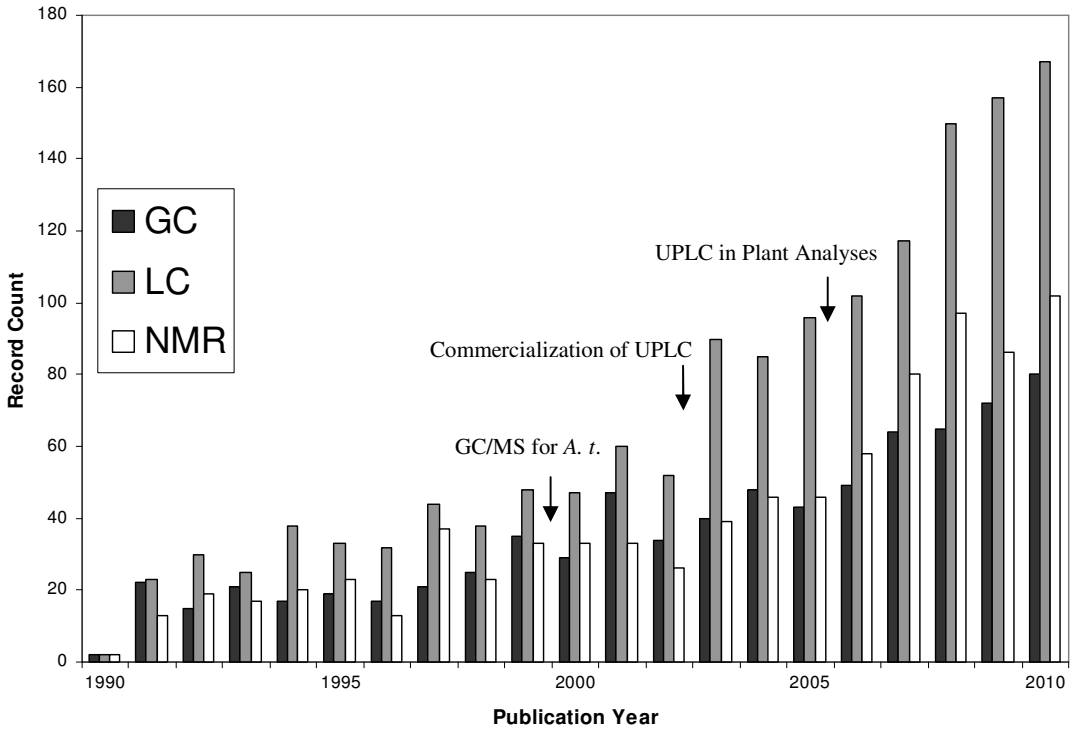


Figure 1.10 Popularity of analytical platforms for plant metabolomics. Record count determined using Web of Knowledge on July 11, 2011 searching by topic (TS = X AND metabol* AND plant) where X was gas chromatography, liquid chromatography or NMR.

for plant-specific analyses boosted popularity of GC-MS particularly for those in the Arabidopsis community [106, 107] facilitating rapid dereplication. Dereplication is the process of determining which observed signals of known molecules, potentially novel compounds, including those that have not yet been described in the literature or have not yet been reported for the organism under study. Retention indices and databases cataloging observed mass-to-charge ratios of plant extracts shorten the length of time it takes to assign known metabolites to observed signals in raw data. The effect of metabonomics as a terminology may result in NMR (white bars) being underrepresented in this figure. Notably, ultra high pressure liquid chromatography (commercialization of

UPLC)[108] may further popularize LC. Already, methods for non-targeted and targeted analyses by UPLC-MS have been biologically informative.[64, 73, 109-111]

1.3.2.1 Chromatography of Plant Extracts

Optimal LC conditions for efficient separation of structurally similar compounds is often not possible using a single type of chromatographic column. Peak widths between 0.5 and 1.5 minutes have been reported using reversed phase (RP) HPLC with a column 15 cm in length (4.6 mm diameter) packed with particles of 3 μm diameter operating at a flow rate of 0.4 mL/min. Because UPLC peaks are narrow, a fast MS analyzer like TOF must be used so that the peaks can sufficiently sampled. Peak widths of 20-30 s allowed for at least 15 datapoints across each chromatographic peak (baseline-to-baseline) to ensure reliable characterization of target compounds.[53] It is not unusual to use three columns with orthogonal separation characteristics in parallel for metabolomics studies.[112] Attention should be given to the effect of mobile phase properties such as buffer concentration, organic modifier, and gradient slope on resolution of metabolites.

For HPLC isocratic separations at 5, 10, 20, 30, and 50% organic can be a starting point if little is known about the behavior of the desired compounds. Chromatographic methods typically involve a linear gradient. The use of a guard column and adequate washing and equilibration between subsequent injections is of critical importance to reduce run-to-run carryover and to preserve column lifetime. One UPLC column used for method development withstood 399 injections of 372 plant samples before it was compromised.

In metabolomic or metabolic profiling studies employing liquid chromatography, UPLC is recommended to minimize peak widths, thereby enhancing chromatographic resolving power.[63] UPLC gradient method development is faster than HPLC by a factor of 2-8 and dropped the average analysis time of one analyte by 1/3.[113] Solvent consumption is minimized (5-18 times lower), owing to shorter equilibration times between runs, proving more ecological and cost effective analyses.[113] UPLC has demonstrated superior reproducibility in retention time and peak area, increasing sensitivity by a factor of 1.5-2. A method developed based on the work of Wolfender[109] using UPLC for this work was ported to HPLC and shown in Figure 1.11 are the resulting UV chromatograms (254 nm) of Arabidopsis extracts enriched by mixed-mode hydrophilic-lipophilic balanced (HLB) SPE in the content of secondary metabolites. Note that total experiment time increases approximately three-fold when the method is migrated from UPLC to HPLC. HPLC, although more time intensive, allows for eight-fold higher loading, ideal for loop storage in our LC-NMR exploratory study (Chapter 5). UPLC was used for method development and rapid screening for a biological study (Chapter 6).

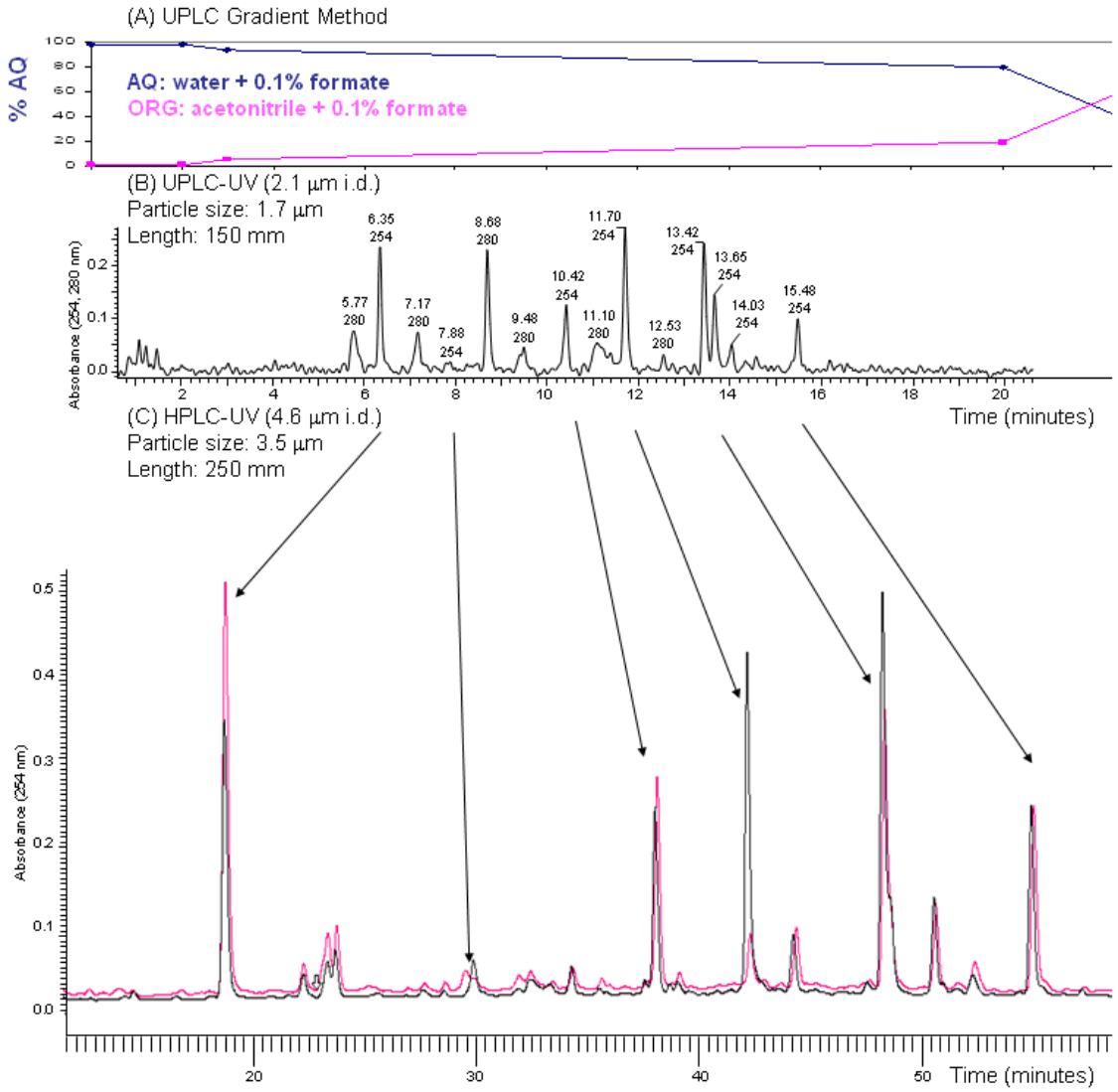


Figure 1.11 LC separation of *Arabidopsis thaliana* secondary metabolites detected by UV absorbance at 254 nm. Method migration does not cause loss of chromatographic resolution. (A) UPLC gradient method, (B) UPLC chromatogram, (C) HPLC-NMR chromatograms (black). Collection and reinjection (pink) reveals instability of plant secondary metabolites, underscoring the utility of on-line data collection.

1.3.2.2 Chromatographic Detectors

Typically LC eluent flows post-column through a non-destructive detector (UV/Vis absorbance or refractive index) before encountering a destructive detector (MS or CLND). UV and/or visible light absorbance can be monitored at one or multiple wavelengths. Ideally, for plant metabolomics a diode-array detector (DAD) records full spectral information, which produces a 2D chromatogram. Taking a trace along the y-axis, an absorption spectrum at any given point in time is extracted. Taking a trace along the x-axis yields an extracted wavelength chromatogram, which shows the absorbance at a given wavelength as the various components of the original mixture are eluted. Figure 1.12 shows an example for molecules secreted by *Xiphidium caeruleum,* a plant related to the ornamental kangaroo paw, separated on a C18 column following enrichment on a C18 SPE cartridge. When a DAD is employed prior to MS/MS or NMR characterization, it gives the researcher a "handle" with which to enter the secondary metabolome in a targeted way, including diverse molecules with a conserved structural motif. From this information we can approach structure elucidation with the hypothesis that the molecules contain one of the following groups: caffeic, ferulic, sinapic, p-coumaric, and isoferulic.[114] In secondary metabolism, because it covers such a large part of chemical space, it is useful to restrict the analysis in one dimension. While UV-Vis absorption is one approach for classification of metabolites, single-ion monitoring can also provide this information, and a more careful examination of the mass spectrometric considerations and recommendations for plant analyses are provided in the following section.

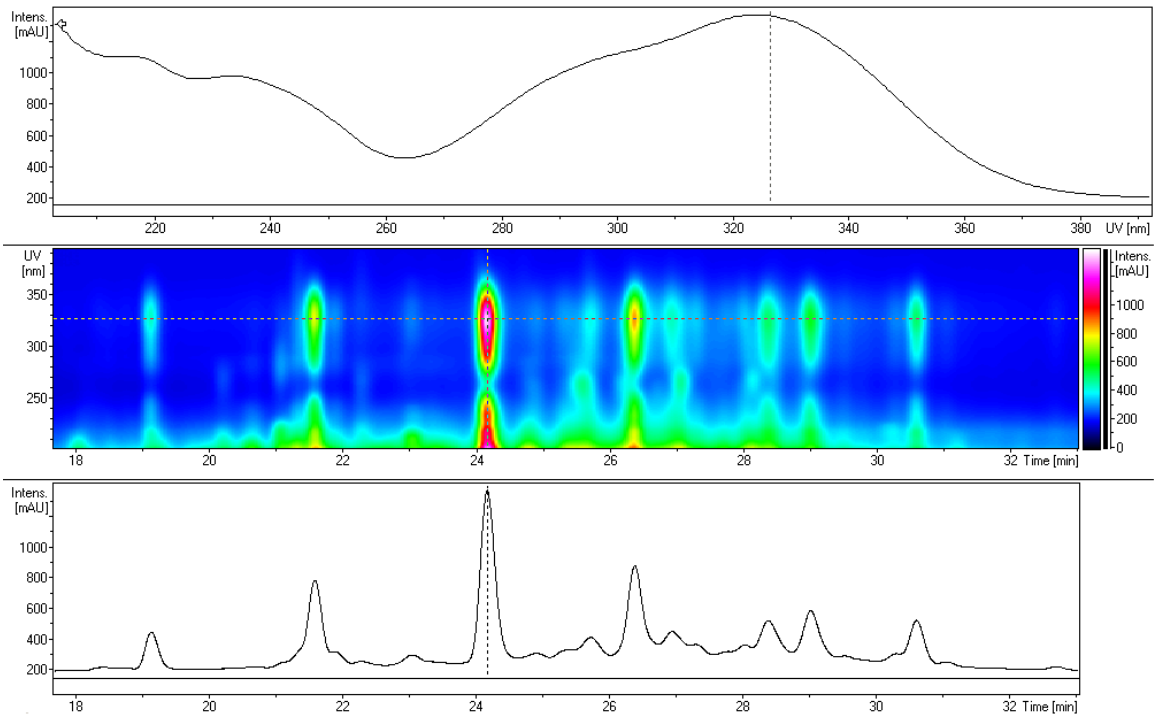


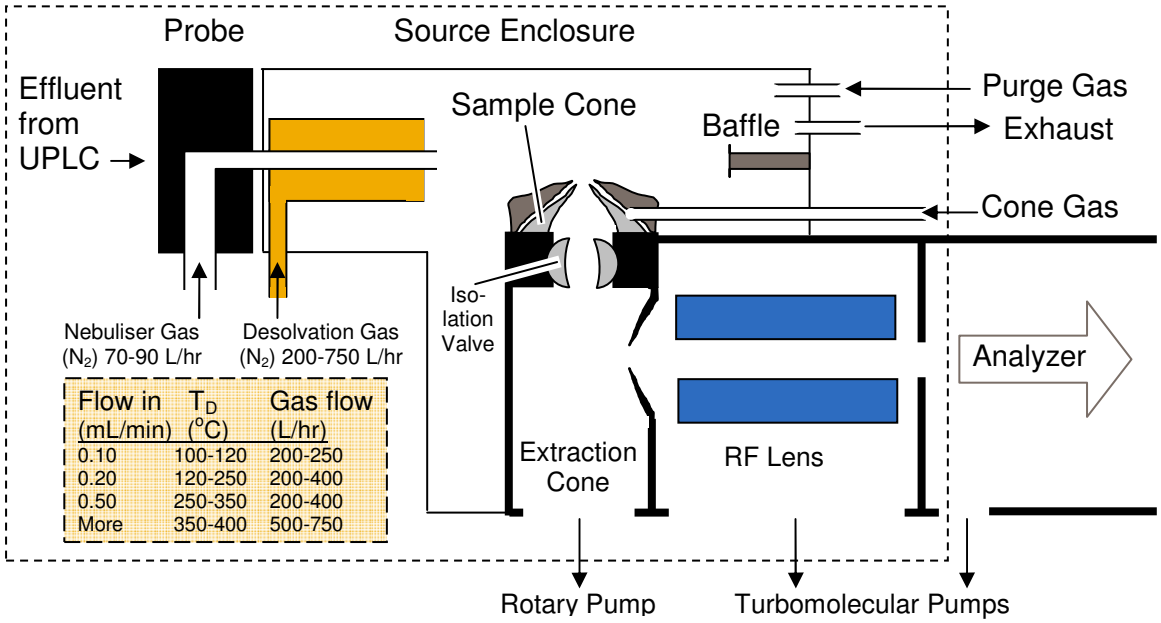
Figure 1.12 HPLC-DAD of C18-SPE-enriched liquid culture medium of *Xiphidium caeruleum* root. UV absorbance between 323-326 nm is considered a hallmark of hydroxycinnamic acids, a class of diverse secondary metabolites commonly attributed with antifungal activity. Detection by DAD facilitates class discrimination in complex mixtures.

Coupling LC to MS requires desolvation and ionization. Ionization efficiency was compared between electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), atmospheric pressure photoionization (APPI), and ESI in negative ion mode was found to provide the highest sensitivity for flavonoids, one type of secondary metabolite synthesized *in planta* that will be the focus of later chapters (5, 6). The material covered here will involve ESI only.

1.3.2.3 Electrospray Ionization Mass Spectrometry

Electrospray mass spectrometers have as many as 14 tuning parameters. If adjusted through a range of 10 values, optimum tune settings must be found out of 10^{14} possible combinations.[37] Since it is practically impossible to perform such an exhaustive search to produce the ideal set of parameters, general recommendations for plant metabolites are presented in this section.

In electrospray ionization (Figure 1.13), mobile phase from the LC column enters through the probe and is pneumatically converted to an electrostatically charged aerosol spray. The solvent is evaporated from the spray by means of the desolvation heater. The resulting ions are drawn through the sample cone aperture into the ion block, from which they are extracted into the mass analyzer. Generally, compounds of less than



Flow in (mL/min)	T_D (°C)	Gas flow (L/hr)
0.10	100-120	200-250
0.20	120-250	200-400
0.50	250-350	200-400
More	350-400	500-750

Figure 1.13 Schematic of UPLC-ESI-MS interface, called the source, site of rapid soft ionization of plant metabolites. Sensitivity and quality of metabolite information can be tuned by adjusting temperatures, voltages, and flow rates as labeled. Embedded table adapted from Waters, Inc.

1000 Da produce singly charged molecular ions ($[M-H]^-$). The source (enclosed in Figures 1.13 and 1.14 by a dotted line) can be tuned to fragment ions within the ion block or to form primarily molecular ions. Positive ion formation may be enhanced by introducing 0.1 to 1% formic acid in the sample solution. Negative ion operation may be enhanced by addition of 0.1% to 1% ammonia to the sample solution. Waters recommends acid should not be used in negative ion mode, however some have found that 0.1% formic acid is preferable.[115]

The cone gas reduces the intensity of solvent cluster ions and solvent adduct ions. Flow rates are typically 100-300 L per hour. The purge gas is not necessary for most ESI applications, but it can help reduce acetonitrile adducts. The source temperature is typically 100 ^oC for 50/50 ACN/H_2O at flow rates up to 50 μL/min. Higher temperatures, up to 150 ^oC are necessary for solvents at higher flow rates and greater water content.

The capillary voltage is usually optimal at 3.0 kV. Positive ESI is usually within 2.5-4.0 kV while negative is usually best between 2.0-3.5 kV. High flow rates may optimize this voltage as low as 1 kV. A sample cone voltage between 25-70 V will produce ions for most samples. The cone voltage should be optimized between the range 15-150 V. The extractor cone voltage should be between 0-5 V. Higher voltages may induce fragmentation.

A schematic of the quadrupole time-of-flight (Q-TOF) instrument used in this research is presented in Figure 1.14. Ions generated in the source are transferred to the quadrupole analyzer MS1 via the independently pumped RF lens. After leaving the

quadrupole analyzer the ions flow into the orthogonal time-of-flight analyzer MS2. The ion beam is focused into the pusher by the acceleration, focus, steer and tube lenses. The pusher directs a section of the beam towards the reflectron, which then focuses ions back to the detector. As ions travel from pusher to detector they are separated in mass according to their flight times, since they enter the field-free region of the reflectron with similar kinetic energies, causing ions of highest mass to charge ratio (m/z) to arrive last.

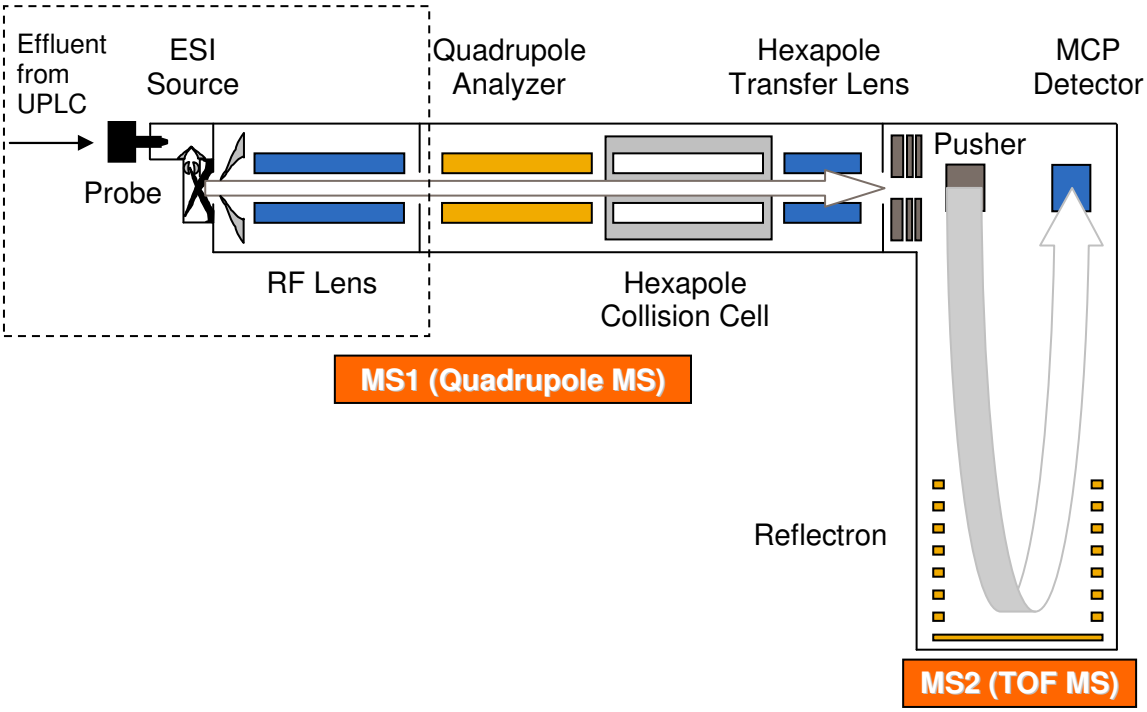


Figure 1.14 Schematic of Q-TOF micro ion optics used in this research. After exiting the UPLC, ions are formed in the ESI source and guided by the RF lens to the ion guide or selector (quadrupole analyzer) connected by a hexapole collision cell to a mass analyzer (TOF) and MCP detector.

The pusher operates up to 30 kHz, full spectra are recorded every 33 microseconds. If the user requests an acquisition rate of 1 spectrum per second, each spectrum viewed on

the host PC will be the sum of 30,000 individual spectra recorded by the MCP. The TOF performs parallel detection of all masses at very high sensitivity and acquisition rates. This is of particular advantage when coupled to fast chromatography, like UPLC. This configuration can be quite compact, a benchtop model has a 0.86 m^2 footprint.

Although LC/MS is rarely used for full structure characterization, LC-MS/MS can be used to determine the occurrence of known compounds as well as being suitable for quantitative analysis. An experiment employing collision induced dissociation (CID) using extract of whole Arabidopsis seedlings was used to generate structural hypotheses regarding the identity of the substance in Figure 1.15, tentatively assigned as a phenolic choline ester.[116]

It is possible to get structural information from the fragmentation pattern of compounds by tandem MS.[117] Selectivity can be achieved in complex mixture analysis by employing multiple reaction monitoring (MRM) for coeluting analytes of different m/z ratios. Single reaction monitoring (SRM) can reduce the dwell time to 10 ms per scan, with a pause of 5 ms necessary for adjustments from one transition to another. Switching between positive and negative mode can be employed in conjunction with reaction monitoring with a cycle of 2-3 seconds for one cycle being the time necessary to measure all transitions including settling times and pauses between transitions.

The advantages of NMR and MS, as well as considerations when planning a plant metabolomic (nontargeted) or metabolic profiling (targeted) experiment, have been presented. At this point, the experiment would be carried out in accordance with the

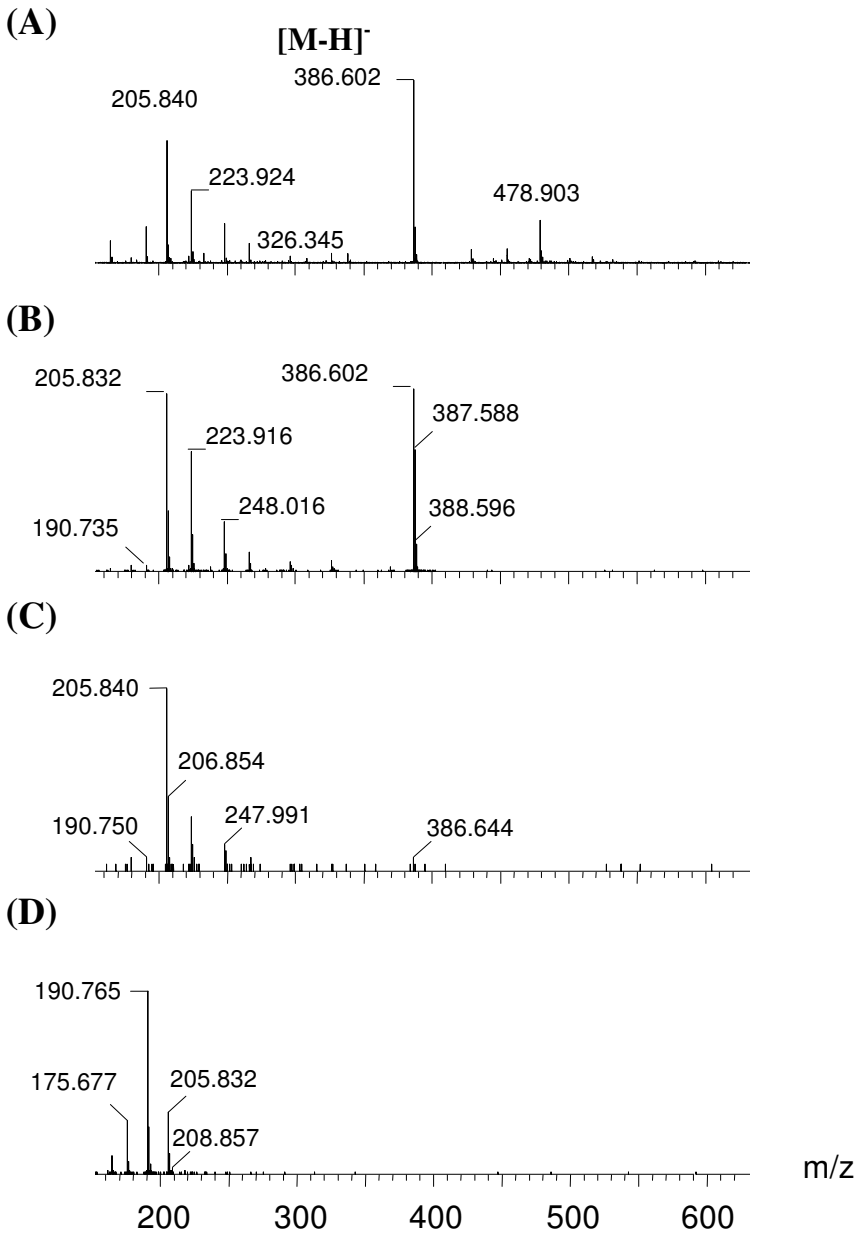


Figure 1.15. Interrogation of a single chromatographic peak by CID MS/MS on subsequent injections of an Arabidopsis seedling extract using collision voltages of (A) 0 eV (B) 10 eV (C) 15 eV (D) 30 eV.

issues addressed in section 1.2.3 regarding replicates and variability. Data processing steps such as normalization are common in all omics studies. Indeed, some of the protocols mentioned here were established in early microarray-based transcriptome

analyses.[118-120] In this work, such techniques were applied to an NMR-based metabolomics dataset acquired with comparison of metabolic snapshots acquired at various time points of low oxygen stress and recovery described in Chapter 3, Section 3.1.2. Our goal was to translate raw data into statistically robust metabolite information through post-acquisition processing as will be described in the following section. This can be thought of as a step to facilitate pattern recognition or dimensionality reduction

1.4 Post-Acquisition Processing

In NMR-based metabolomics, after careful collection of a dataset having a sufficient number of replicates (Figure 1.2) there are many processing steps including apodization, phasing, baseline correction, alignment, integration, scaling and normalization, that can be applied to aid in the extraction of quantitative and metabolically relevant information. LC-MS processing involves many of the same steps; baseline correction, retention time alignment or retention index calculation, mass-to-charge alignment or exact mass correction, peak integration, scaling and normalization. Processing routines should be employed in a context-dependent manner and there is no single method that can be used for all metabolic profiling and metabolomics experiments.[121]

Although some software packages incorporate these processing steps into automated routines, to achieve the best quantitative data possible it is useful to consider the effect of each step on the integrity of the dataset. As discussed in earlier sections, the NMR spectrum or LC-MS chromatogram of most biological samples is complex,

containing many overlapping and/or broad signals. If sampling and acquisition strategies for improving resolution have not been employed, or have been less successful than anticipated, some degree of deconvolution may be applied post-acquisition.

This section will be devoted to a discussion of post-acquisition processing based on 1D ^{1}H NMR dataset as the simplest example. LC-MS post-processing is still experiencing rapid growth, with the open-source platform XCMS becoming one of the best resources for academics.[122] Commercially-available databases (DBc), such as Know-It-All (Bio-Rad, NMR DBc 1,060 Cmpds, MS DBc 199,000 Cmpds) and Chenomx (NMR DBc 300+ Cmpds) are available, which provide spectra for comparison with samples.[123] Fully automated signal deconvolution and assignment of signals to metabolites has not yet been realized. Multiplatform approaches involving a large number of samples can be outsourced for nonbiased data handling.[123]

1.4.1 Apodization

Theoretically, the NMR FID has an infinite duration in the time domain, although in practice signals are collected for a finite time. Often the acquisition time is set to allow the signal to decay into the noise (roughly 3 x T_2*, where T_2* is the time constant for loss of transverse magnetization in an inhomogeneous magnetic field). However, there are circumstances in which acquisition times shorter than 3 x T_2* are used, for example in 2D NMR experiments or when the goal is to enhance the spectral S/N. In such cases, apodization of the FID by multiplication with a function (often called a window function)

can be used to force the FID intensity to zero and reduce distortions that would be produced by the Fourier transformation of a truncated FID.

One of the most commonly used apodization functions is an exponential decay which can enhance S/N at the cost of increased line width and poorer resolution. In the best case, the exponential decay function is on the same order of the resonance T_2^*, and in this case no real loss of resolution occurs. Resolution can become an issue in NMR spectra of complex biological samples, consisting of thousands of resonances, where overlapping peaks are already evident. In a study in which the NMR acquisition and processing parameters including pulse flip angle, number of scans, and amount of exponential apodization were optimized for the metabolomic analysis of rat urine, it was found that the best compromise between increased S/N and broadened resonances was achieved by minimizing the use of apodization and maximizing the number of scans.[105] For ^{1}H NMR spectra, the line broadening factor applied should generally not exceed natural linewidths of the metabolites of interest, 1.0 Hz was most often used in this work.

1.4.2 Phasing and Baseline Correction

Zero and first-order phase errors are introduced into the NMR spectrum by mismatch between the receiver and reference signal phases and by precession of magnetization during the short delay required for pulse ring down prior to the start of acquisition. These errors are corrected by a process called phasing which can be carried out manually by visual inspection of the results of the phase correction or using an automated routine. Although automated phase correction is preferred for analysis of

large data sets and may be more reproducible than manual adjustment, the results obtained should be carefully evaluated especially for spectra that do not have flat baselines or contain a significant residual solvent resonance, as these features can hamper automated phasing routines.

Baseline correction can also be applied manually by inspection of the baseline in a region free of resonances or using automated routines. The goal of this correction is a flat spectral baseline with a zero offset so that resonance integrals reflect the true peak intensity without distortion from integration of the underlying baseline. The digital signal processors and digital filters in most modern NMR spectrometers avoid the baseline distortions created by analog filters and have diminished, but not completely eliminated, the need for post-acquisition baseline correction.[124]

Another important consideration specific to metabolic profiling is that perceived baseline distortions may actually result from residual protein resonances present in the biological sample.[125] One classical solution to this problem is to subtract a spectrum measured for a sample containing only protein from the spectrum of a solution containing the protein and the analytes of interest.[126] Although it is not really practical to prepare an identical protein-only blank for most biological samples, deGraaf et al. effectively accomplished this using a process called "diffusion-sensitized" ^{1}H NMR spectroscopy.[127] In this approach, spectra of rat plasma are measured using an NMR diffusion pulse sequence with gradient amplitudes large enough to eliminate the rapidly diffusing small molecules and generate a reference spectrum containing only the resonances of the macromolecules present in the sample. This spectrum was subtracted from the standard

single pulse NMR spectrum measured for the same plasma sample, yielding a baseline-corrected spectrum which could subsequently be used for quantitation.[127] Excellent correlation was obtained between the two methods for most of the major metabolites including acetate, alanine, creatinine, β-hydroxybutyrate, glucose, lactate and valine. Regression analysis of the results produced by both methods yielded a straight line with a correlation coefficient of 0.9989, a slope of 0.978 and an intercept of 0.054 mM.

A purported advantage of NMR-based metabolomic experiments is that there is little interlaboratory variability.[16] The combined effect of apodization, phasing and baseline correction is a determining factor of whether this statement is true. On behalf of the German Federal Institute of Materials Research and Testing, Malz and Jancke[105] attempted interlaboratory validations of quantitative NMR data. In a 1999 trial, it was found that the results differed by up to 100% between labs when analyzing a simple five component reference mixture. The poor agreement was found to be caused by differences in the acquisition and processing procedures used in each laboratory. In the 2005 trial, the cooperative mandated the use of a 30^{o} pulse, collection of 32k data points, and a relaxation delay of 20 s. The data were then zero filled to 64k points and apodized by line broadening equivalent to 0.3 Hz. Phase and baseline corrections were performed manually, as was the definition of integral regions. In the second trial, the maximum uncertainty was found to be 1.5%, corresponding to a confidence interval of 95%. A similar interlaboratory comparison was conducted[56] using rat urine, although the goal of the experiment was a more qualitative identification of biomarkers for hydrazine-induced toxicity. Interestingly, it was found in this study that the biomarkers identified were not

different whether baseline correction and phasing were conducted manually or automatically.

1.4.3 Selection of Integral Regions

It is well known that the accuracy of integrals depends on digital resolution as well as on the appropriate selection of integration limits.[128] It is relatively simple to achieve adequate digital resolution either by acquisition of sufficient data points or by extending the number of points in the FID through zero-filling or linear prediction. In practice, selection of integral regions can be more of an art. Since NMR resonances have Lorentzian line shape, much larger integral regions are required than for data sets with Gaussian peaks. For a Lorentzian peak with a width at half-height of 0.5 Hz, integration regions set at 3.2 Hz or 16 Hz on either side of the resonance would include approximately 95% or 99% of the peak area, respectively. Note that this analysis does not include the ^{13}C satellites which account for an additional 1.1% of the intensity of carbon-bound protons in samples containing ^{13}C at natural abundance. Use of a fixed integration interval is impractical since line widths vary from resonance to resonance and selection of integral regions sufficiently wide to capture >95% of the resonance intensity is not feasible in most metabolic profiling experiments because of the complexity of the spectra produced by biological samples. For actual samples resulting from tissue extracts, such wide integral regions may not be possible due to resonance overlap.

One approach to the problem of resonance overlap is to use curve fitting to obtain peak areas.[129] Methods for automated peak recognition and metabolite identification have

been developed by industry: AstraZeneca,[130] GlaxoSmithKline,[131] Merck,[125] Hoffman La Roche,[132] Bristol-Myers Squibb,[66] Chemomx,[65, 133] Bio-Rad,[134] Varian,[135] and Bruker.[136] Calculation of peak integrals by fitting procedures has been said to solve overlap problems, however some claim that the precision of the resulting integral rapidly decreases as peak separation becomes smaller.[137] S/N is a critical factor to optimize to achieve high quality results from automated fitting routines.

In a study conducted by Martin,[137] results determined by unsupervised peak fitting showed superior precision to integration by summation if line width was optimized in the peak-fitting integration. An alternative is to use supervised peak fitting, as described by Weljie et al.[133] who employed a predictive metabolite library to generate a Lorenzian peak shape model of each reference. This approach, called targeted profiling, is expected to become an increasingly common method for spectral deconvolution in datasets acquired for purposes of metabolic profiling and it is likely that a complete set of reference standards comprising all available characterized metabolites will be incorporated into the analysis at the outset of an experiment.[131] Other attempts have been made to use mathematical approaches to decompose NMR spectra into individual components, called independent component analysis[138] and molecular factor analysis, but these methods have so far failed to produce biochemically relevant results, as the biomarkers identified rarely match known metabolites.[131]

Many metabolomic experiments employ equidistant binning (Figure 1.6), also known as bucketing, in which the NMR spectrum is divided into many small integral regions of an arbitrary width; 0.02 ppm was used in this work. Binning does not

necessarily yield direct quantitative information about specific metabolites because bins may include intensity from more than one resonance or because a single peak may be divided into two or more integral segments. An advantage to bucketing lies in the observation that some resonances display changes in their chemical shift as a result of small differences in the sample pH or ionic strength. The use of relatively wide bins allows a resonance to have some deviation in its chemical shift and still lie within the same integration region. Although this strategy can simplify the statistical treatment of the dataset, the use of larger bins translates into a reduction of the effective spectral resolution. Dieterle et al.[132] describe a nonequidistant binning approach that takes into account resonance line widths. In this method an average spectrum is calculated using all of the measured spectra. The first derivative of the average spectrum is calculated and the borders of the bins defined based on regions of maximum slope. The application of this binning algorithm resulted in more than 3 times the number of integral regions than the previously used equidistant binning, with a corresponding increase in effective spectral resolution.

In LC-MS investigations, the data is converted into retention time-mass pairs which are in effect "binned" by a supervised step where a mass window and retention time window are defined to account for drift in both mass accuracy and chromatography, respectively. There exist automated routines data cleaning for GC-MS and LC-MS which have been of particular use in plant biology.[139]

1.4.4 Peak Alignment

^{1}H NMR spectra are typically aligned through the use of a chemical shift reference signal, as discussed in section 1.2.5. However, deviations may arise due to variations in the sample matrix and instrumental instabilities that can produce erroneous integrals if automated integration routines are employed.

Peak alignment has been employed by Forshed et al.[130] using two recently developed genetic algorithms, segment-wise peak alignment (SWA) and peak alignment by reduced set mapping (PARS), which were compared to classical bucketing of unaligned spectra using principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) regression analysis.[140] The two dedicated peak alignment methods produced better class separation in the PCA or PLS-DA scores than could be obtained with the bucketed datasets. The incorporation of a peak alignment procedure was also shown to improve the outcomes of a metabolite fingerprinting experiment carried out on a set of green tea samples of different type, grade and geographical origin.[141] The effect of these peak alignment methods on the quantitative accuracy of metabolic profiling datasets has not been shown, but it is thought that errors would be introduced into peak integrals if the integration is achieved by summation since these alignment methods are known to reduce spectral resolution.[140]

1.4.5 Normalization

After apodization, phasing, baseline correction, integration, and possibly alignment have been carried out, the experimenter is left with a table in which each row

represents the data from a given sample and each column corresponds to the integrated area of a single resonance or a spectral bin, as discussed in section 1.4.3. To be able to perform quantitative analysis on even a relative basis, most datasets must be subjected to normalization.[121] Normalization allows data to be directly compared in cases where the biological samples are either of unknown/non-uniform amount or unknown/non-uniform concentration. The most commonly used method of normalization, applied in cases where there is no *a priori* knowledge of the biological sample mass or volume, is normalization to a constant sum.[142] In this method, the sum of each row is calculated (all integrals from a sample) and each integral is then expressed as a fraction of the total spectral integral. This normalization approach is often used in conjunction with binning integration and is not recommended for metabolic profiling of selected, specific integral regions as the sum of all integrals is likely to have more variation when less of the spectrum is sampled.

In cases where the grams of wet or dry weight of solid[136] or volume of fluid constituting each sample is known, normalizing to these factors can lend more biological relevance to the assay. It has been pointed out that this approach is less biologically relevant for metabolic profiling of urine due to the variable nature of urinary volume, and it may be more clinically useful to normalize to the time interval between urinary excretions and body weight of the organism under study.[121] In a comparison using our data between sum and mass normalization (Figure 1.16), normalization to a constant sum was found to have 26% RSD when trying to quantify metabolites using all integral areas

without solvent 40% RSD when trying to quantify metabolites normalizing to tissue dry weight (DW).

Some normalization techniques employed recently in metabolic profiling have been influenced by state-of-the-art quantitative biological techniques such as mRNA microarrays, perhaps because metabolic profiling datasets are now being combined with datasets generated from these types of experiments.[143, 144] In this approach the integrated area of each chosen metabolite resonance is referenced to the integrated area of a "housekeeping" metabolite.[121] This method has been applied in gel electrophoresis for protein quantitation since the early 1990s[145] and more recently in reverse transcription-polymerase chain reaction (RT-PCR) and gene chip array technologies for quantitation of mRNA. Alternatively, the concentration of a specific metabolite can be determined by an independent means, for example by enzymatic assay,[146] which can provide a scaling factor for the entire spectrum.

It has been stated that non-weighted linear multivariate pattern-recognition techniques, such as PCA, favor more intense resonances and fail to identify biologically important compounds that are present in relatively low quantities. Gipson et al.[131] propose a non-uniform weighting scheme to tease out significant biomarkers which may be present at low levels or which have resonances deeply buried in crowded spectral regions. Non-uniform weighting involves a different normalization factor to be used in each column and in each row of the table of data as described earlier. It has been stated that non-uniform weighting can aid in binary classification of samples based on

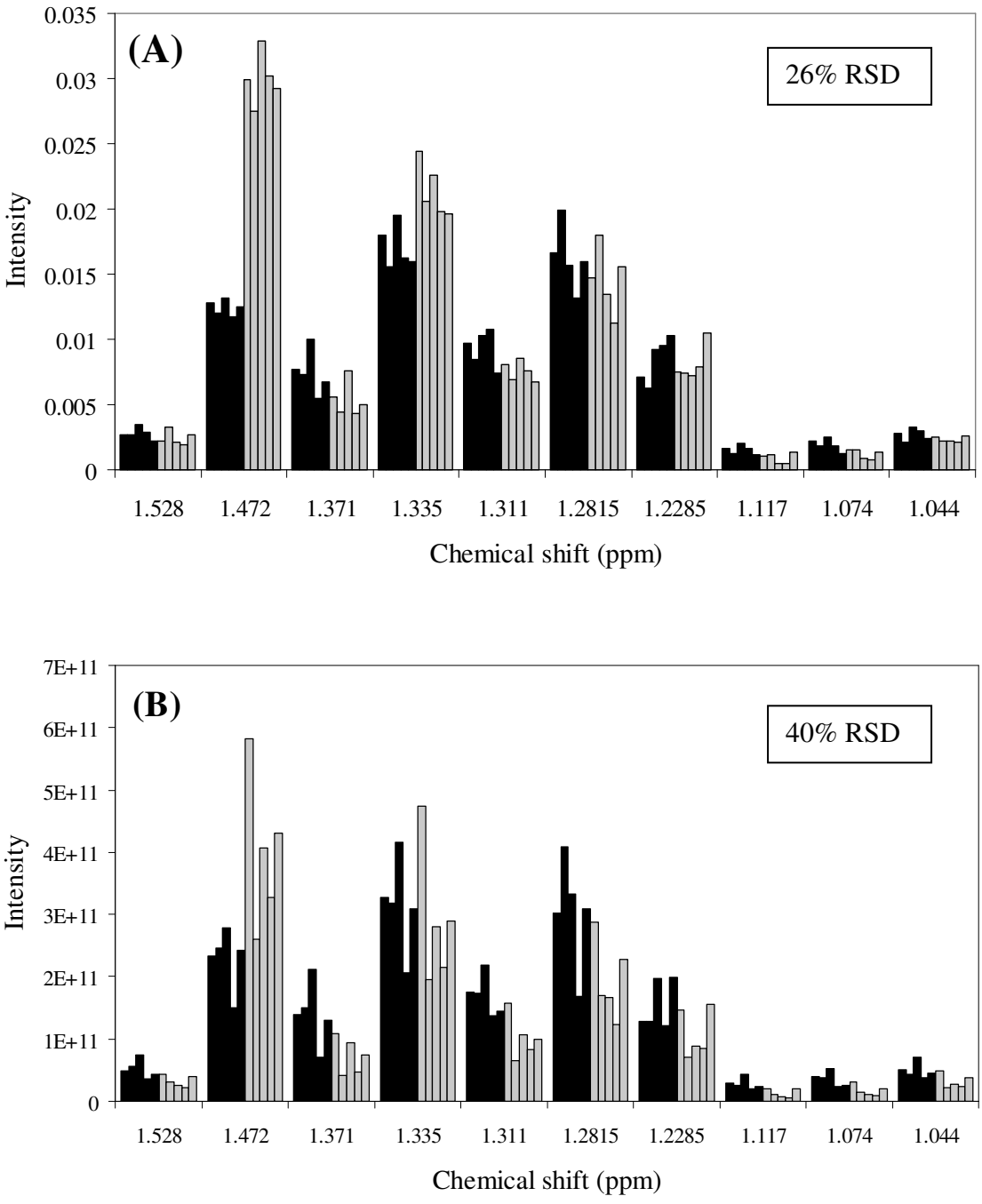


Figure 1.16 Normalization to (A) constant sum or (B) tissue dry weight. Bioreplicates are represented by individual bars (N = 5). Arabidopsis seedling root control (RC) tissue extracts are in black, hypoxia-treated root extracts (RH) are in grey. Shown is the aliphatic region of NMR spectrum occupied by organic and amino acid β and γ proton resonances using integrals for targeted quantitative analysis.

differences in low-abundance metabolites, however the potential for absolute quantitation is lost, therefore it is not recommended for metabolic profiling datasets.

Intrabatch drifting and batch-to-batch differences are quite common in LC-MS and are caused by a variety of factors, including ion source contamination, instrument maintenance between batches, and time between sample preparation and analysis. In large-scale studies (taking months to years to complete), drift can present a significant roadblock to achieving good reproducibility. In targeted compound analysis, calibration standards and stable-isotope-labeled internal standards can be employed to achieve the required precision and accuracy, however in non-targeted studies the observed responses are often of unidentified substances. Calibration with a QC, as recommended and diagrammed by Figure 1.3, is designed to overcome this problem by performing a quantitative scaling of responses based on repeated analysis of a biological calibration sample (not a standard).[4]

1.4.6 Centering, Scaling and Transformation

Centering and scaling are performed on columns of data, represented in this case by the integrated area of one resonance or one spectral bin across all samples. Centering, in general, adjusts for differences in the offset between high and low abundance metabolites. Scaling can adjust for differences in the magnitude of the metabolite fold change by converting the data into differences relative to a scaling factor. Transformations, in general, are nonlinear conversions of the data which aim to correct for nonlinear variance or to make skewed distributions more symmetric.[147] Mean

centering is achieved by subtracting the column mean from each value in the column, giving each column of the table a mean of zero. This is done so that all components found by multivariate analyses have their origin at the centroid of the data.

The most commonly applied scaling technique goes by many names including variance scaling, unit scaling, unit variance scaling and autoscaling. Variance scaling is achieved by dividing all the values for a certain variable by the standard deviation for that variable. After variance scaling, all metabolites have a standard deviation of one and therefore the data is analyzed on the basis of correlations instead of covariances, as is the case with centered data. Many times, mean centering and variance scaling are performed on the same dataset. Pareto scaling[147] uses the square root of the standard deviation as the scaling factor. This has advantages over variance scaling because large fold changes are decreased more than small fold changes. Also, the data does not become dimensionless as after mean centering and variance scaling.

Two types of transformations in common use are the log transformation and the power transformation. Both of these techniques reduce large values more than small values, resulting in a pseudo scaling effect. It can be useful to perform a transformation prior to one of the aforementioned scaling methods because these transformations do not use a scaling factor with any biological or analytical relevance. Unfortunately it is not clear how transformation and scaling methods influence each other when applied to complex metabolic profiling datasets.[147] Parsons et al.[148] compared variance scaling, Pareto scaling, log transformation and mean centering using statistical approaches to quantify the effects of the various scaling, centering and transformation processes. The

data compared resulted from metabolic profiling of canine urine, mussel muscle extracts, and fish liver extracts analyzed by one- and 2D ^{1}H NMR. Multivariate statistical methods (PCA and ANOVA) were utilized to evaluate the effects of scaling techniques upon each data set as a whole instead of treating each resonance or spectral bucket individually. It was found that logarithm transformation improved the binary classification accuracy for the mussel and fish tissue extracts, but the use of the log transformation and variance scaling had approximately equal classification accuracies for the canine urine samples.[148]

1.5 Turning Information/Data into Knowledge

Metabolic profiling, metabolomics, and metabonomics have been characterized as a five-stage process: experimental design and meta-data capture, data pre-processing, cleaned data, turning data to knowledge, and databasing.[37, 149] Thus far, this chapter has described how to arrive at cleaned data. Once a clean data matrix has been constructed, the data can be queried and/or deposited in a repository for public access.

1.5.1 Multivariate Data Analysis

Chemometrics, defined as the analysis of chemical measurements with the aim of pattern recognition, has played a central role in the development of metabolomics.[150] One way of dealing with overlapping peaks in NMR or LC-MS data, besides spectral subtraction and deconvolution by peak fitting, is to spread the data across more dimensions by transforming the data matrix into a new coordinate space using a

multivariable data modeling technique such as PCA.[112] Using this approach, the dimensionality of the problem is reduced without losing much of the information. Using a simple example, two observations can be X_1 and X_2 under conditions 1 and 2. Assuming we have a random sample of N observations on X_1 and X_2, we can find the means $\bar{x}$ and mean-center each observation ($x_i = X_i - \bar{x}$). Two new variables C_1 and C_2 can be created, called principal components, which are linear functions of the original variables X_1 and X_2 as follows:

$$C_1 = a_{11}\, x_1 + a_{12}\, x_2 \qquad (1.4)$$

$$C_2 = a_{21}\, x_1 + a_{22}\, x_2 \qquad (1.5)$$

Weighting coefficients (a_{11}, a_{12}, a_{21} and a_{22}) are chosen such that the variance in C_1 is maximized, C_1 and C_2 are orthogonal (uncorrelated), and the sum of the squares of the coefficients for each principal component is one. This can be thought of as a rotational operation or matrix transformation such that the new axes are drawn in terms of the variance in the dataset rather than just raw intensity values.

The new variables C_1 and C_2 are also known as eigenvectors. The variance explained by C_1 and C_2 are known as the eigenvalues. In this new coordinate space, the sum of the variance is preserved, and the lengths of the new axes are proportional to standard deviations. Focus can be made on the column (Q-mode) or row (R-mode) space. The Q-mode eigenvectors are named scores while the R-mode eigenvectors are named loadings.[151, 152] Often, most of the variability in the data can be explained by a few (2-5) components. In these cases, multidimensional data can be reduced to two dimensions by a scatterplot of the first two eigenvectors. The transformed data matrix can

be visualized in many ways.[153] Clusters of samples according to treatment are generated using "scores" and confidence in metabolite assignments are supported by "loadings." Using multivariate data modeling, covarying regions of the NMR spectrum or total ion chromatogram can be identified as they relate to biological treatment. Hypotheses can be developed in the way of chemical structures which would satisfy the conditions of covarying regions. Multivariate data analysis is useful as a global way to examine clustering and correlations within the entire data set.

An additional advantage of unsupervised data exploration is that outliers are identified. Loading lengths (corresponding to the magnitude of variance) can indicate those rows or columns input to the model which possess excessive variance, for example due to poor water suppression or when a contaminant has been introduced to some (but not all) samples through the use of shared lab space and equipment, for example the speedvacuum. The position of a sample within a scores plot can reveal if it contains covarying integral regions among similar or different treatments.

Partial least squares (PLS) is an ideal technique for analysis of data sets in which the variables are highly correlated and greatly outnumber the samples, as can be the case with data generated by NMR-based metabolomics. In some cases, samples with excessive time between extraction and analysis can be identified by PLS, particularly by plotting leverages versus standardized residual. The algorithm we used is based on nonlinear iterative PLS algorithm developed by Wold. The algorithm is designed to reduce the number of variables to a set of uncorrelated components.[154, 155]

Hierarchical clustering is another unsupervised method that can be applied to samples or variables.[156] This analysis is useful to separate treatments or variables into groups and examine the overall characteristics of the resulting groups. For example, identical treatments should cluster together. This is another way to identify outlying samples. In any case, multiple regions occupied by the same metabolite should cluster in the same group, provided that there is no overlap of metabolite resonances. Coregulated metabolites will cluster together if a small number of groups (10) is defined at the outset of the calculation.

1.5.2 Assigning Metabolites to Peaks

The problem of assigning chemical signals to metabolites is one of the most time consuming aspects of a metabolic profiling experiment. Even experts in the field are reaching the limit at 40-50 metabolites per experiment, adding at a rate of ten per year once the obvious signals have been assigned.[157] 1D NMR spectra or LC-MS chromatograms of metabolite standards in a buffered solution designed to imitate the tissue extract under investigation are used to build an in-house database specific to each plant/tissue/biofluid/extract under study in our lab. The chemical shifts, multiplicities, and J-couplings observed in standards are matched against the molecular data generated by the biological samples until each observed NMR resonance and MS fragment is assigned. The limitation of this approach is that it is based on a survey of metabolite standards which must be known *a priori*.

Comparison of experimental spectra with literature values is extremely important in establishing confidence in assignments. The development of publicly available databases has been slow, but the information shared in these web resources is powerful and growing in spurts as new versions are released and new data is uploaded.[158] Listings of metabolites for different organisms can be found in the KEGG[159] and BioCyc[160] databases, which connect individual metabolites to pathways. General physical and chemical properties of metabolites can be found in chemocentric databases such as PubChem or SciFinder, the chemical abstracts service (CAS) database. The human metabolite database (HMDB)[161] contains more than 7,900 metabolites found in the human body and is annotated with chemical, clinical, and molecular biology data. The Golm metabolome database (GMD)[162] contains mass spectra and retention index metabolite (MSRI) libraries, for 3,362 unique derivatized analytes via GC-MS based metabolic profiling. Another resource for MS-based analyses is the metabolite and tandem MS database (METLIN)[163] where searches can be compared against a 45,160 compound library based on mass, charge, and MS/MS constraints. The most valuable resource for NMR-based metabolomics is the Madison Metabolomics Consortium Database (MMCD)[164] because it is possible to search by NMR chemical shift. Combinations of chemical shifts can be inputs, limiting the number of false positive hits. The search feature also accommodates searching a range of chemical shift values, limiting the number of false negative hits.

1.5.3 Univariate Data Analysis

Since the days of Charles Darwin (1809-1882), biological experimentation and pairwise comparisons have been a useful approach for unraveling the mysteries of living systems. We can obtain an unbiased estimate of population variance from the sample variance by adjusting the number of degrees of freedom lost computing the mean as follows.[165]

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} \qquad (1.6)$$

The sample standard deviation[166] (s) is calculated by subtracting each value (x) by the average of all the values ($\bar{x}$) for integrated metabolites in plant extracts from bioreplicates treated identically. The denominator contains the number of samples (N) minus one to correct the degrees of freedom.

After taking a global snapshot of the metabolic state of an organism under control and stress conditions, with the proper number of observations to provide adequate statistical power, a one-by-one treatment of identified and interesting unidentified chemical signals logically follows. When comparing the average metabolite level between two biological treatments, it is likely that the metabolites do not have equal variance. Therefore it is more robust to consider the Student's *t*-test with two independent small normal samples with unequal variances.[120] When all available values are used, a population standard deviation is calculated, when only a subset of available values are used a sample standard deviation is indicated. In our case, we sample only some bioreplicates, therefore population standard deviation is not appropriate.

By extension, the pooled sample standard deviation (s_{pooled})[166] between two distributions can be calculated for each integral region under investigation

$$s_{pooled} = \sqrt{\frac{s_s{}^2(N_s - 1) + s_c{}^2(N_c - 1)}{N_s + N_c - 2}} \tag{1.7}$$

using the sample standard deviation of a sampling of stressed treatments (s_s) and the sample standard deviation of a sampling of control treatments (s_c) with some number of bioreplicate measurements of each treatment, N_s and N_c respectively. A Student's *t*-test[165] is performed to evaluate significant differences between control and stressed populations for each variable (metabolite). The calculated *t* (t_{calc}) is compared with a table of values

$$t_{calc} = \frac{|\bar{x}_s - \bar{x}_c|}{s_{pooled}} * \sqrt{\frac{N_s N_c}{N_s + N_c}} \tag{1.8}$$

according to the number of samples. Using Excel 2003, t_{calc} can be converted to p-value to establish precise differences for comparisons taking into account the cumulative sample variances. Changes in metabolite levels are ultimately reported analogously to microarray experiments, either as fold change or signal-log ratio

$$\text{Fold change (FC)} = \bar{x}_s / \bar{x}_c \tag{1.9}$$

$$\text{Signal-log ratio (SLR)} = \log_2(\bar{x}_s) - \log_2(\bar{x}_c) \tag{1.10}$$

where $\bar{x}_s$ is the average of the integrated intensity of the metabolite x among samples (i.e. plants subjected to the stress treatment) and $\bar{x}_c$ is the average of the integrated intensity of metabolite x among controls (i.e. plants maintained under a control treatment). SLR is applied to reduce the impact of large changes and enhance the impact of more subtle changes, which may both be biologically important.

1.5.4 Pathway Visualization

Orthogonal measurements (transcriptomics *vs*. metabolomics) can confirm biological hypotheses and lead to more complete understanding of biological networks.[167] Visualization tools such as MapMan (Figure 1.17),[168, 169] Pathway Tools Omics Viewer,[160] KaPPA-view,[170] rely on pathways obtained from KEGG[159] or MetaCyc[171] coordinating transcriptome data in the framework of a functional metabolic pathway display allowing comparison between datasets and between organisms.[172] One drawback to this type of representation is that the standard deviations are not represented, only average intensities painted into boxes and color-coded to express the magnitude of the variable at a single timepoint.[173] It is more challenging to illustrate the effect of a treatment over time or a dose dependence, although sometimes multiple boxes are shown by each node where each box represents one treatment.

In some cases is it useful to build a metabolic pathway map containing only relevant pathways to emphasize those most central to the biological story that emerges from the omic datasets. This format can be less distracting for the reader because there is effectively less noise represented in the same space, but it does not translate as easily from one study to another for direct comparisons. There exist software programs for building pathways from scratch, reviewed by Gehlenborg et al.,[174] but none of these were utilized in this work. Metabolomic investigations are particularly useful for hypothesis generation. For example, results of a non-targeted investigation of dynamic changes in

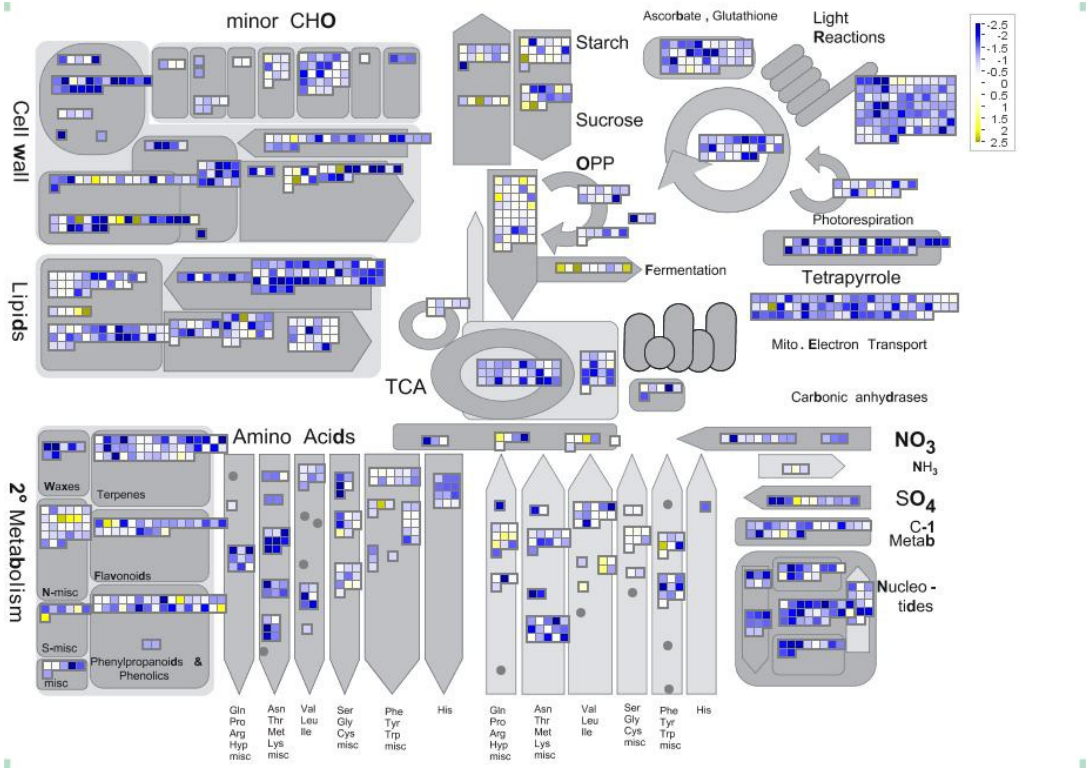


Figure 1.17. Output from MapMan can express each transcript as a different box. Boxes are arranged according to pathways, with each box representing an individual gene. Blue indicates the transcript level decreases, yellow indicates the transcript level increases, white indicates no change in transcript level in response to biological treatment or no transcript detected. A color bar with corresponding $\log_2$ fold change is at upper right.

metabolite levels in rice foliage over a 24-hour period suggested that bottleneck enzymatic steps exist in both carbon and nitrogen metabolism.[175]

Comparison between transcript data and metabolite data can also be performed manually. Signal intensities corresponding to each transcript and selected metabolites can be compared across sample treatments by dividing the intensity of the signal in the stressed sample by the intensity of the signal in its control, yielding a fold change. The transcript fold changes are typically $\log_2$ transformed to prevent small fold changes from being masked by larger fold changes. This was not done to the metabolite data since the

response of metabolite levels to the stress was less dramatic. Transcripts encoding metabolic enzymes received special consideration and consensus was sought in pathways where corresponding enzyme and substrate/products demonstrated significant fold changes in response to biological treatment.

1.6 Motivation

1.6.1 Population Growth, Food Security

Manuel Benítez was a bullfighter from humble origins who was nearly illiterate. On May 20, 1964 he was nearly gored in a fight and became famous overnight in Spain. A reporter questioned whether he feared being stabbed by the bull's horns. His answer was this: “Más cornás da el hambre.” Roughly translated, he said "Hunger stabs you worse." This statement reveals the connection between food, health, and security. If an impoverished human being is willing to risk death to provide for his family, it can be seen that hunger can be a powerful motivating force.

Human population increases, particularly in flood-prone geographical areas, has presented considerable challenges for food producers in Haiti, Bangladesh, Egypt, India, Australia, Pakistan, China, Cuba, Brazil, Mexico, South Africa, and even in the United States.[176] The incidence of flooding has increased 30% in the past 20 years, 2-5% in the past year alone.[177] By 2020, the United Nations predicts that the human population will exceed 7.7 billion,[178] nearly a 12% increase over the current population,[179] with ninety-eight percent of population growth occurring in developing countries.[180] As many as 450 million poor farmers, defined as those whose resources (land, water, labor, capital) do not

permit a secure livelihood, support 1.25 billion poor hungry people. In developing countries, poor farmers are responsible for producing up to 90% of domestically consumed food. The reality that these farmers are facing is decreased yield due to global warming and associated weather instability coupled with a shortage of arable land. Most of the soil with high productivity potential is already under cultivation. In westernized societies, rising social standards increase demand for high-quality foodstuffs. Dietary habits are shifting towards consumption of meat and milk products, causing the demand for grain to more than double. This places even greater need for bolstered crop yields for food stability.

Transgenic crops with improved nutrient profiles may address deficiencies that contribute to mortality or poor health in impoverished countries where malnourishment is a critical issue. Health benefits have been attributed to a variety of phytochemicals,[181] which will likely continue to be of interest in the USA, Japan and Europe where the market for so-called nutraceuticals is predicted to be large and growing. Genetically engineered plants may serve as factories for pharmaceuticals, an issue which has been recently highlighted as a solution for delivery of medicine to "bridge the gap" between developed and developing countries.[182] A primary bottleneck is quality control, representing an opportunity for analytical chemistry to facilitate development.

Corporations with an interest in agricultural biotechnology are currently exploiting metabolic profiling for profit. On the order of 400,000 chromatograms are recorded by BASF per year,[183] using both GC and LC coupled to MS. More than 1.5 million metabolic profiles have been recorded and interpreted as of September 2009 with

annotations accompanying 500-700 of 9000 signals (5-8%), underscoring the need for further attention in this area. Annotation of the metabolome will require expansion of analytical methodologies into regions of chemical space previously uncovered. This calls for methods aimed at both discovery (non-targeted) and careful follow-up (targeted) studies to elucidate structure, function, and biochemical relationships in plant metabolic pathways. Using tagged libraries of model plant species can facilitate the task of annotation, as well as generating leads for agribusiness applications, where the main focus of research is to improve yield.

1.6.2 Climate Change and Sustainability

The "green revolution" of the 1950s, 60s and 70s, which saw increased use of irrigation, synthetic fertilizer, hybrid and other improved seeds, and synthetic pesticides can no longer compensate for rising demand. In the 21st century, "sustainable" approaches are becoming the standard. Biotechnology in the form of genetically-modified organisms has not found acceptance in all countries, however "traditional" breeding has begun to yield varieties of cultivated crop plants able to withstand drought, flooding, salinity and other abiotic stress, expanding utilization of otherwise undesirable land areas.

Plants possess a great power to harness the sun's energy, building nutrients and essential vitamins out of oxygen, sunlight, carbon dioxide, and minerals in the rainwater and soil. Plants serve as feed for our animals, fuel for our machines, food for our families and sources of pharmaceuticals.[184] Our quality of life is enhanced by plant-derived

perfumes and flavorings, cosmetics and toiletries, materials for sporting equipment and transportation, musical instruments and art (pigments, pencils, and photographic film). Many of our dwellings are manufactured and decorated with plant-based materials. The clothing we wear for comfort and fashion is made from plants.

Plants have evolved in parallel with humans for a hundred million years.[185] By learning about plants, we are learning about ourselves. Mammals and land plants, like all eukaryotic organisms, utilize oxygen in cellular respiration. Taking a bird's eye view of cellular metabolism in eukaryotic cells[186] (Figure 1.18) it is clear that a small number of key intermediates are generated via catabolism, which can drive work or become incorporated into cellular building blocks, ultimately supporting growth and reproduction. In oxygen-limiting conditions, severe hypoxia or anoxia, the cell is handicapped in its ability to produce ATP for energy-dependent reactions because of the requirement for oxygen as the final electron acceptor in the mitochondrial electron transport chain that is coupled with ATP production. Under situations of oxygen deprivation, anaerobic energy metabolism can only be generated through substrate level phosphorylation requiring key glycolytic intermediates: phosphoenolpyruvate (PEP) and pyruvate. Oxygen deprivation may induce auxillary metabolic processes to augment the small pool of coupling intermediates formed by glycolysis and the TCA cycle.[186] We aim to understand specifically what can and cannot change in metabolism when oxygen availability changes. These changes may include pathways that generate different kinds and amounts of end products as well as mechanisms for controlling the numbers and kinds of coupling metabolites.

Low oxygen conditions in plants arise naturally by a variety of circumstances; ice encasement of leaves and fruits causes annual crop losses in citrus and other warm-weather fruits where a physical barrier of frozen water prevents gas exchange, flooding (waterlogging of root systems to complete plant submergence) of wheat, corn, soybean and rice at a critical point in development can decimate yields by restricting diffusion of

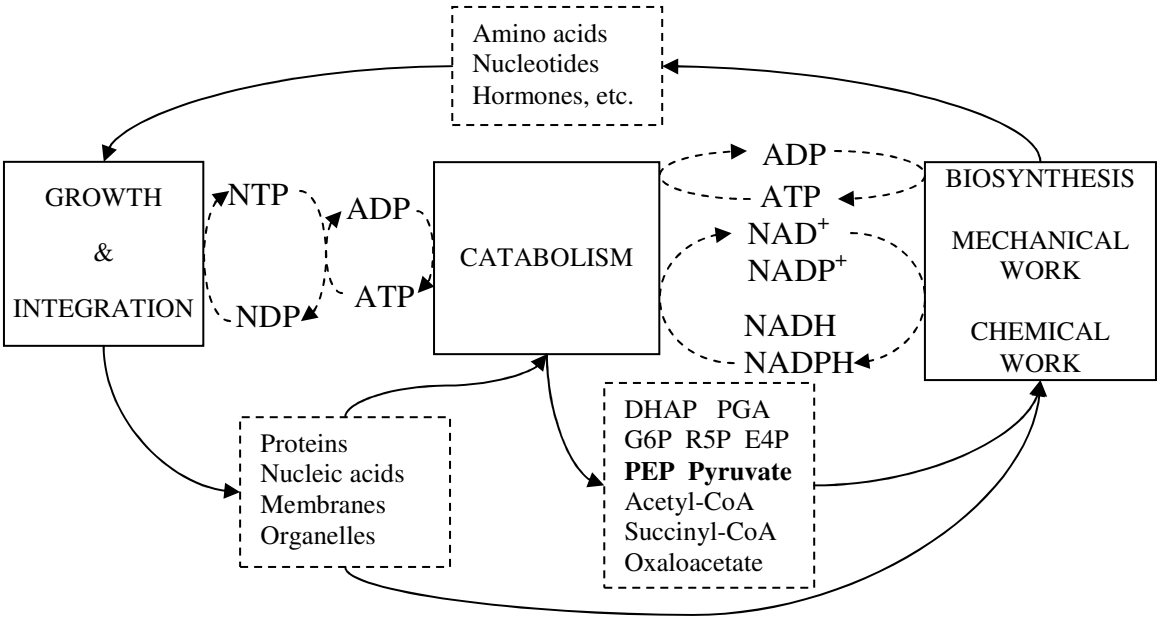


Figure 1.18. A functional block diagram of metabolism in a eukaryotic cell. Although metabolic details are omitted, the diagram emphasizes the small number of connections between the blocks.

gases to inundated organs. In some soils available oxygen is consumed by microbes inducing hypoxic conditions within root systems. In cases of contamination, when tar or other viscous media are dumped onto soil, plants will experience hypoxia or anoxia similar to or more severe than inundation by water. Other cases of low-oxygen arise naturally for burrowing mammals who survive despite poor air flow in tunnels, alpine plants and animals that live at high altitudes, breath-hold divers such as whales and seals,

and hibernating species such as ground squirrels that can survive while taking only a few breaths per hour.

One strategy to prevent spoilage of fruits and flowers under storage is to maintain hypoxic conditions; in particular the study of apple has been revelatory regarding metabolic reconfiguration in oxygen-limited environments.[187] Anoxia-induced metabolic rate depression accounts for the long shelf-life of fresh shellfish in seafood markets, via mechanisms that also help shellfish survive in poorly oxygenated, contaminated, murky or dehydrated tidepools.[186, 188]

Metabolism under oxygen-limited conditions has been an area of active research driven from a clinical perspective. Two of the most common causes of death are stroke and cardiac arrest.[186] In both cases, anaerobiosis occurs in various tissues of the body. In the growth and proliferation of cancer, tissue metabolism becomes anaerobic when sufficient blood supply has not permeated local areas. Studying metabolic adaptations to survive periods of anoxia, which have arisen as a result of evolutionary pressure, has been a useful strategy in plant biology as well.

A consortium of researchers in operation for over a decade is the International Society for Plant Anaerobiosis (ISPA), which aims to cooperatively understand plant adaptations to low oxygen environments. Members from a wide cross-section of scientific disciplines and who work with a diversity of plants (some adapted to grow in low-oxygen environments and some of agricultural importance) include ecologists (who study life and interactions between life forms within ecosystems) and agronomists (who develop plants for specific purposes including food, fuel, feed and fiber). Taking a cross-

disciplinary approach to problems at the interface of biology and chemistry will bring new developments of practical importance and will inform our basic understanding of cellular metabolism.

Metabolic profiling experiments have been shown to be of great use to researchers to determine the impact of genetic modifications on metabolism.[48, 136, 189] The inherent quantitative nature of NMR and its ability to deal with relatively unadulterated biofluids and tissues makes it an ideal analytical technique for metabolic profiling. Many in the field have labored to investigate the effects of acquisition and processing parameters, as shown in this chapter, and such studies have been validated using biological samples. However, the future use of NMR and LC-MS for metabolic profiling and metabolomic experiments will depend on the degree to which the results can be modeled onto biochemical pathways, and generate meaningful biological conclusions that are ultimately supported by the data.

One barrier to the use of metabolic profiling data to draw biological conclusions is that often there is a lack of thorough *metadata* reporting, meaning data about data, such as details about the experimental design, acquisition, and processing parameters. A large team of researchers with experience in the field of metabolic profiling[190] provided a full list of recommendations in an attempt to standardize both the degree of disclosure and the terminology used. Another helpful resource offers kingdom-specific reporting guidelines (i.e. mammals, plants, microbes, etc.) as well as specific standards for organisms studied as they exist freely in nature.[191] If metadata is included with metabolic profiling experiments when they are published, as some journals are beginning to require, there is a

greater likelihood that the results will become part of the larger "omic" field including genomics, transcriptomics, and proteomics. Several studies have already begun to combine datasets from different omics approaches,[143, 144] which allow more confident biological conclusions to be reached.

In this chapter (Chapter 1), analytical chemistry contributions to the growing field of metabolic profiling of plant cells, tissues and organs were introduced. Historical perspectives and challenges were addressed. The details involved in acquiring and processing datasets were articulated. Biological problems motivate the iterative cycle of turning data into knowledge. The focus of this dissertation is to unravel plants' adaptive responses that enable survival of transient low oxygen stress. 1D ^{1}H NMR spectra of plant tissue extracts, the predominant type of data collected in this dissertation, proved particularly useful in plant metabolic profiling to establish the identity and quantity of molecules measured directly in this complex mixture. When signal overlap precluded resonance assignment to a metabolite, a second dimension was employed. For comparison, LC-MS was used to record profiles of hydrophilic metabolites involved in secondary metabolism.

Chapter 2 addresses through experimentation the challenges associated with controlling sources of variance, sampling and number of replicates in experimental design and metadata collection. Using a single pool of tissue harvested from a mature model plant, we identified and optimized factors that influence analytical reproducibility and quality of ^{1}H NMR spectra to represent abundant small molecule metabolites.

The experimental data presented was collected with the aim of obtaining representative and reproducible chemical and biological information. Chapters 3 and 4 provide a time course and organ-specific analysis, respectively, in the study of low oxygen stress. Chapter 5 formalizes our approach to dereplication, employing a multiplatform hyphenated approach combining UPLC-UV-MS and HPLC-UV-NMR to gain a structural foothold to establish molecular identity. Chapter 6 utilizes UPLC-MS exclusively to record metabolite profiles for comparison the advantages and disadvantages as well as relevant factors for such experimental platform. In closing, Chapter 7 presents conclusions and future opportunities at the nexus of chemistry in biology in metabolomics.

1.7 References

1. Moing, A.; Maucourt, M.; Renaud, C.; Gaudillere, M.; Brouquisse, R.; Lebouteiller, B.; Gousset-Dupont, A.; Vidal, J.; Granot, D.; Denoyes-Rothan, B.; Lerceteau-Kohler, E.; Rolin, D., Quantitative metabolic profiling by 1-dimensional ^{1}H NMR analyses: application to plant genetics and functional genomics. *Funct. Plant Biol.* **2004,** 31, 889-902.

2. Feldberg, L.; Venger, I.; Malitsky, S.; Rogachev, I.; Aharoni, A., Dual labeling of metabolites for metabolome analysis (DLEMMA): a new approach for the identification and relative quantification of metabolites by means of dual isotope labeling and liquid chromatography-mass spectrometry. *Anal. Chem.* **2009,** 81, 9257-9266.

3. Hirai, M. Y.; Sugiyama, K.; Sawada, Y.; Tohge, T.; Obayashi, T.; Suzuki, A.; Araki, R.; Sakurai, N.; Suzuki, H.; Aoki, K.; Goda, H.; Nishizawa, O. I.; Shibata, D.; Saito, K., Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Nat. Acad. Sci.* **2007,** 104, (15), 6478-6483.

4. Schilmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E., Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Nat. Acad. Sci.* **2009,** 106, (26), 10865-10870.

5. Rischer, H.; Oresic, M.; Seppanen-Laasko, T.; Katajamaa, M.; Lammertyn, F.; Ardiles-Diaz, W.; Montagu, M. C. E. V.; Inze, D.; Oksman-Caldenty, K.-M.; Goossens, A., Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *Proc. Nat. Acad. Sci.* **2006,** 103, 5614-5619.

6. Greef, J. v. d.; Martin, S.; Johasz, P.; Adourian, A.; Plasterer, T.; Verheij, E. R.; McBurney, R. N., The art and practice of systems biology in medicine: mapping patterns of relationships. *J. Proteome Res.* **2007,** 6, 1540-1559.

7. Carrari, F.; Baxter, C.; Usadel, B.; Urbanczyk-Wochniak, E.; Zanor, M.-I.; Nunes-Nesi, A.; Nikiforova, V.; Centero, D.; Ratzka, A.; Pauly, M.; Sweetlove, L. J.; Fernie, A. R., Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Phys.* **2006,** 142, 1380-1396.

8. Sauveplane, V.; Kandel, S.; Kastner, P.-E.; Ehlting, J.; Compagnon, V.; Werck-Reichhart, D.; Pinot, F., *Arabidopsis thaliana* CYP77A4 is the first cytochrome P450 able to catalyze the epoxidation of free fatty acids in plants. *FEBS Journal* **2009,** 276, 719-735.

9. Penn, K.; Jenkins, C.; Nett, M.; Udwary, D. W.; Gontang, E. A.; McGlinchey, R. P.; Foster, B.; Lapidus, A.; Podell, S.; Allen, E. E.; Moore, B. S.; Jensen, P. R., Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *The ISME Journal* **2009,** 3, (10), 1193-1203.

10. Lane, A. L.; Nyadong, L.; Galhena, A. S.; Shearer, T. L.; Stout, E. P.; Parry, R. M.; Kwasnik, M.; Wang, M. D.; Hay, M. E.; Fernandez, F. M.; Kubanek, J., Desorption electrospray ionization mass spectrometry reveals surface-mediated antifungal chemical defense of a tropical seaweed. *Proc. Nat. Acad. Sci.* **2009,** 106, (18), 7314-7319.

11. Humphry, M.; Bednarek, P.; Kemmerling, B.; Koh, S.; Stein, M.; Goebl, U.; Stuber, K.; Pislewska-Bednarek, M.; Loraine, A.; Schulze-Lefert, P.; Somerville, S.; Panstruga, R., A regulon conserved in monocot and dicot plants defines a functional module in antifungal plant immunity. *Proc. Nat. Acad. Sci.* **2010,** 107, (50), 21896-21901.

12. Bednarek, P.; Schneider, B.; Svatos, A.; Oldham, N. J.; Hahlbrock, K., Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism of indolic and phenylpropanoid secondary metabolism in Arabidopsis roots. *Plant Phys.* **2005,** 138, 1058-1070.

13. Gershenzon, J.; Mabry, T. J., Secondary metabolites and the higher classification of angiosperms. *Nord. J. Bot.* **1983,** 3, 5-34.

14. Maloney, V., Plant Metabolomics. *BioTeach Journal* **2004,** 2, 92-99.
15. Zvi, M. M. B.; Negre-Zakharov, F.; Masci, T.; Ovadis, M.; Shklarman, E.; Ben-Meir, H.; Tzfira, T.; Dudareva, N.; Vanstein, A., Interlinking showy traits: co-engineering of scent and colour biosynthesis in flowers. *Plant Biotechnol. J.* **2008,** 6, 403-415.

16. Robertson, D. G., Metabonomics in toxicology: A review. *Toxicol. Sci.* **2005,** 85, (2), 809-822.

17. Tian, C.; Chikayama, E.; Tsuboi, Y.; Kuromori, T.; Shinozaki, K.; Kikuchi, J.; Hirayama, T., Top-down phenomics of *Arabidopsis thaliana*: metabolic profiling by one- and two-dimensional nuclear magnetic resonance spectroscopy and transcriptome analysis of albino mutants. *J. Biol. Chem.* **2007,** 282, (25), 18532-18541.

18. Griffin, J. L.; Scott, J.; Nicholson, J. K., The influence of pharmacogenetics on fatty liver disease in the wistar and kyoto rats: a combined transcriptomic and metabonomic study. *J. Proteome Res.* **2007,** 6, 54-61.

19. Widarto, H. T.; Meijden, E. V. D.; Lefeber, A. W. M.; Erkelens, C., Metabolomic differentiation of *Brassica rapa* following herbivory by different insect instars using two-dimensional nuclear magnetic resonance spectroscopy. *J. Chem. Ecol.* **2006,** 32, 2417-2428.

20. Hendrawati, O.; Yao, Q.; Kim, H. K.; Linthorst, H. J. M.; Erkelens, C.; Lefeber, A. W. M.; Choi, Y. H.; Verpoorte, R., Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Sci.* **2006,** 170, 1118-1124.

21. Rao, S. R.; Ford, K. L.; Cassin, A. M.; Roessner, U.; Patterson, J. H.; Bacic, A., Proteomic and metabolic profiling of rice suspension culture cells as a model to study abscisic acid signaling response pathways in plants. *J. Proteome Res.* **2010,** 9, 6623-6634.

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

23. Bansal, S.; DeStefano, A., Key elements of bioanalytical method validation for small molecules. *AAPS J.* **2007,** 9, (1), E109-E114.

24. Fiehn, O.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Lee, D. Y.; Lu, Y.; Moon, S.; Nikolau, B., Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J.* **2008,** 53, 691-704.

25. Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006,** 2, (4), 171-196.

26. Connor, S. C.; Gray, R. A.; Hodson, M. P.; Clayton, N. M.; Haselden, J. N.; Chessell, I. P.; Bountra, C., An NMR-based metabolic profiling study of inflammatory pain using the rat FCA model. *Metabolomics* **2007,** 3, (1), 29-39.

27. Bottcher, C.; Roepenack-Lahaye, E. v.; Schmidt, J.; Schmotz, C.; Neumann, S.; Schell, D.; Clemens, S., Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant Phys.* **2008,** 147, 2107-2120.

28. Bing, F. C., The history of the word "metabolism". *J. Hist. Med. All. Sci.* **1971,** 26, 158-180.

29. Dakin, H. D., Studies on the intermediary metabolism of amino-acids. *J. Biol. Chem.* **1913,** 14, (4), 321-333.

30. Kornberg, H., Krebs and his trinity of cycles. *Nat. Rev. Mol. Cell Bio.* **2000,** 1, 225-228.

31. Dejongh, D. C.; Radford, T.; Hribar, J. D.; Hanessia.S; Bieber, M.; Dawson, G.; Sweeley, C. C., Analysis of trimethylsilyl derivatives of carbohydrates by gas chromatography and mass spectrometry. *J. Am. Chem. Soc.* **1969,** 91, (7), 1728-1740.

32. Brown, F. F.; Campbell, I. D.; Kuchel, P. W.; Rabenstein, D. L., Human erythrocyte metabolism studies by H-1 spin echo NMR. *FEBS Lett.* **1977,** 82, (1), 12-16.

33. Rabenstein, D. L., H-1 NMR methods for the noninvasive study of metabolism and other processes involving small molecules in intact erythrocytes. *J. Biochem. Bioph. Meth.* **1984,** 9, (4), 277-306.

34. Gatherer, D., So what do we mean when we say that systems biology is holistic? *BMC Syst. Biol.* **2010,** 4, 22.

35. Allwood, J. W.; Ellis, D. I.; Goodacre, R., Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiol. Plantarum* **2008,** 132, 117-135.

36. Mattoli, L.; Cangi, F.; Maidecchi, A.; Ghiara, C.; Ragazzi, E.; Tubaro, M.; Stella, L.; Tisato, F.; Traldi, P., Metabolomic fingerprinting of plant extracts. *J. Mass Spectrom.* **2006,** 41, 1534-1545.

37. Brown, M.; Dunn, W. B.; Ellis, D. I.; Goodacre, R.; Handl, J.; Knowles, J. D.; O'Hagan, S.; Spasic, I.; Kell, D. B., A metabolome pipeline: from concept to data to knowledge. *Metabolomics* **2005,** 1, (1), 39-51.

38. Rouhi, A. M., Rediscovering natural products. *Chem. Eng. News* **2003,** 81, (41), 77-91.

39. Molyneux; Schieberle, Compound identification: a journal of agricultural and food chemistry perspective. *J. Agric. Food Chem.* **2007,** 55, (12), 4625-4629.

40. Holmes, E.; Antti, H., Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. *Analytst* **2002,** 127, 1549-1557.

41. Soga, T.; Ohashi, Y.; Ueno, Y.; Naraoka, H.; Tomita, M.; Nishioka, T., Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* **2003,** 2, (5), 488-494.

42. Pauli, G. F.; Jaki, B. U.; Lankin, D. C., A routine experimental protocol for qHNMR illustrated with taxol. *J. Nat. Prod.* **2007,** 70, (4), 589-595.

43. Moing, A.; Maucourt, M.; Renaud, C.; Gaudillere, M.; Brouquisse, R.; Lebouteiller, B.; Gousset-Dupont, A.; Vidal, J.; Granot, D.; Denoyes-Rothan, B.; Lerceteau-Kohler, E.; Rolin, D., Quantitative metabolic profiling by 1-dimensional H-1 NMR analyses: application to plant genetics and functional genomics. *Funct. Plant Biol.* **2004,** 31, (9), 889-902.

44. Preinerstorfer, B.; Schiesel, S.; Lammerhofer, M.; Linder, W., Metabolic profiling of intracellular metabolites in fermentation broths from β-lactam antibiotics production by liquid chromatography-tandem mass spectrometry methods. *J. Chromatogr. A* **2010,** 1217, 312-328.

45. Albrechtova, J. T. P.; Vervliet-Scheebaum, M.; Normann, J.; Veit, J.; Wagner, E., Metabolic control of transcriptional-translational control loops (TTCL) by circadian oscillations in the redox- and phosphorylation state of cells. *Biol. Rhythm Res.* **2006,** 37, (4), 381-389.

46. Touma, C.; Palme, R., Measuring fecal glucocorticoid metabolites in mammals and birds: The importance of validation. *Ann. NY Acad. Sci.* **2005,** 1046, 54-74.

47. Bollard, M. E.; Holmes, E.; Lindon, J. C.; Mitchell, S. C.; Branstetter, D.; Zhang, W.; Nicholson, J. K., Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution H-1 NMR spectroscopy of urine and pattern recognition. *Anal. Biochem.* **2001,** 295, (2), 194-202.

48. Choi, H. K.; Choi, Y. H.; Verberne, M.; Lefeber, A. W. M.; Erkelens, C.; Verpoorte, R., Metabolic fingerprinting of wild type and transgenic tobacco plants by H-1 NMR and multivariate analysis technique. *Phytochemistry* **2004,** 65, (7), 857-864.

49. Maniara, G.; Rajamoorthi, K.; Rajan, S.; Stockton, G. W., Method performance and validation for quantitative analysis by H-1 and P-31 NMR spectroscopy. Applications to analytical standards and agricultural chemicals. *Anal. Chem.* **1998,** 70, (23), 4921-4928.

50. Dumas, M. E.; Maibaum, E. C.; Teague, C.; Ueshima, H.; Zhou, B. F.; Lindon, J. C.; Nicholson, J. K.; Stamler, J.; Elliott, P.; Chan, Q.; Holmes, E., Assessment of analytical reproducibility of H-1 NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. *Anal. Chem.* **2006,** 78, (7), 2199-2208.

51. Govindaraju, V.; Young, K.; Maudsley, A. A., Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **2000,** 13, (3), 129-153.

52. Rischer, H.; Oksman-Caldentey, K. M., Unintended effects in genetically modified crops: revealed by metabolomics? *Trends Biotechnol.* **2006,** 24, (3), 102-104.

53. Guo, B.; Huang, Z.; Wang, M.; Wang, X.; Zhang, Y.; Chen, B.; Li, Y.; Yan, H.; Yao, S., Simultaneous direct analysis of benzimidazole fungicides and relevant metabolites in agricultural products based on multifunction dispersive solid-phase extraction and liquid chromatography-mass spectrometry. *J. Chromatogr. A* **2010,** 1217, 4796-4807.

54. Rochfort, S.; Caridi, D.; Stinton, M.; Trenerry, V. C.; Jones, R., The isolation and purification of glucoraphanin from broccoli seeds by solid phase extraction and preparative high performance liquid chromatography. *J. Chromatogr. A* **2006,** 1120, 205-210.

55. Jagemeier, J.; Schneider, B.; Oldham, N. J.; Hahlbrock, K., Accumulation of soluble and wall-bound indolic metabolites in *Arabidopsis thaliana* leaves infected with virulent or avirulent *Pseudomonas syringae* pathovar tomato strains. *Proc. Nat. Acad. Sci.* **2001,** 98, (2), 753-758.

56. Keun, H. C.; Ebbels, T. M. D.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Analytical reproducibility in H-1 NMR-based metabonomic urinalysis. *Chem. Res. Toxicol.* **2002,** 15, (11), 1380-1386.

57. Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L., Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000,** 18, (11), 1157-1161.

58. Dobbin, K.; Simon, R., Sample size determination in microarray experiments for class comparison and prognostic classification *Biostatistics* **2005,** 6, (2), 348-348.

59. Dobbin, K. K.; Simon, R. M., Sample size planning for developing classifiers using high-dimensional DNA microarray data. *Biostatistics* **2007,** 8, (1), 101-117.

60. Hendrawati, O.; Yao, Q. Q.; Kim, H. K.; Linthorst, H. J. M.; Erkelens, C.; Lefeber, A. W. M.; Choi, Y. H.; Verpoorte, R., Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Sci.* **2006,** 170, (6), 1118-1124.

61. Ward, J. L.; Harris, C.; Lewis, J.; Beale, M. H., Assessment of H-1 NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **2003,** 62, (6), 949-957.

62. Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G., Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **2005,** 56, (410), 255-265.

63. Mohamed, R.; Varesio, E.; Ivosev, G.; Burton, L.; Bonner, R.; Hopfgartner, G., Comprehensive analytical strategy for biomarker identification based on liquid chromatography coupled to mass spectremetry and new candidate confirmation tools. *Anal. Chem.* **2009,** 81, (18), 7677–7694.

64. Rogachev, I.; Aharoni, A., UPLC-MS-based metabolite analysis in tomato. http://www.weizmann.ac.il/plants/aharoni/PDFs/b3.pdf, (accessed Dec 31, 2012).

65. Saude, E. J.; Slupsky, C. M.; Sykes, B. D., Optimization of NMR analysis of biological fluids for quantitative accuracy. *Metabolomics* **2006,** 2, (3), 113-123.

66. Arabinar, N.; Ott, K.-H.; Roongta, V.; Mueller, L., Metabolomic analysis using optimized NMR and statistical methods. *Anal. Biochem.* **2006,** 355, 62-70.

67. Averna, T. A.; Kline, E. E.; Smith, A. Y.; Sillerud, L. O., A decrease in H-1 nuclear magnetic resonance spectroscopically determined citrate in human seminal fluid accompanies the development of prostate adenocarcinoma. *J. Urology* **2005,** 173, (2), 433-438.

68. Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K., Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: Application in metabonomic toxicology studies. *Anal. Chem.* **2006,** 78, (2), 363-371.

69. Rundlof, T.; Mathiasson, M.; Bekiroglu, S.; Hakkarainen, B.; Bowden, T.; Arvidsson, T., Survey and qualification of internal standards for quantification by H-1 NMR spectroscopy. *J. Pharmaceut. Biomed.*52, (5), 645-651.

70. Holscher, D.; Brand, S.; Wenzler, M.; Schneider, B., NMR-based metabolic profiling of Anigozanthos floral nectar. *J. Nat. Prod.* **2008,** 71, (2), 251-257.

71. Swanson, M. G.; Zektzer, A. S.; Tabatabai, Z. L.; Simko, J.; Jarso, S.; Keshari, K. R.; Schmitt, L.; Carroll, P. R.; Shinohara, K.; Vigneron, D. B.; Kurhanewicz, J., Quantitative analysis of prostate metabolites using H-1 HR-MAS spectroscopy. *Magn. Reson. Med.* **2006,** 55, (6), 1257-1264.

72. Larive, C. K.; Jayawickrama, D.; Orfi, L., Quantitative analysis of peptides with NMR spectroscopy. *Appl. Spectrosc.* **1997,** 51, (10), 1531-1536.

73. Hanhineva, K.; Rogachev, I.; Kokko, H.; Mintz-Oron, S.; Venger, I.; Karenlampi, S.; Aharoni, A., Non-targeted analysis of spatial metabolite composition in strawberry (Fragaria x ananassa) flowers. *Phytochemistry* **2008,** 69, 2463-2481.

74. Ryan, D.; Robards, K., Analytical chemistry considerations in plant metabolomics. *Sep. Purif. Rev.* **2006,** 35, 319-356.

75. Sekiyama, Y.; Chikayama, E.; Kikuchi, J., Profiling polar and semipolar plant metabolites throughout extraction processes using a combined solution-state and high-resolution magic angle spinning NMR approach. *Anal. Chem.* **2010,** 82, 1643-1652.

76. Lyng, H.; Sitter, B.; Bathen, T. F.; Jensen, L. R.; Sundfor, K.; Kristensen, G. B.; Gribbestad, I. S., Metabolic mapping by use of high-resolution magic angle spinning ^{1}H MR spectroscopy for assessment of apoptosis in cervical carcinomas. *BMC Cancer* **2007,** 7, 11.

77. Le Belle, J. E.; Harris, N. G.; Williams, S. R.; Bhakoo, K. K., A comparison of cell and tissue extraction techniques using high-resolution ^{1}H-NMR spectroscopy. *NMR Biomed.* **2002,** 15, (1), 37-44.

78. Usenius, J.-P.; Vainio, P.; Hernesniemi, J.; Kauppinen, R. A., Choline-containing compounds in human astrocytomas studied by ^{1}H NMR spectroscopy *in vivo* and *in vitro*. *J. Neurochem.* **1994,** 63, (4), 1538-43.

79. Kaiser, K. A.; Barding, G. A.; Larive, C. K., Metabolic profiling of plants by ^{1}H-NMR: A comparison of metabolite extraction strategies using rosette leaves of the model plant *Arabidopsis thaliana*. *Magn. Reson. Chem.* **2009,** 47, (S1), S147-S156.

80. Svatos, A., Mass spectrometric imaging of small molecules. *Trends in Biotechnology* **2010,** 28, (8), 425-434.

81. Boernsen, K. O.; Gatzek, S.; Imbert, G., Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Anal. Chem.* **2005,** 77, (22), 7255-7264.

82. Lowe, R. G. T.; Allwood, J. W.; Galster, A. M.; Urban, M.; Daudi, A.; Canning, G.; Ward, J. L.; Beale, M. H.; Hammond-Kosack, K. E., A combined ^{1}H nuclear magnetic resonance and electrospray ionization-mass spectrometry analysis to understand the basal metabolism of plant-pathogenic *Fusarium* spp. *Mol. Plant Microbe In.* **2010,** 23, (12), 1605-1618.

83. Colquhoun, I. J., Use of NMR for metabolic profiling in plant systems. *J. Pestic. Sci.* **2007,** 32, 200-212.

84. Rabenstein, D. L.; Keire, D. A., Quantitative chemical analysis by NMR. In *Modern NMR Techniques and Their Applications in Chemistry*, Popov, A. I.; Hallenga, K., Eds. Marcel Dekker, Inc.: New York, NY, 1990; Vol. 11, p 323.

85. Sanders, J. K. M.; Hunter, B. K., *Modern NMR spectroscopy: a guide for chemists*. 2nd ed.; Oxford University Press, Inc.: New York, NY, 1993; p 314.

86. Lambert, J. B.; Shurvell, H. F.; Lightner, D. A.; Cooks, R. G., *Organic Structural Spectroscopy*. Prentice-Hall, Inc.: Upper Saddle River, NJ, 1998; p 568.

87. Keeler, J., *Understanding NMR Spectroscopy*. John Wiley & Sons, Inc.: Hoboken, NJ, 2005; p 459.

88. Gipson, G. T.; Tatsuoka, K. S.; Sweatman, B. C.; Connor, S. C., Weighted least-squares deconvolution method for discovery of group differences between complex biofluid 1H NMR spectra. *J. Magn. Reson.* **2006,** 183, 269-277.

89. Xu, Q.; Sachs, J. R.; Wang, T.-C.; Schaefer, W. H., Quantification and identification of components in solution mixtures from 1D proton NMR spectra using singular value decomposition. *Anal. Chem.* **2006,** 78, 7175-7185.

90. Gromova, M.; Roby, C., Toward *Arabidopsis thaliana* hydrophilic metabolome: assessment of extraction methods and quantitative ^{1}H NMR. *Physiol. Plantarum* **2010,** 140, 111-127.

91. Chmurny, G. N.; Hoult, D. I., The ancient and honourable art of shimming. *Concept. Magnetic Res.* **1990,** 2, 131-149.

92. Weiger, M.; Speck, T.; Fey, M., Gradient shimming with spectrum optimisation. *J. Magn. Reson.* **2006,** 182, 38-48.

93. Rabenstein, D. L., Sensitivity enhancement by signal averaging in pulsed/Fourier transform NMR spectroscopy. *J. Chem. Ed.* **1984,** 61, 909-913.

94. Dai, H.; Xiao, C.; Liu, H.; Tang, H., Combined NMR and LC-MS analysis reveals the metabonomic changes in *Salvia miltiorrhiza* Bunge induced by water depletion. *J. Proteome Res.* **2010,** 9, 1460-1475.

95. Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L., Method for determining molar concentrations of metabolites in complex solutions from two-dimensional ^{1}H-^{13}C NMR Spectra. *Anal. Chem.* **2007,** 79, 9385-9390.

96. Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R., High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal. Biochem.* **2008,** 372, 204-212.

97. Lin, C. Y.; Wu, H.; Tjeerdema, R. S.; Viant, M. R., Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics* **2007,** 3, (1), 55-67.

98. Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protoc.* **2007,** 2, (11), 2692-2703.

99. Barjat, H.; Morris, G. A.; Smart, S.; Swanson, A. G.; Williams, S. C. R., High-Resolution Diffusion-Ordered 2D Spectroscopy (HR-DOSY)-A New Tool for the Analysis of Complex Mixtures. *J. Magn. Reson. B* **1995,** 108, 170-172.

100. Nilsson, M.; Duarte, I. F.; Almeida, C.; Delgadillo, I.; Goodfellow, B. J.; Gil, A. M.; Morris, G. A., High-Resolution NMR and Diffusion-Ordered Spectroscopy of Port Wine. *J. Agric. Food Chem.* **2004,** 52, 3736-3743.

101. Rodrigues, E. D.; Silva, D. B. d.; Olivera, D. C. R. d.; Silva, G. V. J. d., DOSY NMR applied to analysis of flavonoid glycosides from *Bidens sulphurea*. *Magn. Reson. Chem.* **2009,** 47, (12), 1095-1100.

102. Tang, H. R.; Wang, Y. L.; Nicholson, J. K.; Lindon, J. C., Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Anal. Biochem.* **2004**, 325, (2), 260-272.

103. Price, K. E.; Lunte, C. E.; Larive, C. K., Development of tissue-targeted metabonomics. Part 1. Analytical considerations. *J. Pharmaceut. Biomed.* **2008**, 46, 737-747.

104. Bertram, H. C.; Malmendal, A.; Petersen, B. O.; Madsen, J. C.; Pedersen, H.; Nielsen, N. C.; Hoppe, C.; Molgaard, C.; Michaelsen, K. F.; Duus, J. O., Effect of magnetic field strength on NMR-Based metabonomic human urine data comparative study of 250, 400, 500, and 800 MHz. *Anal. Chem.* **2007,** 79, (18), 7110-7115.

105. Malz, F.; Jancke, H., Validation of quantitative NMR. *J. Pharmaceut. Biomed.* **2005,** 38, (5), 813-823.

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

107. Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L., Metabolite profiling for plant functional genomics. *Anal. Chem.* **2000,** 72, 3573-3580.

108. Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M., HPLC-MS-based methods for the study of metabonomics. *J. Chromatogr. B* **2005,** 817, 67-76.

109. Grata, E.; Boccard, J.; Glauser, G.; Carrupt, P.-A.; Farmer, E. E.; Wolfender, J.-L.; Rudaz, S., Development of a two-step screening ESI-TOF-MS method for rapid determination of significant stress-induced metabolome modifications in plant leaf extracts: The wound response in *Arabidopsis thaliana* as a case study. *J. Sep. Sci.* **2007,** 30, 2268-2278.

110. Umehara, M.; Hanada, A.; Yoshida, S.; Akiyama, K.; Arite, T.; Takeda-Kamiya, N.; Magome, H.; Kamiya, Y.; Shirasu, K.; Yoneyama, K.; Kyozuka, J.; Yamaguchi, S., Inhibition of shoot branching by new terpenoid plant hormones. *Nature* **2008,** 455, 195-201.

111. Hanhineva, K.; Kokko, H.; Siljanen, H.; Rogachev, I.; Aharoni, A.; Karenlampi, S. O., Stilbene synthase gene transfer caused alterations in the phenylpropanoid metabolism of transgenic strawberry (Fragaria x ananassa). *J. Exp. Bot.* **2009,** 60, (7), 2093-2106.

112. Issaq, H. J.; Van, Q. N.; Waybright, T. J.; Muschik, G. M.; Veenstra, T. D., Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* **2009,** 32, 2183-2199.

113. Spacil, Z.; Novakova, L.; Solich, P., Analysis of phenolic compounds by high performance liquid chromatography and ultra performance liquid chromatography. *Talanta* **2008,** 76, 189-199.

114. Harborne, J. B., *Phytochemical Methods*. John Wiley & Sons: New York, NY, 1973; p 278.

115. Cuyckens, F.; Claeys, M., Mass spectrometry in the structural analysis of flavonoids. *J. Mass Spectrom.* **2004,** 39, 1-15.

116. Bottcher, C.; Roepenack-Lahaye, E. v.; Schmidt, J.; Clemens, S.; Scheel, D., Analysis of phenolic choline esters from seeds of *Arabidopsis thaliana* and Brassica napus by capillary liquid chromatography/electrospray-tandem mass spectrometry. *J. Mass Spectrom.* **2008,** 44, 466-476.

117. Sawada, Y.; Akiyama, K.; Sakata, A.; Kuwahara, A.; Otsuki, H.; Sakurai, T.; Saito, K.; Hirai, M. Y., Widley targeted metabolomics nased on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol.* **2009,** 50, (1), 37-47.

118. DeRisi, J. L.; Iyer, V. R.; Brown, P. O., Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **1997,** 278, 680-686.

119. Marton, M. J.; DeRisi, J. L.; Bennett, H. A.; Iyer, V. R.; Meyer, M. R.; Roberts, C. J.; Stoughton, R.; Burchard, J.; Slade, D.; Dai, H. Y.; Bassett, D. E.; Hartwell, L. H.; Brown, P. O.; Friend, S. H., Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **1998,** 4, (11), 1293-1301.

120. Pan, W., A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **2002,** 18, (4), 546-554.

121. Craig, A.; Cloareo, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C., Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal. Chem.* **2006,** 78, (7), 2262-2267.

122. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Suizdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification. *Anal. Chem.* **2006,** 78, (3), 779-787.

123. Suhre, K.; Meisinger, C.; Doring, A.; Altmaier, E.; Belcredi, P.; Gieger, C.; Chang, D.; Milburn, M. V.; Gall, W. E.; Weinberger, K. M.; Mewes, H.-W.; Angelis, M. H. d.; Wichmann, H.-E.; Kronenberg, F.; Adamski, J.; Illig, T., Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PloS One* **2011,** 5, (11), e13953.

124. Moskau, D., Application of real time digital filters in NMR spectroscopy. *Concept. Magnetic Res.* **2002,** 15, (2), 164-176.

125. Xu, Q. W.; Sachs, J. R.; Wang, T. C.; Schaefer, W. H., Quantification and identification of components in solution mixtures from 1D proton NMR spectra using singular value decomposition. *Anal. Chem.* **2006,** 78, (20), 7175-7185.

126. Derrick, T. S.; McCord, E. F.; Larive, C. K., Analysis of protein/ligand interactions with NMR diffusion measurements: The importance of eliminating the protein background. *J. Magn. Reson.* **2002,** 155, (2), 217-225.

127. de Graaf, R. A.; Behar, K. L., Quantitative H-1 NMR spectroscopy of blood plasma metabolites. *Anal. Chem.* **2003,** 75, (9), 2100-2104.

128. Lindon, J. C.; Ferrige, A. G., Digitization and data-processing in Fourier-transform NMR. *Prog. Nucl. Mag. Res. Sp.* **1980,** 14, 27-66.

129. Griffiths, L., Towards the automatic analysis of H-1 NMR spectra: Part 2. Accurate integrals and stoichiometry. *Magn. Reson.Chem.* **2001,** 39, (4), 194-202.

130. Forshed, J.; Torgrip, R. J. O.; Aberg, K. M.; Karlberg, B.; Lindberg, J.; Jacobsson, S. P., A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *J. Pharmaceut. Biomed.* **2005,** 38, (5), 824-832.

131. Gipson, G. T.; Tatsuoka, K. S.; Sweatman, B. C.; Connor, S. C., Weighted least-squares deconvolution method for discovery of group differences between complex biofluid H-1 NMR spectra. *J. Magn. Reson.* **2006,** 183, (2), 269-277.

132. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Metabolite projection analysis for fast identification of metabolites in metabonomics. Application in an amiodarone study. *Anal. Chem.* **2006,** 78, (11), 3551-3561.

133. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M., Targeted profiling: quantitative analysis of H-1 NMR metabolomics data. *Anal. Chem.* **2006,** 78, (13), 4430-4442.

134. Abshear, T.; Banik, G. M.; D'Souza, M. L.; Nedwed, K.; Peng, C., A model validation and consensus building environment. *SAR QSAR Environ. Res.* **2006,** 17, (3), 311-321.

135. Saito, T.; Nakaie, S.; Kinoshita, M.; Ihara, T.; Kinugasa, S.; Nomura, A.; Maeda, T., Practical guide for accurate quantitative solution state NMR analysis. *Metrologia* **2004,** 41, (3), 213-218.

136. Mounet, F.; Lemaire-Chamley, M.; Maucourt, M.; Cabasson, C.; Giraudel, J.-L.; Deborde, C.; Lessire, R.; Gallusci, P.; Bertrand, A.; Gaudillère, M.; Rothan, C.; Rolin, D.; Moing, A., Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics* **2007,** 3, (3), 273-288.

137. Martin, Y. L., A global approach to accurate and automatic quantitative-analysis of NMR spectra by complex least-squares curve-fitting. *J. Magn. Reson. A* **1994,** 111, (1), 1-10.

138. Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J., Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **2004,** 20, (15), 2447-2454.

139. Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2005,** 56, 219-243.

140. Forshed, J.; Schuppe-Koistinen, I.; Jacobsson, S. P., Peak alignment of NMR signals by means of a genetic algorithm. *Anal. Chim. Acta* **2003,** 487, (2), 189-199.

141. Defernez, M.; Colquhoun, I. J., Factors affecting the robustness of metabolite fingerprinting using H-1 NMR spectra. *Phytochemistry* **2003,** 62, (6), 1009-1017.

142. Martin, F. P. J.; Dumas, M. E.; Wang, Y. L.; Legido-Quigley, C.; Yap, I. K. S.; Tang, H. R.; Zirah, S.; Murphy, G. M.; Cloarec, O.; Lindon, J. C.; Sprenger, N.; Fay, L. B.; Kochhar, S.; van Bladeren, P.; Holmes, E.; Nicholson, J. K., A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Mol. Syst. Biol.* **2007,** 3, 112.

143. Roncalli, J.; Smih, F.; Desmoulin, F.; Dumonteil, N.; Harmancey, R.; Hennig, S.; Perez, L.; Pathak, A.; Galinier, M.; Massabuau, P.; Malet-Martino, M.; Senard, J. M.; Rouet, P., NMR and cDNA array analysis prior to heart failure reveals an increase of unsaturated lipids, a glutamine/glutamate ratio decrease and a specific transcriptome adaptation in obese rat heart. *J. Mol. Cell. Cardiol.* **2006,** 42, 526-539.

144. Fan, T. W. M.; Higashi, R. M.; Lane, A. N., Integrating metabolomics and transcriptomics for probing Se anticancer mechanisms. *Drug Metab. Rev.* **2006,** 38, (4), 707-732.

145. Thellin, O.; Zorzi, W.; Lakaye, B.; De Borman, B.; Coumans, B.; Hennen, G.; Grisar, T.; Igout, A.; Heinen, E., Housekeeping genes as internal standards: use and limits. *J. Biotechnol.* **1999,** 75, (2-3), 291-295.

146. Gartland, K. P. R.; Bonner, F. W.; Nicholson, J. K., Investigations into the biochemical effects of region-specific nephrotoxins. *Mol. Pharmacol.* **1989,** 35, (2), 242-250.

147. van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,** 7, 142.

148. Parsons, H. M.; Ludwig, C.; Gunther, U. L.; Viant, M. R., Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics* **2007,** 8, 234.

149. Delitala, M. A., Metabolomics: a five-phase process for turning metabolomic data into accessible knowledge. http://ejournal.vudat.msu.edu/index.php/mmg445/article/viewFile/156/114, (accessed Dec 31, 2012).

150. van der Greef, J.; Smilde, A. K., Symbiosis of chemometrics and metabolomics: past, present, and future. *J. Chemometr.* **2005,** 19, (5-7), 376-386.

151. Malinowski, E. R., Theory of error applied to factor loadings resulting from combination target factor analaysis. *Anal. Chim. Acta* **1980,** 122, 327-330.

152. Roscoe, B. A.; Hopke, P. K., Error estimates for factor loadings and scores obtained with target transformation factor analysis. *Anal. Chim. Acta* **1981,** 132, 89-97.

153. Harlow, L. L., *The essence of multivariate thinking: basic themes and methods*. Lawrence Erlbaum Assoc.: Mahwah, NJ, **2005**; p 264.

154. Boccard, J.; Grata, E.; Thiocone, A.; Gauvrit, J.-Y.; Lanteri, P.; Carrupt, P.-A.; Wolfender, J.-L.; Rudaz, S., Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding. *Chemom. Intell. Lab. Syst.* **2007,** 86, 189-197.

155. Winter, W.; Deubner, R.; Holzgrabe, U., Multivariate analysis of nuclear magnetic resonance data--characterization of critical drug substance quality of gentamicin sulfate. *J. Pharmaceut. Biomed.* **2005,** 38, (5), 833-839.

156. Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G., Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.* **2006,** 78, 743-752.

157. Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S.; Raftery, D., Chemoselective ^{15}N tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. *Anal. Chem.* **2009,** 81, 4882-4888.

158. Hegeman, A. D., Plant metabolomics-meeting the analytical challenges of comprehensive metabolite analysis. *Brief. Funct. Genomics* **2010,** 9, (2), 139-148.

159. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K. F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Kirakawa, M., From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **2006,** 34, (34), 354-357.

160. Paley, S. M.; Karp, P. D., The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res.* **2006,** 34, (13), 3771-3778.

161. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; Souza, A. D.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009,** 37, D603-D610.

162. Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmuller, E.; Dormann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.; Fernie, A. R.; Steinhauser, D., GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* **2005,** 21, 1635-1638.

163. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** 27, 747-751.

164. Cui, Q.; Lewis, I. A.; Hegeman, A.; Anderson, M. E.; Li, J.; Schulte, C. F.; Wesfler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the Madison metabolomics consortium database. *Nat. Biotechnol.* **2008,** 26, 162-164.

165. Allen, M. P., *Understanding Regression Analysis*. Plenum Press: New York, NY, 1997; p 216.

166. Skoog, D. A.; West, D. M.; Holler, F. J., *Fundamentals of Analytical Chemistry*. 7th ed.; Saunders College Pub.: Ft. Worth, TX, 1996; p 870.

167. Rischer, H.; Oresic, M.; Seppanen-Laakso, T.; Katajamaa, M.; Lammertyn, F.; Ardiles-Diaz, W.; Van Montagu, M. C. E.; Inze, D.; Oksman-Caldenty, K.-M.; Goossens, A., Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *P. Natl. Acad. Sci. USA* **2006,** 103, (14), 5614-5619.

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

169. Osuna, D.; Usadel, B.; Morcuende, R.; Gibon, Y.; Blasing, O. E.; Hohne, M.; Gunter, M.; Kamlage, B.; Trethewey, R. N.; Scheible, W.-R.; Stitt, M., Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J.* **2007,** 49, 463-491.

170. Sakurai, N.; Shibata, D., KaPPA-View for integrating quantitative transcriptomic and metabolomic data on plant metabolic pathway maps. *J. Pestic. Sci.* **2006,** 31, 293-295.

171. Ipera, T. I.; Tropsha, A.; Faulon, J.-L.; Rintoul, M. D., Systems chemical biology. *Nat. Chem. Biol.* 3, (8), 447-450.

172. Mustroph, A.; Lee, S. C.; Oosumi, T.; Zanetti, M. E.; Yang, H. J.; Ma, K.; Yaghoubi-Masihi, A.; Fukao, T.; Bailey-Serres, J., Cross-kingdom comparison of transcriptomic adjustments to low-oxygen stress highlights conserved and plant-specific responses. *Plant Physiol.* **2010,** 152, (3), 1484-1500.

173. Zhang, J.; Zhang, Y.; Du, Y.; Chen, S.; Tang, H., Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res.* **2011,** 10, 1904-1914.

174. Gehlenborg, N.; O'Donoghue, S. I.; Baliga, N. S.; Goesmann, A.; Hibbs, M. A.; Kitano, H.; Kohlbacher, O.; Neuweger, H.; Schneider, R.; Tenenbaum, D.; Gavin, A.-C., Visualization of omics data for systems biology. *Nat. Methods* **2010,** 7, (3), S56-S68.

175. Sato, S.; Masanori, A.; Soga, T.; Nishioka, T.; Tomita, M., Time-resolved metabolomics reveals metabolic modulation in rice foliage. *BMC Syst. Biol.* **2008,** 2, 51.

176. Normile, D., Reinventing Rice to Feed the World. *Science* **2008,** 321, 330-333.

177. Blunden, J.; Arndt, D. S.; Baringer, M. O., State of the Climate in 2010. *Bull. Amer. Meteor. Soc.* **2011,** 92, (6), S1-S266.

178. Oerke, E.-C.; Dehne, H.-W., Safeguarding production - losses in major crops and the role of crop protection. *Crop Prot.* **2004,** 23, 275-285.

179. U.S. Census Bureau, http://www.census.gov/ (accessed Jul 11, 2011).

180. Love, B. E.; Spaner, D. *Improvement of the agricultural sustainability and livelihoods of poor farmers through biotechnology: reality or speculation?*; International Development Research Centre: Alberta, Canada, 2008; p 26.

181. Dillard, C. J.; German, J. B., Phytochemicals: neutraceuticals and human health. *J. Sci. Food Agric.* **2000,** 80, 1744-1756.

182. Cordell, G. A., Plant medicines key to global health. *Chem. Eng. News* **2011,** 89, (26), 52-56.

183. Renz, A.; Jones, T., Phenotyping and metabolite profiling for improved crops. In CePCeB IGERT Seminar, Riverside, CA: September 1, 2009.

184. Royal Horticultural Society, Plants in our daily life. http://apps.rhs.org.uk/schoolgardening/uploads/documents/2010_Plants_in_our_Daily_Life_1176.pdf (accessed Jul 19, 2011).

185. Brown, B., Abiogenesis. In *Evolution: A Historical Perspective*, Baigrie, B., Ed. Greenwood Press: Westport, CT, 2007; pp 129-148

186. Hochachka, P. W., *Living without Oxygen: Closed and Open Systems in Hypoxia Tolerance*. Harvard University Press: Cambridge, MA, 1980; p 181.

187. Ho, Q. T.; Verboven, P.; Verlinden, B. E.; Herremans, E.; Wevers, M.; Carmeliet, J.; Nicolaï, B. M., A three-dimensional multiscale model for gas exchange in fruit. *Plant Physiol.* **2011,** 155, 1158-1168.

188. Storey, K. B., *Functional metabolism: regulation and adaptation*. John Wiley & Sons: Hoboken, NJ, 2004; p 594.

189. Ratcliffe, R. G.; Shachar-Hill, Y., Measuring multiple fluxes through plant metabolic networks. *Plant J.* **2006,** 45, (4), 490-511.

190. Amorim, L. C. A.; Cardeal, Z. D. L., Breath air analysis and its use as a biomarker in biological monitoring of occupational and environmental exposure to chemical agents. *J. Chromatogr. B* **2007,** 853, (1-2), 1-9.

191. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Keun, H. C.; Craig, A.; Pearce, J. T. M.; Bruce, S. J.; Hardy, N.; Sansone, S. A.; Antti, H.; Jonsson, P.; Daykin, C.; Navarange, M.; Beger, R. D.; Verheij, E. R.; Amberg, A.; Baunsgaard, D.; Cantor, G. H.; Lehman-McKeeman, L.; Earll, M.; Wold, S.; Johansson, E.; Haselden, J. N.; Kramer, K.; Thomas, C.; Lindberg, J.; Schuppe-Koistinen, I.; Wilson, I. D.; Reily, M. D.; Robertson, D. G.; Senn, H.; Krotzky, A.; Kochhar, S.; Powell, J.; van der Ouderaa, F.; Plumb, R.; Schaefer, H.; Spraul, M., Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol.* **2005,** 23, (7), 833-838.

Chapter Two

Tissue Extraction for Comparative Metabonomics in *Arabidopsis*

This chapter contains some text and figures from "A comparison of metabolite extraction strategies for ^{1}H NMR-based metabolic profiling using mature leaf tissue from the model plant *Arabidopsis thaliana"* Kayla A. Kaiser, Gregory A. Barding, Jr. and Cynthia K. Larive. Magnetic Resonance in Chemistry. (2009) Volume 47 Issue S1, Pages S147-S156.

KAK prepared 65 matched samples of *Arabidopsis* leaves harvested at growth stage 6.00 to investigate the effect of post-harvest and sample preparation manipulations on the resulting ^{1}H NMR metabolic profile. GAB kindly contributed by performing all methanolic extractions and enabling evaluation of acetonitrile/water extraction/^{1}H NMR metabolic profiling robustness by replicating the work of KAK. This work was supported by National Science Foundation (CHE-0616811) to CKL.

This chapter also contains material from Kim, J. D.; Kaiser, K. A.; Larive, C. K.; and Borkovich, K. "Use of ^{1}H NMR to measure intracellular metabolite levels during growth and asexual sporulation in *Neurospora crassa*" Eukaryotic Cell. Volume 10, Number 6, Pages 820-831, **2011.**

JDK prepared cultures of *Neurospora* under conditions of high or low sucrose for a comparative study between wild-type and a mutant carrying a G-protein α subunit gene deletion. Leading up to this work, a comparison of homogenization techniques (glass mini-bead vs. micropestle) revealed higher extraction efficiency using a micropestle. In a follow-up experiment, prompted by reviewer's helpful suggestions, it was established that the sample preparation protocol produces a stable extract reflecting intracellular endogenous metabolites of *N. crassa* over a time period greatly exceeding the duration of the nominative sample preparation time. This work was supported by Public Health Service grants GM048626 and GM086565 from the National Institute of General Medical Sciences to K.A.B. C.K.L. gratefully acknowledges support from National Science Foundation grant CHE 0848976.

KAK, JDK and GAB gratefully acknowledge support from the National Science Foundation Integrative Graduate Education Research and Training (IGERT) program in Chemical Biology, which supported both studies in this chapter. (DGE-0504249).

2.1 Introduction

Metabolomics is a field which aims for the identification and relative quantification of all metabolites present in an organism.[1, 2] Metabolomic investigations are particularly useful for hypothesis generation. For example, results of a non-targeted investigation of dynamic changes in metabolite levels in rice foliage over a 24-hour period suggested that bottleneck enzymatic steps exist in both carbon and nitrogen metabolism.[3] Regardless of the terminology used, metabolite analysis has become an essential component of systems biology and attention to this field continues to increase.[4-6]

Systematic comparison of sample preparation methods for metabolite analysis in yeast,[7] *Escherichia. coli*,[8] filamentous fungi,[9, 10] chickpea,[11] snowberry,[12] a variety of herbs,[13] earthworm,[14] fish muscle and liver,[15, 16] bovine liver,[17] rat brain,[18] and human serum[19] have been reported. At the time this work was undertaken, limited studies had been performed using the plant *Arabidopsis thaliana*,[20, 21] targeted to detection by mass spectrometry. Because of its importance as a model system, a more extensive evaluation was necessary to determine the advantages and pitfalls of several common sample preparation techniques as they affect the outcomes of ^{1}H NMR methods. In a review article by Fukusaki and Kobayashi,[3] the authors emphasize that "among the many steps [in a plant metabolomic experiment], the extraction procedure is the most important one." With this in mind we were motivated to improve metabolite extraction procedures prior to ^{1}H NMR analyses for *A. thaliana* and the filamentous fungus model, *Neurospora crassa*.

2.1.1 *Arabidopsis thaliana*: A Useful Weed for Systems Biology

Arabidopsis thaliana, often simply called Arabidopsis, is a model system for basic plant research. It has been studied intensively for over 50 years.[22] Its short generation time (5-6 weeks), small size, tolerance to indoor growth conditions, and the ease of large scale mutagenesis make it ideally suited for laboratory work. Its genome is small (120 Mb) relative to most vascular plants and its nucleotide sequence was completed in December 2000.[23] There exists an extensive collection of indexed insertional mutants that are publicly available for experimentation. *Agrobacterium tumefaciens* can be used to transform *A. thaliana*, whereby non-native genes with the left and right borders of *A. tumefaciens* T-DNA sequences can be readily introduced into the nuclear genome, contributing daily to the number and variety of mutants available to researchers. Commercially available DNA microarrays, such as the Affymetrix ATH1 GeneChip, can provide rapid, quantitative analysis regarding the steady state gene expression levels of the over 24,000 recognized genes. Currently about 5,000 genes have been functionally annotated based on experimental evidence, and the number is growing each day. With all that is known about its genome, great interest in Arabidopsis metabolomic analysis has been generated,[24] yet a detailed analysis of the effect of sample preparation steps on the outcome of a ^{1}H NMR metabolic profiling experiment has not been reported for this plant model.

An exhaustive discussion of all plant compounds and their properties is beyond the scope of this work, however basic features contributing to the solubility of relatively abundant compounds will be useful for the discussion that follows.[2, 25, 26] Carbohydrates

including sucrose, glucose, fructose, arabinose, fucose, galactose, mannose, rhamnose and xylose, as well as members of the hexose-phosphate pool are metabolites that may be found in the water-soluble fraction of *Arabidopsis* extracts, as well as the standard amino acids and related amines such as choline, ornithine, and citrulline. Plant lipids and waxes are highly soluble in organic solvents such as chloroform.[17] Plant waxes are structurally similar to fatty acids, except that they are built upon a longer alcohol backbone bound to more than three fatty acid chains. In the ^{1}H NMR spectrum, lipids and waxes give rise to intense broad resonances in the alkane region (1.0 ppm, 1.3 ppm, 1.6 ppm) which obscure sharp resonances of many compounds of interest.[14-16] Therefore, the removal of these resonances became a focus of our method development.

2.1.2 Motivation and Goals for an Extraction Comparison

Factors to optimize throughout method development included sensitivity, robustness, precision, speed, and ease. Consideration was given to providing better information, improving experimental outcomes for either metabolomic discovery-type experiments or to facilitate determination of chemical identity and quantity as in metabolic profiling. Already this study has been useful to a number of researchers exploring new applications[27] and new methods of analysis.[28-31] Using the model plant *A. thaliana*, the effect of the extraction method on the ^{1}H NMR metabolic profile was evaluated with respect to reproducibility and quality. Extract refinement steps were appraised for their ability to improve ^{1}H NMR spectral quality through the removal of lipid resonances. Solid-phase extraction (SPE) sorbents were compared for their

selectivity and recovery of primary and secondary metabolites, which will be discussed further in Chapter 5. Although this study was performed using a single species and genotype at a particular developmental stage, the results can be extended to other systems.

NMR-based metabonomics experiments are implemented practically by sampling (as non-invasively as possible) biological systems, applying minimal sample preparation, and interrogation of the intact mixture by NMR. Overlapping signals in one-dimensional proton spectra are deconvoluted via chemometric multivariable data modeling. Biomarkers are assigned to a particular state of the biological system through correlation and covariance analysis of variables, which represent regions of the NMR spectra. Researchers in plant biology have been slow to implement metabonomic approaches, perhaps because the number of metabolites plants biosynthesize is proposed to surpass the size of the human metabolome (~1,000s of molecules) by two orders of magnitude. Nevertheless, for abundant primary metabolites cellular physiology is largely conserved across all eukaryotic organisms. Therefore common approaches from the literature were applied to *A. thaliana* to explore its potential to inform studies of cellular adjustments to survive low oxygen stress using NMR-based metabonomics (applications highlighted in Chapters 3, 4).

Although many extraction protocols rely on methanol in combination with water and/or chloroform,[32] our laboratory began optimization of a sample preparation method using acetonitrile, as suggested by Gullberg et al.,[20] and applied it successfully for the analysis of oxygen-deprived Arabidopsis seedlings.[33] Extraction with CD_3CN/D_2O

proved useful, enabling metabolomic classification of seedlings exposed to different oxygen environments, [33] which is fully presented in Chapter 3. Changes in the levels of metabolites such as succinate, lactate, alanine, and ethanol were expected to be involved in the plant's hypoxic/anoxic response and those detected from this set were in fact responsible for segregation of ^{1}H NMR spectra of extracts of control, 2 hour, and 9 hour stressed seedlings using a principal component analysis (PCA) model. Interestingly, γ-aminobutyric acid (GABA), glutamate and glutamine were also found to play a role in sample segregation according to duration of hypoxic stress, illustrating the advantage of nonspecific NMR-based metabolomic analysis. Furthermore, the spectral quality of the CD_3CN/D_2O extracts was sufficient to allow peak integration and calculation of quantitative fold changes for all metabolites mentioned above except pyruvate and ethanol. Although this metabolite analysis confirmed the induction of expected stress responses, a more refined extraction strategy was desirable for quantitation of greater numbers of small molecule metabolites to reduce the interference from broad resonances observed in the spectra.

In this chapter, several popular extraction methods are compared.[24, 32, 34-36] In any case where aqueous or part-aqueous solvents were employed, a 100 mM deuterated ammonium acetate buffer was introduced to reduce variance in the chemical shifts of pH-sensitive metabolites, as recommended by Colquhoun.[37] Additionally, 0.5 mM NaN_3 was added as a bacteriostat[38, 39] and 100 μM TMSP-d_4 was utilized as a chemical shift reference. Following perchloric acid treatment, extracts were titrated to pD 5. All other extractions were carried out in a solvent buffered to either pD 5 (low pH) or a pD 7

(neutral pH). The terms low and neutral pD will be utilized throughout the remainder of this chapter. For each sample treatment, multiple replicates (N = 3-5) were examined. For clarity a single NMR spectrum is shown in most figures, which is representative of that treatment. It should be emphasized that the intensities of the spectra presented have not been mass-normalized and are presented for qualitative interpretation only.

Not investigated in this study were the use of high or low temperature extractions,[13, 17, 19, 21, 24, 34] the use of ultrafiltration to remove high molecular weight impurities,[3] the effect of repeated extraction on the same tissue pellet,[24, 40] microwave-assisted extraction,[11] sonication during extraction,[12, 17] duration of extraction/homogenization,[16] duration of centrifugation,[7] and the removal of metal ions to facilitate absolute quantification by NMR,[35] however the reader is directed to the existing literature on these topics.

2.2 Experimental Methods

2.2.1 Reagents and Solutions

All deuterated solvents were purchased from Cambridge Isotope Laboratories (Andover, MA). Murashige and Skoog (MS) salts (1x),[41] Phytagel (0.4%) and sucrose (1%) which constitute solid growth medium for plant cultivation were purchased from Sigma-Aldrich (St. Louis, MO). MS salts contain macronutrients (NH_4NO_3, $CaCl_2$, $MgSO_4$, K_3PO_4, KNO_3) between 170-1650 mg/L. Myo-inositol is present at an intermediate amount, 100 mg/L. Micronutrients (H_3BO_3, $CoCl_2$, $CuSO_4$, $FeSO_4$, $MnSO_4$, KI, $NaMoO_4$, $ZnSO_4$, Na_2EDTA,) are present at lower concentrations, between 0.025-

37.2 mg/L. B vitamins (niacin or B3, pyridoxine or B6, and thiamine or B1) are included in amounts ranging from 0.1-0.5 mg/L. Growth medium was prepared in double distilled water (ddH_2O), adjusted to pH 5.7-5.8 using NaOH and sterilized at 250 ^oC for 25 min in an autoclave.

2.2.2 Sampling and Sample Preparation

2.2.2.1 Equipment

Seeds of Arabidopsis (ecotype Columbia-0, Arabidopsis Biological Resource Center) were germinated on Petri dishes, 10 cm in diameter (Fisherbrand, cat. 08-757-13, 100 x 15 mm). Sterilized growth medium and seeds were handled in a positive flow hood. Plates were placed vertically in a growth chamber (Percival Scientific, Inc., Perry, IA; model CU36L5C8) under long day conditions (16 h at 50 $\mu M\ s^{-1}\ m^{-2}$ light: 8 h dark) at 23 ^oC. Seedlings were transferred to soil and allowed to mature to growth stage 6.0[22] under long day conditions, watering once daily. Rosette leaves were cut with scissors quickly into a porcelain mortar (Coors, Golden, CO) immediately snap-frozen and homogenized by pestle until the tissue resembled a fine powder. Liquid nitrogen was delivered to the mortar by dewar flask, tissue was collected in clear plastic falcon tubes (15 or 50 mL). Tissues were dehydrated on a lyophilizer (Labconco, Kansas City, MO; Cat. No. 77520-00). Tissue dry or wet weight was established for each sample using an analytical balance (Mettler AE 260 Deltarange, Hayward, CA). Extraction was carried out in plastic microcentrifuge tubes (Eppendorf, Hauppauge, NY) using a micropestle (Eppendorf) to grind tissues under varying extraction conditions. Extracts were

concentrated by speedvacuum (Savant SC110, RVT 400, GP 110) and resolubilized by vortex mixing (Thermolyne Maxi Mix Plus M63215, 60 Hz, Asheville, NC). When necessary, tissues were stored at -86^{o}C in an ULT freezer (Thermoforma, model 726, Asheville, NC). Extracts were clarified by centrifugation (Eppendorf 5415R) prior to ^{1}H NMR analysis by a superconducting Bruker BioSpin 14.098 Tesla magnet system fitted with a 5 mm inverse broadband probe and capable of digital quadrature detection. Topspin and AMIX software were used on a Linux host workstation.

2.2.2.2 Plant Cultivation

A. thaliana (ecotype Columbia-0) seeds were surface sterilized by the dry method as follows. No more than 400 seeds were placed in 1.5 mL Eppendorf microcentrifuge tubes, labeled with pencil regarding genotype. In a test tube rack with lids open, microcentrifuge tubes were placed in a dessicator in a fume hood. In a 200 mL beaker, 50 mL of commercial bleach was spiked with 1.5 mL of concentrated HCl. The chlorine gas generated is effective after 3-16 h, depending on the severity of contamination.

Wet sterilization can be used with smaller or larger numbers of seeds. In this pre-treatment, 95% ethanol is introduced (10 mL in 15 mL Falcon tube for 0.5 mL of dry seeds). After 10 min of shaking and draining, a solution containing 1.2% sodium hypochlorite (20% commercial bleach) and 0.1% TWEEN 20 (a polysorbate surfactant) is introduced. This bleach/tween solution (250 mL total volume) must be prepared freshly and replaced at least every 3 d. After 10 min of shaking in bleach/tween, seeds

are rinsed 3 times with ddH_2O and allowed to imbibe in 5-6 mL ddH_2O. Hydrating seeds are allowed to vernalize at 4 ^oC for 3-4 d.

Solid growth medium is prepared by combining 4.3 g MS salts, 8-10 g agar (0.4% Phytagel, Sigma), 10 g (1 % w/v) sucrose in 1 L of distilled H_2O, adjusted to pH 5.7-5.8 with 1 M NaOH (~8 drops). Growth medium is sterilized by autoclave (25 min at 121 ^oC) in a 2 L plastic container, allowed to cool for 30-60 min covered in aluminum foil and placed on a stir plate. Once media is poured into plates, it takes about 1 hour to solidify and can be stored under refrigeration for up to 3 weeks. Depending on the amount poured into each Petri dish (plate) this recipe makes 20-32 plates.

Seeds are plated (approximately 100 seeds per plate) via sterile Pasteur pipet, delivered in a stream of ddH_2O, driven by pressure from a new pipet bulb. The solvent (water) is allowed to reach equilibrium within the growth support while in a horizontal position (~1 hour) after which excess water is removed from the plate by pipet prior to positioning plates vertically in the growth chamber. Seedlings were transferred from Petri dishes to soil containing plots (10 cm x 10 cm x 8 cm) on day 7 after plating and placed in a growth facility under long day conditions at 40% humidity, 22 ^oC, and watered once per day to bulk seed for future experiments or to bulk tissue for method development.

2.2.2.3 Harvesting

Rosette leaves from multiple plots were harvested and combined into a single pool of tissue at growth stage 6.00.[22] Quenching of enzymatic activity was achieved by

flash freezing in liquid N_2 (-77 ^oC). This tissue was homogenized thoroughly by mortar and pestle and stored in a capped Falcon tube sealed with Parafilm until use.

2.2.2.4 Preservation

To obtain tissue dry weight, tissue in Falcon tubes (15 or 50 mL) under liquid N_2 were thermally equilibrated in a bath of dry ice and acetone (-78 ^oC) prior to placing on lyophilizer overnight (-40 ^oC). Tissues were stored at -20 ^oC or -86 ^oC for 3, 10, 35 or 75 d to determine if metabolite stability depends on storage conditions (Figure 2.1, Storage comparison). Tissue samples stored for an equivalent length of time (10 d) were used to observe the effect of lyophilization of the tissue on the resulting NMR-based metabolic profile (Figure 2.1, Lyophilization or Not). Tissues stored for an equivalent amount of time were also used to compare the effect of extraction solvent on the quality of metabolomic data produced (Figure 2.1, Aq. vs. Org, and Org). The utility of lipid removal treatments was also investigated (Figure 2.1, Lipid elimination).

2.2.2.5 Establishment of Tissue Extraction Methods

Extraction methods were taken from commonly cited metabolic profiling and metabonomics literature[20, 34] as well as a method previously established by our laboratory.[26] Methods compared are summarized in Figure 2.1. Approximately 100 mg frozen pulverized tissue constituted each sample, to which 700 µL of extraction solvent was added. For all methods, the matrix-tissue suspension was manually homogenized directly by hand in a microcentrifuge tube using a micropestle (Eppendorf) for 4 min,

twisting the pestle clockwise on each downward thrust with a frequency equivalent to 2 Hz at room temperature. All extraction and reconstitution solvents contained 100 µM TMSP-d_4 as a chemical shift reference.

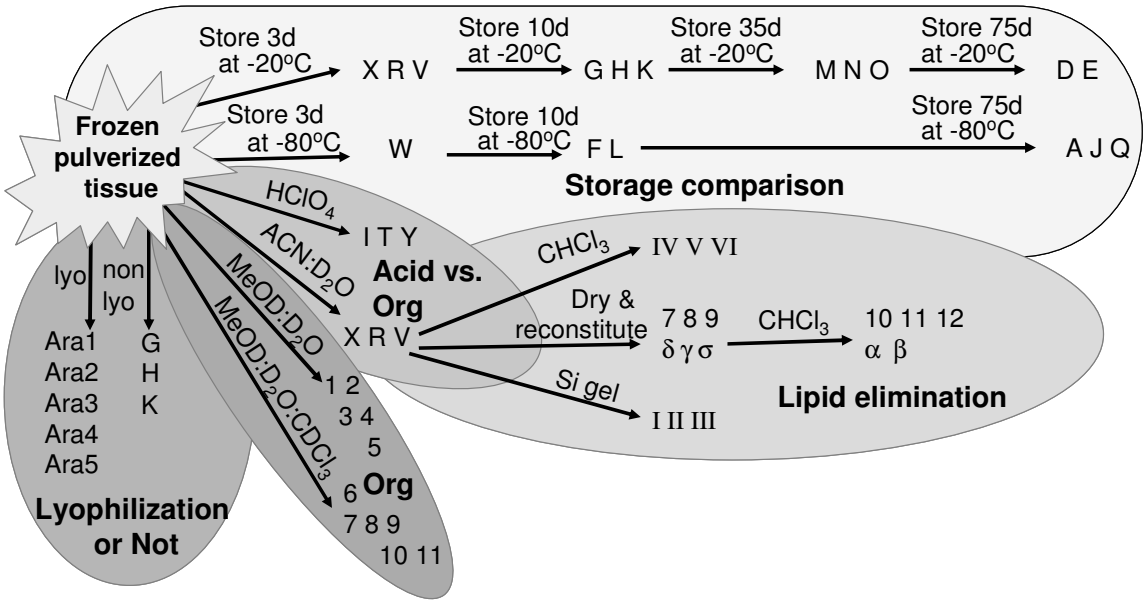


Figure 2.1. A single pool of five-week-old Arabidopsis rosette leaves was used for optimization of storage conditions, extraction solvents, and refinement (clean-up) steps. The effect of lyophilization was compared against wet tissue. Wet tissue was stored at -20 °C or -80 °C to evaluate the effect of storage on ^{1}H NMR metabolome. Primary extraction solvents included acidified D_2O, D_2O/CD_3CN (50/50), D_2O/CD_3OD (75/25), and $D_2O/CD_3OD/CDCl_3$ (1/2.5/1). Secondary sample clean-up steps included secondary extraction with $CDCl_3$, silica gel, or a simple dehydration and reconstitution in D_2O buffer. This last step can be followed with addition/removal of $CDCl_3$ for additional lipid removal in increasingly complex mixtures.

Based on a method for metabolite profiling in mammalian serum, an original extraction solvent, 50/50 acetonitrile/aqueous, yielded stable solutions with no evidence of residual enzymatic activity, i.e. structural changes over time, in our NMR-based metabolite profiles (data not shown). Acetonitrile (50%) has been shown to precipitate proteins without small molecule loss, e.g. Boernsen et al., 2005.[42] Buffered aqueous (D_2O) solution was prepared containing 50 mM CD_3COOD and 500 µM NaN_3 as a

biocide.[39] Deuterated aqueous solutions were titrated by DCl to the desired pH meter reading. Samples extracted using 0.83 M perchloric acid were neutralized following a 15 min digestion and 4 min centrifugation at 2,300 x *g*. Neutralization brought these extracts from a pH meter reading of -0.24 to pD 5 using NaOD, and these were prepared in a series of five analytical replicates for comparison.

The effect of solvent pH on metabolite extraction efficiency was examined by comparing an extract prepared at low pH with an extract prepared at neutral pH. Measurement of pH was carried out in deuterated aqueous solutions, not after mixing with organic modifier, a common practice in chromatography.[43, 44] Only after recording pH were organic modifiers added to extraction solvents. Aqueous pH can be converted to pD using Equation 2.1.

$$pD = \text{pH meter reading} + 0.40 \qquad (2.1)\ ^{45}$$

The neutral pD extraction buffer contained 500 µM NaN_3 to compensate for the antimicrobial effect of low pD extraction solvents.

The amount and identity of organic component was varied at neutral pD using 50 mM CD_3COOD. The original extraction solvent (50/50 ACN/D_2O) was compared with 25/75 MeOD/D_2O and 25/50/25 D_2O/MeOD/$CDCl_3$. Studies by Want et al. and Carlson[19, 46, 47] reported that both methanolic and acetonitrile extracts contain a small amount of residual protein (2% and 6%, respectively). As metabolomics is still a relatively new field, there is no absolute consensus yet on the ideal approach to sample preparation.[8, 48]

Strategies for lipid elimination included the following; deuterated chloroform extraction, evaporation to dryness and reconstitute, and treatment with silica gel. For the deuterated chloroform extraction a liquid-liquid extraction involved the addition of 100 µL of $CDCl_3$ to the extracted sample in 50/50 CD_3CN/D_2O and metabolites were allowed to partition between phases while mixing by vortexer for 60 s. For the procedure in which samples are taken to dryness and then reconstituted, the supernatant from the extracted sample was transferred to a new centrifuge tube, speed-vacuumed to dryness and reconstituted, allowing the dried metabolites to redissolve in an aqueous-only reconstitution buffer while mixing by vortexing at 60 Hz for 60 s. Only the aqueous layer was selected for NMR analysis. To test the effectiveness of silica gel for lipid removal between 25-30 mg of silica gel was lyophilized in 200 µL D_2O after being vortexed 20 s. Supernatant from the extracted sample was directly added to tubes containing silica gel. These were vortexed together for 60 s and centrifugation for 2 min pelleted gel particles with adsorbate (Silica gel 60, Sigma-Aldrich).

After vortexing, all extracts were clarified by centrifugation for 4 min at 12,200 x *g* and 650 µL of supernatant was transferred directly to a 5 mm NMR tube (Wilmad, Buena, NJ) for analysis. Samples were prepared and analyzed in a random order immediately after extraction or subsequent refinement steps were completed. The pH meter reading of each extract was recorded post-analysis. The resulting pH meter reading of the neutral pD extracts were 6.67 ± 0.28, while those at low pD were 4.57 ± 0.22. Using Equation 2.1, these can be called pD 7 (neutral) and pD 5 (low), respectively.

2.2.2.6 NMR Data Acquisition

Proton NMR spectra were acquired with selective saturation of the residual water resonance using a Bruker Avance spectrometer operating at 600 MHz. Free induction decays (FIDs) were acquired into 25860 time points and zero filled to 65536 points. A spectral width of 7716 Hz was excited using a 90^{o} pulse. A relaxation delay of 1.5 s was used, and 1200 scans were coadded following 16 dummy scans for a total experiment time of 64 min. The temperature of the sample was maintained at 25 ^{o}C. Manual shimming was performed on each sample, and the TMSP line width was 1.77 $\pm$ 0.67 Hz across the dataset, following application of 1.0 Hz line broadening.

A comparison between two probes, broadband inverse (BBI) and triple-resonance inverse (TXI) probe and two different users (KAK and GAB) was undertaken to evaluate the robustness of the ^{1}H NMR metabonomics pipeline. Samples were individually extracted by KAK, and transferred to six different NMR tubes. These were randomly analyzed on the same day on two different spectrometers fitted with two different probes. Our goal was to determine which, if any, metabolites are susceptible to differences in stability over time, whether there is a systematic advantage of one probe type over another, and whether manually-defined parameters that vary between samples such as tuning/matching and shimming have a significant effect on the measured metabolome.

2.2.2.7 Post-Acquisition Processing

Spectra were processed using Topspin 2.0 (Bruker). Phasing was applied automatically, however with such complex spectra the phasing was always verified, and

in some cases corrected, manually. Spectra were integrated using equidistant integral regions of 0.02 ppm width over the range of 0.50 to 9.00 ppm, excluding the regions containing the resonances of solvent impurities: CHD_2CN and CHD_3COOD (1.87 – 2.06 ppm), CHD_2OD (3.30 – 3.37 ppm), $CHCl_3$ (7.46 – 7.89 ppm), and HOD (4.20 – 5.20 ppm). A simple DC baseline correction was carried out in Excel. Integral regions were normalized to the total integrated area of each spectrum. No further post-acquisition manipulations (mean-centering, scaling, transformation) were carried out on the dataset. Spectral assignments were made by comparison with spectra of approximately 100 authentic metabolite standards and by cross-referencing the metabolomic literature.[49]

2.2.2.8 Multivariate Statistical Data Modeling

Multivariate data analysis was carried out using Minitab® 14 Statistical Software (Minitab Inc., State College, PA). PCA was used to examine clustering and correlations within the entire data set. The PCA model was constructed using all samples, summarized in Table 2.1. A sample identifier (ID), the extraction conditions, storage temperature, storage time, time between extraction and ^1H NMR analysis, and pH meter reading are provided for all samples. Cross-validation was performed to evaluate model robustness.

2.3 Results and Discussion

2.3.1 Stability of Extracts

2.3.1.1 Matrix-Free Estimations of Metabolite Stability

Sucrose solutions were prepared by dissolving 24 mg of pure sucrose standard (Sigma) directly in 1000 μL of either perchloric acid in D_2O or 50/50 CD_3CN/D_2O containing 100 μM of TMSP-d_4 as a chemical shift reference. The 50/50 CD_3CN/D_2O containing 100 μM of TMSP-d_4 as a chemical shift reference. The 50/50 CD_3CN/D_2O was buffered to a pH meter reading of 4.53 with 50 mM ammonium-d_4 acetate-d_3. These solutions were immediately transferred to a 5 mm NMR tube for analysis. For the sucrose solution in 50/50 CD_3CN/D_2O, NMR spectra were recorded at 10 min intervals over a two h period, while the NMR spectra for the sucrose solution in perchloric acid were recorded at 5 min intervals over a one h period. The time between acquisitions was controlled by running an experiment consisting of only 164 or 68 dummy scans, respectively. Selective saturation of the residual water resonance was applied in both cases. FIDs were acquired into 25860 time points and zero filled to 65536 points. A spectral width of 7716 Hz was excited using a 90^o pulse. A relaxation delay of 1.5 s was used, and 16 scans were coadded following 4 dummy scans for a total experiment time of 1.1 min. The temperature of the sample was maintained at 25 ^oC. Tuning and matching of the probe and shimming were performed manually. FIDs were not apodized prior to Fourier transformation. Phasing was applied manually and no baseline correction was necessary. The resonances of β-glucose (4.64 ppm), fructose (4.09 ppm), and sucrose (5.40 ppm) were integrated using integral regions equivalent to at least 50 * the width at

Table 2.1 Sample identification (ID), extraction conditions, temperature tissue was stored until extraction (^{o}C), duration of storage (d), time between extraction and NMR analysis (h), and extract pH meter reading.

ID	extraction conditions	Storage Temp (^{o}C)	Storage time (d)	Extr to Analy (h)	pH
A	ACN/D2O	-80	75	1	4.62
D	ACN/D2O	-20	75	1	4.65
E	ACN/D2O	-20	75	1	4.67
F	ACN/D2O	-80	10	1	4.63
G	ACN/D2O	-20	10	1	4.65
H	ACN/D2O	-20	10	1	4.70
I	HClO4	-20	3	12	4.43
J	ACN/D2O	-80	75	1	4.65
K	ACN/D2O	-20	10	1	4.67
L	ACN/D2O	-80	10	1	4.63
M	ACN/D2O	-20	35	4	4.72
N	ACN/D2O	-20	35	1	4.67
O	ACN/D2O	-20	35	1	4.60
Q	ACN/D2O	-80	75	1	4.78
R	ACN/D2O	-20	3	5	4.45
T	HClO4	-20	3	2	4.55
V	ACN/D2O	-20	3	10	4.43
W	ACN/D2O	-80	3	2	4.63
X	ACN/D2O	-20	3	6	4.35
Y	HClO4	-20	3	5	4.50
F_1	ACN/D2O	-80	30	5	4.63
L_1	ACN/D2O	-80	30	7	4.63
W_1	ACN/D2O	-80	35	3	4.63
7	extract lyophilized after being extracted with ACN/D2O	-80	46	1	4.83
8	extract lyophilized after being extracted with ACN/D2O	-80	46	1	4.76
9	extract lyophilized after being extracted with ACN/D2O	-80	46	1	4.65
11	extract lyophilized after being extracted with ACN/D2O	-80	46	1	5.01
12	extract lyophilized after being extracted with ACN/D2O	-80	46	1	5.01
I_1	ACN/D2O (Si gel)	-80	61	1	4.64
II	ACN/D2O (Si gel)	-80	61	1	4.64
IV	ACN/D2O (CHCl3)	-80	61	1	4.45
V_1	ACN/D2O (CHCl3)	-80	61	1	4.43
VI	ACN/D2O (CHCl3)	-80	61	1	4.30
Ara1	tissue lyophilized first then reconstituted ACN/D2O	-80	41	1	4.63
Ara2	tissue lyophilized first then reconstituted ACN/D2O	-80	41	1	4.09
Ara3	tissue lyophilized first then reconstituted ACN/D2O	-80	41	1	4.63
Ara4	tissue lyophilized first then reconstituted ACN/D2O	-80	41	1	4.01
Ara5	tissue lyophilized first then reconstituted ACN/D2O	-80	41	1	4.63
α	vacuumed after ACN/D2O, reconsti in aq only, LLE	-80	73	1	4.45
β	vacuumed after ACN/D2O, reconsti in aq only, LLE	-80	73	1	4.61
δ	speedvacuumed after ACN/D2O, reconst aq only	-80	73	1	4.63
γ	speedvacuumed after ACN/D2O, reconst aq only	-80	73	1	4.00
σ	speedvacuumed after ACN/D2O, reconst aq only	-80	73	1	4.64
Ara6	extracted with aq only buffer 1st	-80	108	1	4.69
Ara7	extracted with aq only buffer 1st	-80	108	1	4.70
Ara8	extracted with aq only buffer 1st	-80	108	1	4.71
Ara9	extracted with aq only buffer 1st	-80	108	1	4.70
Ara10	extracted with aq only buffer 1st	-80	108	1	4.71
GB1	MeOD/D2O/CDCl3	-80	442	1	6.75
GB2	MeOD/D2O/CDCl3	-80	442	1	7.09
GB3	MeOD/D2O/CDCl3	-80	442	1	7.01
GB4	MeOD/D2O/CDCl3	-80	442	1	6.89
GB5	MeOD/D2O/CDCl3	-80	442	1	7.05
GB6	MeOD/D2O	-80	442	1	6.26
GB7	MeOD/D2O	-80	442	1	6.78
GB8	MeOD/D2O	-80	442	1	6.81
GB9	MeOD/D2O	-80	442	1	6.68
GB10	MeOD/D2O	-80	442	1	6.89
GB11	MeOD/D2O	-80	442	1	6.63

half-height ($w_{½}$) of each resonance, defined manually. The plot shown in Figure 2.7c was generated using Origin 7.5 (Northampton, MA).

The stability of metabolic intermediates succinate and pyruvate was evaluated in aqueous reconstitution buffer (containing 50 mM ammonium-d_4 acetate-d_3, 500 µM NaN_3, and 100 µM TMSP-d_4) under neutral conditions (pD 7) and the same solution titrated to low pD (pD 5) using 1 M DCl. NMR measurements were recorded on days 1, 21, and 38 using water presaturation (50 dB), four dummy scans were followed by 16 scans which were coadded to give sufficient signal-to-noise for quantitative interrogation, relative integrals versus time are presented in Figure 2.2. Standards in solution were stored at 4°C between analyses. No appreciable loss of succinate or pyruvate was noted.

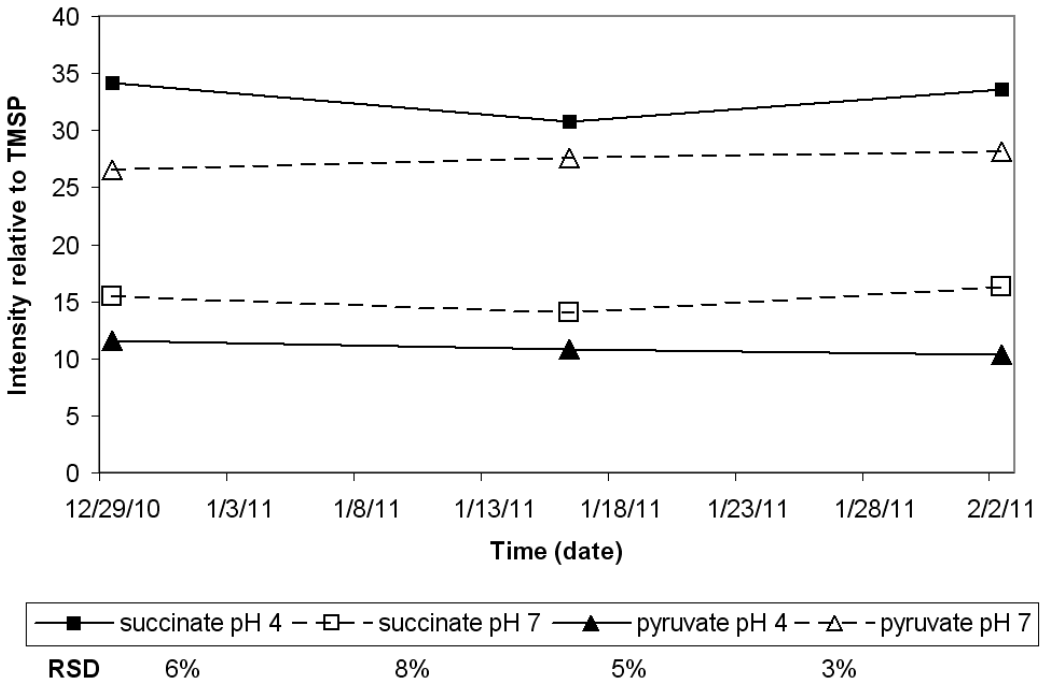


Figure 2.2 Stability of succinate and pyruvate in aqueous reconstitution buffer at low or neutral pH. No appreciable loss of signal intensity was noted for either metabolite over the observation period. Relative standard deviation (RSD) of replicate measurements (N=3) is shown below each condition of the legend.

2.3.1.2 Stability Studies Using Extracts

Neurospora crassa is a model filamentus fungus that is not unlike Arabidopsis, in that it is multicellular eukaryote that expresses metabolic and physiologically distinct phases of its reproductive cycle. When the fungus has established itself in a nutrient-rich environment, it will form long filaments with some branching called hyphae. Like other filamentous fungi, Neurospora undergoes asexual reproduction called conidiation, in an attempt to escape from unfavorable growth conditions. The structures formed are called conidia spores and are similar to sexually reproduced spores in that they contain both a metabolic reserve of energy designed for long-term stable storage and the genetic material required to reestablish growth and continue asexual or sexual reproduction when/if favorable conditions become available. Extracts of biological material grown on Vogel's Medium (VM) and isolated by filtration, were prepared as recommended by Kaiser et al.[50] The stability of fungal extracts was evaluated over a 2 h period. One spectrum was acquired every 10 min to examine extract stability. Figure 2.3 is a comprehensive representation of all observable metabolites cataloged in fungal whole cultures and conidia. The spectra in Figure 2.3a, for metabolites isolated from a whole culture (primarily hyphae), demonstrate that over time, we observed no change in the abundance of the metabolites. Integration and statistical evaluation of this dataset revealed no trend detectable by PCA or univariate analysis (data not shown) as a function of time. The spectra shown in Figure 2.3b, illustrate the stability of an extract prepared from conidia and reveal three key metabolites that change significantly in intensity over the course of the experiment.

A detailed examination of Figure 2.3 reveals many differences in the stability of the extracts prepared from whole tissue and conidia only. Conidia are isolated by partitioning from aqueous Vogel's liquid medium and an immiscible isoparaffinic hardrocarbon fluid, Soltrol 170.[51] Conidia of *N. crassa* are hydrophobic, and mycelium (hyphae) are not, allowing conidia to be readily isolated from whole cultures in this manner. Soltrol 170 has additionally been shown to prevent spore germination.

First, the predominant sugars observed in our fungal extracts are glucose and trehalose. The reproducibility of the spectra containing the anomeric resonances of these two sugars (Carbohydrate Region, 5.2 ppm) in extracts of hyphal tissue (Figure 2.3a, Whole Culture) indicates that no residual enzymatic activity remains in the solution post-extraction. In the extracts of Soltrol-isolated conidia (Figure 2.3b, Conidial Isolate) the conversion of trehalose to glucose is clearly observed. An isoform of trehalase is stored in conidia to facilitate rapid resumption of respiration post-dessication. Trehalase is known to operate stably in the conversion of trehalose to glucose at pH 5.0, with an isoelectric point of 3.7, it is likely to retain activity even under our acidified extraction conditions (pD 5.0).[52]

Second, it can be clearly seen in Figure 2.3b that over time citrate appears, but this trend does not correlate with any other metabolite changes. The effect may be due to decomplexation by paramagnetic metal ions which mask the citrate resonances. When the intracellular environment of the fungus is rapidly disrupted by extraction, equilibra which were maintained *in vivo* by compartmentalization or other active mechanisms are lost. One strategy to overcome such effects is to include a competitive ligand with

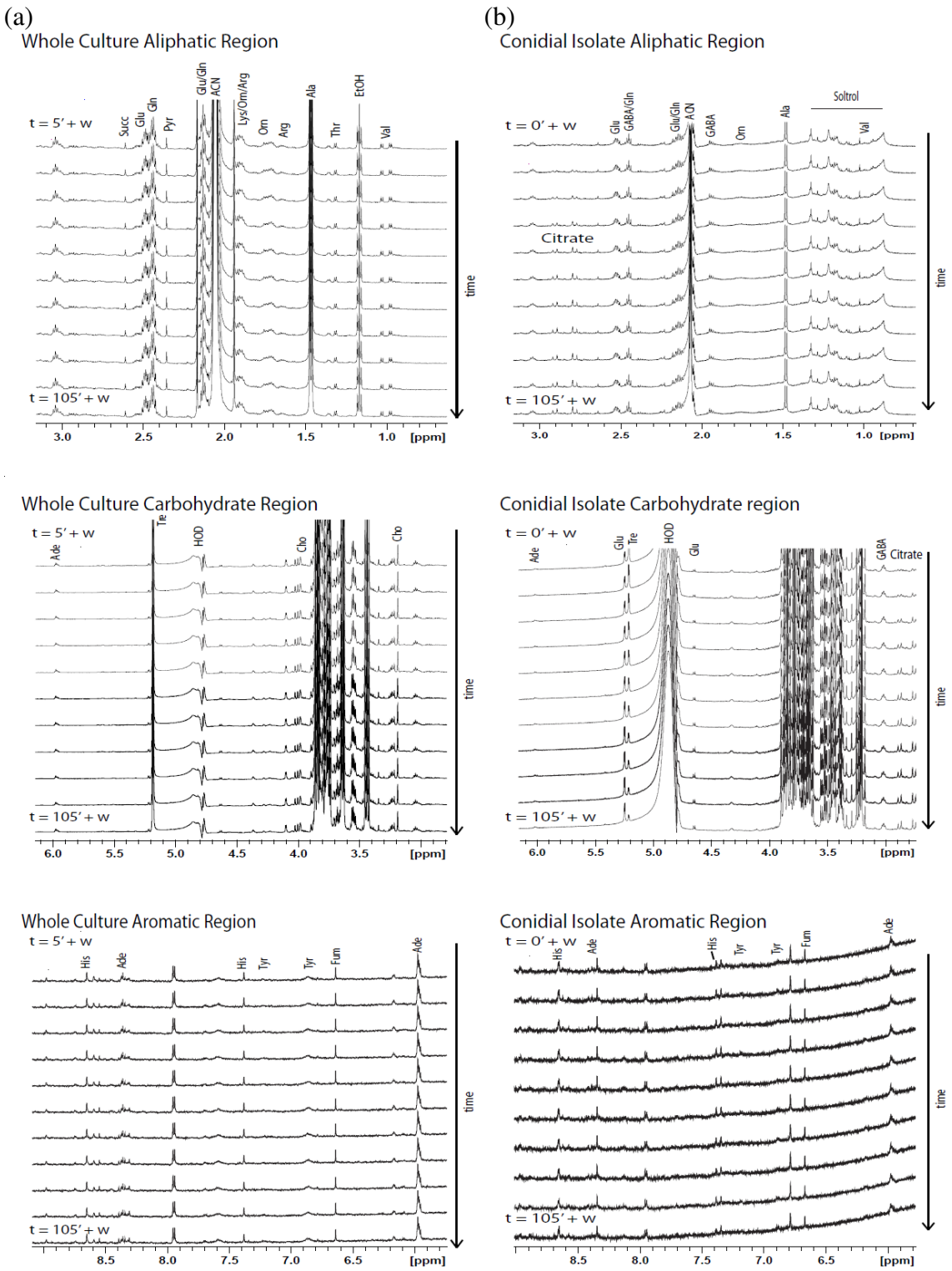


Figure 2.3 Stability of *N. crassa* extracts at 25 °C in our extraction solvent (caption continued next page)

Figure 2.3 Stability of *N. crassa* extracts at 25.2 ^{o}C in our extraction solvent (continued previous page) [acetate-buffered 1:1 (v/v) acetonitrile-d_3:deuterium oxide at pD 5] over a 140 min period reveals differences between whole-culture and conidial isolates. Presented are expansions of three regions of our NMR spectra; Aliphatic (1-3 ppm), Carbohydrate (3-6 ppm), and Aromatic (6-9 ppm). Metabolite abbreviations: valine (Val), ethanol (EtOH), threonine (Thr), alanine (Ala), arginine (Arg), ornithine (Orn), lysine (Lys), glutamate (Glu), glutamine (Gln), pyruvate (Pyr), succinate (Succ), γ-aminobutyric acid (GABA), choline (Cho), glucose (Glu), trehalose (Tre), adenosine (Ade), fumarate (Fum), tyrosine (Tyr), histindine (His). Solvent abbreviations: acetonitrile-d_2 (ACN), water-d (HOD), Spectra are stacked with the earliest recorded time point on top for each group. Because of the greater abundance of metabolites in whole culture samples, one spectrum was acquired every 5 min over a period of 140 min. Signal-to-noise for trehalose resonances exceeded 300. Shown here is one spectrum measured every ten min. For hyphal isolates, less material was presented per sample and therefore each spectrum shown here represents signal averaging for 10.5 min over a period of 140 min. The factor w in each equation for time (t) represents the time for the sample to defrost from -80^{o}C and for manual shimming on the sample prior to executing the first experiment, typical values for w were 15 min.

affinity for metal complexation into the buffer used to record NMR metabolite profiles. EDTA has been used by others,[53] and it should be considered for inclusion in future metabolic profiling experiments. Most importantly, the metabolic profile of the hyphal extracts was stable, and other than the glucose/trehalose ratio and citrate equilibria, no other differences were observed in the metabolite fingerprints of the conidia extracts over time.

While it is useful to discuss extract stability using model organisms such as *N. crassa*, it is clear that plant metabolomes represent greater molecular diversity than bacteria, humans or fungi (Karakok et al.[54] utilized 551 bacterial-, 825 fungal-, 1102 human-, and 2351 plant metabolites for a representative comparison). Therefore, to further establish the validity of our sample preparation, plant tissue was the source of metabolites for the remainder of this work. Figure 2.4 shows the overall scheme used for comparison of metabolite extraction strategies from mature (rosette) leaf tissue of *A. thaliana*. By comparing the extract produced from 1 (dry tissue) and 3 (wet tissue), the effect of lyophilization can be observed. Comparing the extract produced by 2 and 3, the

effect of extraction pD is highlighted. Aqueous (4) and acidified aqueous (5) were compared with acetonitrile/aqueous (3) at low pD. Organic identity and composition (2, 6, 7) were varied at neutral pD. Additional steps to minimize lipids (8, 9) were compared with the original method (3).

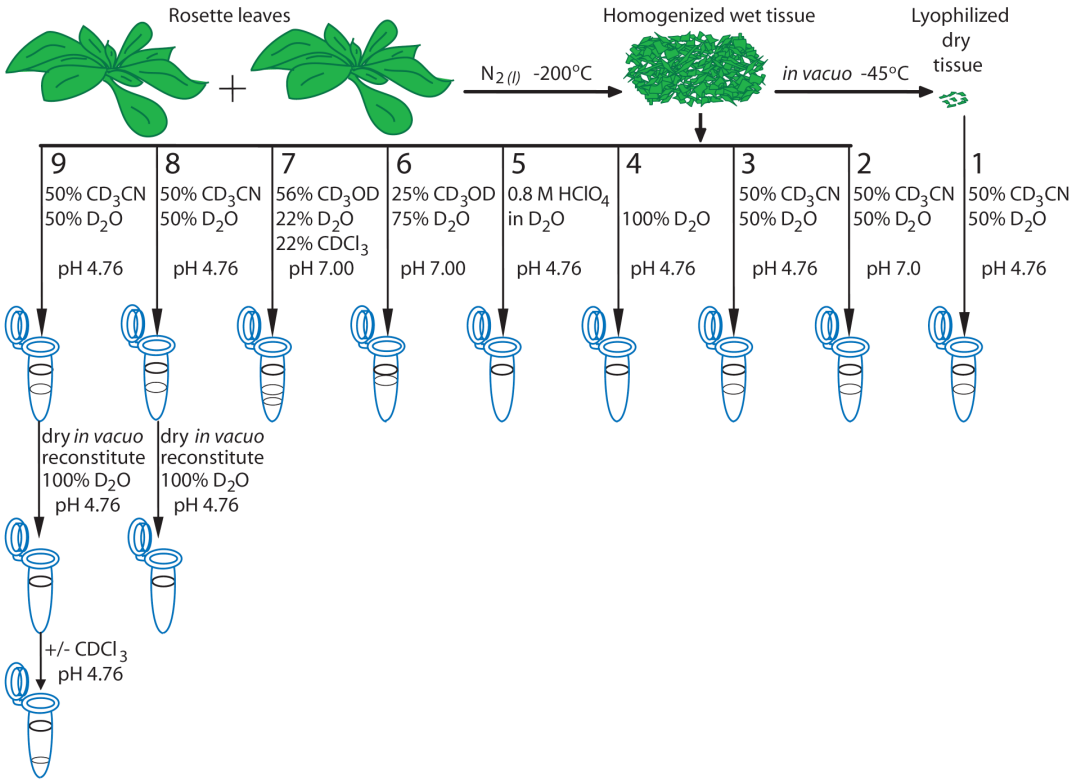


Figure 2.4 Schematic of the metabolite extraction strategies compared in this work. The rosette leaves of two 10 cm x 10 cm plots of *Arabidopsis thaliana* (growth stage 6.00) were pooled and homogenized under liquid nitrogen to quench intracellular enzymatic activity. Each analytical replicate was derived from 100 mg of either wet (2-9) or lyophilized tissue (1), homogenized by micropestle directly in a microcentrifuge tube in the presence of 700 µL of the extraction solvent indicated containing 100 µM TMSP-d_4. Extracts were clarified by centrifugation, after which 650 µL of supernatant was analyzed by ^{1}H NMR for strategies 1-7. In the strategies labeled 8 and 9, the primary extraction solvent was evaporated under vacuum and metabolites were reconstituted in 700 µL of secondary solvent as indicated. After centrifugation, the extract was transferred to either (8) an NMR tube for immediate analysis or (9) a new centrifuge tube. A tertiary refinement step was added to strategy 9, where a liquid-liquid extraction was performed by addition, agitation and removal of $CDCl_3$ prior to centrifugation, transfer to an NMR tube, and analysis.

2.3.2 Effect of Initial Tissue State

While the simplest approach to analysis of plant tissue is direct homogenization and extraction, it is also possible to subject the homogenized tissue to lyophilzation prior to extraction. An advantage of lyophilization is that it permits normalization of metabolites to sample dry weight and reduces the intensity of the H_2O resonance in the extract spectra. Another advantage of lyophilization is that it lends long-term stability to tissue that must remain in storage for an extended duration prior to analysis. For example, dry tea leaf tissue stored for 10-20 years showed no significant change in its metabolite composition.[2] Figure 2.5 provides a comparison of ^{1}H NMR spectra of *Arabidopsis* tissue with (2.5a) and without (2.5b) lyophilization prior to extraction in 50% CD_3CN/50% D_2O at low pD. A large resonance due to acetonitrile-d_2 appears around 2.0 ppm. The expansion of these spectra highlighting the aliphatic region (0.75-1.5 ppm) is presented in Figures 2.5c and 2.5d to emphasize the contribution of the broad lipid/wax resonances in the spectrum of the lyophilized tissue extract. These lipid/wax peaks obscure the resonances of several important metabolites. A possible explanation for the increase in resonance intensity of lipids/waxes in Figures 2.5a and c is that disruption of the cellular membranes during lyophilization increases their solvent availability. These results suggest that in a study evaluating changes in the plant's lipid or wax profile, the efficiency of extraction of these components may be improved by lyophilization of the tissue prior to extraction.

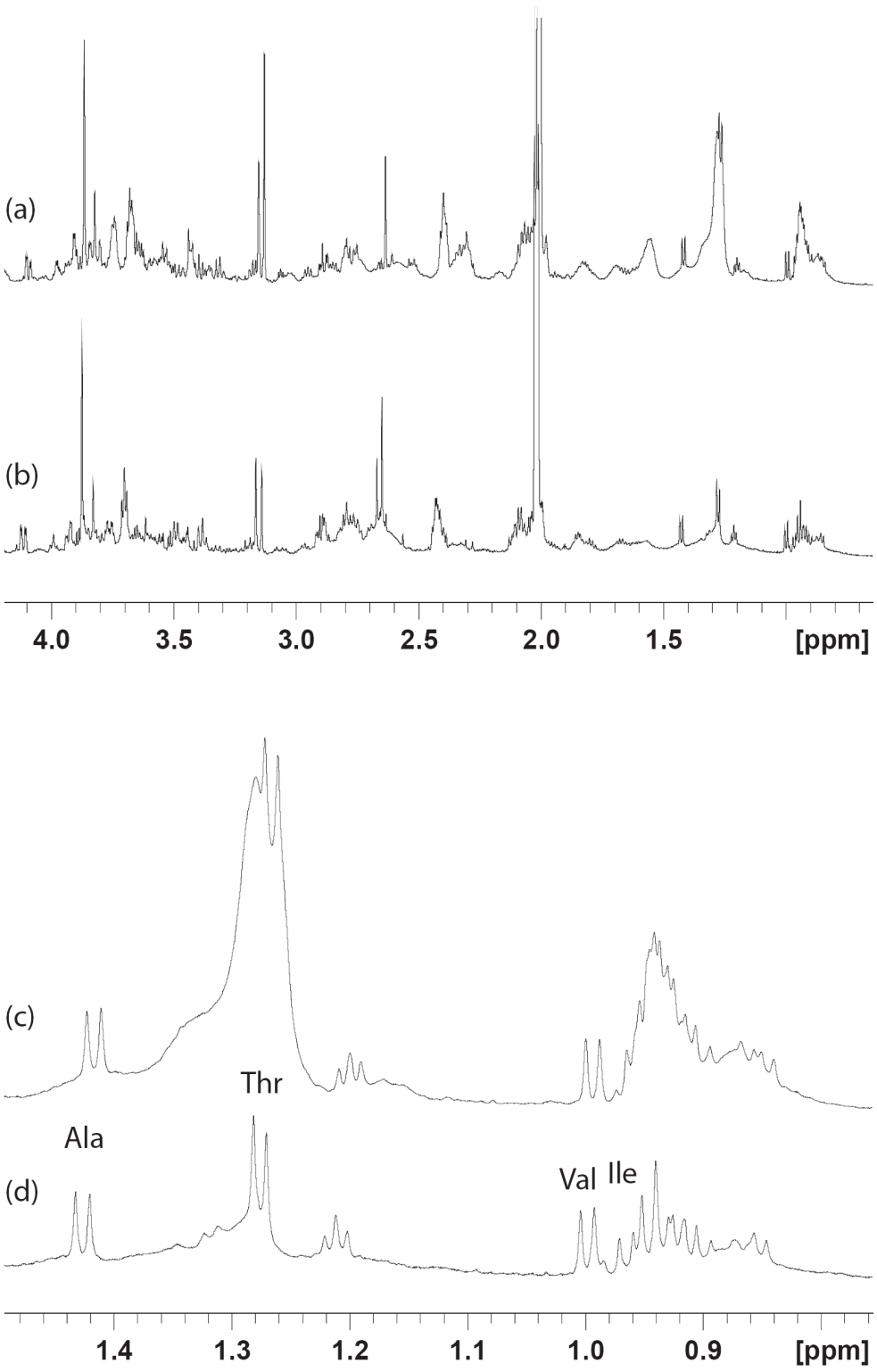


Figure 2.5 Effect of initial tissue state: lyophilized tissue (a,c) is compared with wet tissue (b,d) when extracted under constant conditions.

2.3.3 Effect of Solvent pH

Sample pH is an important parameter in metabolic profiling experiments because the extraction efficiencies of some metabolites can be affected by the pH of the extraction solvent. In addition, the chemical shifts of many metabolites depend on pD, especially near the pK_a of the functional group.[55] Figure 2.6 compares results obtained for different extraction solvents at neutral (2.6a) or low pD (2.6b). We hypothesized that extraction at neutral pD, which is between most of the pK_a values of the ionizible groups of amino acids, would yield more reproducible chemical shifts despite small variances in pD between analytical replicates. Interestingly, using a neutral pD buffer did not seem to increase reproducibility and, in fact, in some spectral regions resulted in a diminished spectral quality due to the increased presence of lipid/wax resonances.

Control of solvent pH is also an interesting approach for resolving spectral overlap between the metabolites of interest and interfering compounds, as shown in the spectral expansions presented in Figures 2.6c (neutral pD) and f (low pD). For example, because of its adjacent carboxylate group, the γ proton resonance of glutamate shows a pronounced shift upfield as the pD is raised, permitting quantitation of the well resolved glutamate and glutamine resonances in Figure 2.6c. In contrast, in Figure 2.6f the resonances of succinate (2.562 ppm) and pyruvate (2.306 ppm) are clearly resolved and can be quantified by peak integration.[33] However, at neutral pD, these low-intensity resonances were no longer visible due to overlap with other peaks. Therefore, the choice of pH for a given extraction may be governed by the need to resolve the signals of particular metabolites of interest.

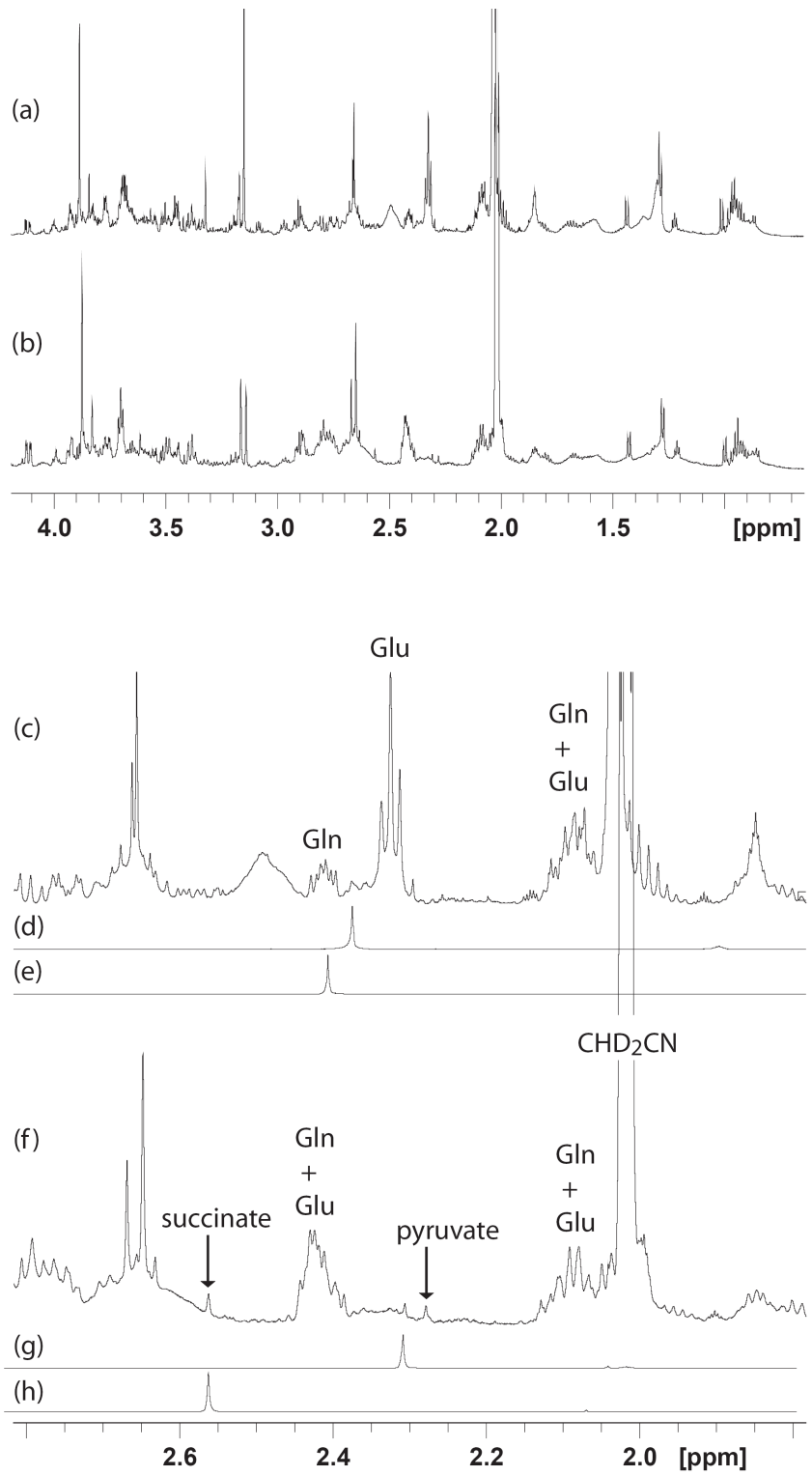


Figure 2.6 ^{1}H NMR spectra of primary extracts in 50/50 CD_3CN/D_2O obtained using (a) neutral or (b) low pH solvents. An expansion reveals (c) neutral pD improves resolution of Gln and Glu, whereas (f) low pD improves resolution of glycolytic end-product pyruvate (d, g) and TCA cycle intermediate succinate (e, h).

2.3.4 Effect of Solvent Composition: Acetonitrile/Aqueous, Aqueous, Acidified Aqueous

Because many important primary metabolites are water-soluble, CD_3CN/D_2O extraction (Figure 2.7a) was compared with an aqueous-only extraction buffer (Figure 2.7b) as well as extraction using perchloric acid (Figure 2.7c). Direct extraction with a buffered aqueous solvent was reported as the favored technique of Brown et al., when compared with acetonitrile, benzene, chloroform, methanol, and DMSO.[14] The D_2O and CD_3CN/D_2O extractions were carried out by homogenization of tissue in the presence of the extraction solvent for 4 min. In both cases, the extraction solvent was maintained at low pD by a 50 mM ammonium-d_4 acetate-d_3 buffer. The spectrum in Figure 2.7c resulted from tissue homogenized in a 0.83 M perchloric acid solution for 15 min then partially neutralized to a pH meter reading of 4.49 ± 0.06 with sodium deuteroxide. This procedure was adapted from the widely-used protocol of Stitt et al.[56]

Comparison of the aliphatic regions of the spectra in Figure 2.7 shows that the aqueous and perchloric acid extracts are much cleaner, with essentially no interference from resonances of the lipid/wax components. A disadvantage of mixed acteonitrile-aqueous solvent systems is that acetonitrile does not associate strongly with the aqueous phase, and this effect is more pronounced in deuterated phases. Initially homogeneous acetonitrile-aqueous mixtures slowly resume partial binary character, observable as loss of magnetic field homogeneity along the z-axis due to partial segregation along a vertical solvent gradient.[57] An additional advantage of a simple D_2O buffer extraction is that the

spectra are simplified compared with mixed organic/aqueous systems since the residual proton resonance of the organic component (i.e. acetonitrile-d_2 or methanol-d_3) is eliminated. However, the increased water content of perchloric acid extracts means that a larger spectral region will be affected by solvent suppression, making it more difficult to reliably detect nearby resonances.

Although the D_2O extract spectrum shown in Figure 2.7b is of high quality, with sharp signals and minimal contamination by lipids and waxes, the expansion shown in Figure 2.7e highlights a problem with this method. Because sucrose is the primary storage carbohydrate in most plants, we expect it to be the dominant sugar in our extract spectra,[58] and indeed a prominent resonance for the sucrose anomeric proton is observed in the spectrum of the CD_3CN/D_2O extract in Figure 2.7d. At first glance, we found the complete loss of the sucrose anomeric signal, expected at 5.4 ppm in D_2O solution, and the proliferation of resonances in the carbohydrate-rich spectral region from 4.7 to 3.35 ppm to be quite surprising. An important drawback to aqueous extractions is that contamination by soluble enzymes can reduce sample stability, in this case resulting in degradation of sucrose, possibly by sucrose synthase or invertase present in the extract. For example, in a study comparing aqueous extractions, including microwave-assisted aqueous extraction and extraction with boiling water, with alcoholic extractions using methanol and ethanol, Kim et al. found that high temperature was necessary to denature enzymes in the aqueous-only extracts.[12] Additionally, aqueous environments, especially those containing sugars, are ideally suited for microbial growth. Because these microorganisms consume, produce, and excrete a range of metabolites, their presence can

complicate interpretation of the metabolite profile of the organism of interest. This problem has been noted for some time by the metabolomic community with regard to urine samples, and addition of NaN_3 is recommended to prevent microbial growth.[39] Since our extraction solvents all included this biocide, it is reasonable to discount the effects of microbial growth in these experiments.

Although many researchers have adopted perchloric acid as an extraction solvent,[9, 21, 32] others have determined that its drawbacks outweigh its advantages.[7, 8, 18] Perchloric acid extraction has been reported to be less reproducible compared with other extraction methods.[15, 18] Sources of error for this method include pH adjustment following homogenization, the time dependence of extraction efficiency or yield, and the potential for chemical modification of metabolites. Many authors report that desalting the extract, critical if derivatization is to be performed,[7] and pH adjustment in unbuffered extracts adds error and decreases throughput.[8] Maharjan and Ferenci compared ethanol, hot methanol, cold methanol, aqueous potassium hydroxide, methanol/chloroform and perchloric acid. Using yeast fed uniformly [^{14}C]-glucose, it was determined that the perchloric acid extracts had the highest relative standard deviation (14%), while methanolic extracts had the lowest (5%).[8] Subsequent perchloric acid extractions on the same tissue pellet have been shown to have high yield whereas methanol/chloroform extracts do not, meaning that several extraction steps (or longer extraction times) must be used with perchloric acid to achieve equivalent extraction efficiency, ultimately leading to decreased throughput.[18] Perchloric acid extracts from yeast had a relatively low

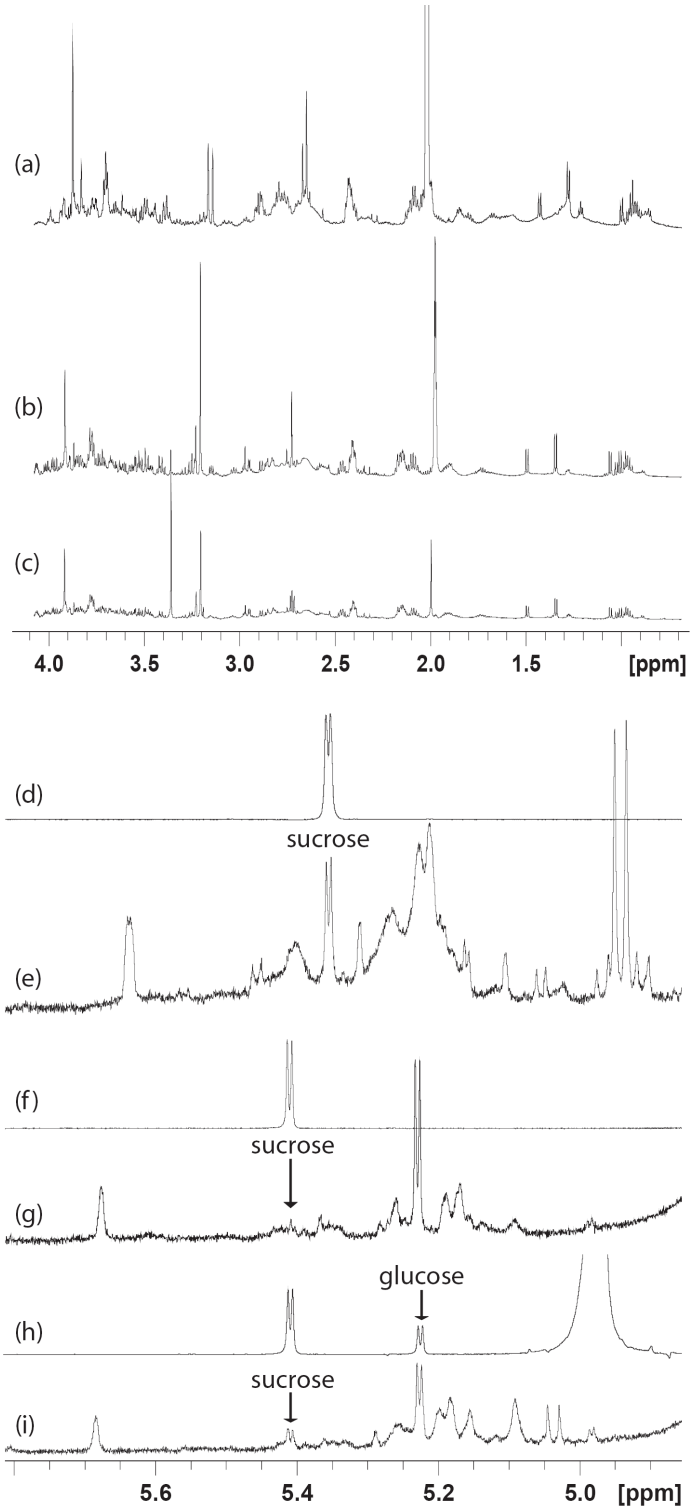


Figure 2.7 Comparison of 50/50 CD_3CN/D_2O (a,g), buffered D_2O (b,g), or perchloric acid in D_2O (c,i) as extraction solvents for ^1H NMR metabonomics of mature *A. thaliana* leaf tissue. Spectra of a sucrose standard spiked into the corresponding extraction buffer (d, f, h) are presented for comparison.

recovery of amino acids, organic acids, fatty acids, and nucleotides, but a relatively high recovery of peptides, sugars and sugar alcohols[7] when contrasted against chloroform/methanol/water, ethanol, aqueous potassium hydroxide, methanol/water and methanol alone.

One of the advantages of perchloric acid is that it is a strong denaturant, rapidly inactivating enzymes and therefore quenching enzymatic activity. However, a related drawback of perchloric acid extraction is the potential for chemical modification of metabolites by this strong acid. Destruction of pyruvate, NAD, $NADH^+$, and indole compounds (such as nucleotides) has been noted under extreme pH.[7, 8] Examination of the spectral expansion in Figure 2.7f shows evidence of acid hydrolysis of sucrose during the perchloric acid extraction. Although a small sucrose resonance is detected at 5.4 ppm, it is much less intense than was observed in the CD_3CD/D_2O extract. Prominent resonances of fructose and the α and β glucose anomers are also observed in the spectra of the perchloric acid extracts. PCA of CD_3CN/D_2O extracts at low pD and perchloric acid extracts partially neutralized to pH meter reading 4.62 ± 0.18 among 3 replicates revealed segregation along the first principal component axis (PC1). Inspection of PC1 loadings with respect to the original variables revealed that the perchloric acid extracts were significantly different from the CD_3CN/D_2O extracts in all spectral regions occupied by sugars (data not shown).

To further investigate the stability of sucrose to perchloric acid extraction, ^1H NMR spectra were measured as a function of time for sucrose spiked solutions in

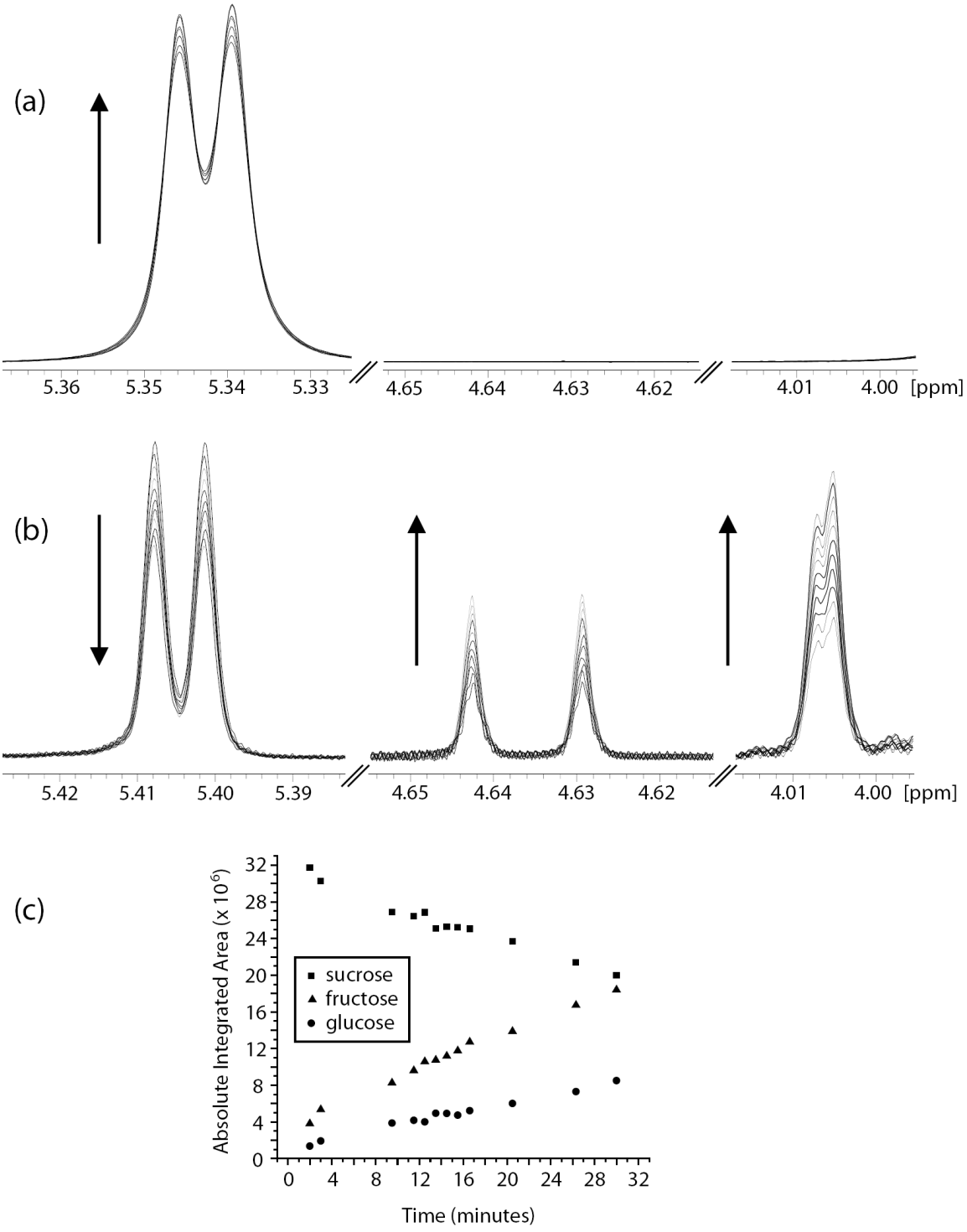


Figure 2.8 Stability of sucrose in (a) acetonitrile-buffered D_2O and (b) perchloric acid. (c) Over a duration as brief as 30 min, acid-catalyzed cleavage of sucrose (■) and corresponding production of monosaccharides glucose (●) and fructose (▲) can be clearly observed and quantified with one-dimensional ^1H NMR.

CD_3CN/D_2O in a solution buffered by 100 mM acetate-d_4 at a pH meter reading of 4.53 (Figure 2.8a) and 0.8 M $HClO_4$ in D_2O (Figure 2.8b), studied in duplicate. No hydrolysis of the sucrose is observed in CD_3CN/D_2O over two hours. Instead, a small increase in both line width and peak height of the sucrose anomeric resonance is observed, which we attribute to partial segregation of the mixed solvent.

As observed in Figure 2.8b, sucrose was steadily degraded by acid-catalyzed cleavage to glucose and fructose in perchloric acid over the same two hour time period. Therefore, a danger of perchloric acid extraction of plant tissue is that it can result in erroneously high levels of glucose and fructose and artificially low levels of sucrose. The sucrose, glucose and fructose resonances in Figure 2.8b were integrated and plotted as a function of time (Figure 2.8c). The kinetics plot in Figure 2.8c suggests that during a 15 min perchloric acid extraction,[56] the level of sucrose could be reduced by as much as 25% of its initial value. The different apparent rates of formation of glucose and fructose arise from the measurement of the resonance of only the glucose β anomer because of overlap of the α-glucose resonance with the HOD signal of the solvent.

2.3.5 Effect of Solvent Composition: Methanol/Aqueous, Acetonitrile/Aqueous, Methanol/Aqueous/Chloroform

To evaluate the suitability of methanol-based extraction for the analysis of plant tissue, we compared the results obtained with 50/50 CD_3CN/D_2O, 25/75 CD_3OD/D_2O and 56/22/22 $CD_3OD/D_2O/CDCl_3$. For each solvent system, the D_2O portion contained deuterated ammonium acetate buffer at neutral pD. It was demonstrated by Maharjan and

Ferenci that methanolic extracts had a significantly higher number of metabolites detected (close to 100) while the other extraction techniques they compared (ethanol, aqueous potassium hydroxide, methanol/chloroform, and perchloric acid) yielded between 50 and 80 metabolites.[8] According to Wu et al.,[16] the popularity of methanol/water/chloroform extractions is due to a 1959 paper[59] introducing what is now called "Bligh-Dyer lipid extraction." Adaptations of the original protocol have been employed since, and a chloroform/methanol/water combination has been reported to form a single mixture.[20] This type of extraction is particularly useful for acid-labile metabolites.[21]

Figure 2.9 shows the ^{1}H NMR spectra of extracts produced by 25/75 CD_3OD/D_2O (2.9a), 50/50 CD_3CN/D_2O (2.9b), and 56/22/22 $CD_3OD/D_2O/CDCl_3$ (2.9c). Although the aliphatic regions of Figures 2.9a and b both show interferences from resonances of lipids/waxes, the CD_3OD/D_2O spectrum in Figure 2.9a contains several broad features that prohibit accurate measurements of many metabolite signals throughout the spectrum, as well as an intense resonance at 1.285 ppm that is not observed in Figure 2.9b. A high quality spectrum is obtained for the mixed $CD_3OD/D_2O/CDCl_3$ extract shown in Figure 2.9c. This single-step extraction for the fractionation of hydrophilic and hydrophobic metabolites has been widely used.[21, 36] In our experiments, we were interested only in the hydrophilic primary metabolites, which are resolved in Figure 2.9c with little interference from the underlying lipid resonances.

Although methanol has been widely used as an extraction solvent, it has been shown to enolize aldehyde groups in plant extracts.[12] Using methanol as an extraction

solvent can also run the risk of transmethylation of sugar esters.[20] Additionally, it has been known since 1964 that some enzymes, such as phosphatases,[20] can retain appreciable activity in the presence of methanol.[32] It can also be seen in Figure 2.9b and c that the resonance from CHD_2OD obscures a portion of the spectrum between 3.30 – 3.37 ppm. Despite the aforementioned drawbacks, alcoholic extractions have several advantages (1) no salts are added, (2) they can easily be removed by evaporation, and (3) they minimally effect pH.[8] However, for ^{1}H NMR metabolomics experiments, the presence of numerous lipid resonances precludes the use of CD_3OD/D_2O.

Incomplete separation of the chloroform layer can be a problem when chloroform is employed to remove lipids and the extract under analysis is not sufficiently ionic. Although not tested in this study, potassium chloride can be added to increase the polarity of the aqueous phase, thereby improving partitioning of polar lipids into the chloroform layer.[15] Chloroform creates additional concerns because of its toxicity,[7] and its ability to extract additives such as plasticizers from plastic labware.[16] Additionally, the $CHCl_3$ impurity in deuterated chloroform introduces a singlet with noticeable ^{13}C satellites obscuring metabolite resonances between 7.46 – 7.89 ppm.

Although chloroform was effective in removing the lipid resonances in Figure 2.9c, in our hands the $CD_3OD/D_2O/CDCl_3$ extraction solvent did not give reproducible results. Figure 2.10 shows spectra obtained for five replicate extractions with this solvent. While variations are observed throughout the spectra, the intensities of the unassigned singlets at 2.69 and 2.70 ppm aptly illustrate the difficulties we encountered with this solvent system.

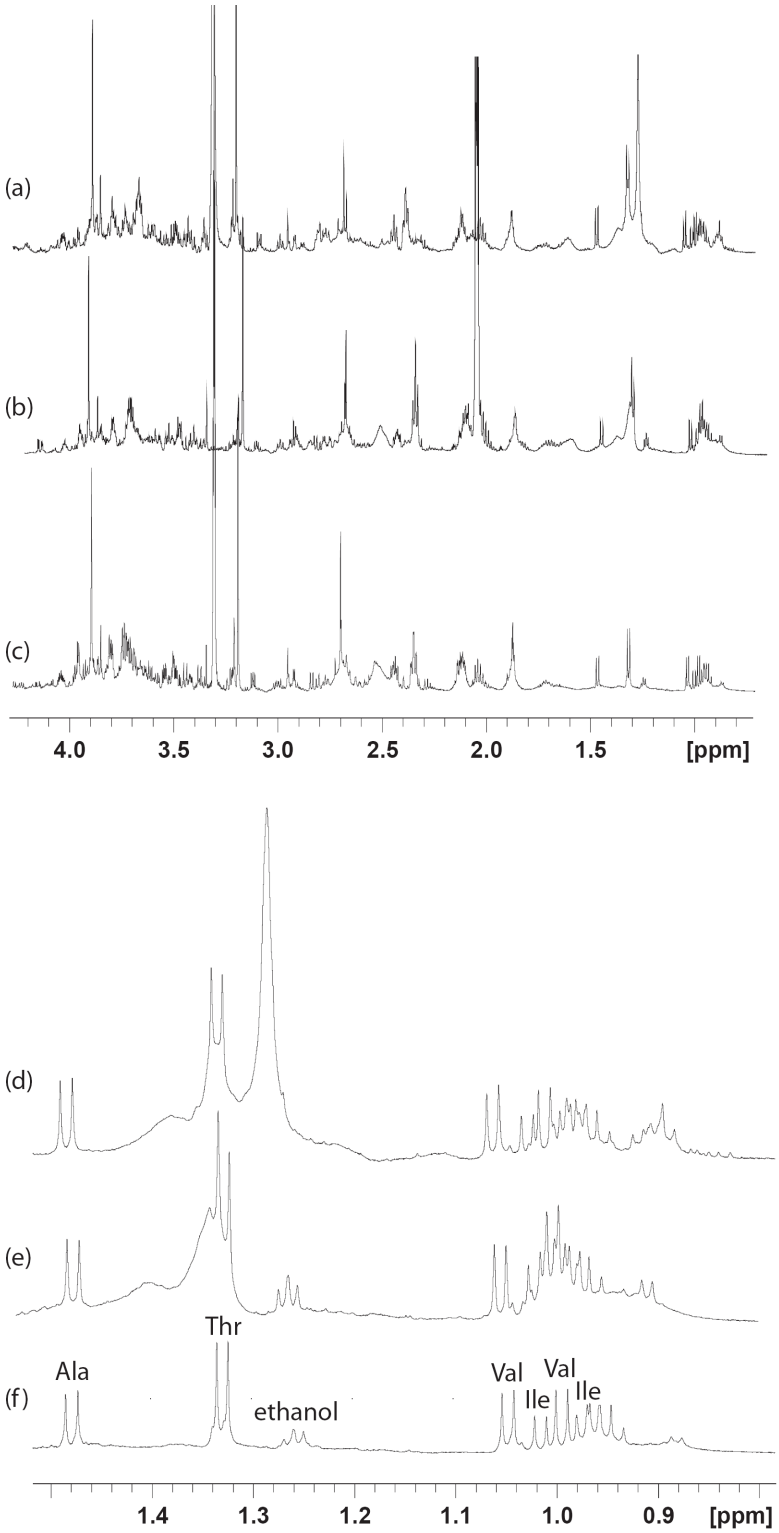


Figure 2.9 Comparison of mixed solvent systems (a,d) CD_3OD/D_2O, (b,e) CD_3CN/D_2O, and (c,f) $CD_3OD/D_2O/CDCl_3$ for extraction of metabolites from leaves of mature *A. thaliana*.

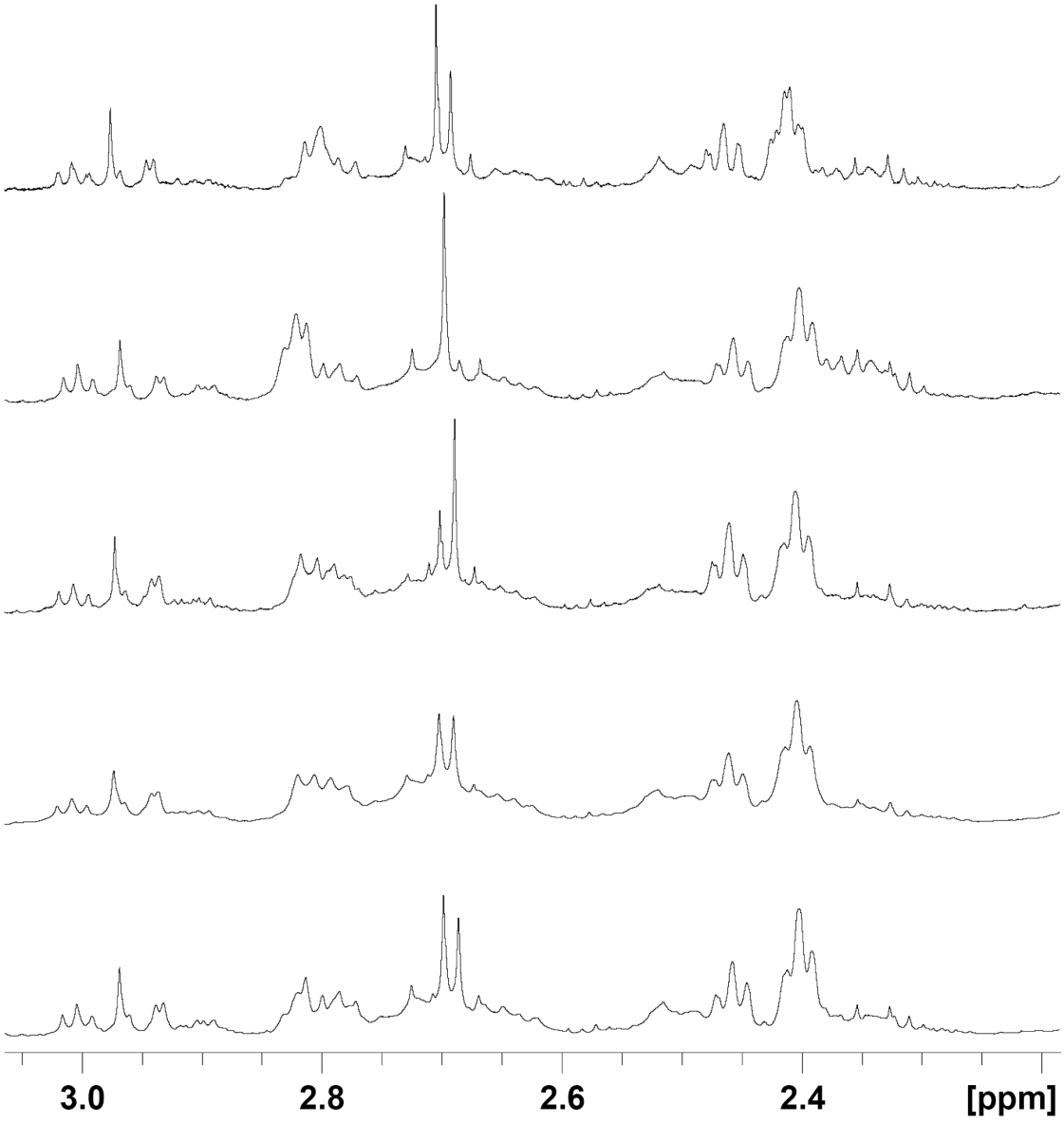


Figure 2.10 ^{1}H NMR spectra of primary extracts resulting from a biphasic solvent system containing CD_3OD, D_2O, and $CDCl_3$. Although these spectra are from analytical replicates of a single *A. thaliana* rosette tissue sample, the extracts do not have a uniform composition demonstrated by the different intensities of the resonances at 2.69 and 2.70 ppm.

2.3.6 Refinement of a Protocol to Enhance Lipid Removal

Unfortunately all the extraction solvents discussed thus far were found to be lacking in some respect. The advantage to working with aqueous extracts is that they

gave sharp resonances and were free of interfering signals due to lipid and waxes. A disadvantage is that we observed significant degradation of some important metabolites, for example sucrose, in these solvents. The mixed $CD_3OD/D_2O/CDCl_3$ extraction solvent also produced clean, lipid-free spectra however replicate spectra suffered from poor reproducibility. Because we were able to achieve excellent reproducibility in our CD_3CN/D_2O extractions and because the samples produced had acceptable stability, we decided to explore additional refinement steps to reduce the lipid/wax content in the ^1H NMR spectra obtained with this solvent. Although there are claims in the metabolic profiling literature that additional extract refinement steps are time consuming and difficult to automate,[18, 20] as shown in Figure 2.9 they can vastly improve the quality of the data used for quantitative analysis.

Figure 2.11a shows the ^1H NMR spectrum obtained using 50/50 CD_3CD/D_2O at low pH without additional refinement steps. The expansion of the aliphatic region in Figure 2.11d illustrates the interference of the lipid resonances with several signals of the aliphatic amino acids, for example Thr, the doublet centered around 1.275 ppm. The spectrum in Figure 2.11b and its expansion in Figure 2.11e were produced by a two-step process, corresponding to strategy 8 in Figure 2.4. In step (1) the tissue was homogenized in the CD_3CN/D_2O extraction solvent used to produce Figure 2.11a. Following extraction, the supernatant was dried by speedvacuum, and reconstituted in aqueous buffer at low pD. Although the solvent could have been removed by either lyophilization or speedvacuum, a study of sample preparation methods for global metabolite analysis of yeast by Villas-Boas et al. found solvent evaporation under reduced pressure (i.e.

speedvacuum) to be preferable, providing excellent recovery for all amino acids, organic acids, nucleotides and sugars.[7] It is obvious, although worth mentioning, that both sample lyophilization and solvent evaporation are unsuitable if the biological question under study involves volatile metabolites.[7] Although our motivation was reduction of the lipid/wax content of our samples, there are additional advantages of a drying step, including a reduction in the magnitude of the HOD resonance and removal of the CD_2HCN peak.[35] Reconstitution in D_2O produces extracts that retain good line shape during long acquisition times unlike the degradation of homogeneity we experienced due to partial phase separation of mixed acetonitrile-d_3/D_2O extracts. In addition, the process of drying the sample allows concentration of metabolites by using a smaller volume of the reconstitution solvent, although this was not employed in the current study.[3, 24, 60]

The spectrum in Figure 2.11c and its expansion in Figure 2.11f were produced by further refinement of the protocol used for Figure 2.11b, shown as strategy 9 in Figure 2.4. For this sample after CD_3CN/D_2O extraction and drying, the sample was reconstituted in aqueous buffer at low pD and a liquid-liquid extraction performed with $CDCl_3$. Compared with the $CD_3OD/CDCl_3/D_2O$ extraction which provided spectra of similar quality, this multistep extraction procedure was much more reproducible in our hands. Although the additional $CDCl_3$ extraction step did not provide a significant advantage for these samples, we have seen cases, such as lyophilized *Arabidopsis* seedlings, for which $CDCl_3$ extraction was necessary to achieve adequate removal of the lipid resonances.[33] Lipid resonances (based on a pure standard of stearic acid) were

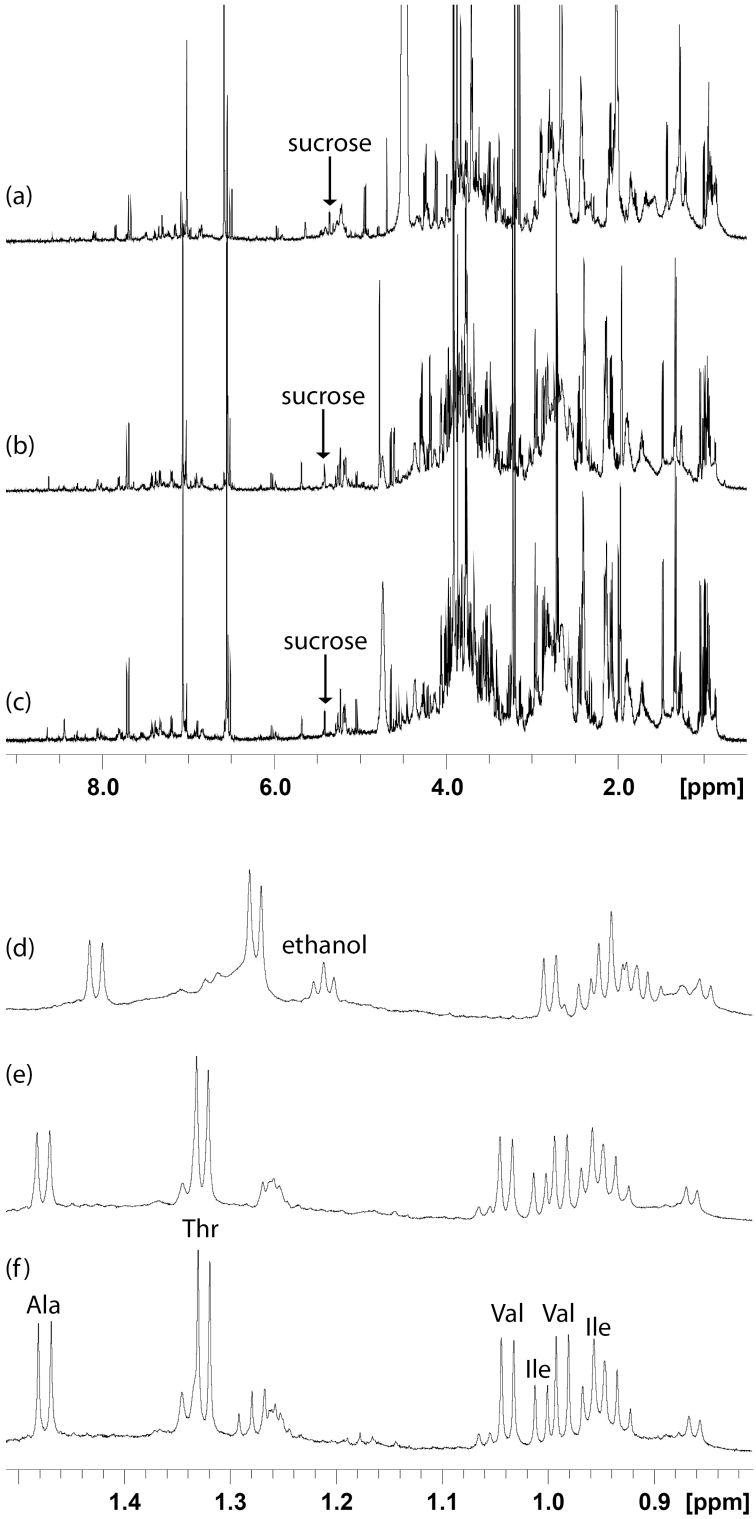


Figure 2.11. ^{1}H NMR spectra of refined extracts resulting from (a,d) 50/50 CD_3CN/D_2O extraction was performed at low pD and compared to (b,e) drying of the primary extract under vacuum with reconstitution in D_2O, and (c,f) liquid-liquid extraction with $CDCl_3$ following drying and reconstitution in buffered D_2O.

found to obscure broad regions around 1.26 ppm and 3.13 ppm. Metabolites which may have been observable in these regions include Thr, lactate, acetaldehyde, Ile, Leu, oxaloacetate upfield and polyamines choline, betaine, arginosuccinate, and arginine downfield, respectively.

2.3.7 Principal Component Analysis of Extraction Reproducibility

Although careful examination of representative spectra was fruitful for qualitative comparison of metabolite extraction strategies, method reproducibility must also be assessed. One approach to evaluating reproducibility is to employ a multivariate statistical technique, such as PCA, which can simultaneously account for all resonances exhibited in a global metabolite profile. The scores plot in Figure 2.12 represents each NMR spectrum as a single point in principal component space, reducing the dimensionality from 370 independent variables to two scores along principal component axes (PC1 and PC2). Because the tissue was sampled from a single homogeneous bioreplicate, the variance observed in the PCA scores plot can be attributed to differences between the extraction methods. The first two principal components together accounted for 66.8% of the variance in the entire dataset, and the model required greater than 10 principal components to explain over 95% of the variance.

Figure 2.12 demonstrates that an unsupervised multivariate statistical technique (PCA) is capable of classifying the samples according to the nature of the solvent in which the tissue was extracted (D_2O, CD_3OD/D_2O, $CD_3OD/D_2O/CDCl_3$, or CD_3CN/D_2O). Within these four groups, some samples cluster according to the strategy

originally employed in the extraction. For example, the extracts in CD_3CN/D_2O at low pD (◆, strategy 2) can be distinguished from those at neutral pD (■, strategy 3). Additionally, strategies 2 and 3 are distinguishable from tissue that was lyophilized (●, strategy 1). In the region of the scores plot containing the extracts in D_2O, a similar segregation can be observed. The tissue directly extracted using buffered D_2O (△, strategy 4) is distinct from all others in the cluster however the other extracts in D_2O (▷, strategy 5; ◇, strategy 8; ○, strategy 9) do not segregate in the scores plot of PC1 versus PC2. Closer examination of higher order PCs revealed that the scores along PCs 5-8 distinguish the samples extracted with perchloric acid (▷, strategy 5) from the other strategies arriving at a sample in D_2O (data not shown). Interestingly, PCs 1-8 do not distinguish strategies 8 (◇), or 9 (○), suggesting that additional chloroform extraction following drying and reconstitution in D_2O was not necessary for this pool of tissue (data not shown).

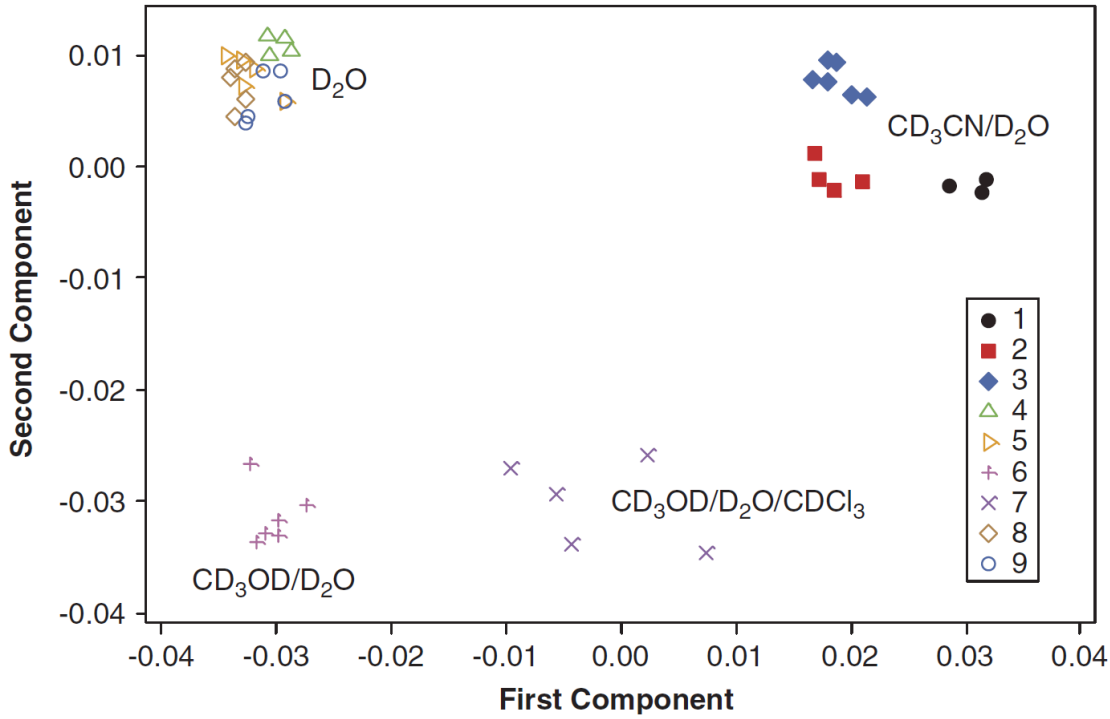


Figure 2.12 PCA scores plot of selected ^{1}H NMR spectra recorded in this study reveals unsupervised classification of samples according to the nature of the solvent in which the NMR spectrum was acquired. Each NMR spectrum is represented as a single point in principal component space.

It can be inferred that analytical replicates in close proximity to each other on this plot show good method reproducibility and *vice versa*. Most notably, the five replicates extracted using a single mixture of 56/22/22 $CD_3OD/D_2O/CDCl_3$ (×, strategy 7) occupy a large area of component space reinforcing the conclusion that, at least in our hands, extraction with $CD_3OD/D_2O/CDCl_3$ is not as reproducible as the other strategies examined.

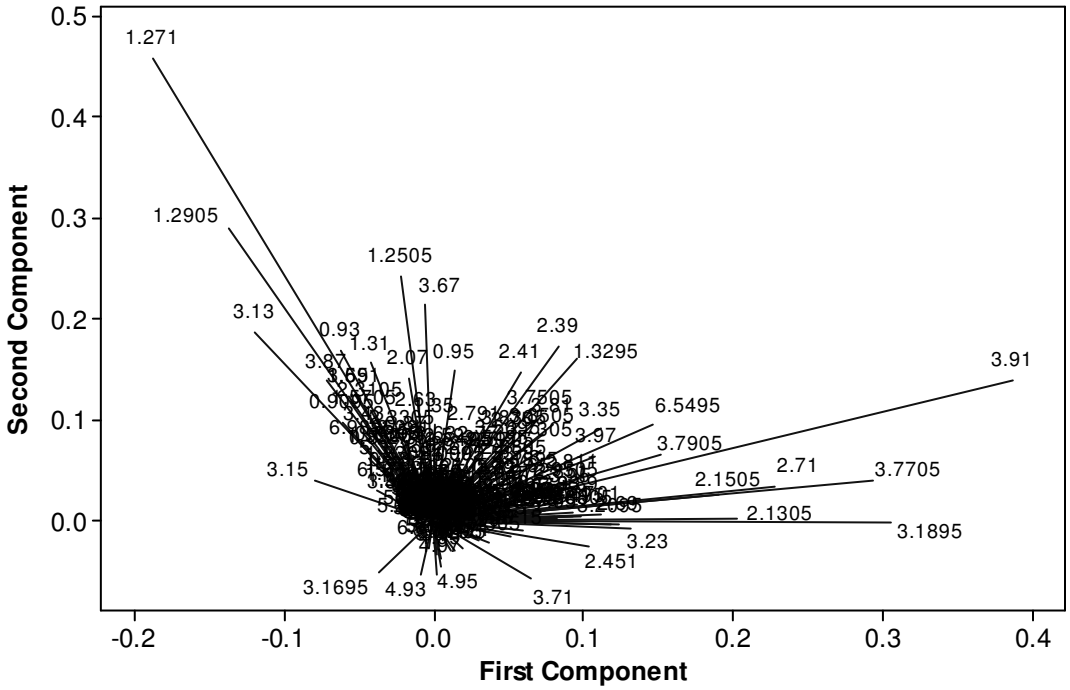


Figure 2.13 Loadings plot corresponding to clusters observed in scores (Figure 2.12). It can be seen that regions of high variability such as 1.3 ppm and 3.1 ppm are responsible for segregating samples. Regions that distinguish samples according to extraction solvent occupy the same integral regions as glutamine (α, 3.79 ppm; β, 2.13 ppm; δ, 2.7 ppm) and citrate (2.71 ppm) as well as fumaric acid (6.55 ppm). [61]

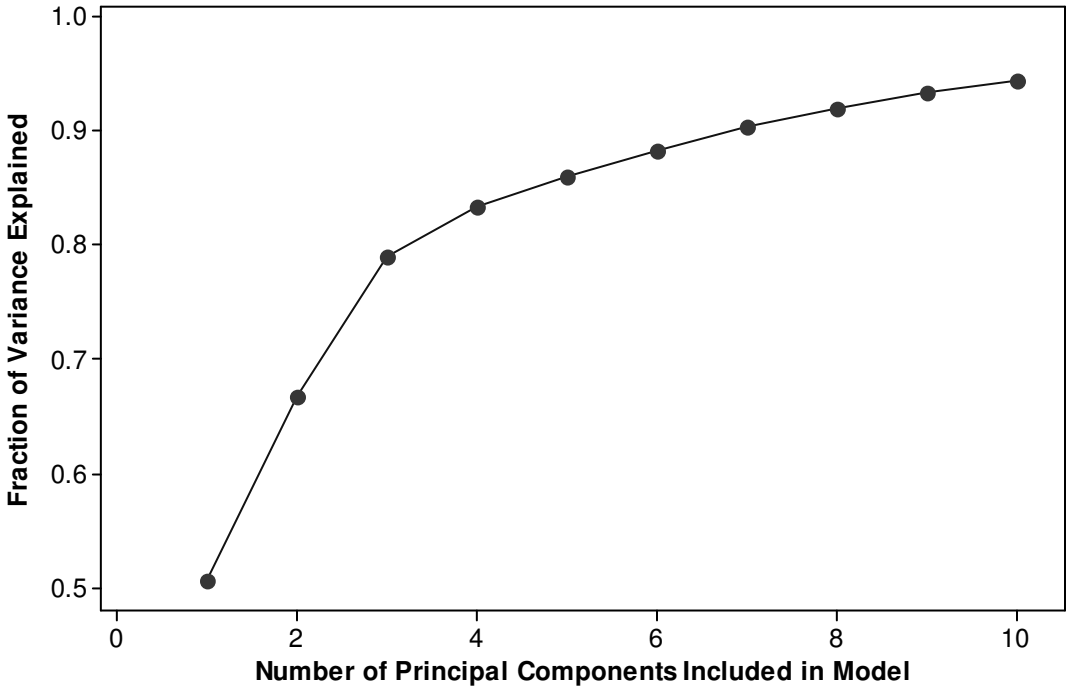


Figure 2.14 An explained variance plot corresponding to the scores (Figure 2.12) and loadings (Figure 2.13) plots presented for the storage/extraction comparison dataset. The first principal component explains 50.6% of the variance. Two principal components are sufficient to explain 66.8% of the variance.

2.4 Conclusions

NMR is well-suited for quantitative analysis and is a well-established tool for metabolite fingerprinting in plants.[32] All metabolite profiling and metabolomic experiments involving extraction of metabolites from tissue begin with a two-step treatment: (1) immediate cessation of enzymatic activity and (2) disruption of tissue to release intracellular metabolites. It is the our recommendation and that of other authors[1] that

"the optimum preparation protocol should be developed on a case-by-case basis...

depending on the characteristics of target metabolites, the number of metabolites being examined and their respective quantities."

Our findings indicate that perchloric acid is not appropriate for plant metabolomic studies because it facilitated acid-catalyzed cleavage of sucrose, complicating biological interpretation of the resulting metabolic profile. Extraction in simple D_2O buffer proved unsatisfactory because it did not quench all enzymatic activity.

Methanol/D_2O/chloroform showed the greatest degree of analytical variability in our study. However, the tolerance of a given experiment with respect to the analytical variance depends on the magnitude of change in the metabolic profiles of control and treated organisms and the amount of biological variability, and this solvent system may prove desirable for some studies. Although acetonitrile-D_2O extracts had good reproducibility and stability, other disadvantages including partial phase separation and contamination by lipids led us to explore additional refinements to this extraction method. The combination of strategies 1, 2, 8 and 9 has been adopted in our laboratory for further plant metabolic profiling studies, described in Chapters 3-4 of this thesis.

By exploration of the data using multivariate statistics, planning and execution of careful follow-up studies revealed the effect of common methods for metabolic extractions on mature leaves from the model plant *A. thaliana*. This underscores the "hypothesis generating" aspect of metabolomic studies, one critical experiment is often not sufficient to establish metabolite identity and behavior. Compromises between experiment time/sample throughput and sensitivity are an issue in NMR-based metabolomics. A method for preparing plant extracts for ^1H NMR analysis, amenable to

integration and quantitation of small abundant primary metabolites, was established and will be applied to investigate questions of a biological nature.

2.5 References

1. Fukusaki, E.; Kobayashi, A., Plant metabolomics: potential for practical operation. *J. Biosci. Bioeng.* **2005,** 100, 347-354.

2. Ryan, D.; Robards, K., Analytical chemistry considerations in plant metabolomics. *Sep. Purif. Rev.* **2006,** 35, 319-356.

3. Sato, S.; Masanori, A.; Soga, T.; Nishioka, T.; Tomita, M., Time-resolved metabolomics reveals metabolic modulation in rice foliage. *BMC Syst. Biol.* **2008,** 2, 51.

4. Hall, R. D.; Brouwer, I. D.; Fitzgerald, M. A., Plant metabolomics and its potential application for human nutrition. *Physiol. Plantarum* **2008,** 132, 162-175.

5. Guy, C.; Kopka, J.; Moritz, T., Plant metabolomics coming of age. *Physiol. Plantarum* **2008,** 132, 113-116.

6. Hegeman, A. D., Plant metabolomics-meeting the analytical challenges of comprehensive metabolite analysis. *Brief. Funct. Genomics* **2010,** 9, (2), 139-148.

7. Villas-Boas, S. G.; Hojer-Pedersen, J.; Akesson, M.; Smedsgaard, J.; Nielsen, J., Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **2005,** 22, 1155-1169.

8. Maharjan, R. P.; Ferenci, T., Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Anal. Biochem.* **2003,** 313, 145-154.

9. Jernejc, K., Comparison of different methods for metabolite extraction from *Aspergillus niger* mycelium. *Acta Chim. Slov.* **2004,** 51, 567-578.

10. Hajjaj, H.; Blanc, P. J.; Goma, G.; Francois, J., Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi. *Fems Microbiol. Lett.* **1998,** 164, (1), 195-200.

11. Kerem, Z.; German-Shashoua, H.; Yarden, O., Microwave-assisted extraction of bioactive saponins from chickpea (*Cicer arietinum* L). *J. Sci. Food Agric.* **2005,** 85, 406-412.

12. Kim, H. K.; Choi, Y. H.; Luijendijk, T. J. C.; Rocha, R. A. V.; Verpoorte, R., Comparison of extraction methods for secologanin and the quantitative analysis of secologanin from *Symphoricarpos albus* using ^{1}H-NMR. *Phytochem. Anal.* **2004,** 15, 257-261.

13. Mattoli, L.; Cangi, F.; Maidecchi, A.; Ghiara, C.; Ragazzi, E.; Tubaro, M.; Stella, L.; Tisato, F.; Traldi, P., Metabolomic fingerprinting of plant extracts. *J. Mass Spectrom.* **2006,** 41, 1534-1545.

14. Brown, S. A. E.; Simpson, A. J.; Simpson, M. J., Evaluation of sample preparation methods for nuclear magnetic resonance metabolic profiling studies with *Eisenia fetida*. *Environ. Toxicol. Chem.* **2008,** 27, 828-836.

15. Lin, C. Y.; Wu, H.; Tjeerdema, R. S.; Viant, M. R., Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics* **2007,** 3, (1), 55-67.

16. Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R., High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal. Biochem.* **2008,** 372, 204-212.

17. Ametaj, B. N.; Bobe, G.; Lu, Y.; Young, J. W.; Beitz, D. C., Effect of sample preparation, length of time, and sample size on quantification of total lipids from bovine liver. *J. Agric. Food Chem.* **2003,** 51, 2105-2110.

18. LeBelle, J. E.; Harris, N. G.; Williams, S. R.; Bhakoo, K. K., A comparison of cell and tissue extraction techniques using high-resolution ^{1}H-NMR spectroscopy. *NMR Biomed.* **2002,** 15, 37-44.

19. Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G., Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.* **2006,** 78, 743-752.

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

21. Hausler, R. E.; Fischer, K. L.; Flugge, U. I., Determination of low-abundant metabolites in plant extracts by NAD(P)H fluorescence with a microtiter plate reader. *Anal. Biochem.* **2000,** 281, 1-8.

22. Boyes, D. C.; Zayed, A. M.; Ascenzi, R.; McCaskill, A. J.; Hoffman, N. E.; Davis, K. R.; Gorlach, J., Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* **2001,** 13, 1499-1510.

23. Initiative, T. A. G., Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000,** 408, 796-815.

24. Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L., Metabolite profiling for plant functional genomics. *Anal. Chem.* **2000,** 72, 3573-3580.

25. Last, R. L.; Jones, A. D.; Shachar-Hill, Y., Towards the plant metabolome and beyond. *Nat. Rev. Mol. Cell Biol.* **2007,** 8, 167-174.

26. D'Auria, J. C.; Gershenzon, J., The secondary metabolism of Arabidopsis thaliana: growing like a weed. *Curr. Opin. Plant Biol.* **2005,** 8, 308-316.

27. Zhang, J. T.; Zhang, Y.; Du, Y. Y.; Chen, S. Y.; Tang, H. R., Dynamic metabonomic tesponses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res.* **2011,** 10, (4), 1904-1914.

28. Sekiyama, Y.; Chikayama, E.; Kikuchi, J., Evaluation of a semipolar solvent system as a step toward heteronuclear multidimensional NMR-based metabolomics for C-13-labelled bacteria, plants, and animals. *Anal. Chem.* **2011,** 83, (3), 719-726.

29. Alvarez, M. D.; Donarski, J. A.; Elliott, M.; Charlton, A. J., Evaluation of extraction methods for use with NMR-based metabolomics in the marine polychaete ragworm, Hediste diversicolor. *Metabolomics* **2010,** 6, (4), 541-549.

30. Gromova, M.; Roby, C., Toward *Arabidopsis thaliana* hydrophilic metabolome: assessment of extraction methods and quantitative 1H NMR. *Physiol. Plantarum* **2010,** 140, (2), 111-127.

31. Martineau, E.; Tea, I.; Loaec, G.; Giraudeau, P.; Akoka, S., Strategy for choosing extraction procedures for NMR-based metabolomic analysis of mammalian cells. *Anal. Bioanal. Chem.* **2011,** 401, (7), 2133-2142.

32. Kruger, N. J.; Tronscoso-Ponce, M. A.; Ratcliffe, R. G., ^{1}H NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues. *Nature Protoc.* **2008,** 3, 1001-1012.

33. Branco-Price, C.; Kaiser, K. A.; Jang, C. J. H.; Larive, C. K.; Bailey-Serres, J., Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana. *Plant J.* **2008,** 56, (5), 743-755.

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

35. Moing, A.; Maucourt, M.; Renaud, C.; Gaudillere, M.; Brouquisse, R.; Lebouteiller, B.; Gousset-Dupont, A.; Vidal, J.; Granot, D.; Denoyes-Rothan, B.; Lerceteau-Kohler, E.; Rolin, D., Quantitative metabolic profiling by 1-dimensional ^{1}H-NMR analyses: application to plant genetics and functional genomics. *Funct. Plant Biol.* **2004,** 31, 889-902.

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

37. Colquhoun, I. J., Use of NMR for metabolic profiling in plant systems. *J. Pestic. Sci.* **2007,** 32, (3), 200-212.

38. Saude, E. J.; Sykes, B. D., Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics* **2007,** 3, 19-27.

39. Lauridsen, M.; Hansen, S. H.; Jaroszewski, J. W.; Cornett, C., Human urine as test material in ^{1}H NMR-based metabonomics: recommendations for sample preparation and storage. *Anal. Chem.* **2007,** 79, 1181-1186.

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

41. Murashige, T.; Skoog, F., A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **1962,** 15, (3), 473-497.

42. Boernsen, K. O.; Gatzek, S.; Imbert, G., Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Anal. Chem.* **2005,** 77, (22), 7255-7264.

43. Canals, I.; Portal, J. A.; Bosch, E.; Roses, M., Retention of ionizable compounds on HPLC. 4. mobile-phase pH measurement in methanol/water. *Anal. Chem.* **2000,** 72, 1802-1809.

44. Espinosa, S.; Bosch, E.; Roses, M., Retention of ionizable compounds on HPLC. 5. pH scales and the retention of acids and bases with acetonitrile-water mobile phases. *Anal. Chem.* **2000,** 72, 5193-5200.

45. Glasoe, P. K.; Long, F. A., Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **1960,** 64, 188-190.

46. Carlson, E., Integrating proteomics and metabolomics to map bacterial Encystment. In Riverside, CA, 2010.

47. Carlson, E. Indiana University, Bloomington, IN. Personal communication., 2010.

48. Gromova, M.; Roby, C., Toward *Arabidopsis thaliana* hydrophilic metabolome: assessment of extraction methods and quantitative ^{1}H NMR. *Physiol. Plantarum* **2010,** 140, 111-127.

49. Cui, Q.; Lewis, I. A.; Hegeman, A.; Anderson, M. E.; Li, J.; Schulte, C. F.; Wesfler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the Madison metabolomics consortium database. *Nat. Biotechnol.* **2008,** 26, 162-164.

50. Kaiser, K. A.; Barding, G. A.; Larive, C. K., Metabolic profiling of plants by ^{1}H-NMR: A comparison of metabolite extraction strategies using rosette leaves of the model plant *Arabidopsis thaliana*. *Magn. Reson. Chem.* **2009,** 47, (S1), S147-S156.

51. Bonnen, A.; Brambl, R., Germination physiology of *Neurospora crassa* conidia. *Exp. Mycol.* **1983,** 7, 197-207.

52. Thevelein, J. M., Regulation of trehalose mobilization in fungi. *Microbio.l Rev.* **1984,** 48, 42-59.

53. Asiago, V. M.; Gowda, G. A. N.; Zhang, S.; Shanaiah, N.; Clark, J.; Raftery, D., Use of EDTA to minimize ionic strength dependent frequency shifts in the ^{1}H NMR spectra of urine. *Metabolomics* **2008,** 4, (4), 328-336.

54. Karakoc, E.; Sahinalp, S. C.; Cherkasov, A. In *Comparative QSAR analysis of bacterial-, fungal-, plant- and human metabolites*, Pacific Symposium on Biocomputing, Maui, Hawaii, 2007; Maui, Hawaii, 2007; pp 133-144.

55. Fan, T. W.-M.; Higashi, R. M.; Lane, A. N.; Jardetzky, O., Combined use of ^{1}H-NMR and GC-MS for metabolite monitoring and in vivo ^{1}H-NMR assignments. *Biochem. Biophys. Acta* **1986,** 882, 154-167.

56. Stitt, M.; Lilley, R. M. C.; Gerhard, R.; Heldt, H. W., Metabolite levels in specific cells and subcellular compartments of plant leaves. *Methods Enzymol.* **1989,** 174, 518-552.

57. Szydlowski, J.; Szykula, M., Isotope effect on miscibility of acetonitrile and water. *Fluid Phase Equilib.* **1999,** 154, (1), 79-87.

58. American Society of Plant Biologists, *Biochemistry & Molecular Biology of Plants;* John Wiley & Sons: Somerset, NJ, 2000; p 1367.

59. Bligh, E. G.; Dyer, W. J., A Rapid Method of Total Lipid Extraction and Purification. *Can. J. Biochem.* **1959,** 37, 911-917.

60. Price, K. E.; Lunte, C. E.; Larive, C. K., Development of Tissue-Targeted Metabonomics: Part 1. Analytical Considerations. *J. Pharm. Biomed. Anal.* **2008,** 46, 737-747.

61. Dao, T. T. H.; Puig, R. C.; Kim, H. K.; Erkelens, C.; Lefeber, A. W. M.; Linthorst, H. J. M.; Choi, Y. H.; Verpoorte, R., Effect of benzothiadiazole on the metabolome of Arabidopsis thaliana. *Plant Physiol.Biochem.* **2009,** 47, 146-152.

Chapter Three

Oxygen Deprivation in Whole *Arabidopsis* Seedlings:

Inquiry into Primary Metabolism

Based in part on the research paper "Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*" Cristina Branco-Price, Kayla A. Kaiser, Charles J. H. Jang, Cynthia K. Larive, and Julia Bailey-Serres. The Plant Journal (2008) Volume 56, Issue 5, Pages 743-755.

In this work, CB-P used *Arabidopsis thaliana* expressing a dual epitope tag, His_6-FLAG (HF), on its Ribosomal Protein L18 (RPL18) driven by the constitutive cauliflower mosaic virus *35S* promoter.[1] CP-B performed polyribosome immunopurification, mRNA isolation and analysis, and enzymatic assays. Matched samples were given to KAK for metabolite extraction and ^{1}H NMR metabonomic analysis. CJHJ carried out Robust Multi-chip Average (RMA) normalization and analysis of microarray data using the open-source software package R. This work was supported by National Science Foundation grants IBN-0420152 and CHE-0616811 to JBS and CKL, respectively; National Science Foundation Integrative Graduate Education Research and Training Program fellowships (DGE-0504249) to KAH and CJHJ; and a Portuguese Foundation for Science and Technology graduate fellowship (SRFH/BD/9165/2002) to CB-P.

Systems biology focusing on nontargeted small molecule (< 400 Da) analysis employing NMR is called metabonomics (or metabolomics). This chapter describes such an investigation undertaken in the model plant *Arabidopsis thaliana.* In this work, seedling extracts were interrogated by one-dimensional ^{1}H NMR. Multivariate data analysis was applied to examine clustering and correlations on a global level. Targeted univariate statistical analysis revealed significant changes in the relative concentration of alanine (Ala), asparagine (Asn), γ-aminobutyric acid (GABA), glutamate (Glu), glutamine (Gln), lactate, succinate and valine (Val) after 9 hours of oxygen deprivation stress. Evidence from paired transcriptomic experiments supports the induction of

mRNAs encoding transaminases (for protein catabolism) and glyoxylate cycle (for lipid mobilization) in addition to the expected fermentation pathways account for plant survival of prolonged oxygen deprivation stress.

3.1 Introduction: Metabolic Adjustments under Low Oxygen

To interpret the results of a biologically centered study, it became necessary to delve into the current state of knowledge in low oxygen metabolism in higher plants. This introductory section is divided into a general commentary on the importance of oxygen in living systems, a justification for our choice of test system, and experimental parameters such as age, and timing of stress and recovery treatments.

3.1.1 Importance of Oxygen for Life on Earth

Oxygen is the third most abundant element in the universe, comprising 20% of the Earth's atmosphere, and is critical to sustaining life on earth. The utilization of oxygen in aerobic respiration is conserved across all eukaryotic organisms. Land plants, algae and cyanobacteria carry out photosynthesis using sunlight, carbon dioxide, and water, to produce oxygen. Oxygen is consumed in the final step of the mitochondrial electron transport chain, producing adenosine triphosphate (ATP), which is the energy currency within the cell. If oxygen becomes unavailable, the consequence is an energy crisis. Plant cells cope with shortages of ATP caused by insufficient oxygen by a diverse range of strategies.[2] At the metabolic level, a key response is production of ATP at the substrate level coupled with regeneration of NAD^+ through fermentation of pyruvate.

The goal of this chapter is to investigate how an external cue such as low oxygen invokes the cellular responses in the model plant *Arabidopsis thaliana.* It is useful to use the advanced molecular and genetic tools available, thus the most well-studied model Angiosperm was chosen for our study.[3] Its small size (~10 cm), short life cycle (~6 wk), and relatively small genome (~25,000 genes) provides enough complexity for a meaningful analysis while representing general plant hypoxic responses observed in crop plants such as wheat, rice, and barley.[4-7] One ecotype of Arabidopsis, called Columbia-0, has been particularly well characterized.[8]

Oxygen deprivation and eventual anoxia in plants can arise naturally as a result of flooding, microbial blooms, tissue density, high metabolic activity, and soil compaction. Crop damage due to flooding stress, ranging from root waterlogging to complete submergence, can lead to substantial financial losses. Studying plant responses to oxygen deprivation can answer basic biological questions leading to a better understanding of how energy is used by stressed organisms. This work represented the first study in either plants or animals of the contribution of transcriptional and translational regulation to the initial and prolonged adjustments in metabolism and other biological processes that aid survival of oxygen deprivation stress. The reversal of regulation following a short period of reoxygenation was also evaluated.

In this study, oxygen deprivation was simulated using Argon gas as a substitute for air, displacing both oxygen and carbon dioxide, limiting aerobic respiration and photosynthesis. Arabidopsis seedlings were grown on 10 cm agar plates containing macro- and micronutrients and sucrose which fit neatly into stress chambers.[9] Long-day

conditions (16 h light, 8 h dark), uniform temperature (23^{o}C) and light level (50 μM m^{-2} s^{-1}) ensured efficient seedling germination, with roughly 100 seedlings per plate. At growth stage 1.0 (cotyledons fully open), two hours into the light, seedlings were deprived of air gradually via displacement by argon gas in a low light environment (6.5 μM m^{-2} s^{-1}) simulating turbid flood waters.

Biochemically, primary metabolism refers to the reactions necessary to sustain life. The pathways involved are generally presented in the following order; glycolysis, the tricarboxylic acid (TCA) cycle, mitochondrial electron transport and oxidative phosphorylation, lipid metabolism, photosynthesis, and finally nitrogen metabolism.[10] A general representation of selected pathways is provided by Figure 3.1. This figure was constructed to interrelate small molecule metabolites for which we had obtained pure standards. Some of these molecules were determined to be responsible for signals in our Arabidopsis extracts, others have been identified in metabolomic experiments in other living systems. We observe flux through these pathways by accumulation of an endproduct or depletion of a substrate. We sought out congruous information among transcript abundance overall and enrichment in polysome complexes or sequestration under stress. Some of these pathways will be explored individually regarding its body of literature. Less well-studied is the half-life as well as degree of translation of individual transcripts.

Lipid metabolism will not be discussed here, as we have undertaken the analysis of predominantly hydrophilic metabolites in this work. Evidence for the operation of mitochondrial electron transport and oxidative phosphorylation in this work is somewhat

indirect, although this process is largely impaired by oxygen deprivation, since the final electron acceptor is molecular oxygen. Also, photosynthesis was not studied directly in this work despite its ability to operate, albeit somewhat less efficiently, in the generation of ATP even without oxygen. Overall, statistically significant changes in metabolites involved in nitrogen metabolism were observed in response to oxygen deprivation. Nitrogen metabolism is a route for NO production, which has been proposed as a signaling molecule of low oxygen status in plant tissues.[11] Therefore nitrogen metabolism will be further detailed.

Glycolysis in plants begins with the catabolism of sucrose, and terminates in pyruvate, which is considered a metabolic hub[12] and is probably a primary site of regulation for glycolysis and respiration in mammals, bacteria, and plants.[13] Anaerobic fermentation is responsible for regeneration of NAD^+ for maintaining glycolysis, a "core" hypoxic response in plants.[14-16] When ATP levels are insufficient to sustain the classical TCA cycle as shown in Figure 3.1, alternate routes must be activated for consumption of pyruvate. Statistically significant changes in metabolites involved in the GABA shunt are observed in response to oxygen deprivation along with accumulation of succinate. Disruption of the urea cycle has been observed in rats under oxygen deprivation. Choline represents a precursor to membrane phospholipids, as well as the osmoprotectant glycine-betaine,[17-19] providing metabolic flexibility under environmental stress.

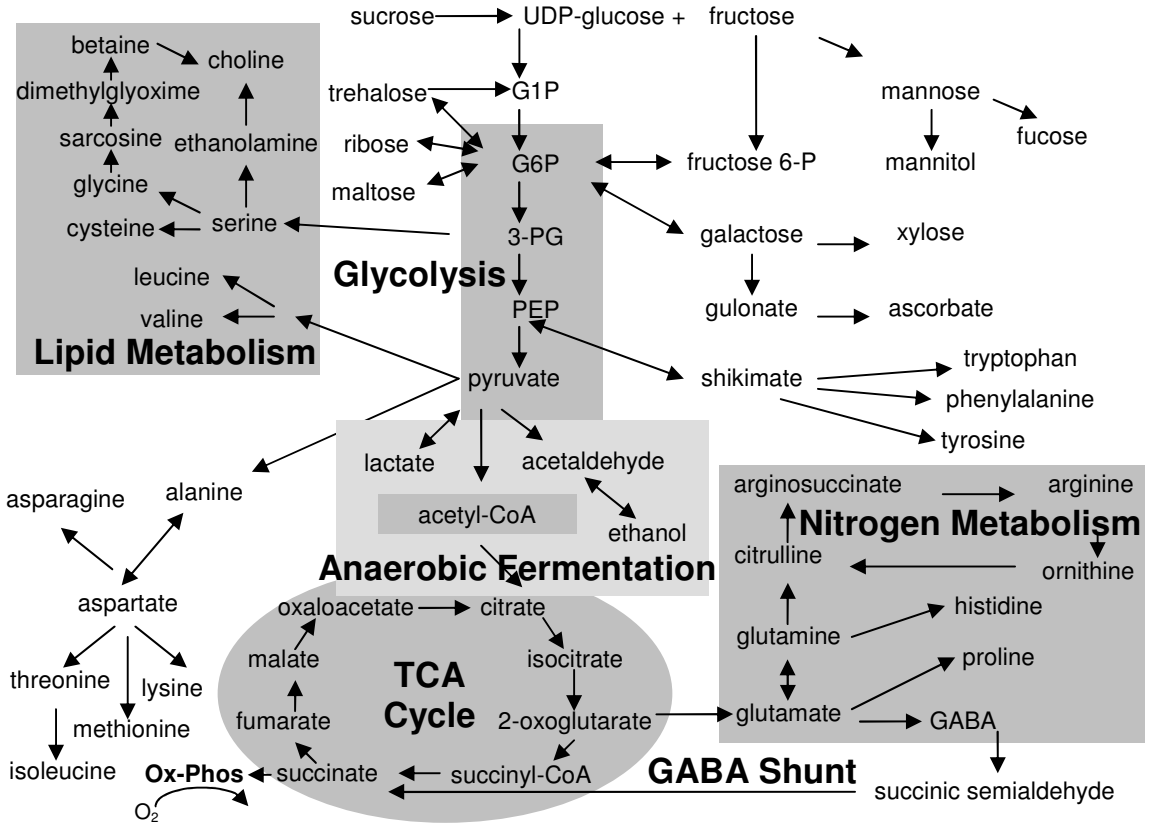


Figure 3.1. Metabolites and pathways that can be probed by ^{1}H NMR-based metabolic profiling. Pathways are highlighted to emphasize the predominant cycles, sources, and sinks.

3.1.2 Arabidopsis and Systems Biology: Cross-Validation

A transgenic Arabidopsis line expressing an engineered FLAG(His_6)-tagged ribosomal protein L18 (RPL18) that allows efficient immunopurification (IP) of ribosome complexes was used because it allowed easy isolation of mRNAs associated with ribosomes, presumably undergoing translation.[1,20] Genes which are selectively translated under stress by ribosome complexes are a subpopulation or in some cases distinct from constitutively expressed genes at the same timepoint. This study aimed to inform about active translational status as well as reveal which messages are sequestered until the plant

is released from stress.[21-23] The aim of this study was to compare plant responses under conditions of stress and non-stress with respect to translation.

The primary rate-limiting step in mRNA translation is the initial recruitment of the 43S-preinitiation complex, requiring multiple molecules of ATP. Translational regulation can be achieved by mRNA sequestration, a process that limits specific mRNAs from polysome complexes. Because mRNAs may be present and not translated, metabolic profiles should be more reflective of mRNA profiles resulting from immunopurified mRNA-ribosome complexes rather than the total (steady state) mRNA pool.

Commercially available GeneChips (Affymetrix, Santa Clara, CA) are capable of reporting the abundance of mRNAs for approximately 24,000 genes. Quantitative readout is accomplished by preparation of fluorescently labeled cRNA followed by hybridization to 25-mer capture oligonucleotides covalently attached in picomolar amounts to a one square centimeter area representing 1.3 million distinct pixels or features. Redundancy within the ATH1 GeneChip is accomplished by the presence of 11 probe pairs per sequence, which can be converted to transcript abundance by accompanying software. This platform is supported by genome annotation by The Arabidopsis Information Resource (TAIR),[24] who maintain and improve the annotation of probe pair sets and genes represented by them. GeneChips are now available for human, mouse, rat and 28 other organisms. About one-third of the available resources for chip-based transcriptome analysis offered by the company Affymetrix, belong to plants.

This work provided a novel approach to efficient immunopurification of actively-translated mRNAs, allowing us to demonstrate that the translatome (ribosome-associated mRNAs isolated by immunopurification) gives a better picture of molecular responses to oxygen deprivation than the transcriptome (total mRNA). Our contribution to this study was to monitor small molecules that are involved in primary metabolic pathways by applying a metabonomic approach with the hypothesis that the stress-induced adjustments in the translatome would more accurately reflect changes in the metabolic state in response to low-oxygen stress, temporary (2 hour) or prolonged (9 hour), compared with respective normoxic controls. Additionally a sample was harvested after 1 hour recovery at normoxia following release from prolonged stress.

Microarray experiments were performed in parallel with NMR-based metabolic profiling (Figure 3.2) to identify and quantify individual transcript levels in total mRNA population under stress and control conditions. Furthermore, we aimed to identify transcripts and thereby cellular processes that are repressed at the translational level.[21, 25] Ten bioreplicates were prepared. All plates were pooled and partitioned into microcentrifuge tubes in 100 mg (wet weight) portions. These samples were used for mRNA extraction, perchloric acid and acetonitrile:water extractions

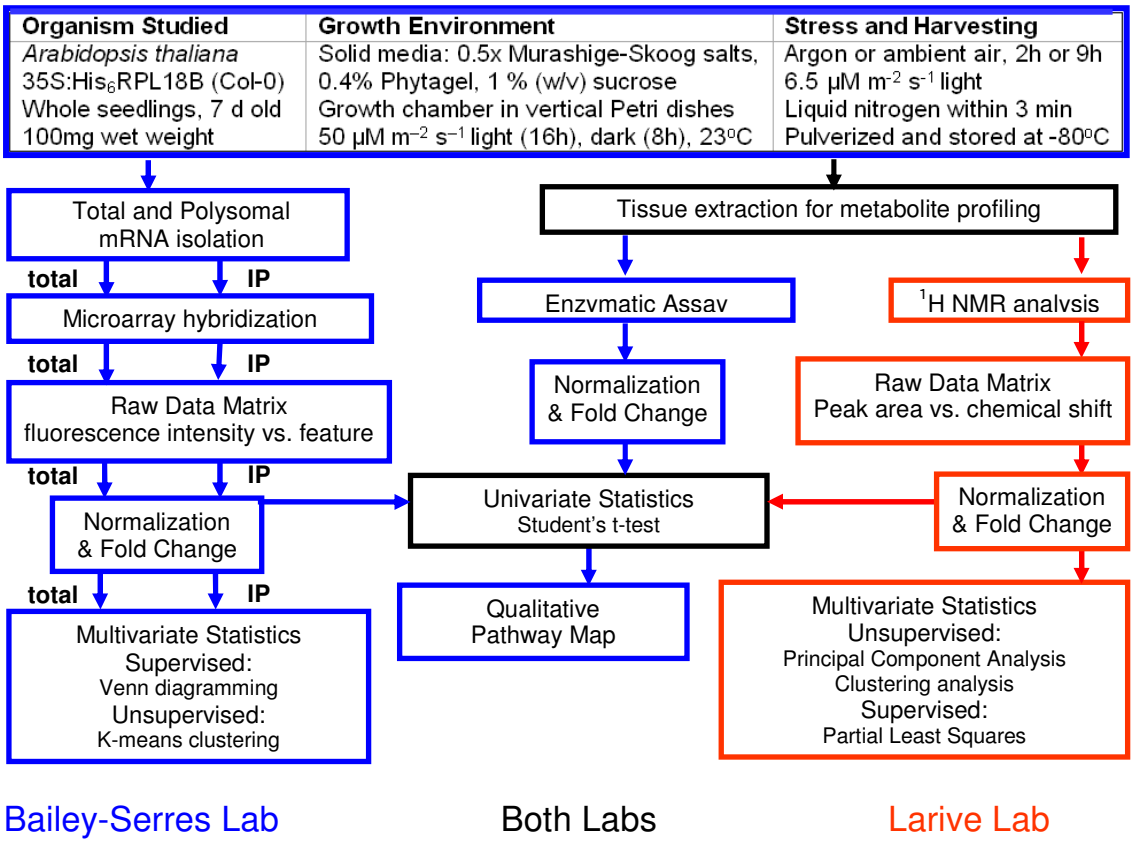


Figure 3.2. Experimental design to study the effect of oxygen deprivation stress and reoxygenation on seedlings of the model plant *Arabiodopsis thaliana* using a combined mRNA and metabolite profiling approach.

for enzyme-based metabolite quantificiation and ^{1}H NMR-based nontargeted metabonomics, respectively.

Arabidopsis is a model organism and its genome is among the first three complete sequences to have been determined.[26] Complementarity between systems biology and classical approaches has been elegantly articulated[27] and both approaches continue to be useful. By taking a unified approach employing simultaneous quantitative determination

of (1) mRNA sequestration dynamics and (2) the far downstream evidence of translation – substrates and products of metabolic reactions, we provide future evidence of how plants have adapted at the cellular level to survive low oxygen stress. For these studies, sophisticated molecular biology techniques were used for handling of large molecules (>8000 Da), involving efficient immunopurification using an epitope tag on a component of a ribosome bound to actively-translated mRNAs.

Enzymatic assay for a handful of molecules, as well as non-targeted direct measurement of abundant small molecules (<400 Da) by ^{1}H NMR, were performed in parallel to establish the validity of this the approach to metabolite measurements (Figure 3.2). Once we had metabolite and transcript profiles of several bioreplicates (N=3-4), we determined fold changes (FC) in transcript and metabolite abundance by dividing normalized signal intensity under stress by normalized signal intensity of untreated controls for each variable. Correlated variables within data matrices can assist with signal assignment and metabolite identification. Pathway analysis between transcript and metabolite data was carried out using publicly-available open-source software (MapMan,[28] OmicsViewer[29]). Qualitative comparisons were also made by analysis of mRNA functional annotation and consensus with corresponding metabolic pathways regulated by the resulting gene product.

3.1.3 Time Points for Growth, Stress, Harvest, and Recovery

When planning a large omics study, experimental design is the first step in the pipeline. For clarification see Chapter 1, Section 1.2 which outlines compounding factors

that impact the analytical data collection. More generally, considerations must be given to any variable which can impact the data. Rigorous reporting standards have been set forth by some of the leaders in our field.[30] With this level of detail in mind, here the process of experimental design is articulated as it relates to the data we collected.

It is known that short-term oxygen deprivation induces a general repression of protein synthesis.[31] In the absence of a steady supply of ATP, such energetically costly processes are reduced. Short-term oxygen deprivation is distinguished by accumulation of the nonvolatile fermentation endproduct, lactic acid, below two-fold relative to control levels. Oxygen deprivation stress is considered long-term when lactate accumulates more than two-fold above the control.[32, 33] A mechanism for lactate formation[32] is shown in Figure 3.3. In the short-term, glucose can be mobilized to pyruvate and aspartate (Asp) can be used as a substrate to regenerate the intermediate NADH. The enzymes catalyzing these processes are optimized to function around neutral pH. As early mechanisms begin to fail, the pH balance within the cell is disrupted and the intracellular environment becomes acidified.[2] At least one enzyme required for ethanolic fermentation is optimized for function at a reduced pH.

By recording observations at short- (2 h) and long-term (9 h) time points, we hoped to observe some component(s) of an acclimation response at early time points that regulated cascades of responses at later time points.[34] These may represent branchpoints

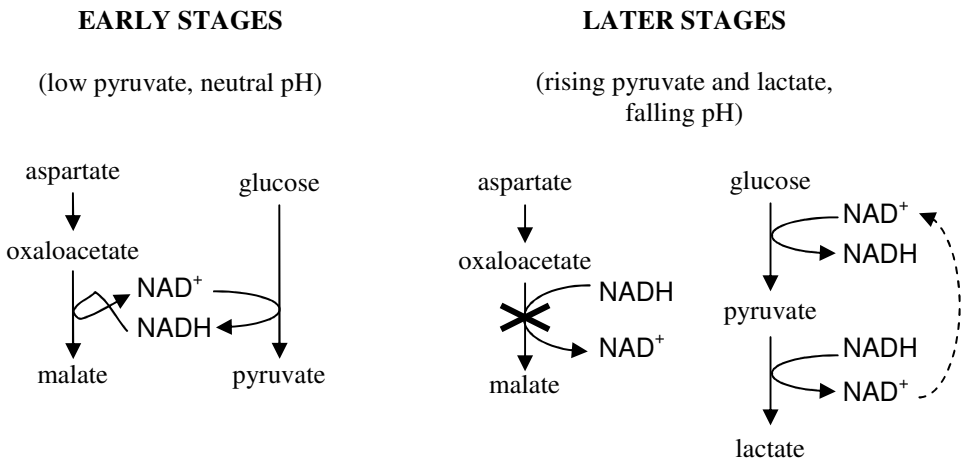


Figure 3.3. Anaerobic glycolysis provides NAD^+ via coupling with malate dehydrogenase or activation of lactate dehydrogenase. In plants, the decreased cytosolic pH following anoxia is correlated with increased pyruvate decarboxylase activity, catalyzing the conversion of pyruvate to acetaldehyde, which is consumed by alcohol dehydrogenase to produce ethanol and simultaneously regenerate NAD^+.

in metabolic pathways or transcription factors responsible for the regulation of downstream processes.[35] For example, it has been shown that factors such as low pH at early time points is responsible for induction of adaptive strategies at later time points. Our goal was to accomplish a quantitative evaluation of translational and metabolic adjustments under hypoxic stress, both under short- and long-term scenarios to determine the importance of selective mRNA translation under ATP-limiting energy conditions. We hypothesized that consumptive and non-vital processes are downregulated or repressed via a mechanism that expedites the return to homeostasis upon reoxygenation.

Prior to modern advances in molecular biology, absorbance spectroscopy was a standard method for translational studies. This study included polysome fractionation to portray global translational status, or polysome loading of mRNAs. Arabidopsis plants were homogenized in an extraction buffer and subjected to fractionation in a sucrose

gradient under ultracentrifugation. Free ribosomal subunits, both small (40S) and large (80S), were resolved from ribosome-mRNA complexes and polysome-mRNA complexes. The UV absorbance profile along the centrifuge tube was integrated for quantification of translation status. So-called polysome profiles resolved under a sucrose gradient can reveal a 35% general reduction in translation due to oxygen deprivation, which is reversed upon 1 hour of recovery.

The Bailey-Serres lab[36, 37] revealed that both short- (2 hours) and long-term (9 hours) treatment with argon gas resulted in global repression of translation of ~50% of cellular mRNAs --consistent with a ~50% reduction in polysome content, as well as the stabilization of ATP reserves at ~50% of pre-stress levels. Although primary root meristems did not survive the stress, seedlings initiated lateral roots upon recovery.

3.1.4 Replicates and Variability

Since variables in large data matrices tend to be highly correlated, a smaller number of bioreplicates is necessary to achieve statistical power. In this study 10 bioreplicates were available for analysis. Only as many were used as was necessary to construct a robust model (N = 4) so that excess tissue could be archived for further study.

For a comparative approach to be successful, many experimental details must be regulated throughout sampling and sample handling prior to analysis. Some uncertainty might be related to temporal variations between samples. Furthermore, variations in temperature, light levels, and microenvironments within the growth chamber contribute to this uncertainty. Moreover, each seed has a unique carbohydrate reserve, which yields

plants of different character even when treated identically. These effects were mitigated by pooling 50 plants to comprise one bioreplicate. The entire experiment was repeated on multiple days, so there may be small differences related to duration and intensity of stress treatment, attributable to variance in time between harvesting and storage among bioreplicates (~20-50%). When all samples were prepared, variance is also introduced during defrosting and extraction of tissue, duration of extraction and time between extraction and analysis (~3-8%). In a study involving over 30 participating laboratories to demonstrate the analytical robustness of quantitation by NMR, the authors reported a maximum combined measurement uncertainty of 1.5% with a confidence interval of 95%. Quantitation by NMR in intralaboratory comparisons[38] was found to have an uncertainty of the result of 0.5% for high resolution ^{1}H and ^{31}P NMR measurements.[39] Overall, the reliability of the NMR experiment is higher than the reproducibility of the extraction and the reproducibility of the extraction is biological treatment, so analytical and technical replicates were not recorded.

3.1.5 Justification for Choice of Analytical Technique

A common alternative technique to ^{1}H NMR-based metabolic profiling is gas chromatography coupled to mass spectrometry (GC-MS),[40-43] and many investigators utilize both methods. Although GC-MS methodology is perhaps more widely used in plant biology, the inherent quantitative nature of NMR and minimal sample preparation were our reasons for choosing this technique. The case for systems biology, in particular NMR-based metabonomics, is succinctly stated by Susan O'Connor of Merck:

> "A strong advantage of using NMR for metabolic profiling is the potential for revealing unexpected changes which otherwise would not be possible to predict without extensive knowledge of the underlying mechanisms of compound efficacy or disease. This advantage is facilitated by the ability of NMR to reproducibly detect structural information for all hydrogen-containing compounds present above the detection limit." [44]

For its inherently quantitative nature and power to determine metabolite identity through nondestructive suite of 2D experiments, NMR was chosen to be the sole analytical technique employed in this work.

Specific molecules that can be quantitatively monitored are selected amino acids, aliphatic acids, and sugars (see pathway diagram in Figure 3.1 and chemical shifts of selected metabolite resonances in Figure 3.4). This NMR-based metabolic fingerprinting approach has proved useful in understanding the mechanisms of survival plants have adapted to withstand herbivory,[45] salt stress,[19] temperature stress,[46, 47] light stress,[4] chemical stress,[48] and nutrient availability.[46, 47, 49, 50]

3.2 Experimental Section

3.2.1 Sample Preparation

Arabidopsis seedlings grown on Petri dishes on solid agar [containing half-strength (0.5x) micro- and macronutrients (MS salts) and 1% (w/v) sucrose], were transferred 2 hours into the light cycle from a growth chamber (after 7 d, 16 h 50 μM m^{-2} s^{-1} light, 8 h dark, 23°C) to stress chamber (6.5 μM m^{-2} s^{-1} light). Durations in stress

chambers were 2 hours (2HS, 2NS), 9 hours (9HS, 9NS) or 10 hours (1R). Treatment with Argon gas resulted in oxygen deprivation stress condition (HS), while chambers left open to laboratory air provided a control condition (NS). A recovery group (1R) was treated with Argon gas for 9 hours and exposed to laboratory air for 1 hour. Plant cultivation, stress treatments, and harvesting were performed by C. B.-P. Whole seedlings were collected and frozen under liquid nitrogen; tissue was pulverized and stored at -80 ^{o}C until use. For more metadata, see Chapter 2. These steps were summarized in Figure 3.2, where clear detail regarding MSI-designated minimum required metadata are provided.

Metabolic profiling was conducted by ^{1}H NMR of 50 whole seedlings by homogenization in 700 μL of 1:1 (v/v) acetonitrile-d_3:deuterium oxide containing 50 mM sodium acetate-d_3, 50 mM acetic acid-d_4, 1 mM of 3-trimethylsilylpropionic acid-d_4 sodium salt (TMSP-d_4), adjusted to pH 5 with deuterium chloride. Centrifugation for 4 minutes at 5000 rpm (~6,000 x *g*) was employed to precipitate proteins[51] and remove cellular debris after 4 minutes of manual homogenization with a micropestle (Eppendorf). An aliquot (600 μL) of supernatant was transferred to a 5 mm NMR tube. Four independent biological replicate experiments were used for each stress treatment or control.

Metabolite standards were also measured, including amino acids, participants in the tricarboxylic cycle, mono- di- and tri- saccharides, metabolic coenzymes/cofactors, and plant hormones. Standards for each metabolite were analyzed in the same buffer as

the tissue samples, adjusted to the average sample pH meter reading (samples: 4.56 $\pm$ 0.04 across the dataset, standards: 4.58 $\pm$ 0.12 out of 48).

3.2.2 NMR Data Acquisition

NMR spectra were recorded using a Bruker Avance Spectrometer operating at 600.06 MHz using a water presaturation pulse sequence. A relaxation delay of 1.5 seconds was followed by selective saturation (50 dB) of the residual water (HOD) resonance. A spectral width of 7716 Hz was excited by a 90^{o} proton pulse (7.82 μsec, -5 dB). Free induction decays (FIDs) were acquired into 25860 time points. An acquisition time of 1.675 sec was sufficient to prevent truncation artifacts. Compromise between adequate signal-to-noise ratio (S/N) for quantitation (>250) and sample throughput necessitated coadding 1200 scans per bioreplicate following 16 dummy scans for a total experiment time of 64 min. The temperature of the sample was maintained at 298 K. Manual shimming was performed for each sample until the TMSP line width at half-height was less than 1.5 Hz.

3.2.3 NMR Data Processing

All NMR spectra were zero filled to 131072 points and apodized by multiplication by an exponential function equivalent to 1.0 Hz line-broadening prior to Fourier transformation. Variance introduced by manual shimming between samples can be partially compensated for by applying exponential multiplication to each spectrum such that the resulting linewidths are similar across the dataset.[52] Reported metabolite ^{1}H

NMR chemical shifts (δ) were determined relative to TMSP-d_4 (0.00 ppm). Using Topspin 2.0 (Bruker), phasing and baseline correction were applied automatically.

3.2.4 Selection of Integral Regions

Once a set of NMR spectra has been recorded for a series of given biological treatments, a data matrix is generated to fit the pattern indicated in Table 3.1. Samples are formatted as rows and responses are organized as columns. The limits of integration fit one of two general patterns depending on the knowledge outcome desired.

3.2.4.1 Nontargeted Metabonomics

In these experiments, each spectrum was integrated using a set of equidistant regions of 0.02 ppm width over the range of 0.5-9.5 ppm. Integral regions containing the resonances of buffer and solvents, acetate and acetonitrile (1.96 - 2.04 ppm), and HOD (4.28 - 4.48 ppm) were excluded. These non-targeted integrals are of a sufficient size to contain one linewidth, therefore each integral region is may be representative of a particular metabolite. Analysis of this type may be considered exploratory, or hypothesis-generating, and is useful in gaining a better understanding of interrelationships among the treatments and variables. Global information was obtained by utilizing bins representing the relevant frequencies of the ^{1}H NMR spectrum.

Table 3.1 Raw data format for multivariable (nontargeted) metabonomics. Samples are listed in rows. Integral regions at intervals of 0.02 ppm contribute 427 variables for each sample. The constant sum normalization factor is calculated by summing all column values.

	Integral region x_1	…	Integral region x_n	Constant Sum $\Sigma x_{1\ldots n}$
2NS2	$x_{2NS2,1}$	…	$x_{2NS2,\,n}$	Σx_{2NS2}
2NS3	.		.	.
2NS7	.		.	.
2NS9	.		.	.
2HS2	.		.	.
2HS4	.		.	.
2HS6	.		.	.
2HS8	.		.	.
9NS1	.		.	.
9NS3	.		.	.
9NS5	.		.	.
9NS9	.		.	.
9HS3	.		.	.
9HS7	.		.	.
9HS8	.		.	.
9HS9	.		.	.
9+1 2	.		.	.
9+1 5	.		.	.
9+1 6	.		.	.
9+1 9	$x_{9+1\ 9,\ 1}$	…	$x_{9+1\ 9,\ n}$	$\Sigma x_{9+1\ 9}$
Avg	μ_1	…	μ_n	

3.2.4.2 Targeted Quantitative Analysis

Once the data had been explored in a non-targeted manner, a second set of integral regions were manually selected, ensuring that peak maxima of biologically important metabolites were not split between separate integral regions, for the purpose of quantifying well-resolved peaks. When possible, multiplets were bucketed into a single integral region, thereby increasing S/N, provided that part of a multiplet was not obstructed by overlap with resonances from other metabolites.

3.2.5 Spectral Editing

Another way of visualizing trends in the data is by visual inspection of a set of NMR spectra. The human mind has an amazing capacity to recognize patterns. Despite our best efforts to remain strictly objective, spending time looking at raw or normalized spectral data can sometimes lead to promising hypotheses which can be refined by further more objective comparison. The data presented in Appendix A, together with Figure 3.4 presents subtracted spectra overlaid with spectra of selected metabolite standards to illustrate how this type of "fingerprinting" is accomplished by visual inspection.

In this work, spectral addition of bioreplicates was used to improve signal to noise and average out sample-to-sample variation, followed by spectral subtraction of the average control spectrum from the average stressed spectrum. In this way, metabolites which accumulate or become depleted under oxygen deprivation can be visualized spectrally when obstructing metabolites are not changing in response to stress. This was performed for visualization of the hypoxic induction of the metabolite lactate, which was obscured by a broad resonance most likely generated by methylene protons on fatty acids. The challenge to spectral editing is that it works best when pH is tightly controlled to prevent pH-dependent chemical shift changes in metabolites with titratable groups having pK_a values around the pH of the samples. The drawback to spectral subtraction is that it is no longer possible to quantitate the magnitude of a fold change from a subtracted spectrum, instead only a rough estimate of the change in metabolite abundance is possible.

Spectral editing was carried out to estimate fold changes for metabolites found in regions where resonance overlap precluded direct integration.[53, 54] Subtraction of an average control spectrum from stressed spectra has been used elsewhere in plant metabolomics to visualize spectrally changes in metabolites as a function of treatment.[55] In sample preparation, both treatments were handled identically to yield aqueous extracts at the same ionic strength and pH. If these conditions have been met, spectral subtraction can be applied to find the difference between treatment and control metabolic profiles. This difference spectrum can more easily reveal metabolites that are affected by biological treatment.

3.2.6 Scaling and Normalization

One advantage of NMR is its inherent quantitative nature; if identical acquisition parameters have been used, the integrated intensity of a resonance will be directly proportional to the number of nuclei responsible for generating the signal. However, in biological studies it is desirable to quantitatively compare the amount of a metabolite in one tissue sample to the amount of that same metabolite in a different tissue sample. This quantitative comparison could be achieved by scaling the detector response to the original tissue mass. However, in these experiments differences in tissue water content as well as contamination of seedlings from the growth media has made such normalization less effective. Efforts to build consensus on a "housekeeping" metabolites to use for internal quantification have so far been unsuccessful for metabolite studies.[56]

In general, metabonomic data cannot be reliably interpreted without cleaning.[57, 58] These issues were addressed in Chapter 1; in Chapters 3-4 metabolite data was normalized to the total integrated area of its NMR spectrum, termed the "constant sum."[59, 60] There are several advantages to constant sum normalization: it accounts for sample-to-sample variability regarding the nature of the original sample (i.e. water content), resonances from contaminants can be excluded from the constant sum calculation, and the method may even account for differences in extraction efficiency. The qualifications for dividing the NMR spectra into integral regions are articulated in the following sub-sections (3.2.7 and 3.2.8) depending on the desired knowledge outcome.

Scaling and centering are common processing steps. In this study, $\log_2$ scaling was applied to calculated fold changes to moderate the effects of dramatic up- or down-regulation of transcripts. Metabolite fold changes were in general less dramatic, and therefore, metabolite data was not scaled. Mean-centering is a technique used to shift the centroid of the entire dataset towards the mean and therefore any resulting comparative plots will be centered around the origin (0,0). Mean-centering does not affect clustering outcomes, but only centers the resulting plot around zero. It is accomplished by calculating the mean of each variable (μ) and subtracting each datapoint by its mean, as shown in Table 3.1.

3.2.7 Unsupervised Data Evaluation

Multivariate data analysis was carried out using Minitab® 14 statistical software. Principal component analysis (PCA) was used to examine clustering and correlations

within the entire data set. Clusters were generated using the scores and the confidence level in NMR spectral assignments were supported by the loadings. The PCA model was constructed using all samples.

PCA was chosen as a statistical method because it provides a global overview of large multivariate datasets. The reduction in dimensionality imposed by this transformation makes the detection of sample groupings very simple as each sample is represented by one point on the scores plot. Samples that appear close together on the graph are said to 'cluster,' which is indicative that they have covarying signals and may be similar in overall biochemical composition.[44]

The loadings plot can be useful to identify covarying integral regions, which may facilitate assignment of signals to metabolites. Most metabolites produce more than one ^{1}H NMR resonance. A significant problem in one-dimensional proton NMR-based metabolic profiling is that metabolite resonances may be directly or partially overlapped. There are ways that one can deconvolve[54, 61, 62] the signal of biological interest from the background by careful design of biological specimen collection with proper negative controls.

3.2.8 Supervised Statistics

A Student's *t*-test[63] was performed to evaluate significant differences between control and stressed populations for each variable chosen to represent a single metabolite. The calculated *t* (t_{calc}) was compared with a table of values[63] using the following formula.

$$t_{calc} = \left[\frac{\left| \bar{x}_s - \bar{x}_c \right|}{s_{pooled}} \right] * \sqrt{\frac{N_s N_c}{N_s + N_c}} \qquad (3.1)$$

Using Excel 2003, t_{calc} was converted to p-value to establish precise differences for comparisons taking into account the cumulative sample variances. Changes in metabolite levels were ultimately reported analogously to microarray experiments, either as fold change or signal-log ratio.

3.2.9 Integrating "-omic" Datasets

Comparison between Affymetrix GeneChip (ATH1) transcript data and ^{1}H NMR metabolite data was performed manually. The software OmicsViewer and MapMan were employed for cross-platform integration, but were found to be of little use. A customized pathway diagram summarizing our findings was determined to be a more insightful and directed approach to reporting. Although this is counter-intuitive to the growing movement of standardization, the dataset is freely available (Supplementary Table S2 in Branco-Price et al.[64]), and anyone is invited to browse it at will along with corresponding metabolite information presented in Table 3.2.

A formidable challenge in metabolite analysis is that metabolomes are complicated. Metabolic reactions are highly branched and interconnected. A steady-state snapshot may not provide enough information about a sum of cellular processes. The aim of these experiments was twofold (1) to observe global metabolic changes upon early and late energy crisis and recovery in a nontargeted way, (2) to quantify specific target molecules involved in biochemical pathways of relevance in oxygen deprivation

stress by interrogating the data in a unique way to follow-up on a hypothesis generated by global analysis. Our hypothesis was that steady state metabolite concentrations captured at early and late time points would inform us about the activation of various pathways as a response to oxygen deprivation. We hoped to infer some knowledge regarding fluxes of carbon, nitrogen, sulfur and phosphorus resource allocation within the plant under energy crisis by showing complementary activity between the metabolome, the transcriptome, and the translatome.

The approach described in this chapter is to identify specific metabolites in a direct extract of wet plant tissue by comparing the spectrum of the extract to that of authentic metabolite standards in the same buffer. Well-resolved peaks were integrated to provide relative quantitation. Spectral subtraction of control from stressed extracts can provide information on metabolites whose resonances are overlapped with metabolites that do not respond to oxygen deprivation treatment. Spectra of this type are shown together with metabolite standards in Appendix A.

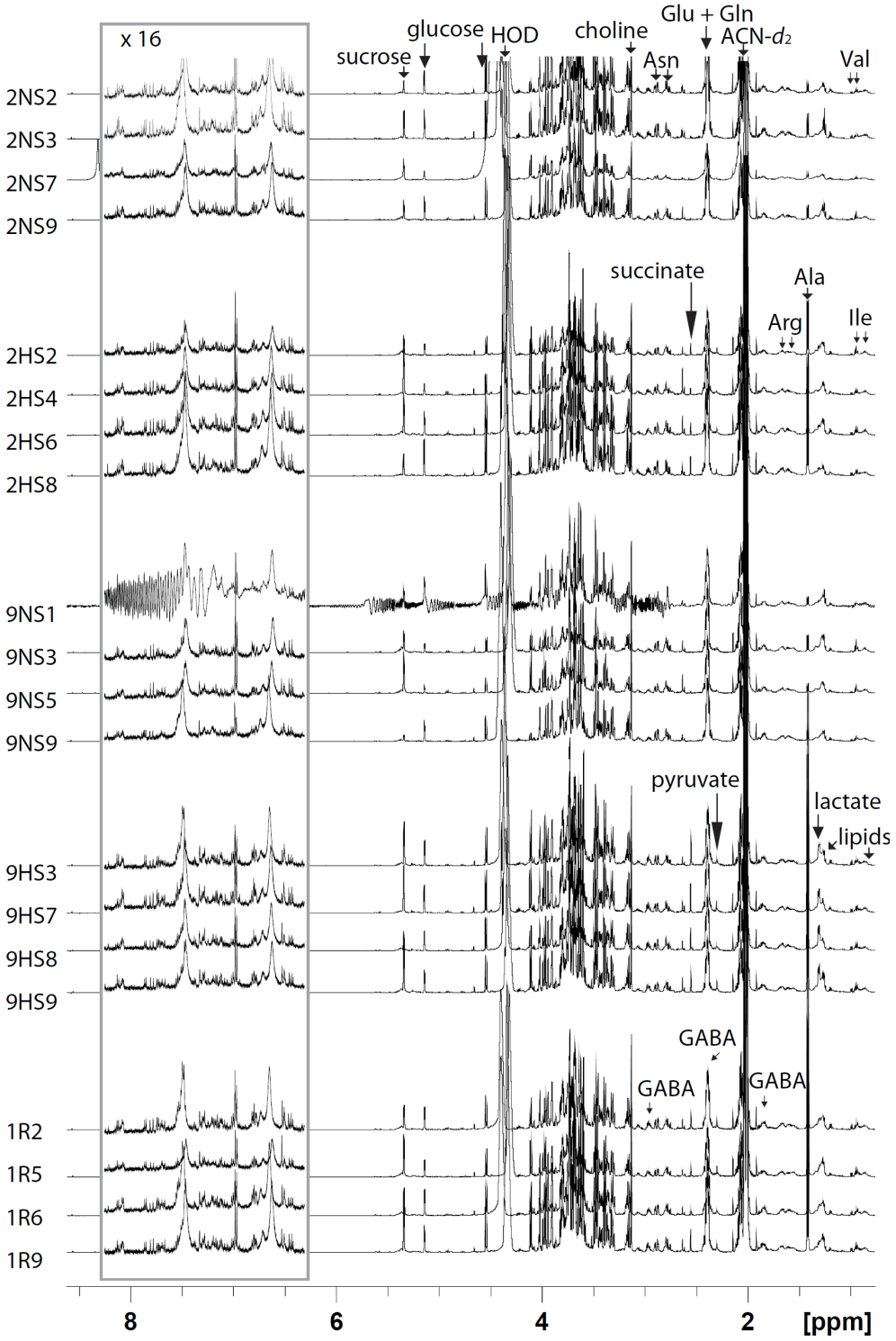


Figure 3.4 ^{1}H NMR spectra of direct tissue extracts. Resonances of sugars, organic and amino acids appear highly overlapped. Two hour control (2NS) and 2 hour hypoxia (2HS) stress are distinct from 9 hour control (9NS), 9 hour stress (9HS), and 9 hour stress after 1 hour recovery (1R).

3.3 Results and Discussion

Figure 3.4 shows representative ^{1}H NMR spectra of direct tissue extracts. Direct extract implies that the metabolic activity of a population of 50 seedlings (7 days old) was quenched with liquid nitrogen. Wet tissue was extracted in 1:1 acetonitrile-d_3:water-d_2 maintained at a pH meter reading of 4.55 by deuterated acetate buffer (100 mM). Treatments shown in Figure 3.4 from bottom to top include 2 hour non-stress whole seedlings (2NS), 2 hours oxygen deprivation-stressed whole seedlings (2HS), 9 hours non-stress whole seedlings (9NS), and 9 hours oxygen deprivation-stressed whole seedlings (9HS). A population allowed to recover for 1 hour after 9 hours oxygen deprivation (1R) was also prepared in parallel, but without a 10 hour negative control.

Resonances of branched-chain amino acids (Val, Ile) are highest upfield. A broad signal is contributed by lipids, but a sharp doublet from lactate is visible in stressed conditions. Ala can be observed to accumulate under stress. Substrates of nitrogen metabolism including arginine (Arg), ornithine, lysine (Lys), Glu, Gln, Asn and choline generally maintain homeostasis even under stress, if anything becoming limited/depleted.

Small organic acids succinate and pyruvate contribute low intensity singlet resonances, but can be seen to accumulate under oxygen deprivation stress. Sucrose, glucose and fructose were the dominant sugars observed in Arabidopsis extracts, however these may have been contributed by growth medium and not representative of endogenous sugar levels. Resonances in the aromatic region, even when enhanced 16-fold, are not of sufficient S/N ratios to be quantified by this method. This much

information we were able to gain by visual inspection, for an unsupervised approach, we turned to multivariate data analysis of our one-dimensional ^{1}H NMR metabolite profiles.

3.3.1 Multivariate Data Analysis of NMR-based Metabolic Profiles

An in-depth discussion of the merits and approaches of chemometric data evaluation are presented in Chapter 1, Section 1.5.1. The highly correlated variables in the data resulting from four bioreplicates via NMR-based metabolic snapshots warranted the application of PCA for unsupervised pattern recognition, dimensionality reduction, and identification of global trends. The scores plot (Figure 3.5) revealed a cluster formed by all control samples. There was good agreement even though one set was collected 4 hours into the day whereas the other was collected 11 hours into the day. This suggests that basic aerobic respiration in the presence of light dominates the metabolite profile of these samples.[4, 65-67]

The next distinct cluster revealed by PCA was the 2 hour stress group. Although these samples were prepared over several weeks in a 2 ½ month period, there is clear agreement and distinction from controls.[68-73] The third cluster is occupied by both 9 hour stressed seedling extracts and those seedlings which were stressed for 9 hours and allowed 1 hour to recover. This suggests that it will take longer than 1 hour for global metabolite profiles to recover to 2 hour stress levels and those of control seedlings. These metabolic differences may be accounted for by the activity of aminotransferases,[74-77] which have been shown to be involved in posthypoxic recovery in both plant and mammalian tissues.

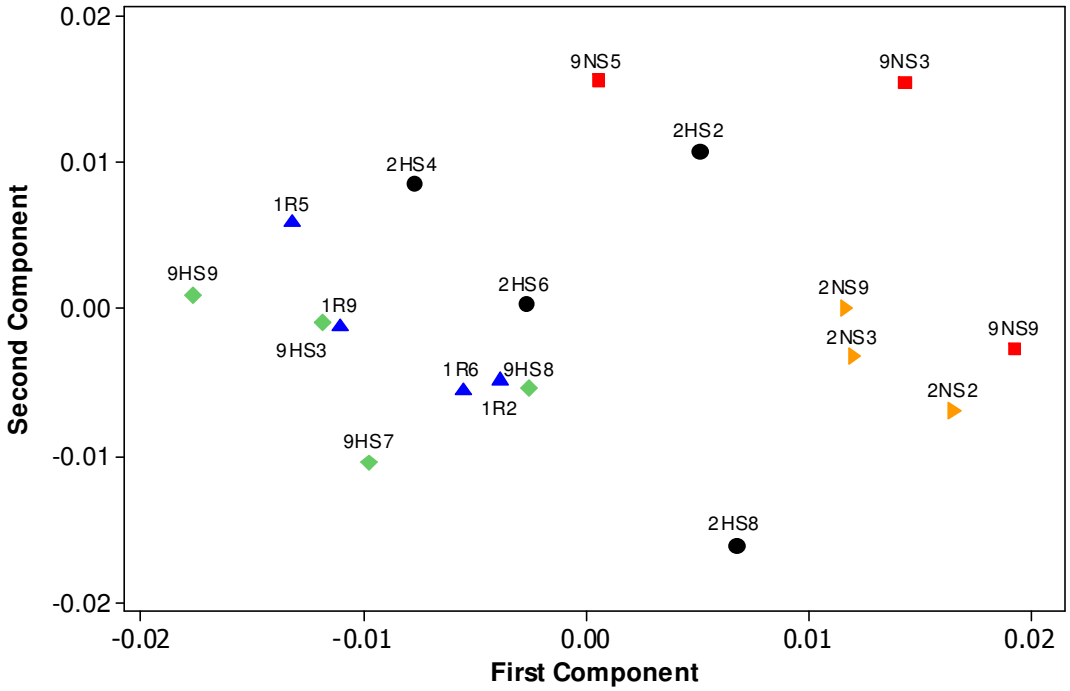


Figure 3.5 PCA of ^{1}H NMR spectra (binned into 427 variables) of 19 plant extracts is reduced to a two-dimensional scores plot, revealing the variables responsible for clustering of samples according to duration of oxygen deprivation stress treatment.

Transcripts and enzymes are responsible for reestablishment of homeostasis upon recovery. Expression of certain "ready to rumble" enzymes which rapidly facilitate recovery is thought to be accomplished under stress, unleashing a barrage of chemical energy upon reoxygenation, although our global metabolite profiles do not support this hypothesis.[71, 78] Recovery from oxygen deprivation has been studied in *C. elegans*,[79] where it was found that only mutants with enhanced translational repression can recover from oxygen deprivation in 1-2 hours but wild-type worms did not survive the 22 hour oxygen deprivation even when given 24 hours of reoxygenation. Translational repression

is an evolutionarily conserved strategy to survive hypoxia/anoxia via decreases in ATP consumption, protein aggregation and an alteration of the proteome. During a recovery period, reactive oxygen species (ROS) are generated,[80] which are thought to function as initiators of a signaling cascade. Other proposed regulators of oxygen deprivation responses include cytokinins, mutants capable of self-regulating cytokinin production under environmental stresses that survive better relative to wild-type plants.[72, 81] Although interesting, the processes occurring in recovery are too poorly understood to cover in detail here. The compounding biological factors making it difficult to interpret metabolic snapshots taken 1 hour after return to normoxia are that simultaneous might include: (1) turnover of enzymes and mRNAs from anaerobic response, (2) restoration of normal protein synthesis, (3) restoration of oxidative phosphorylation, and (4) recycling of Ala to pyruvate.

The loadings plot (Figure 3.6) suggests that the clustering of samples is due to chemical shift regions corresponding to Ala, Gln, lactate, succinate, and glucose. These were confirmed by visual inspection of representative spectra from each sample treatment (Figure 3.4) as well as by comparison with difference spectra (presented for individual metabolites in Appendix A: NMR-based Primary Metabolite Database). Identifications were made by overlaying a spectrum of a metabolite standard on the top of the treatment or difference spectra. Special attention was given to pH effects, chemical shifts of resonances, coupling patterns, and multiple resonances attributed to a single molecule. If all of these pieces of information were not in agreement, the metabolite identity was not established.

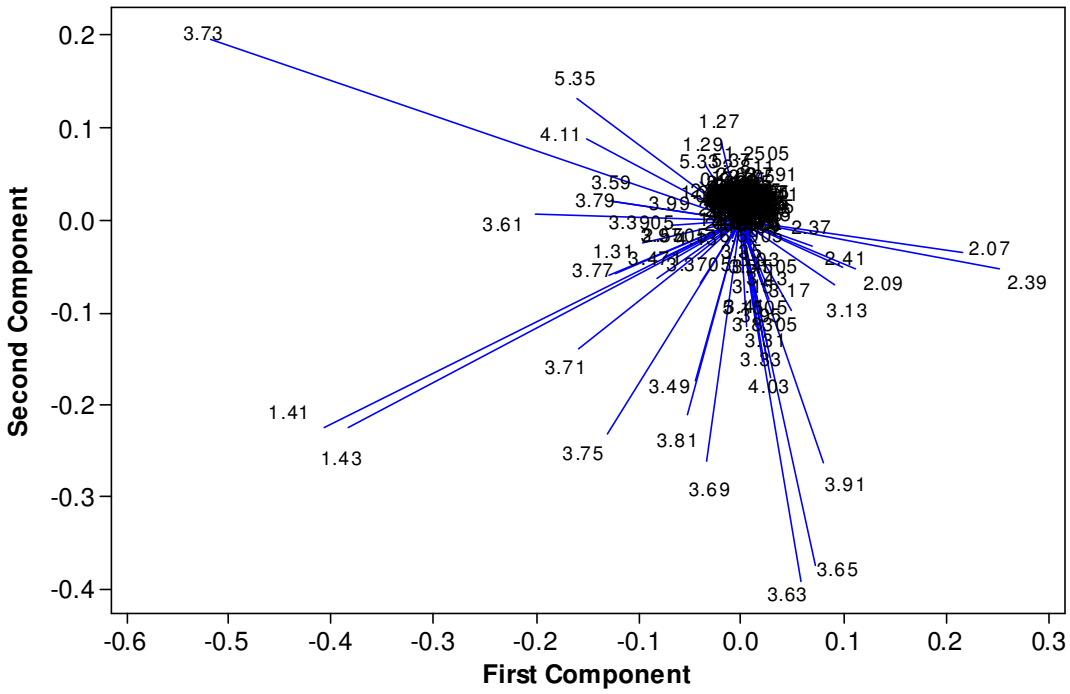


Figure 3.6 PCA of ^{1}H NMR spectra (binned into 427 variables) of 19 plant extracts is reduced to a two-dimensional loadings plot revealing integral regions varying significantly. Directionality indicates covarying resonances.

An explained variance plot (Figure 3.7) reveals that the first principal components explains less than 50% of the variance in the dataset while the first and second components together explain approximately 67% of the variance in the dataset. Plotting eigenvalue on the y-axis and eigenvector number on the x-axis reduced dimensionality from 427 vectors to two vectors; considering the information contained in the first two eigenvectors of our model account for 67.4% of the variance in the original dataset. It is generally considered that the first two principal components should account for greater than 65% of the variance in the dataset. Further cross-validation is warranted, but so far we are confident that the scores and loadings plots provide us with a representative meta-

view of global changes in the metabolome. Although this does not represent all variance, it can reduce the dataset to two dimensions, allowing for first-pass interpretation of which metabolites are responsible for the observed clustering.

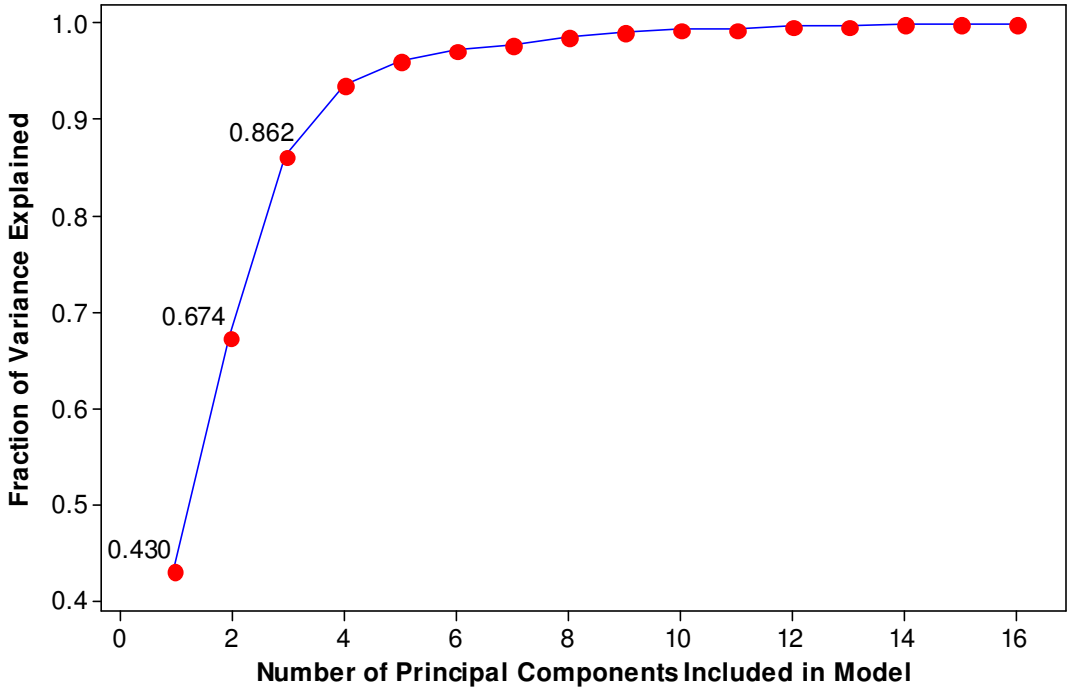


Figure 3.7 Plotting eigenvalue on the y-axis and eigenvector on the x-axis gives an explained variance plot. Our scores and loadings plots account for 67.4% of the variance in the original dataset.

Cross validation was performed as recommended by Broadhurst and Kell[82] to ensure that metabolites indicated by the loadings plot were in fact responsible for the clustering observed in the scores plot. This was carried out in an unsupervised approach by identifying the loadings most important to the model, by calculating "loading length." The PCA model outputs PCs (eigenvectors) in both Q-mode (scores) and R-mode (loadings).[83] Using only R-mode (loadings) and visualizing a vector plot made by using only the first two principal components (as shown in Figure 3.6), it is possible to

calculate the length of the hypotenuse (c) of a triangle formed for each of the original variables and their projection onto the first two principal component axes in the new coordinate space (a representing PC1 on the x-axis, b representing PC2 on the y-axis). Using Pythagorian Theorem, loading length is calculated by solving for c. This approach identified the original variables occupying NMR spectral regions 3.91, 3.81, 3.7505, 3.7305, 3.71, 3.69, 3.6495, 3.63, 2.39, 2.07, 1.429, and 1.409 ppm as having the longest "loading length."

Constructing a new PCA model using only the integral regions listed above (Figure 3.8), we can see that these variables are responsible for the biologically relevant

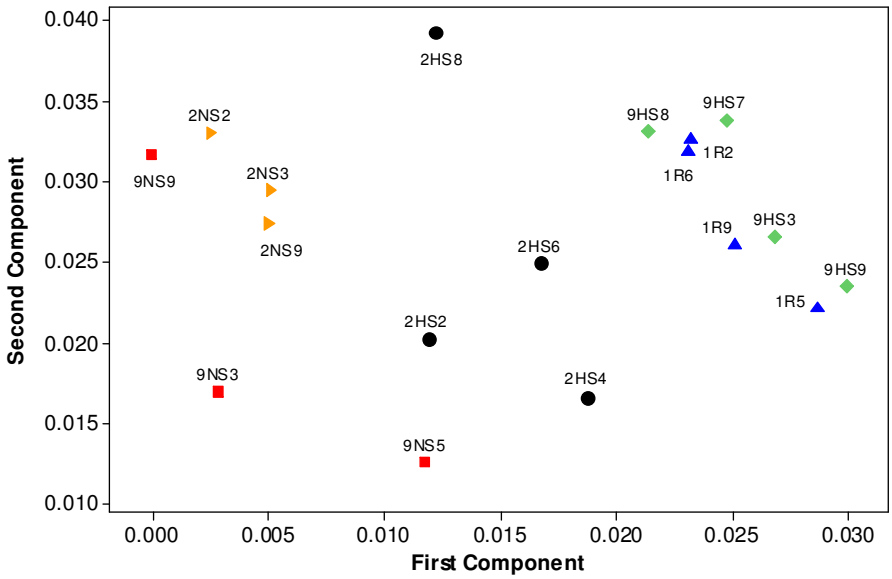


Figure 3.8 The PCA model was cross-validated by recalculating PCs utilizing only 12 variables, those which had been indicated by the loadings plot to explain the variance in the dataset. This is to check for false positives.

clusters discussed previously. If the clustering was not preserved, we would be including unnecessary variables from the original data matrix. Conversely, we can take the remaining data matrix, excluding these 12 integral regions and construct a new PCA model.

As shown in Figure 3.9, PCA is useful for narrowing the scope of NMR-based metabolomic datasets, yet it does not give direct information on specific metabolites.

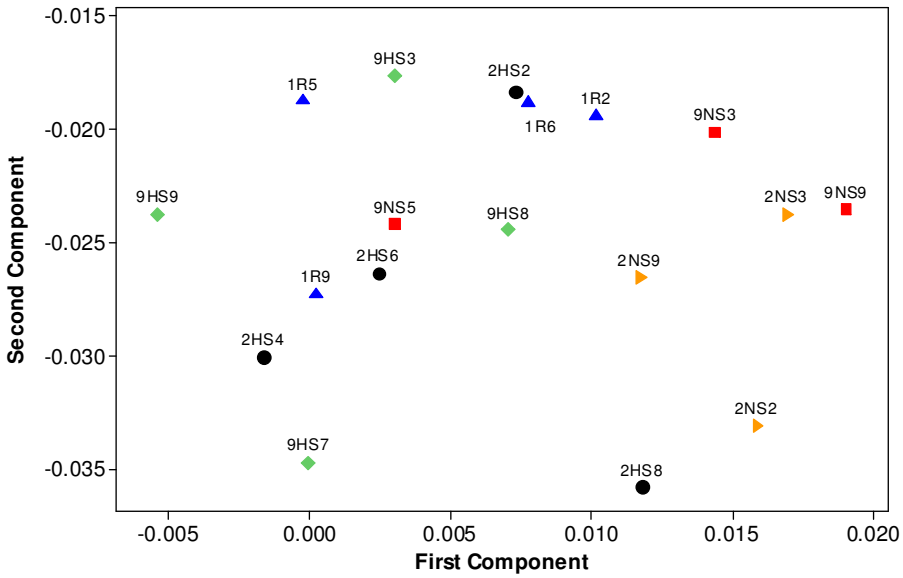


Figure 3.9 The PCA model was cross-validated by recalculating PCs excluding 12 variables, namely those indicated by the loadings plot to explain the variance in the dataset. This is to check for false negatives.

Covariance analysis provides for the identification of metabolites or pathways that are significantly perturbed by low oxygen levels. However, a different type of analysis is needed to establish statistical and biological significance, and the results do not directly translate to a biological interpretation. An alternate representation is to consider intense

metabolite resonances occupying regions of spectral space that do not overlap. By this technique, we assign signals to abundant molecules in plant extracts and quantify the accumulation or depletion of as many metabolites as possible.

3.3.2 Univariate Analysis of NMR-based metabolite profiling data

Statistical evaluation of the metabolic profiles generated by ^{1}H NMR allowed direct integration and quantitation of 12 metabolites, reported in Table 3.2. One additional metabolite, lactate, was determined by spectral subtraction due to overlap with other resonances. The expected production of significant amounts of Ala at the early time points was observed (greater than 3-fold increase above the control), as well as the production of succinate greater than 4-fold above the control values. The production of lactate was observed but it was impossible to calculate a fold change since levels could not be accurately determined in the controls.

Interestingly, we observed an average increase of sucrose, however it was not statistically significant due to large variability among bioreplicates, which might have been an artifact of sampling. When Arabidopsis seedlings are released from oxygen deprivation treatment, it is best to harvest them as quickly as possible. Inevitably, contamination from the media results since the seedlings are beginning to form lateral roots and root hairs, which adhere strongly to the growth support. When medium alone was sampled, it was found to contain an abundance of sucrose and smaller amounts of fructose and glucose.

Table 3.2 Metabolites quantified by ^{1}H NMR of direct tissue extracts. Chemical shifts integrated are listed. Numbers reported for each treatment represent the normalized signal intensity of a metabolite under stress divided by the normalized signal intensity of the same metabolite in the corresponding control condition, yielding a fold change for each metabolite. Fold changes report the effect of 2 hours of stress (2HS/2NS), 9 hours of stress (9HS/9NS), and the effect of 1 hour recovery against 9 hour controls (1R/9NS) and stress (1R/9HS). Significance is denoted by * at the 95% confidence interval and by ** at the 99% confidence interval. For lactate, the fold change could not be quantified because of low levels in control tissue and spectral overlap, therefore + and +++ denote moderate and strong increases, respectively.

Metabolite	Shift (ppm)	2HS/2NS	9HS/9NS	1R/9NS	1R/9HS
Ala	1.42	3.3*	6.5**	6.8**	1.0
Asp	2.79	1.0	0.7*	0.7**	1.1
Choline	3.13	0.8	0.8*	0.8	1.1
Fructose	4.02	0.8	1.1	1.2	1.0
GABA	2.98	1.0	1.3**	1.2**	0.9
Glucose	5.15	0.8	1.2	1.2	1.0
Glu	2.10	1.0	0.7*	0.7**	0.9
Gln	2.06	0.9	0.7**	0.7**	1.0
Lactate	1.13	+	+++	+	---
Ornithine	1.75	0.8	1.0	1.2	1.2
Succinate	2.56	4.2**	11.4**	7.4**	0.7*
Sucrose	5.34	1.9	1.1	1.1	1.0
Val	0.99	1.1	1.3	1.6**	1.2

After nine hours of oxygen deprivation stress, many more metabolites were statistically significantly altered from their values in the control. The production of GABA was observed in addition to an even greater accumulation of Ala and succinate. The amino acids Asn, Glu, and Gln were observed in oxygen deprivation stressed tissue extracts to decrease at the 95% confidence interval to 70% of the amount they would have been present in the control samples. The biogenic amine choline was significantly affected by oxygen deprivation in cell cultures.[17] It is proposed in this work that the changes in these metabolites together represents a consumption of nitrogen reserves and

suggests that the nitrogen cycle has been disrupted in the attempt to maintain homeostasis and ATP production to support critical cellular processes. Selected metabolite average relative abundances are represented graphically in Figure 3.10. Standard deviation is displayed as y-error bars for each average reported (N=3-4) among bioreplicates.

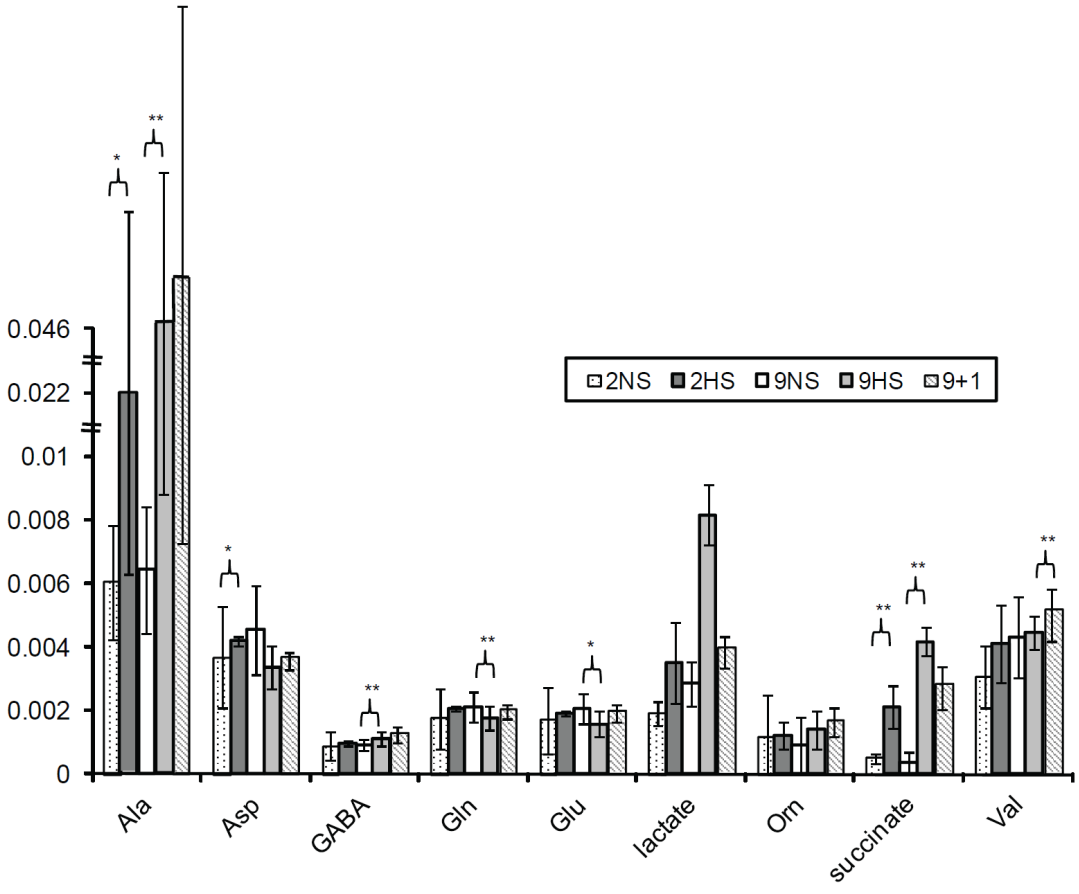


Figure 3.10 Average relative metabolite abundance in extracts representing Arabidopsis seedlings (N=3-4) for hypoxia controls and treatments. Two hour control and stress (2NS, 2HS) are compared with nine hour control and stress (9NS, 9HS) and nine hour stress followed by one hour recovery (9+1). Standard deviations among bioreplicates are shown as y-error bars. Significant differences are clear in lactate and succinate. Ala is shown on a different scale due to its abundance, which would have dwarfed all other observations.

Given one hour of reoxygenation, Ala, GABA and succinate were found to remain elevated over the nine hour control group. Ala levels did not change significantly

upon recovery, however succinate and lactate levels were significantly altered in the recovery period. Glu, Gln and Asn also remain significantly depleted even after one hour of recovery, while the disruption of choline was no longer considered significant. One striking feature of the recovery period is the accumulation of valine. This metabolite is synthesized in an aminotransferase reaction involving both pyruvate and Ala, catalyzed by branched chain aminotransferase (BCAT), so perhaps it is another neutral substrate which can distribute carbon skeletons funneled through glycolysis to prevent buildup of lactate when aerobic respiration resumes. Mutants of valine production[84] could be used to test this hypothesis *in planta*. A chemical genomic approach has recently contributed to the body of knowledge in this area.[85] BCAT, also known as Transaminase B, is responsible for the final step in Ile and Val production in Arabidopsis.[86, 87]

3.3.3 Insights from GeneChip Measurements

Total and polysomal RNA extracted from three independent biological replicate experiments were used for microarray hybridization analysis against Arabidopsis ATH1 GeneChip arrays. Probe-specific and multichip background corrections were carried out for total and polysomal RNA. Further adjustments were made to polysomal RNA data to account for variation in the levels of protein synthesis and hence global polysome levels under the different treatment conditions. The normalization factor was generated from sucrose density gradient absorbance profiles of polysomes purified from tissue.

Following 2 hours of oxygen deprivation (short-term) we observed only a select population (2%) of mRNAs were increased in abundance in immunopurified mRNA,

whereas a larger population (6%) were increased in abundance in total mRNA. A set of 178 transcripts were conserved between the two, increasing in stress conditions compared to control at a significance level of 99.9% or greater.

A similar but more dramatic trend was observed after 9 hours (prolonged) stress, with 20% of mRNAs significantly increased in total mRNA population and only 5% of mRNAs significantly increased in the immunopurified group. A large group (~5000 transcripts) of mRNAs were found to decrease in polysome immunopurified populations under both stress conditions. By comparative analysis, the molecular outcomes of short-term and prolonged oxygen deprivation and subsequent recovery for 1 hour versus controls were evaluated gene-wise. These time points were chosen based on RT-PCR monitoring of transcripts known to be elevated in response to hypoxic stress, namely Alcohol Dehydrogenase (ADH1). RT-PCR results correlated well with microarray data.

3.3.4 Pathway Modeling

By examining significantly altered metabolites and genes corresponding to transcripts showing significant changes in relative abundance in response to oxygen deprivation stress or recovery, we may postulate on flux through pathways. In many cases we had insufficient data to fully map these results onto pre-packaged pathways. Volatile trapping, as well as starch and insoluble carbohydrate analysis, and lipid analysis, would be necessary to ascertain a whole picture of the plant's responses to low oxygen and anoxia. Additionally, although *Arabidopsis thaliana* (Col-0) is one of the top three model organisms, at the time of this work, less than 60% of genes had been

functionally identified. Therefore, some significantly altered transcripts are as yet unidentified. For the range of pathways we were able to confidently probe, we found several significant observable metabolic tendencies, which we now understand are part of the "core" hypoxic response. A model summarizing the main findings of this work is displayed in Figure 3.11. The induction of flux through the pathways observed in our

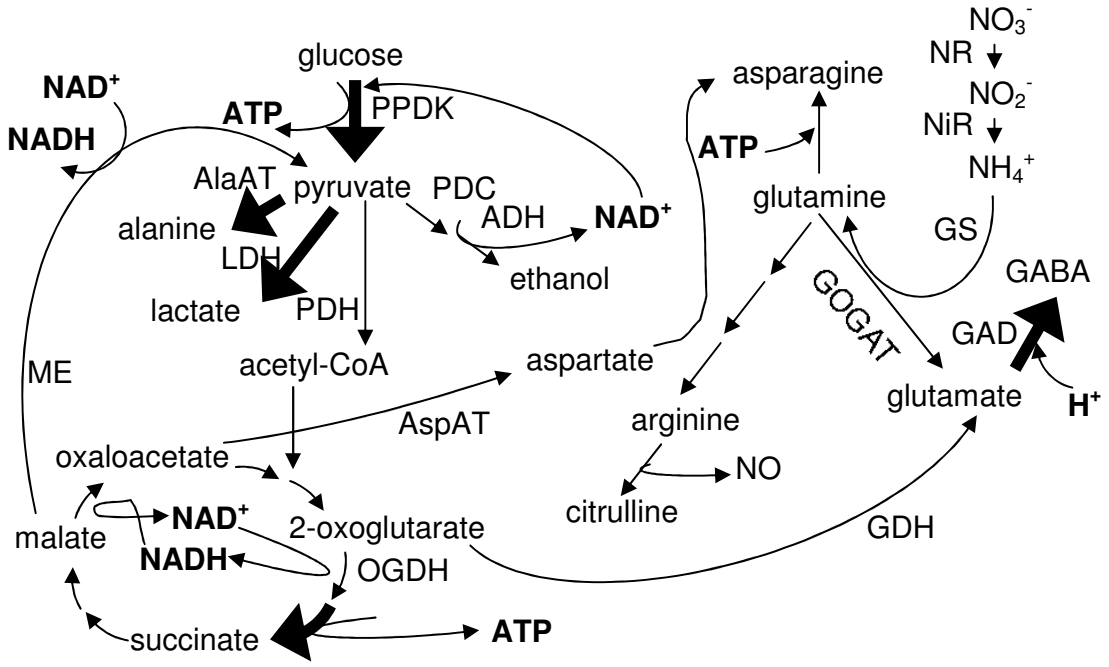


Figure 3.11 Metabolic adjustments under oxygen deprivation. Metabolites, enzymes, and intermediates are shown. The primary rate-limiting step of glycolytic flux occurs at PPDK, the process generates one molecule of ATP. Pyruvate undergoes fermentation via LDH to lactate, transamination to become alanine, dehydrogenated yielding acetyl-CoA which can enter the TCA cycle, or decarboxylated to acetaldehyde and dehydrogenated via ADH to ethanol, regenerating NAD^+ for reentry into glycolysis. The oxoglutarate dehydrogenase and succinate dehydrogenase are highly favored processes for ATP generation, leading to production of succinate. NAD^+ may be regenerated via malate dehydrogenase or malic enzyme. Glutamic acid dehydrogenase can consume excess cellular protons, forming GABA, which may be a key substrate along with alanine for maintaining a balance between carbon and nitrogen metabolism.

study is represented by large arrows. Coupling intermediates (NAD^+, ATP, H^+) are represented where literature and data suggest it is relevant. Enzymes are abbreviated for some reactions. Furthermore, some but not all arrows may function bidirectionally.[87] A summary of the literature in a variety of plants in regards to metabolic networks, metabolomics, and enzymatic quantitation of primary metabolism is provided by Table 3.3. Metabolic network coordination is still being understood in crop and model systems. Conserved mechanisms involve induction of ethanol and lactate fermentation, producing ATP via glycolysis, controlled by cytosolic pH.

3.3.4.1 Sucrose Catabolism, Glycolysis and Fermentation

Catabolism of sucrose is achieved via sucrose synthase, a less energetically costly process than the invertase route. Flux of carbon through glycolysis, which is the main source of energy under aerobic conditions, is blocked under oxygen deprivation. Glycolysis terminates in pyruvate, which is a metabolic hub.[12] However under low-oxygen stress, lower ATP levels are maintained at the substrate level via glycolysis coupled with NAD^+ regeneration through anaerobic fermentation. Fermentation is a key component of the low oxygen response in plants.[14-16] When oxidative phosphorylation provides inadequate ATP, limited by absence of O_2 (the final electron acceptor), to sustain the classical citric acid cycle as shown in Figure 3.1, fermentation provides alternate routes for consumption of pyruvate. The ATP generated is necessary to retain critical processes such as RNA transcription, protein synthesis and the activity of plasma membrane and vacuolar proton pumps that limit acidification of the cytosol. Figure 3.11

Table 3.3 Summary of plant primary metabolism literature. Critical processes for survival are summarized for each abiotic stress.

Plant	Age	Organ	Condition/ Treatment/ Genotype	Major Findings	Ref
Chlamydomonas reinhardtii (algae, plant ancestors)	2 d	culture	5 or 0.035% CO_2, $\pm$ AOA (AlaAT inhibitor)	AlaAT transiently induced by growth under low CO_2 conditions	88
Corynebacterium glutamicum (bacterium)		culture	*in vivo* kinetics of transaminase network after glucose load	Valine production strain has alterered flux through pyruvate (3-fold), metabolite concentrations have subsecond turnover rates	84
Lemna minor L.	7 d	whole plants	7.0 mmol N / dm^3 (N-free, NH_4^+ or NO_3^-)	Nitrogen starvation altered protein contents and Glu, Ser, Ala, glyoxylate, pyruvate	89
Lotus japonicus (ecotype GIFU B-129)	12 wk	root, nodule and leaves	waterlogging (N_2 sparged water) $\pm$ leghemoglobin (via RNAi)	Coupled transaminases are central with minor consideration given to glycolysis and the TCA cycle	90
maize (hybrid Funk 4323) and pea (var. Alaska)	2 d	root tip (2-4 g)	[50 mM glucose, 0.1 mM $CaSO_4$] 10x $\pm$ O_2(g) or 1x $\pm$ N_2(g)	Lactate fermentation responsible for acidification of cytoplasm	91
oat (*Avena sativa* L.)	10 d	leaf slices	0.3 M sorbitol, 2 mM KH_2PO_4, 2 mM $MgCl_2$, 1mM EDTA, 5 mM HEPES/KOH + 3 mM ^{15}N Glu or Ala	Role of Ala in photorespiratory nitrogen metabolism	92
Oryza sativa, cv. M201	3 d	apical shoots	superfused in [0.5mM $CaSO_4$, 5 mM glucose, 5 mM NH_4SO_4, 5 mM $NaHCO_3$, 5 mM $Na^{13}CH_3COOH$] + air or N_2 for 23 h (dark)	Glyoxylate cycle contributes to maintenance of glycolysis for energy production	93
potato (*Solanum tuberosum*) cv Desiree	10 wk	tuber cores	PKc decrease via RNAi	Regulation of AOX by pyruvate *in vivo*	13
radish (*Raphanus sativus* L., var. Champion)	mature	leaf	N_2 or air exposure to leaf halves, dark, room temp	Accumulation of pyruvate shifts equilibrium of AlaAT towards Ala accumulation, at the expense of Glu and Asp, concomitant with the accumulation of GABA	94
soybean (*Glycine max* [L.] Merr. cv Corsoy 79), compared with pea, bean, pea, alfalfa, tomato, radish, barley, foxtail, maize, purslane, pigweed	45 d	detached leaf	$\pm$ light, cold-shock (6°C, 5 min), heat-shock (33°C, 5 min)	Adjustment of Gly, GABA, Ala, Glu and Asp in response to light	95
tomato (*Lycopersicon esculentum* Mill. Cv VFMT) x barley LDH	3 wk	petiole section	transformant root cultures in liquid medium, + barley LDH	Progressive inhibitition of LDH activates PDC, promoting alcohol fermentation	96

is one model demonstrating how survival is achieved, based on our analysis of 9-hour argon-treated 7-day-old seedlings of the model plant *Arabidopsis thaliana* (Col-0) grown on 1% sucrose-enriched 0.5x MS solid media under long day (16 h light) conditions.

In sucrose and starch metabolism, sucrose synthase (links energy-efficient sucrose catabolism to glycolysis) and the carbohydrate kinase family (glucose/G-6-P or F-6-P/F-1,6-P or F-6-P/F-2,6-P) were found to increase with stress. The transcript for *HEXOKINASE2* (glucose/G-6-P) was found to decrease upon stress but increase after recovery. The Student's *t*-test did not identify any fold changes as significant according to the NMR data however, the multivariate statistics indicate that some of the sugar resonances contribute to the variance in the dataset. Indeed, inspection of Figures 3.4 (NMR spectra) and 3.6 (PCA loadings plot) clearly shows integrals in the sugar region varying significantly. Possibly this is related more to normal sugar usage during the diurnal cycle and not directly linked to sample treatment, or contamination with media.

Under conditions of oxygen deficiency, glycolytic flux to pyruvate is needed to boost substrate-level ATP generation. Fermentation of pyruvate to generate lactate or ethanol serves to regenerate NAD^+ allowing glycolysis to continue. There is increasing evidence that under oxygen limiting conditions the TCA cycle operates as two branches rather than as an 8-step cycle, independently consuming and regenerating NAD^+/NADH while yielding additional ATP via conversion of succinyl CoA to succinate catalyzed by succinyl CoA synthase.[2] The production of aspartic acid is inhibited by lack of ATP, while production of GABA is stimulated, potentially a sink for protons generated by the

production of lactate. Nitrogen assimilation may also be stimulated for the purpose of generation of the signaling molecule nitric oxide (NO).[11, 97]

Metabolic profiling by ^{1}H NMR showed good agreement with the transcript profiling by the Affymetrix GeneChip array. In the glycolytic and fermentation pathway,, the transcripts for lactate dehydrogenase (interconversion of pyruvate and lactate), alcohol dehydrogenase (conversion of ethanol to acetaldehyde), phosphofructokinase (F-6-P/F-1,6-P or F-6-P/F-2,6-P), hexokinase (glucose/G-6-P) were found to increase with stress. The resonance at 2.30 ppm for pyruvate shows a visible increase upon stress, but its fold change cannot be accurately calculated due to high background in this region. Fortunately, this metabolite was quantifiable by enzymatic assay, which showed that it remained elevated about 5-fold at 2 and 9 hour time points and after 1 hour recovery.

The signals from ethanol and acetaldehyde are unidentifiable in the sample spectra either due to low abundance or high background in these regions. This is possibly because both of these molecules are volatile and may have evaporated instead of accumulating within the plant tissue. While the tissue was quenched under liquid nitrogen, it was stored for up to 3 months in Eppendorf microcentrifuge tubes at -80 ^{o}C and extracted on the benchtop at room temperature (22 ^{o}C) during which time volatile metabolites could have vaporized, oxidized, or degraded. Enzymatic measurements of ethanol fold changes showed a significant accumulation after 2 hours of oxygen deprivation, but no differences between 9 hour or recovery and controls were observed.

The anomeric protons of sugars are useful for quantitation because they are well-resolved however it is not possible to distinguish resonances of glucose from G-6-P. Standard solutions of glucose and G-6-P can be distinguished in the 3.1 – 3.5 ppm region, but in the sample spectra this region is not useful for quantitation because that is a crowded region. In comparing enzymatic measurements of sugars to the NMR-based quantitation, large discrepancies became apparent. Upon further investigation, it was determined that the perchloric acid used to extract tissue prior to enzymatic assay was in fact cleaving the glycosidic bond of sucrose (Figure 2.8). This provided misleading evidence of low sucrose and higher amounts of glucose and fructose. Altogether the attempt to quantify sugars was severely compromised and became one focus of our method development in preparation for future studies of Arabidopsis under oxygen deprivation (Section 2.3.4).[98]

3.3.4.2 Fermentation Pathways

During anaerobic stress the metabolic hub molecule, pyruvate, can be metabolized to lactate via lactate dehydrogenase (LDH). Alternately, pyruvate can be metabolized to ethanol via pyruvate decarboxylase (PDC) and alcohol dehydrogenase (ADH) to regenerate NAD^+ required to sustain glycolysis, and associated ATP production.[99] Lactate fermentation may contribute to the acidification of the cytoplasm[91, 100] an early response to hypoxic stress, which may progressively inhibit LDH and activate PDC, promoting alcohol fermentation, a later response.[96] These cytosolic enzymes have been well studied in maize. These processes are balanced to minimize carbon losses via

evaporation of acetaldehyde and ethanol while maintaining a working pH in various cellular compartments to support specialized reactions for temporary survival of low oxygen,[15] and if the stress is applied slowly lactate pumps may help maintain homeostasis.

Accumulation of pyruvate under anaerobic conditions may shift the equilibrium of alanine aminotransferases (Table 3.4) towards Ala accumulation, which occurs rapidly in plants under anaerobiosis.[94, 101] The precise function of Ala accumulation under anoxia is unknown, but it is speculated that it may serve as a storage form of pyruvate (perhaps in the vacuole), controlling the supply of pyruvate to LDH and PDC, and indirectly the flux of lactate and ethanol.[5, 102] Ethanol can be recycled to pyruvate upon reoxygenation. The effect of both overexpressing LDH1 and T-DNA knockouts on fermentation metabolism (measuring ethanol and lactate only) revealed the importance of this mechanism in Arabidopsis.[73]

3.3.4.3 Transaminases: Route for Protein Catabolism

When sugar reserves are low, especially under prolonged stress, proteins must be catabolized to amino acids for energy production. Transaminases play a key role in funneling nitrogen to Asp and Glu for nitrogen excretion. Table 3.4 illustrates the large diversity of transaminases. Enzyme classification (EC) numbers and references for further reading are provided. Some of these enzymes are denoted in Figure 3.11.

Changes in abundance and activity of transaminases (also called aminotransferases) and their corresponding metabolic substrates and products have been

reported for a variety of systems under molecular oxygen- and carbon dioxide-deprivation. Alanine aminotransferase (AlaAT) is induced under anaerobic conditions in barley roots.[5, 102] Furthermore, in *Chlamydomonas reinhardtii* alanine:2-oxoglutarate aminotransferase is transiently induced by growth under low CO_2 conditions.[88] The

Table 3.4 Transaminases catalyze the reaction between amino acids and alpha-keto acids and mediate carbon and nitrogen homeostasis under oxygen deprivation.

Ref	Reaction catalyzed	Enzyme Name, Enzyme Classification (EC) Number	
[90]	glutamate + oxaloacetate ↔ 2-oxoglutarate + aspartate	AspAT, Asp aminotransferase,	EC 2.6.1.1
[5, 88]	glutamate + pyruvate ↔ 2-oxoglutarate + alanine	AlaAT, Ala aminotransferase,	EC 2.6.1.2
[89]	2-ketoglutarate + glycine ↔ glutamate + glyoxylate	GGAT, Glu:glyoxylate aminotransferase,	EC 2.6.1.4
[90]	pyruvate + GABA ↔ alanine + succinic semialdehyde	GABA-T, γ-aminobutyric acid transaminase,	EC 2.6.1.19
[89]	2-ketoglutarate + alanine ↔ pyruvate	GPAT, Glu:pyruvate aminotransferase,	EC 2.6.1.21
[92, 104]	alanine + glyoxylate ↔ pyruvate + glycine	AOAT, Ala:oxoglutarate.aminotransferase,	EC 2.6.1.44
[89]	glyoxylate + serine ↔ hydroxypuruvate + glycine	SGAT, Ser:glyoxylate aminotransferase,	EC 2.6.1.45
[90]	2 glutamate ↔ glutamine + 2-ketoglutarate	GOGAT, Glu:oxoglutarate amidotransferase,	EC 2.6.1.53
[84]	valine + pyruvate ↔ 3-methyl-2-oxobutanoate + alanine	BCAT, branched-chain aminotransferase,	EC 2.6.1.66

GABA-dependent production of Ala under hypoxia was studied in mature (40 d) *Arabidopsis thaliana* via glutamate decarboxylase (GAD) and GABA transaminase (GABA-T) insertional knockout mutants under 5% oxygen.[70] The behavior of AlaAT1 overexpression and insertional knockout Arabidopsis mutants was observed under hypoxia, finding that AlaAT plays a critical role in recovery, remobilization of Ala back to pyruvate in (17 d) seedlings following 24 h low-oxygen stress.[76]

Because transaminase-catalyzed reactions are coupled, they are difficult to show in traditional pathway diagrams such as Figures 3.1 and 3.11 here. An alternative representation is afforded by Rocha et al.[90] where coupled transaminase reactions are displayed centrally with minor consideration given to glycolysis and the TCA cycle.

In other studies in plants it was suggested that the synthesis of Ala may occur at the expense of the acidic amino acids, Glu and Asp,[94, 101] and occurs concomitantly with the accumulation of an alpha-keto acid terminating in a free amine, 4-aminobutyrate or γ-aminobutyrate (GABA).[94, 95, 100, 101] Although GABA is a potent neurotransmitter in mammals, its role as a signaling molecule has not been established in plants.[103] study using coleoptiles of germinating rice, extreme accumulation of Ala (eighty-fold), citrate, malate, GABA, succinate, Gln (twenty-fold), and Glu (four-fold), were observed after 24 hours of submergence. Asn was the only metabolite to have shown a depletion.[93] The accumulation of Ala, GABA and succinate were proposed to contribute to ATP production.[2]

Extensive measurements of metabolite accumulation under low oxygen have been made using NMR, even intact rice root tips can be monitored over time, therefore its utility and reliability had been established. In such a study, anaerobic rice coleoptiles did not display the build-up of glycolytic byproducts (pyruvate, NADH and H^+). It was proposed that these substrates were consumed in various reactions leading to the production of ethanol, Ala, Gln and GABA. It was observed that the TCA cycle remains operative to an extent, sustained by diverting intermediates at the bottleneck (succinate) to GABA production and by replenishing the diverted carbon via the glyoxylate cycle reactions.[93]

It is well established that proline accumulates in a wide variety of plants under abiotic stress.[68, 105] Asn accumulates in a wide variety of plants subjected to drought, mineral deficiency, toxic metal and pathogen attack.[106] Neither proline nor aspartic acid

accumulate under oxygen deprivation stress, and almost all nitrogen-15 labelled ammonium accumulates as Ala.[71]

In GABA metabolism, the mRNAs encoding the enzymes glutamate decarboxylase 1 (conversion of Glu to GABA) and GABA transaminase (interconversion between succinic semialdehyde, GABA, and α–ketoglutarate) were found to decrease with stress, yet increase upon recovery. The fold change of α–ketoglutarate, also called 2-oxoglutarate, could not be determined because it was below the detection limits of the NMR method. Metabolic profiling identified significant decreases in Glu and increases in GABA.

Figure 3.11 presents a pathway diagram specifically indicating processes induced by oxygen deprivation. The accumulation of GABA is thought to consume cytosolic protons, limiting the acidification in consort with lactate production. Formation of GABA may also prevent ammonification.[107] Labeling experiments with $^{13}CO_2$, $^{15}NH_4$ and $^{15}NO_3$ in the light found that ammonium assimilation is largely independent of concurrent assimilation of CO_2,[67] therefore pathways such as the GABA shunt may be fully operative under strict anoxia and CO_2 limited conditions. Buildup of GABA has also been proposed to result from inhibition of GABA-T (EC 2.1.6.19, Table 3.4).[108] The link between polyamines and the GABA pool (via catabolism) is gaining increasing attention, and a better understanding of the interplay between primary and secondary metabolism is needed. The GABA shunt may represent a bridge linking secondary metabolism to the center of plant carbon/nitrogen metabolism.[103]

3.3.3.4 TCA Cycle and Related Reactions

The Krebs cycle, citric acid cycle, or tricarboxylic acid (TCA) cycle is evolutionarily conserved throughout organisms that undergo aerobic respiration. It is arguably the origin of life,[109] spinning small molecules such as carbon dioxide and acetate into precursors for sugar and amino acid biosynthesis. This series of reactions is like a power plant within the mitochondrion of eukaryotic cells. Under aerobic conditions each molecule of glucose produces six $NADH_2^+$, two $FADH_2$, and two ATP molecules. $NADH_2^+$ and $FADH_2$ are energy carriers to the electron transport chain in the inner mitochondrial membrane. The electrons delivered to the electron transport system provide energy to "pump" protons across the inner mitochondrial membrane to the outer compartment. Oxygen starvation causes an energy crisis due to inhibition of cytochrome *c* oxidase of the mitochondrial electron transport chain. The resultant decrease in ATP is generally alleviated by increased consumption of stored and/or soluble carbohydrates for substrate level generation of ATP via glycolysis and NAD^+ regeneration via fermentation.[14] If the oxygen supply is cut off, electrons and protons cease to flow through the electron transport system, preventing a proper proton gradient which is necessary to power the synthesis of ATP.

The conventional 8-step TCA cycle (Figure 3.1) is one manner in which the reactions can be organized. Recent evidence from labeling studies and metabolic network models suggest that non-cyclic flux modes occur in leaves in the light and under anoxia. One particular model that has been recently proposed involves two weakly connected branches operating in opposing directions (Figure 3.11).[15, 67, 90] While

glycolysis occurs in the cytosol, pyruvate enters the mitochondrion to enter the TCA cycle, while electron donors such as NADH and NAD(P)H enter into oxidative phosphorylation in the mitochondrial membrane.[110]

The glyoxylate cycle is a modified TCA cycle producing succinate, which is transported into the mitochondrion for entry into respiration. The glyoxylate cycle is localized in the peroxisome.[111] Alternately the glyoxylate cycle can produce malate and oxaloacetate, leading to gluconeogenesis. In seedlings, the glyoxlate cycle is responsible for mobilization of fat stores in the seed into sucrose providing building blocks for anabolism before the photosynthetic apparatus develops.[112] This can be an important strategy for mobilization of lipid reserves for energy production under anoxia.[10]

Having reviewed the current state of knowledge in low oxygen metabolism, acknowledging the importance of oxygen in all living systems on Earth, we proposed that ^{1}H NMR based metabonomics would provide sufficient metabolite data for comparison with established literature, based on studies in model plant systems *Arabidopsis*[113] and tobacco.[114, 115] Metabolic fingerprints by ^{1}H NMR under different treatment conditions combined with supervised and unsupervised modeling of multivariate data have been shown to reveal disease biomarkers and adverse drug effects in humans and animal models.[116]

In the TCA cycle, almost all transcripts encoding associated enzymes showed decrease in polysome association upon stress, except pyruvate kinase (PEP/pyruvate), phosphoenolpyruvate carboxylase (PEP/oxaloacetate). Several flux studies revealed bypass of the TCA cycle where PEP is converted into malate and subsequently into

pyruvate,[117] presumably involving these enzymes. A similar trend was observed in submerged (hypoxic) grey poplar.[69] Upon reoxygenation, almost all TCA cycle enzyme transcripts were found to recover association with polysomes.

The PCA loadings plot identified the chemical shift region of 1.30 ppm which contains resonances of lactate and oxaloacetate. It appears one of these metabolites accumulated in response to low oxygen stress, yet it is unclear if this is due to the high lipid background in this region. The resonance of PEP and many of the TCA cycle metabolites (citrate, isocitrate, fumarate, malate) were unfortunately not detectable in this study. Mattoo et al.[118] were able to measure malate by ^{1}H NMR in 25 mg DW tomato fruit extracts by integrating its resonance at 4.29 ppm, which was very close to the residual water in our extracts precluding integration for quantitation. Carbon-13 labeling was utilized in Arabidopsis cell cultures, leading to 1D ^{13}C-NMR quantification of citrate, isocitrate, fumarate, and malate.[119] NMR-active ^{13}C-2-acetate has been used in feeding under low oxygen conditions.[93] Zhang et al.[107] measured fumarate in tobacco by integrating the resonance at 6.52 ppm again using a sample size of 25 mg DW leaf tissue. Visual inspection of the spectra, scanning regions where these metabolites would have been, leads to the suggestion that the manipulation of extract pH may alleviate spectral overlap in some cases.[98] Further experiments that focus on better characterization of these responses, incorporating measurement of organic acids and related-TCA cycle metabolites by GC-MS is recommended.[120-124]

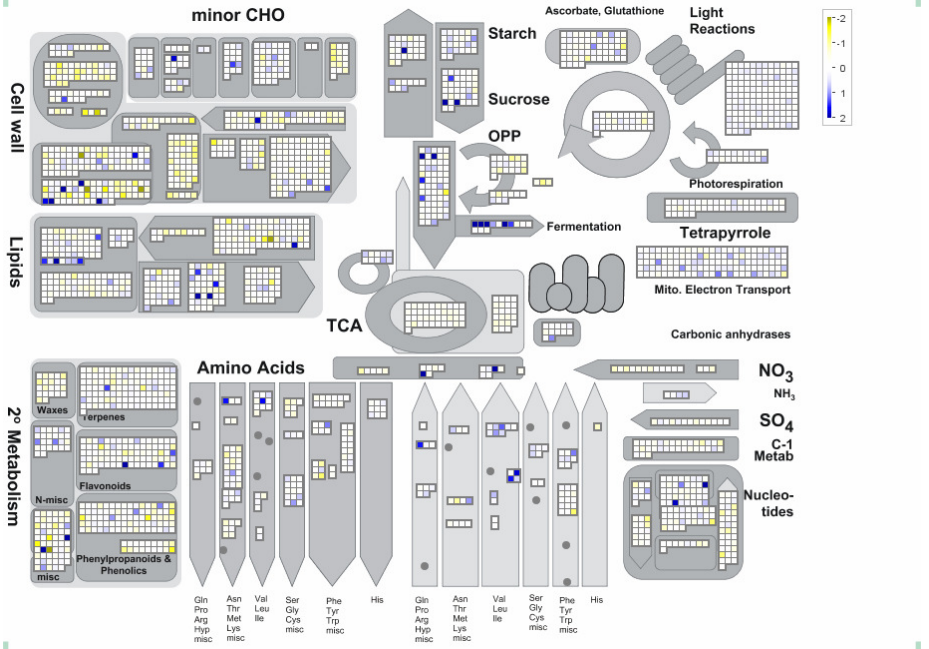
The most dramatic fold change in TCA metabolites observed in response to hypoxia was succinate. This may be because both succinic semialdehyde and

α–ketoglutarate can be subsequently converted to succinate. Labeling studies in which oxygen is severely limited have shown that $^{14}CO_2$ fixation in parasitic helminths, which colonize anoxic environments such as mammalian gastrointestinal tract, leads to labeling of malate as the first stable and strongly labeled intermediate, but carbon-14 ultimately accumulates mainly in succinate.[32] In our study, CO_2 fixation is unlikely to occur since argon is replacing all "air."

Another theory for the buildup of succinate is that it can contribute to substrate level ATP production and maintain mitochondrial redox balance. Experimental studies have observed that succinate accumulation proceeds with molar equivalents of ATP. Succinate is a well-established carbon sink in anoxic fish tissue. Redox coupling of mitochondrial reactions, via fumarate reductase, proceeds from glucose to succinate. In anoxic oyster hearts, ^{14}C-Asp feeding experiments revealed that about half accumulates in succinate, while another 20% is found in malate, presumably *en route* to succinate via a reverse TCA activity. Another possible route to succinate is through a segment of the TCA cycle. Again in anoxic oyster hearts, about 50% of ^{14}C Glu is incorporated into succinate. These reactions (Asp and Glu fermentation to succinate) are thought to provide redox couples for glucose fermentation.[32]

Pathway analysis using MapMan software[28] was carried out to examine the global trends in metabolic adjustments to survive low oxygen stress. Figure 3.12 shows the signal-log-ratio of the fold change between mRNA abundance after two hours of low oxygen stress in (a) total and (b) immunopurified populations. Boxes on this diagram correspond to features of the Affymetrix ATH1 GeneChip, but are here organized with

(a)



(b)

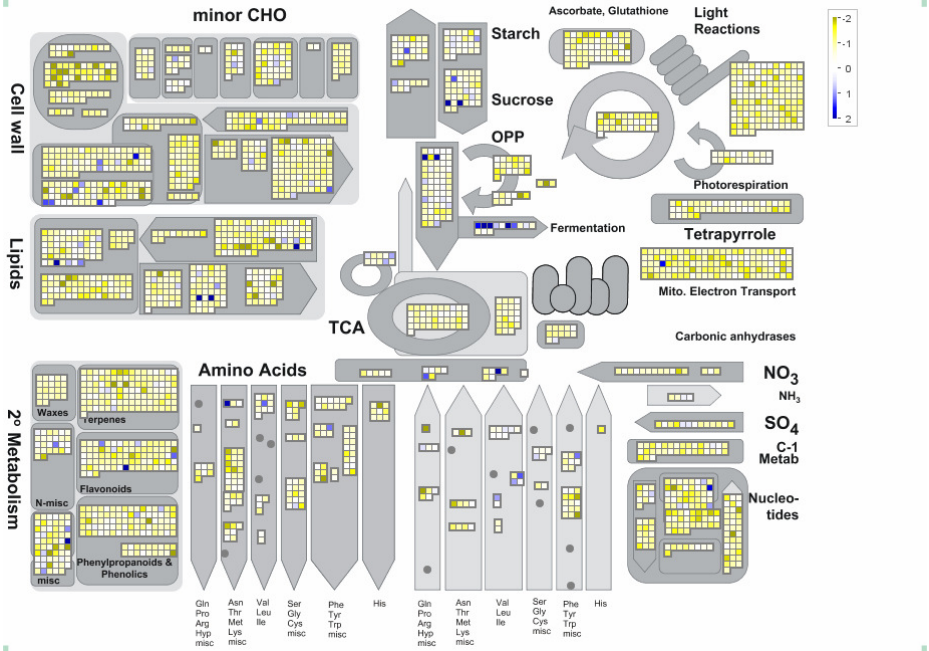
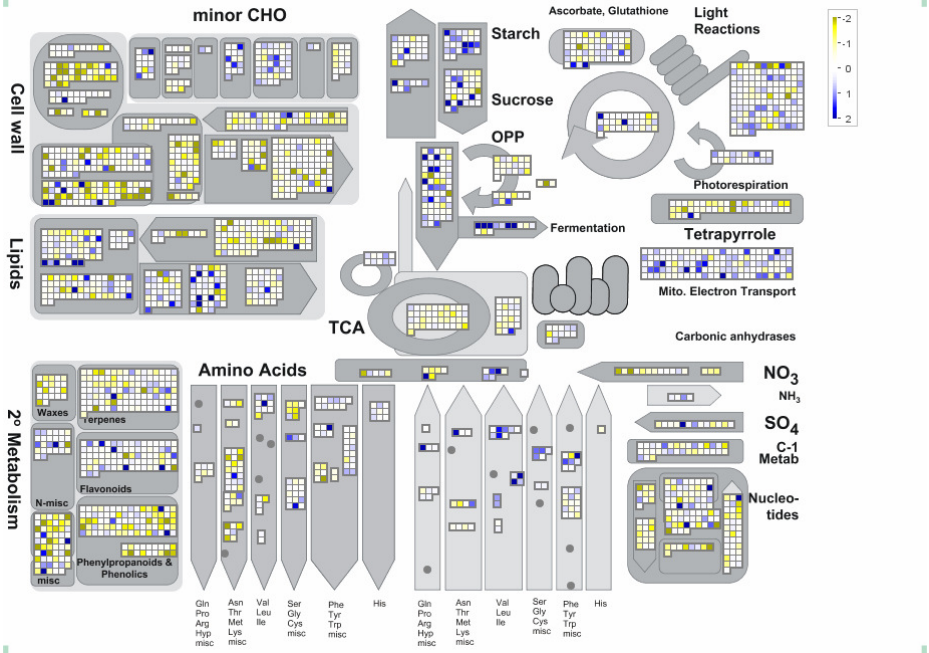


Figure 3.12 MapMan outputs with normalized relative transcript abundance shown as a signal-$\log_2$-ratio (SLR) to scale for dramatic changes in abundance. A SLR of +2, corresponding to a 4-fold induction, is shown in blue while a SLR of -2, corresponding to a 4-fold decrease is a yellow shade.
Panel (a) shows 2HS/2NS whole seedling total mRNA while panel (b) shows 2HS/2NS whole seedling immunuopurified (IP) mRNA (see Figure 3.2) illustrating the strong induction of transcripts encoding proteins associated with fermentative pathways.

other mRNAs associated with similar metabolic pathways. General areas of note are sucrose mobilization, glycolysis, fermentation, and selective catabolism of amino acids.

Figure 3.13 is a similar illustration of the adjustments in response to nine hours of low oxygen stress, when anoxia has been established. In 3.13a the mRNAs which are higher in abundance in the stress condition are shown in blue, whereas mRNAs which are depleted in the stress condition compared with controls are shown in yellow.

(a)



(b)

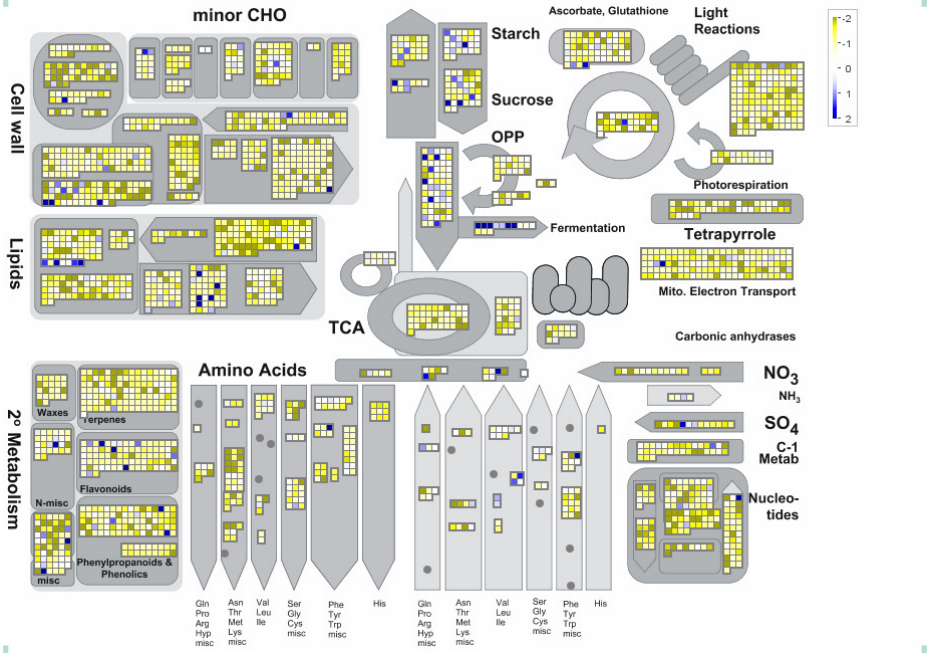


Figure 3.13 MapMan outputs with normalized relative transcript abundance shown as a signal-$\log_2$-ratio (SLR) to scale for dramatic changes in abundance. A SLR of +2, corresponding to a 4-fold induction, is shown in blue while a SLR of -2, corresponding to a 4-fold decrease is a yellow shade.
Panel (a) shows 9HS/9NS whole seedling total mRNA while panel (b) shows 9HS/9NS whole seedling immunuopurified (IP) mRNA (see Figure 3.2) illustrating the strong induction of fermentative pathways, including catabolism of specific aliphatic and aromatic amino acids. The dramatic coverage of yellow indicates translational repression across a range of processes, with induction of specific pathways.

3.4 Conclusions

NMR was chosen for this study as a non-biased, quantitative approach for non-targeted profiling of metabolites extracted from whole seedlings. The time for analysis of each sample, including sample preparation, was less than two hours, but there was no capability for multiplexing the analysis. A large amount of information was recorded with each experiment; ^{1}H NMR resonances were observed for most metabolites present at concentrations in the micromolar regime.

Multivariate statistical treatment accounted for all observable metabolites. Rather than a separate enzymatic assay for each analyte, we observed directly the relative quantities and how they changed as a function of biological treatment. Using binning to cover the entire spectrum, we were able to readily identify spectral regions of high variance and determine covarying regions (Figure 3.6).

Quantitative fold changes were reported for metabolites with a well-resolved resonance with a S/N above 10. This represented 12 metabolites including sugars, amino acids and organic acids (Table 3.2). Concentrations were determined by NMR for Ala, Glu, Gln, Asn, valine, ornithine, glucose, fructose, sucrose, succinate, choline, and GABA. Comparison with similar studies in a diverse array of photosynthetic organisms showed good agreement in a recent paper.[125] Fold changes similar to those we reported for 7 d old Arabidopsis were observed in *Oryza sativum* (rice)*, Solanum tuberosum* (potato)*, Ricinus communis* (castor)*, Populus × canescens* (poplar)*, Lotus japonicus* (model legume)*, and Chlamydomonas reinhardtii* (model algae). This underscores evolutionary conservation in adaptations to survive low oxygen, illustrating the utility of

model organisms and how studies using them can be extended to useful crop (*O. sativa* or cultivated) species.

Both transcript profiling by Affymetrix GeneChip (ATH1) and metabolite profiling by ^{1}H NMR generates a massive amount of information in a relatively short time, requiring months of analysis to gain insight into the biological processes under study. One challenge is the judicious choice of sampling time points. Knowing that mRNA populations can be reconfigured within five minutes,[126] it is unlikely that protein expression and downstream enzymatic activity can be modified to change metabolite levels on the same timescale.

One limitation of these techniques is that differences between sample populations are expressed at relative levels with some confidence that the two populations are different. This is distinctly different from classical experiments to measure the activity of one enzyme or the concentration of one metabolite very accurately and those which report milligram of metabolite per gram fresh weight of tissue. However, integrating data from two intrinsically interrelated processes such as transcription, translated transcripts, and metabolism can provide more insight into the state of an organism than a classical experiment.

It was previously known that pyruvate is converted to lactate and ethanol in hypoxic plant tissues, preventing carbon sources from entering the TCA cycle. It was also observed that pyruvate and Glu are converted to Ala,[70, 90] where Ala is thought to play several roles. Ala accumulation may help control cytoplasmic pH, it may be acting as a pool of three-carbon molecules to provide a recyclable carbon store, [47] or it may be

drawing pyruvate away from the production of acetaldehyde which is a toxic compound to plants. The integration of transcription and metabolite profiling datasets shows promise in revealing the mechanisms of transcriptional and metabolic regulation or adjustments in response to hypoxic stress.[125]

In the area of oxygen deprivation, open questions remain regarding light availability, nutrients availability, temperature and combination stresses and their impact on primary metabolism and energy utilization. The interplay between carbon and nitrogen metabolism is of interest.[103, 106] Efforts to improve plant survival of stress can be mediated by exogenous application of nutrient, in agriculture these are called safeners. The effects of oxygen deprivation stress pretreatment and its enhancement of survivability have not been fully investigated in Arabidopsis,[7, 127] although such work has increased the performance of heart transplants.[127, 128] If farmers could anticipate a oxygen deprivation-inducing condition (i.e. waterlogging or partial to complete submergence or ice-encasement) with an understanding of molecular signaling under oxygen deprivation, survivability and biomass accumulation of crops could be increased by simulating oxygen deprivation pre-treatment by exogenous application of some endogenous substance. Alternately bioengineered plants with inducible promoters driving the expression of survivability-related genes could be planted and treated with an inducing agent in advance of environmental stress such as flood or ice storm. The use of transgenic plants to determine which pathways are critical for survival may shed light on areas of current controversy, such as whether the TCA cycle is operating in a reversed mode or two distinct branches.[103]

Experiments comparing phenotypic changes of roots and shoots under oxygen deprivation in the light and dark[127] reporting changes in volatile metabolites, adenylates and sugars,[4] have been completed but a more detailed look at the molecular mechanisms utilized by Arabidopsis to survive oxygen deprivation could be achieved using our ^{1}H NMR metabolic profiling approach. In chapter 4, we explore differences in these molecular mechanisms in roots and shoots. Photosynthetic capability of shoots as well as the specialized mechanisms for nutrient assimilation in roots would be distinct. This is of relevance to agriculture because some crops are predominantly shoots (broccoli, spinach, lettuce) and some are roots (potato, ginger, carrot).

3.5 References

1. Zanetti, M. E.; Chang, I. F.; Gong, F.; Galbraith, D. W.; Bailey-Serres, J., Immunopurification of polyribosomal complexes of Arabidopsis for global analysis of gene expression. *Plant Physiol.* **2005,** 138, (2), 624-635.

2. Bailey-Serres, J.; Voesenek, L. A. C. J., Flooding stress: acclimations and genetic diversity. *Annu. Rev. Plant Biol.* **2008,** 59, 313-339.

3. Meyerowitz, E. M., Arabidopsis, a useful weed. *Cell* **1989,** 56, 263-269.

4. Mustroph, A.; Boamfa, E. I.; Laarhoven, L. J. J.; Harren, F. J. M.; Pors, Y.; Grimm, B., Organ specific analysis of the anaerobic primary metabolism in rice and wheat seedlings II: Light exposure reduces needs for fermentation and extends survival during anaerobiosis. *Planta* **2006,** 225, 139-152.

5. Good, A. G.; Muench, D. G., Purification and characterization of an anaerobically induced alanine aminotransferase from barley roots. *Plant Physiol.* **1992,** 99, 1520-1525.

6. Stoimenova, M.; Igamberdiev, A. U.; Gupta, K. J.; Hill, R. D., Nitrite-driven anaerobic ATP synthesis in barley and rice root mitochondria. *Planta* **2007,** 226, (2), 465-474.

7. Mustroph, A.; Albrecht, G., Fermentation metabolism in roots of wheat seedlings after hypoxic pre-treatment in different anoxic incubation systems. *J. Plant Physiol.* **2007,** 164, (4), 394-407.

8. Boyes, D. C.; Zayed, A. M.; Ascenzi, R.; McCaskill, A. J.; Hoffman, N. E.; Davis, K. R.; Gorlach, J., Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* **2001,** 13, 1499-1510.

9. Kaiser, K. A.; Bailey-Serres, J., Plants and oxygen. http://youtu.be/JEU1rnxhZsk (accessed Aug 26, 2011).

10. Campbell, M. K.; Farrell, S. O., *Biochemistry*. Fifth ed.; Thompson Brooks/Cole: Belmont, CA, **2006**; p 689.

11. Borisjuk, L.; Macherel, D.; Benamar, A.; Wobus, U.; Rolletschek, H., Low oxygen sensing and balancing in plant seeds: a role for nitric oxide. *New Phytol.* **2007,** 176, (4), 813-823.

12. Grossman, A. R.; Catalanotti, C.; Yang, W. Q.; Dubini, A.; Magneschi, L.; Subramanian, V.; Posewitz, M. C.; Seibert, M., Multiple facets of anoxic metabolism and hydrogen production in the unicellular green alga *Chlamydomonas reinhardtii*. *New Phytol.* **2011,** 190, (2), 279-288.

13. Oliver, S. N.; Lunn, J. E.; Urbanczyk-Wochniak, E.; Lytovchenko, A.; Dongen, J. T. v.; Faix, B.; Schmalzlin, E.; Fernie, A. R.; Geigenberger, P., Decreased expression of a cytosolic pyruvate kinase in potato tubers leads to a decline in pyruvate resulting in an *in vivo* repression of the alternative oxidase. *Plant Physiol.* **2008,** 148, 1640-1654.

14. Fukao, T.; Bailey-Serres, J., Plant responses to hypoxia - is survival a balancing act? *Trends Plant Sci.* **2004,** 9, (9), 449-456.

15. Bailey-Serres, J.; Voesenek, L., Life in the balance: a signaling network controlling survival of flooding. *Curr. Opin. Plant Biol.* **2010,** 13, (5), 489-494.

16. Mustroph, A.; Lee, S. C.; Oosumi, T.; Zanetti, M. E.; Yang, H. J.; Ma, K.; Yaghoubi-Masihi, A.; Fukao, T.; Bailey-Serres, J., Cross-kingdom comparison of transcriptomic adjustments to low-oxygen stress highlights conserved and plant-specific responses. *Plant Physiol.* **2010,** 152, (3), 1484-1500.

17. Hare, P. D.; Cress, W. A.; van Staden, J., Dissecting the roles of osmolyte accumulation during stress. *Plant Cell Environ.* **1998,** 21, 535-553.

18. Sweetlove, L. J.; Fell, D.; Fernie, A. R., Getting to grips with the plant metabolic network. *Biochem. J.* **2008,** 409, 27-41.

19. Kim, J. K.; Bamba, T.; Harada, K.; Fukusaki, E.; Kobayashi, A., Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress treatment. *J. Exp. Bot.* **2007,** 58, 415-424.

20. Tague, B. W.; Mantis, J., *Arabidopsis protocols*. Second ed.; Humana Press: Clifton, NJ, **2006**; Vol. 323, pp 469.

21. Bailey-Serres, J.; Sorenson, R.; Juntawong, P., Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends Plant Sci.* **2009,** 14, (8), 443-453.

22. Fennoy, S. L.; Bailey-Serres, J., Post-transcriptional regulation of gene expression in oxygen-deprived roots of maize. *Plant J.* **1995,** 7, (2), 287-295.

23. Fennoy, S. L.; Nong, T.; Bailey-Serres, J., Transcriptional and post-transcriptional processes regulate gene expression in oxygen-deprived roots of maize. *Plant J.* **1998,** 15, (6), 727-735.

24. The Arabidopsis Information Resource (TAIR). http://www.arabidopsis.org (accessed Mar 10, 2007).

25. Bailey-Serres, J., Selective translation of cytoplasmic mRNAs in plants. *Trends Plant Sci.* **1999,** 4, (4), 142-148.

26. Initiative, T. A. G., Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000,** 408, 796-815.

27. Gatherer, D., So what do we mean when we say that systems biology is holistic? *BMC Syst. Biol.* **2010,** 4, 22.

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

29. Paley, S. M.; Karp, P. D., The pathway tools cellular overview diagram and omics viewer. *Nucleic Acids Res.* **2006,** 34, (13), 3771-3778.

30. Fiehn, O.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Lee, D. Y.; Lu, Y.; Moon, S.; Nikolau, B., Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J.* **2008,** 53, 691-704.

31. Bailey-Serres, J.; Freeling, M., Hypoxic Stress-Induced Changes in Ribosomes of Maize Seedling Roots. *Plant Physiol.* **1990,** 94, 1237-1243.

32. Hochachka, P. W., *Living without Oxygen: Closed and Open Systems in Hypoxia Tolerance*. Harvard University Press: Cambridge, MA, **1980**; p 181.

33. Storey, K. B., *Functional Metabolism: regulation and adaptation*. John Wiley & Sons: Hoboken, NJ, **2004**; p 417.

34. Bailey-Serres, J.; Chang, R., Sensing and Signalling in Response to Oxygen Deprivaton in Plants and Other Organisms. *Ann. Bot-London* **2005,** 96, 507-518.

35. Helbing, D.; Deutsch, A.; Diez, S.; Peters, K.; Kalaidzidis, Y.; Padberg-Gehle, K.; Lammer, S.; Johansson, A.; Breier, G.; Schulze, F.; Zerial, M., Biologistics and the struggle for efficiency: concepts and perspectives. *Adv. Complex Syst.* **2009,** 12, (6), 533-548.

36. Kawaguchi, R.; Bailey-Serres, J., Regulation of translational initiation in plants. *Curr. Opin. Plant Biol.* **2002,** 5, 460-465.

37. Branco-Price, C., Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Ann. Bot-London* **2005,** 96, 647-660.

38. Maniara, G.; Rajamoorthi, K.; Rajan, S.; Stockton, G. W., *Anal. Chem.* **1998,** 70, 4921-4928.

39. Malz, F.; Jancke, H., Validation of quantitative NMR. *J. Pharmaceut. Biomed.* **2005,** 38, 813-823.

40. Kolbe, A.; Olivier, S. N.; Fernie, A. R.; Stitt, M.; Donger, J. T. v.; Geigenberger, P., Combined transcript and metabolite profiling of Arabidopsis leaves reveals fundamental effects of the thiol-disulfide status on plant metabolism. *Plant Physiol.* **2006,** 141, 412-422.

41. Carrari, F.; Baxter, C.; Usadel, B.; Urbanczyk-Wochniak, E.; Zanor, M.-I.; Nunes-Nesi, A.; Nikiforova, V.; Cantero, D.; Ratzka, A.; Pauly, M.; Sweetlove, L. J.; Fernie, A. R., Integrated anaysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.* **2006,** 142, 1380-1396.

42. Rischer, H.; Oresic, M.; Seppanen-Laakso, T.; Katajamaa, M.; Lammertyn, F.; Ardiles-Diaz, W.; Van Montagu, M. C. E.; Inze, D.; Oksman-Caldenty, K.-M.; Goossens, A., Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *P. Nat. Acad Sci. USA* **2006,** 103, (14), 5614-5619.

43. Hirai, M. Y.; Sugiyama, K.; Sawada, Y.; Tohge, T.; Obayashi, T.; Suzuki, A.; Araki, R.; Sakurai, N.; Suzuki, H.; Aoki, K.; Goda, H.; Nishizawa, O. I.; Shibata, D.; Saito, K., Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *P. Nat. Acad Sci. USA* **2007,** 104, (15), 6478-6483.

44. Connor, S. C.; Gray, R. A.; Hodson, M. P.; Clayton, N. M.; Haselden, J. N.; Chessell, I. P.; Bountra, C., An NMR-based metabolic profiling study of inflammatory pain using the rat FCA model. *Metabolomics* **2007,** 3, (1), 29-39.

45. Widarto, H. T.; Meijden, E. V. D.; Lefeber, A. W. M.; Erkelens, C., Metabolomic differentiation of *Brassica rapa* following herbivory by different insect instars using two-dimensional nuclear magnetic resonance spectroscopy. *J. Chem. Ecol.* **2006,** 32, 2417-2428.

46. Rivasseau, C.; Seemann, M.; Boisson, A. M.; Streb, P.; Gout, E.; Douce, R.; Rohmer, M.; Bligny, R., Accumulation of 2-C-methyl-d-erythritol 2,4-cyclodiphosphate in illuminated plant leaves at supraoptimal temperatures reveals a bottleneck of the prokaryotic methylerythritol 4-phosphate pathway of isoprenoid biosynthesis. *Plant Cell Environ.* **2009,** 32, (1), 82-92.

47. Deborde, C.; Maucourt, M.; Baldet, P.; Bernillon, S.; Biais, B.; Talon, G.; Ferrand, C.; Jacob, D.; Ferry-Dumazet, H.; de Daruvar, A.; Rolin, D.; Moing, A., Proton NMR quantitative profiling for quality assessment of greenhouse-grown tomato fruit. *Metabolomics* **2009,** 5, (2), 183-198.

48. Hendrawati, O.; Yao, Q.; Kim, H. K.; Linthorst, H. J. M.; Erkelens, C.; Lefeber, A. W. M.; Choi, Y. H.; Verpoorte, R., Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Sci.* **2006,** 170, 1118-1124.

49. Hoefgen, R.; Nikiforova, V. J., Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiol. Plantarum* **2008,** 132, 190-198.

50. Kim, S. W.; Koo, B. C.; Kim, J.; Liu, J. R., Metabolic discrimination of sucrose starvation from Arabidopsis cell suspension by ^{1}H NMR Based Metabolomics. *Biotechnol. Bioprocess Eng.* **2007,** 12, 653-661.

51. Boernsen, K. O.; Gatzek, S.; Imbert, G., Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Anal. Chem.* **2005,** 77, (22), 7255-7264.

52. Rabenstein, D. L., Sensitivity enhancement by signal averaging in pulsed/Fourier transform NMR spectroscopy. *J. Chem. Ed.* **1984,** 61, 909-913.

53. Govindaraju, V.; Young, K.; Maudsley, A. A., Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **2000,** 13, 129-153.

54. Xu, Q.; Sachs, J. R.; Wang, T.-C.; Schaefer, W. H., Quantification and identification of components in solution mixtures from 1D proton NMR spectra using singular value decomposition. *Anal. Chem.* **2006,** 78, 7175-7185.

55. Fumagalli, E.; Baldoni, E.; Abbruscato, P.; Piffanelli, P.; Genga, A.; Lamanna, R.; Consonni, R., NMR techniques coupled with multivariate statistical analysis: tools to analyse *Oryza sativa* metabolic content under stress conditions. *J. Agron. Crop Sci.* **2009,** 195, (2), 77-88.

56. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. application in ^{1}H NMR metabonomics. *Anal. Chem.* **2006,** 78, 4281-4290.

57. Brown, M.; Dunn, W. B.; Ellis, D. I.; Goodacre, R.; Handl, J.; Knowles, J. D.; O'Hagan, S.; Spasic, I.; Kell, D. B., A metabolome pipeline: from concept to data to knowledge. *Metabolomics* **2005,** 1, (1), 39-51.

58. Fukusaki, E.; Kobayashi, A., Plant Metabolomics: Potential for Practical Operation. *J. Biosci. Bioeng.* **2005,** 100, 347-354.

59. Martin, F.-P. J.; Dumas, M.-E.; Wang, Y.; Legido-Quigley, C.; Yap, I. K. S.; Tang, H.; Zirah, S.; Murphy, G. M.; Cloarec, O.; Lindon, J. C.; Sprenger, N.; Fay, L. B.; Kochhar, S.; Bladeren, P. v.; Holmes, E.; Nicholson, J. K., A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Mol. Syst. Biol.* **2007,** 3, 112.

60. Zhang, S.; Zheng, C.; Lanza, I. R.; Nair, K. S.; Raftery, D.; Vitek, O., Interdependence of signal processing and analysis of urine ^{1}H NMR spectra for metabolic profiling. *Anal. Chem.* **2009,** 81, (15), 6080-6088.

61. Gipson, G. T.; Tatsuoka, K. S.; Sweatman, B. C.; Connor, S. C., Weighted least-squares deconvolution method for discovery of group differences between complex biofluid ^{1}H NMR spectra. *J. Magn. Reson.* **2006,** 183, 269-277.

62. Gromova, M.; Roby, C., Toward *Arabidopsis thaliana* hydrophilic metabolome: assessment of extraction methods and quantitative ^{1}H NMR. *Physiol. Plantarum* **2010,** 140, 111-127.

63. Allen, M. P., *Understanding Regression Analysis*. Plenum Press: New York, NY, 1997; p 216.

64. Branco-Price, C.; Kaiser, K. A.; Jang, C. J. H.; Larive, C. K.; Bailey-Serres, J., Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* **2008,** 56, (5), 743-755.

65. Gupta, K. J.; Zabalza, A.; van Dongen, J. T., Regulation of respiration when the oxygen availability changes. *Physiol. Plantarum* **2009,** 137, (4), 383-391.

66. Plaxton, W. C.; Podesta, F. E., The functional organization and control of plant respiration. *Crit. Rev. Plant Sci.* **2006,** 25, (2), 159-198.

67. Sweetlove, L. J.; Beard, K. F. M.; Nunes-Nesi, A.; Fernie, A. R.; Ratcliffe, R. G., Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci.* **2010,** 15, 462-470.

68. van Dongen, J. T.; Frohlich, A.; Ramirez-Aguilar, S. J.; Schauer, N.; Fernie, A. R.; Erban, A.; Kopka, J.; Clark, J.; Langer, A.; Geigenberger, P., Transcript and metabolite profiling of the adaptive response to mild decreases in oxygen concentration in the roots of arabidopsis plants. *Ann. Bot-London* **2009,** 103, (2), 269-280.

69. Kreuzwieser, J.; Hauberg, J.; Howell, K. A.; Carroll, A.; Rennenberg, H.; Millar, A. H.; Whelan, J., Differential Response of Gray Poplar Leaves and Roots Underpins Stress Adaptation during Hypoxia. *Plant Physiol.* **2009,** 149, (1), 461-473.

70. Miyashita, Y.; Good, A. G., Contribution of the GABA shunt to hypoxia-induced alanine accumulation in roots of *Arabidopsis thaliana*. *Plant Cell Physiol.* **2008,** 49, (1), 92-102.

71. Limami, A. M.; Glevarec, G.; Ricoult, C.; Cliquet, J. B.; Planchet, E., Concerted modulation of alanine and glutamate metabolism in young *Medicago truncatula* seedlings under hypoxic stress. *J. Exp. Bot.* **2008,** 59, (9), 2325-2335.

72. Liu, F.; VanToai, T.; Moy, L. P.; Bock, G.; Linford, L. D.; Quackenbush, J., Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. *Plant Physiol.* **2005,** 137, 1115-1129.

73. Dolferus, R.; Wolansky, M.; Carroll, R.; Miyashita, Y.; Ismond, K.; Good, A., Functional analysis of lactate dehydrogenase during hypoxic stress in *Arabidopsis*. *Funct. Plant Biol.* **2008,** 35, 131-140.

74. Rocha, M.; Sodek, L.; Licausi, F.; Hameed, M. W.; Dornelas, M. C.; van Dongen, J. T., Analysis of alanine aminotransferase in various organs of soybean (*Glycine max*) and in dependence of different nitrogen fertilisers during hypoxic stress. *Amino Acids* **2010,** 39, (4), 1043-1053.

75. Miyashita, Y.; Dolferus, R.; Ismond, K.; Good, A. G., Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in *Arabidopsis thaliana*. *Plant J.* **2007,** 49, 1108-1121.

76. Loreti, E.; Poggi, A.; Novi, G.; Alpi, A.; Perata, P., A genome-wide analysis of the effects of sucrose on gene expression in Arabidopsis seedlings under anoxia. *Plant Physiol.* **2005,** 137, 1130-1138.

77. Allan, W. L.; Simpson, J. P.; Clark, S. M.; Shelp, B. J., γ-Hydroxybutyrate accumulation in Arabidopsis and tobacco plants is a general response to abiotic stress: putative regulation by redox balance and glyoxylate reductase isoforms. *J. Exp. Bot.* **2008,** 59, (9), 2555-2564.

78. Drew, M. C., Oxygen deficiency and root metabolism: Injury and acclimation under hypoxia and anoxia. *Ann. Rev. Plant Phys.* **1997,** 48, 223-250.

79. Anderson, L. L.; Mao, X.; Scott, B. A.; Crowder, C. M., Survival from hypoxia in *C. elegans* by inactivation of aminoacyl-tRNA synthetases. *Science* **2009,** 323, 630-633.

80. Subbaiah, C. C.; Sachs, M. M., Molecular and cellular adaptations of maize to flooding stress. *Ann. Bot-London* **2003,** 91, S119-127.

81. Zhang, J.; Toai, T. V.; Huynh, L.; Preiszner, J., Development of flooding-tolerant *Arabidopsis thaliana* by autoregulated cytokinin production. *Mol. Breed.* **2000,** 6, 135-144.

82. Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006,** 2, (4), 171-196.

83. Faber, K. Reliability of principal component analysis. http://www.chemometry.com/Research/PCA.html (accessed Nov 14, 2011)

84. Magnus, J. B.; Hollwedel, D.; Oldiges, M.; Takors, R., Monitoring and modeling of the reaction dynamics in the valine/leucine synthesis pathway in *Corynebacterium glutamicum*. *Biotechnol. Prog.* **2006,** 22, 1071-1083.

85. Zabalza, A.; Orcaray, L.; Igal, M.; Schauer, N.; Fernie, A. R.; Geigenberger, P.; Dongen, J. T. v.; Royuela, M., Unraveling the role of fermentation in the mode of action of acetolactate synthase inhibitors by metabolic profiling. *J. Plant Physiol.* **2011,** 168, 1568-1575.

86. Knill, T.; Schuster, J.; Reichelt, M.; Gershenzon, J.; Binder, S., Arabidopsis branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. *Plant Physiol.* **2008,** 146, 1028-1039.

87. Plant Metabolic Network (PMN) http://pmn.plantcyc.org/ARA/, on www.plantcyc.org (accessed Nov 14, 2011)

88. Chen, Z.-Y.; Burow, M. D.; Mason, C. B.; Moroney, J. V., A low-CO_2-inducible gene encoding alanine:alpha-ketoglutarate aminotransferase in *Chlamydomonas reinhardtii*. *Plant Physiol.* **1996,** 112, 677-684.

89. Rutter, J. C.; Erismann, K. H., Influence of Nitrogen Availability on Aminotransferases in Lemna minor L. *J. Exp. Bot.* **1985,** 36, (165), 583-589.

90. Rocha, M.; Licausi, F.; Araujo, W. L.; Nunes-Nesi, A.; Sodek, L.; Fernie, A. R.; van Dongen, J. T., Glycolysis and the tricarboxylic acid cycle are linked by alanine aminotransferase during hypoxia induced by waterlogging of *Lotus japonicus*. *Plant Physiol.* **2010,** 152, (3), 1501-1513.

91. Roberts, J. K. M.; Callis, J.; Jardetsky, O.; Walbot, V.; Freeling, M., Cytoplasmic acidosis as a determinant of flooding intolerance in plants. *P. Nat. Acad Sci. USA* **1984,** 81, 6029-6033.

92. Betsche, T.; Eisling, R., Refixation of photorespiratory ammonia and the role of alanine in photorespiration: Studies with ^{15}N. *Plant Soil* **1986,** 91, 367-371.

93. Fan, T. W.-M.; Lane, A. N.; Higashi, R. M., *In vivo* and *in vitro* metabolomic analysis of anaerobic rice coleoptiles revealed unexpected pathways. *Russ. J. Plant Physil.* **2003,** 50, (6), 787-793.

94. Streeter, J. G.; Thompson, J. F., Anaerobic accumulation of gamma-aminobutyric acid and alanine in radish leaves (*Raphanus sativus* L.). *Plant Physiol.* **1972,** 49, 572-578.

95. Wallace, W.; Secor, J.; Schrader, L. E., Rapid accumulation of gamma-aminobutyric acid and alanine in soybean leaves in response to an abrupt transfer to lower temperature, darkness, or mechanical manipulation. *Plant Physiol.* **1984,** 75, 170-175.

96. Rivoal, J.; Hanson, A. D., Metabolic control of anaerobic glycolysis. Overexpression of lactate dehydrogenase in transgenic tomato roots supports the Davies-Roberts hypothesis and points to a critical role for lactate secretion. *Plant Physiol.* **1994,** 106, 1179-1185.

97. Blokhina, O.; Fagerstedt, K. V., Reactive oxygen species and nitric oxide in plant mitochondria: origin and redundant regulatory systems. *Physiol. Plantarum* **2010,** 138, (4), 447-462.

98. Kaiser, K. A.; Barding, G. A.; Larive, C. K., Metabolic profiling of plants by ^{1}H-NMR: A comparison of metabolite extraction strategies using rosette leaves of the model plant *Arabidopsis thaliana*. *Magn. Reson. Chem.* **2009,** 47, (S1), S147-S156.

99. Davies, D. D., Anaerobic metabolism and the production of organic acids. In *The Biochemistry of Plants*, Davies, D. D., Ed. Academic Press: New York, **1980**; Vol. 2, pp 581-611.

100. Ratcliffe, R. G., Metabolic aspects of the anoxic response in plant tissue. In *Environment and Plant Metabolism: Flexibility and Acclimation*, Smirnoff, N., Ed. Bios Scientific: Oxford, **1995**; pp 111-127.

101. Stewart, G. R.; Larher, F., Accumulation of amino acids and related compounds in relation to environmental stress. In *The Biochemistry of Plants*, Miflin, B. J., Ed. Academic Press: New York, **1980**; Vol. 5, pp 609-635.

102. Good, A. G.; Crosby, W. L., Anaerobic induction of alanine aminotransferase in barley root tissue. *Plant Physiol.* **1989,** 90, 1305-1309.

103. Fait, A.; Fromm, H.; Walter, D.; Galili, G.; Fernie, A. R., Highway or byway: the metabolic role of the GABA shunt in plants. *Trends Plant Sci.* **2007,** 13, (1), 14-19.

104. Lea, P. J.; Blackwell, R. D.; Murray, A. J. S.; Joy, K. W., The use of mutants lacking glutamine synthetase and glutamate synthetase to study their role in plant nitrogen metabolism. In *Plant Nitrogen Metabolism*, Poulton, J. E.; Romero, J. T.; Conn, E. E., Eds. Plenum Press: New York, **1989**; pp 157-189.

105. Zhang, J. T.; Zhang, Y.; Du, Y. Y.; Chen, S. Y.; Tang, H. R., Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res.* 10, (4), 1904-1914.

106. Lam, H.-M.; Coschigano, K. T.; Oliveira, I. C.; Melo-Oliveira, R.; Coruzzi, G. M., The molecular-genetics of nitrogen assimilation into amino acids in higher plants. *Ann. Rev. Plant Phys.* **1996,** 47, 569-593.

107. Zhang, J.; Zhang, Y.; Du, Y.; Chen, S.; Tang, H., Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res.* **2011,** 10, 1904-1914.

108. Ratcliffe, R. G., *In vivo* NMR studies of the metabolic response of plant tissues to anoxia. *Ann. Bot-London* **1997,** 79, (Supplement A), 39-48.

109. Brown, B., Abiogenesis. In *Evolution: A Historical Perspective*, Baigrie, B., Ed. Greenwood Press: Westport, CT, **2007**; pp 129-148.

110. Rasmusson, A. G.; Fernie, A. R.; Dongen, J. T. v., Alternative oxidase: a defense against metabolic fluctuations? *Physiol. Plantarum* **2009,** 137, 371-382.

111. Wurtele, E. S.; Li, L.; Berleant, D.; Cook, D.; Dickerson, J. A.; Ding, J.; Hofmann, H.; Lawrence, M.; Lee, E. K.; Li, J.; Mentzen, W.; Miller, L.; Nikolau, B. J.; Ranson, N.; Wang, Y., MetNet: systems biology software for Arabidopsis. In *Concepts in Plant Metabolomics*, Springer.: **2007**; pp 145-158.

112. Beeckmans, S., Glyoxylate cycle. *P. Nat. Acad Sci. USA* **1980,** 77, (3), 1521-1525.

113. Ward, J. L.; Harris, C.; Lewis, J.; Beale, M. H., Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **2003,** 62, 949-957.

114. Choi, H.-K.; Choi, Y. H.; Verberne, M.; Lefeber, A. W. M.; Erkelens, C.; Verpoorte, R., Metabolic fingerprinting of wild type and transgenic tobacco plants by ^{1}H NMR and multivariate analysis technique. *Phytochemistry* **2004,** 65, 857-864.

115. Nicholson, J. K.; Lindon, J. C.; Holmes, E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999,** 29, (11), 1181-1189.

116. Lindon, J. C.; Holmes, E.; Nicholson, J. K., So What's the Deal with Metabonomics? *Anal. Chem.* **2003**, 384A-391A.

117. Schwender, J., Metabolic flux analysis as a tool in metabolic engineering of plants. *Curr. Opin. Biotech.* **2008,** 19, (2), 131-137.

118. Mattoo, A. K.; Sobolev, A. P.; Neelam, A.; Goyal, R. K.; Handa, A. K.; Segre, A. L., Nuclear magnetic resonance spectroscopy-based metabolite profiling of transgenic tomato fruit engineered to accumulate spermidine and spermine reveals enhanced anabolic and nitrogen-carbon interactions. *Plant Physiol.* **2006,** 142, 1759-1770.

119. Williams, T. C. R.; Miguet, L.; Masakapalli, S. K.; Kruger, N. J.; Sweetlove, L. J.; Ratcliffe, R. G., Metabolic network fluxes in heterotrophic Arabidopsis cells: stability of the flux distribution under different oxygenation conditions. *Plant Physiol.* **2008,** 148, 704-718.

120. Roessner, U.; Luedemann, A.; Brust, D.; Fiehn, O.; Linke, T.; Willmitzer, L.; Fernie, A. R., Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **2001,** 13, (1), 11-29.

121. Carrari, F.; Nunes-Nesi, A.; Gibon, Y.; Lytovchenko, A.; Loureiro, M. E.; Fernie, A. R., Reduced expression of aconitase results in an enhanced rate of photosynthesis and marked shifts in carbon partitioning in illuminated leaves of wild species tomato. *Plant Physiol.* **2003,** 133, (3), 1322-1335.

122. Villas-Boas, S. G.; Hojer-Pedersen, J.; Akesson, M.; Smedsgaard, J.; Nielsen, J., Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **2005,** 22, 1155-1169.

123. Farre, E. M.; Fernie, A. R.; Willmitzer, L., Analysis of subcellular metabolite levels of potato tubers (*Solanum tuberosum*) displaying alterations in cellular or extracellular sucrose metabolism. *Metabolomics* **2008,** 4, 161-170.

124. Fait, A.; Angelovici, R.; Less, H.; Ohad, I.; Urbanczyk-Wochniak, E.; Fernie, A. R.; Galili, G., Arabidopsis Seed Development and Germination Is Associated with Temporally Distinct Metabolic Switches. *Plant Physiol.* **2006,** 142, 839-854.

125. Narsai, R.; Rocha, M.; Geigenberger, P.; Whelan, J.; Dongen, J. T. v., Comparative analysis between plant species of transcriptional and metabolic responses to hypoxia. *New Phytol.* **2011**, 190, 472-487.

126. Fernie, A. R.; Geigenberger, P.; Stitt, M., Flux an important, but neglected, component of functional genomics. *Curr. Opin. Plant Biol.* **2005,** 8, (2), 174-182.

127. Ellis, M. H.; Dennis, E. S.; Peacock, W. J., Arabidopsis roots and shoots have different mechanisms for hypoxia stress tolerance. *Plant Physiol.* **1999,** 119, 57-64.

128. Oppegard, S. C.; Nam, K. H.; Carr, J. R.; Skaalure, S. C.; Eddington, D. T., Modulating Temporal and Spatial Oxygenation over Adherent Cellular Cultures. *Plos One* **2009,** 4, (9), 8.

Chapter Four

Micrometabolomics of hypoxic seedling roots and shoots of *Arabidopsis*

Arabidopsis thaliana was used to study the effect of 2 h of hypoxic stress on the central carbon and amino acid metabolism (primary metabolic pathways) of roots and shoots of 1-wk-old seedlings. Seedlings were grown by A.M., under observation by K.A.K. Seedling harvests were carried out by A.M. and K.A.K. together. Immunoprecipitation and mRNA isolation for transcriptome and translatome profiling by use of Affymetrix ATH1 GeneChips was done by A.M. Metabolite extraction for ^{1}H NMR and corresponding analyses were done by K.A.K.

4.1 Introduction

The goals of this work were to quantify metabolites in seedling extracts, using a nontargeted approach spanning primary metabolism. We aimed to determine if the metabolites present in the root are different from those in the shoot under control conditions. In addition, we were interested in examining whether the root and shoot respond differently to a decrease in oxygen availability (hypoxic stress). Taking lessons learned from the work presented in Chapter 3, we were limited by what changes can be measured, but with improvements in sample preparation we hoped to ascertain a clearer picture of primary metabolites involved in low oxygen stress.

Previous organ-specific analysis of the anaerobic primary metabolism in rice and wheat seedlings left open questions.[1, 2] Emission of ethanol and acetaldehyde by roots and shoots of intact seedlings, carbon dioxide and oxygen were monitored simultaneously. In those studies, tissues were collected for metabolite profiling after 4 and 24 h of anaerobic treatment in light and darkness. Glucose, fructose, sucrose and

starch were determined by enzymatic assays.[3] Adenylates were quantified. Activities of pyruvate decarboxylase (PDC, EC 4.1.1.1) and alcohol dehydrogenase (ADH, EC 1.1.1.1) were also measured.[4]

To understand the nature of the data presented in this chapter, it is necessary to understand some basic principles underlying the interplay between plant organs. Specifically, in this chapter we investigate separately seedling roots and shoots under oxygen deficiency. Plant roots become exposed to transient flooding, or irrigation followed by slow drainage, which can lead to oxygen deficiency. In soils saturated with water, O_2 availability is very low because the diffusion of molecular oxygen through water is approximately 10,000 times slower than in air.[5] Soil microorganisms can also proliferate and cause oxygen shortages, forcing roots to endure anaerobic conditions. Shoots and other above-ground organs can experience anoxia in cases where ice has encased leaves.[6] Because of the functional adaptations that distinguish roots and shoots, we investigated hypoxic responses of each tissue type separately.

The aim of this study was to investigate the differential metabolomic response between Arabidopsis (7-d-old) seedling roots and shoots when subjected to short-term (2 h) oxygen deprivation. To investigate this, we placed a population of seedlings grown vertically on agar (0.5x MS salts, 1% sucrose) plates in chambers flooded with Ar gas or exposed to laboratory air. We measured metabolite concentrations by one-dimensional ^{1}H NMR in extracts of lyophilized tissue harvested within less than 3 min following release from argon gas treatment. The results are interpreted in relation to paired translatome analyses using ribosome immunopurification and the Affymetrix ATH1

GeneChip. We hypothesized that by pooling tissue from hundreds of seedlings treated identically, with enough biological replicates, we would discover distinct responses to oxygen deprivation corresponding to the seedling's first organs: root and shoot.

4.2 Experimental Methods

Arabidopsis (35S:HF-RPL18, Columbia-0 ecotype) seeds were surface sterilized by incubation for 5 min in 95 % (v/v) ethanol followed by 10 min in 20 % (v/v) bleach with 0.1 % (v/v) Tween-20, rinsed in sterile water, and imbibed at 4 °C for 3 d. Seeds (~25 per plate) were transferred to 10 cm Petri dishes with solid MS media (0.43 % (w/v) Murashigi Skoog (MS) salts (Sigma, St. Louis, MO), 0.4 % (w/v) Phytagel (Sigma), 1 % (w/v) sucrose, pH 5.7), and placed at a vertical orientation in a growth chamber (Percival Scientific, Inc, model CU36L5C8) under long day conditions (16 h light at ~80 µmol photons m^{-2} s^{-1}/ 8 h darkness) at 23 °C. After 7 d, stress treatments were commenced after the end of the 16-h-day. Oxygen deprivation (hypoxia stress, HS) was imposed exactly as described by Branco-Price et al.[7] Briefly, 16 plates per chamber were placed vertically in Lucite chambers into which 99.998 % (v/v) argon gas was pumped and ambient air was allowed to exit under positive pressure. The time required to purge the chambers of air was about 1.5 h, determined empirically using the sublimation of solid CO_2 to fill chambers with "air" which was evacuated by Ar, bubbling through Erlenmeyer flasks containing distilled water to safeguard against desiccation stress. This treatment gradually deprives the plants of oxygen and carbon dioxide, thereby limiting both photosynthesis and aerobic respiration. For control (NS) treatment, plates were placed in

identical chambers open to ambient air. Both treatments were carried out under dim light (5 to 7 µmol photons m^{-2} s^{-1}) at room temperature (23 to 25 °C). After 2 h of treatment, tissues were harvested into liquid N_2. Tissue harvest was accomplished within 3 min of removal from the treatment chamber.

Roots and shoots were separated by razor blade directly on each plate. Root tissue was transferred to a mortar containing liquid nitrogen. Shoot tissue was transferred to a second mortar containing liquid nitrogen. Tissues were pulverized until reaching the consistency of a fine powder. The root tissue from all 16 plates within each chamber were pooled into the first mortar and transferred to a single 15 mL centrifuge tube, as was the shoot tissue. Each set of tissue was lyophilized immediately following harvesting in the tube provided. The duration of lyophilization was 3 d to ensure complete dryness. Dry tissue was stored at -86 ^{o}C until extraction. Tubes were labeled as follows: RC1_Feb01 = Root tissue, Control, chamber 1, harvested 2/1/08.

The first set of seedlings was planted January 25, 2008 and the last set of NMR experiments were performed the week of June 2, 2008. Measurement of adenylates was carried out separately from ^{1}H NMR-based metabonomics. Translatomes (ribosome-associated mRNA populations) were isolated by immunopurification and analyzed by hybridization to Affymetrix ATH1 GeneChips.

Building on our previous work (Chapter 3), this study included microdissection to separate the root from shoot. Steps were added to the extraction protocol to decrease the abundance of interfering lipids as well as to resuspend metabolites in an aqueous buffer not susceptible to instability over the course of an NMR experiment (Chapter 2). In the

data acquisition, adjustment to pH 7 was carried out to reduce spectral overlap in the region between 2-3 ppm.

4.3 Results and Discussion

The combination of two complementary systems biological approaches proved capable of generating large datasets. There are so-called "oceans of data, rivers of information, and only small puddles of knowledge" to be explored after such an investigation.[8] The first-pass survey of our NMR data was carried out using unsupervised multivariate statistics to reduce the dimensionality of the dataset. Clustering by covariance among variables (samples, treatments, NMR chemical shift regions) was observed by applying a principal component analysis (PCA) algorithm, carried out by the statistical software package Minitab, which was more carefully described in Chapter 1, Section 1.5.1. In this analysis, each NMR spectrum is represented by a single data point (Figure 4.1). A global perspective of the data is taken into consideration by examining it in this way. Outlier detection and errors in sample labeling can be addressed before any further manipulations to the dataset are undertaken.

4.3.1 Multivariate Data Analysis of NMR-based Metabolic Profiles

Examining the dataset qualitatively involves the following initial conditions: five experimental harvests, each containing two bioreplicates, totaling N=10 observations for each biological treatment. Treatments included control conditions (2 h of exposure to lab air under low light) and oxygen deprivation (2 h of exposure to Ar treatment under low

light). The extracts were buffered using acetate-d_3 to a pH meter reading of 4.5. Following extraction at low pH, samples were titrated to neutral pH using ND_4OD.

In a first-pass examination of the dataset using PCA (data not shown), no clustering was observed when all spectral regions (excluding solvent resonances) were included in the model (data not shown). The largest variance was introduced by chemical shift regions occupied by resonances of the sugar protons in the region 3.4 – 4.2 ppm. The hypothesis-generating multivariate data analysis led us to propose that instead of large variations in sugar contents due to biological variance, an easier explanation is that sugar from the growth medium was present in different amounts in all samples. Due to the nature of the harvesting (accomplished rapidly to avoid reoxygenation following release from oxygen deprivation), it is likely that sucrose in the growth medium was havested along with seedlings, and most likely more prominently in roots since root hairs serve to secure roots to the media. Removing all spectral regions occupied by sugar resonances (glucose, fructose, sucrose) prior to PCA afforded a clearer picture of the endogenous metabolome of seedlings (Figure 4.1).

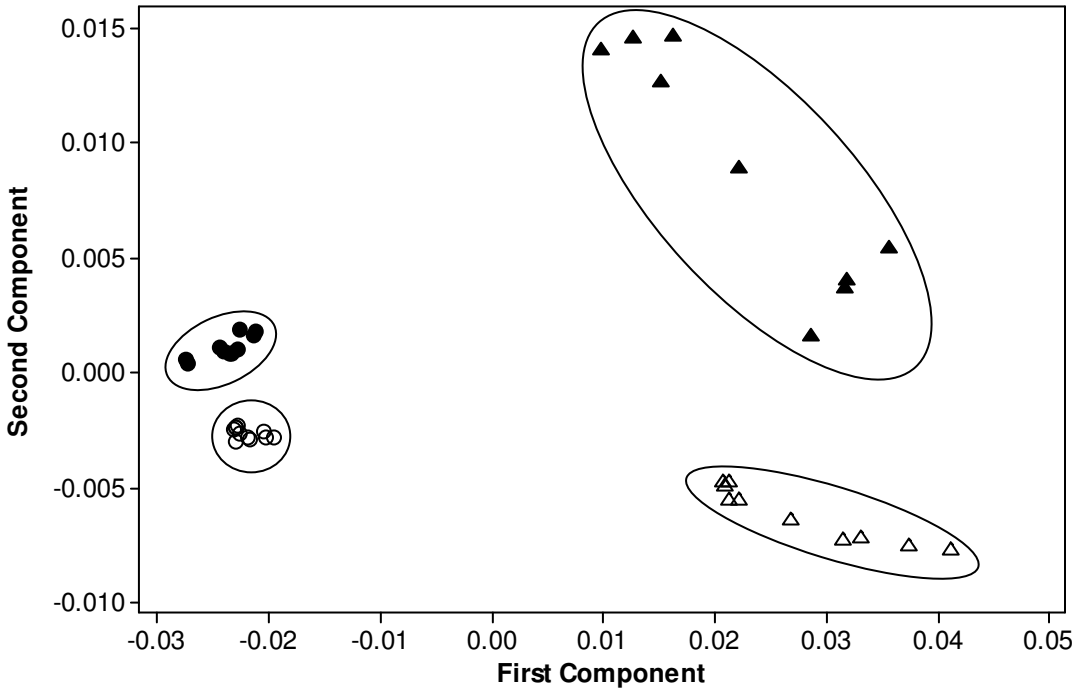


Figure 4.1 Scores plot resulting from PCA using a covariance matrix. Inputs to the model were NMR spectra binned to 0.02 ppm bin widths, excluding regions occupied by solvents and carbohydrate constituents of growth medium. Open symbols are control conditions and closed symbols show the effect of 2 h of Ar treatment on roots (circles) and shoots (triangles).

PCA, detailed in Chapter 1, reveals which samples covary among all spectral regions by their closeness in principal-component space, a new coordinate system arising from a transformation of the original data matrix in which variance is maximized by the first component and more components are calculated in orthogonal directions until all the variance in the dataset is explained.[8-10] In Figure 4.1, four clusters of points enclosed by large ovals are clearly observable in the scores plot, distinguishing root (circles) from shoot (triangles) along the first principal component axis. Treatments can be distinguished along the second principal component axis; open symbols represent control

treatments and closed symbols represent hypoxia-stressed tissue extracts. This description is useful for examination of the corresponding loadings plot (Figure 4.2).

Not unexpectedly, the metabolic fingerprint of roots is distinct from shoots.[11-13] In Figure 4.2, the variables making the largest contributions to the difference between roots and shoots are the β -CH_2- resonance of glutamine (Gln) which occupies 2.13 and 2.15 ppm and the α -CH_2- resonance which occupies 2.45 and 2.47 ppm. Other nearby regions contributed in smaller ways (2.17, 2.43, and 2.49 ppm) because they contain less signal intensity. The variable that makes the largest contribution to differences based on treatment is the β -CH_3 resonance of alanine (Ala) which occupies 1.47 and 1.49 ppm.[14, 15] The next most prominent feature along the second principal component is contributed by the β -CH_3 of lactate (1.33 ppm).

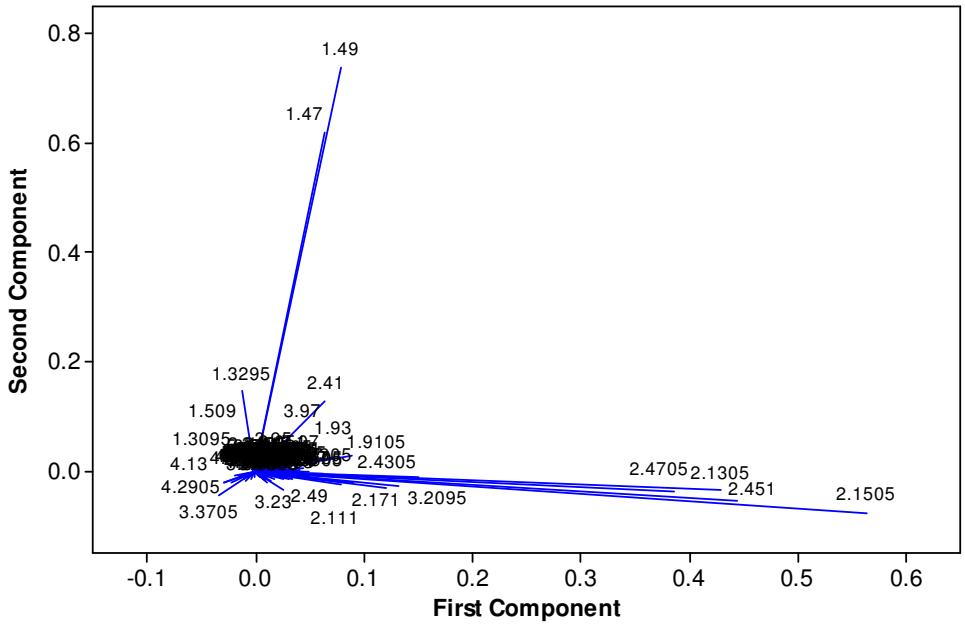


Figure 4.2 Loadings plot indicates variables that contribute to the segregation observed in the scores plot in the comparison of the root and shoot metabolome.

Multivariate data modeling, using PCA or other unsupervised techniques, must be done with care.[16, 17] A model should be verified before biological interpretations are inferred.[18] An explained variance plot (Figure 4.3) reveals that the first principal component accounts for 73% of the variance of this dataset. Taken together, the first and second principal components account for greater than 84% of the variance. The model was cross-validated as in Chapter 3 by removing highly loaded variables until observed clusters give way to biological noise. No fewer than 20 variables were needed to distinguish between clusters.

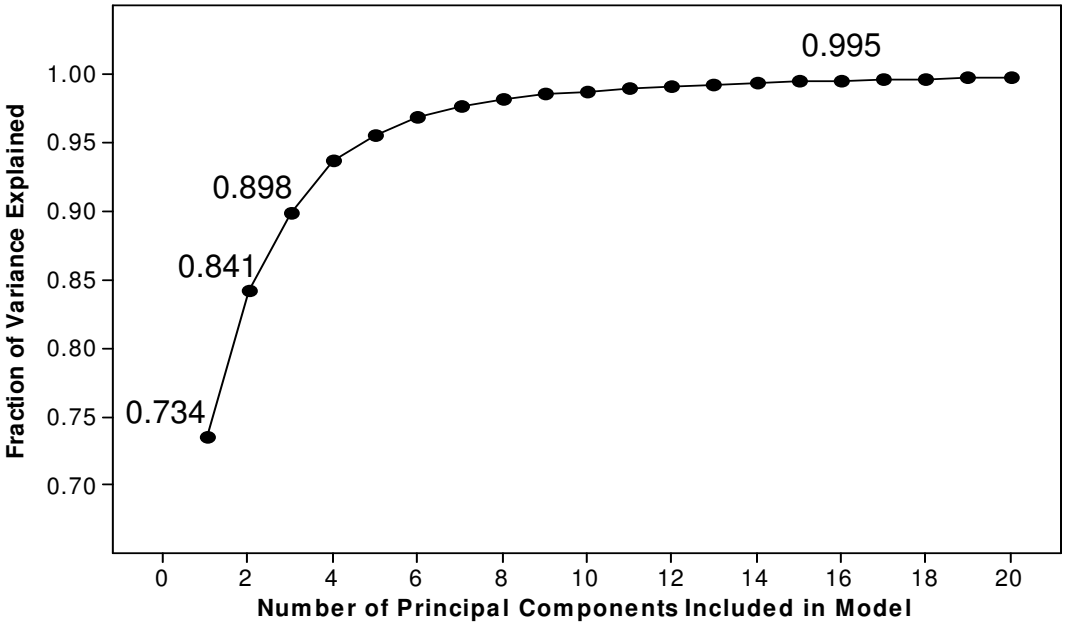


Figure 4.3 Explained variance plot demonstrating that a single component accounts for 73.4% of the variance in the dataset, while the first two components together explain 84.1% of the variance.

After subjecting the dataset to multivariable modeling, clustering of samples according to biological treatment and organ was observed due to relatively low variability among bioreplicates. Further investigations were carried out to address issues related to our original research question. It was clear that the metabolites present in the root are different from those in the shoot, even under control conditions, so the dataset was interrogated to determine whether the root and shoot respond to hypoxic stress differently. To achieve this aim, hypoxia-stressed shoot extracts were directly compared with control root extracts (Figures 4.4, 4.5 and 4.6) and analogously for shoot tissue extracts (Figures 4.7, 4.8 and 4.9). Spectra are presented for comparison in Figure 4.11.

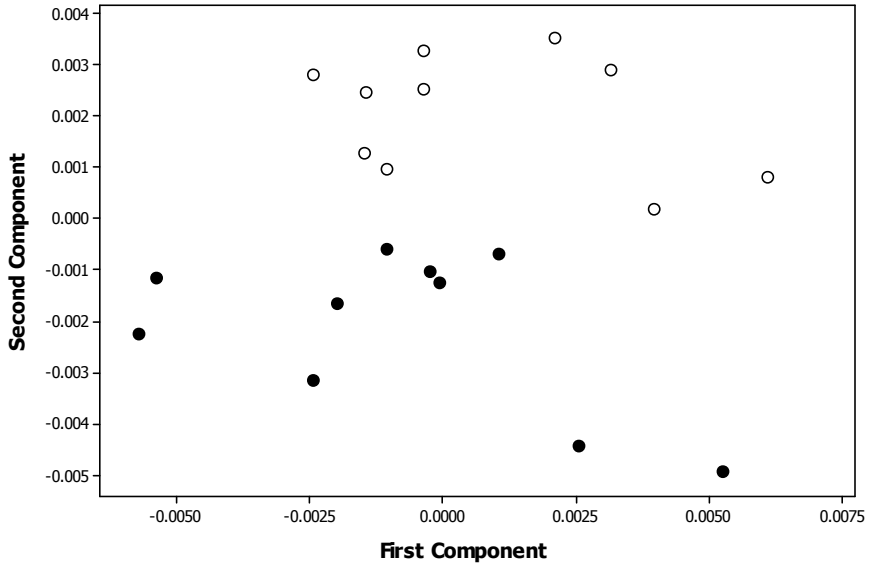


Figure 4.4 Scores plot resulting from PCA using a covariance matrix for roots under hypoxia stress and control conditions. Inputs to the model were NMR spectra binned to 0.02 ppm bin widths, excluding regions occupied by solvents and constituents of growth medium. Open symbols are control conditions and closed symbols show the effect of 2 h of Ar treatment on roots (circles).

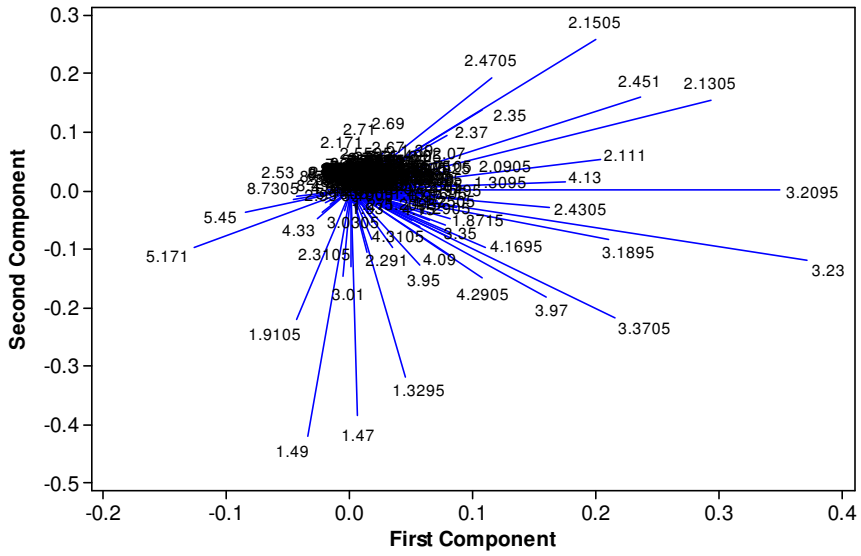


Figure 4.5 Loadings plot indicates the variables that contribute most to the segregation observed in the scores plot for roots under hypoxia stress and control conditions. PC1 represents mostly the biological variance whereas the treatments are distinguished along PC2.

Comparing control and hypoxia-stressed roots using unsupervised PCA (Figure 4.4), a clear distinction according to treatment emerges along the second component axis. The first component seems to describe biological variance unrelated to treatment. Examining the corresponding loadings (Figure 4.5) NMR spectral regions occupied by GABA (β -CH_2- 1.91 ppm, α -CH_2- 2.3 ppm, γ -CH_2- 3.03 ppm) become significant contributors along the second component axis in a negative direction.

Compared to the results presented in Figure 4.1, focusing on a single tissue type gives rise to a dataset that is more complex to model using principal components. Among root extracts, three components are necessary to explain greater than 77% of the variance in the dataset, (Figure 4.6). A value of 65% explained variance is a generally accepted cutoff that can alternately be determined by including only variables of

statistical significance.[19] Softer rules, such as "the beginning of the plateau," "first local maximum," or "bend in the knee" are often applied.[20]

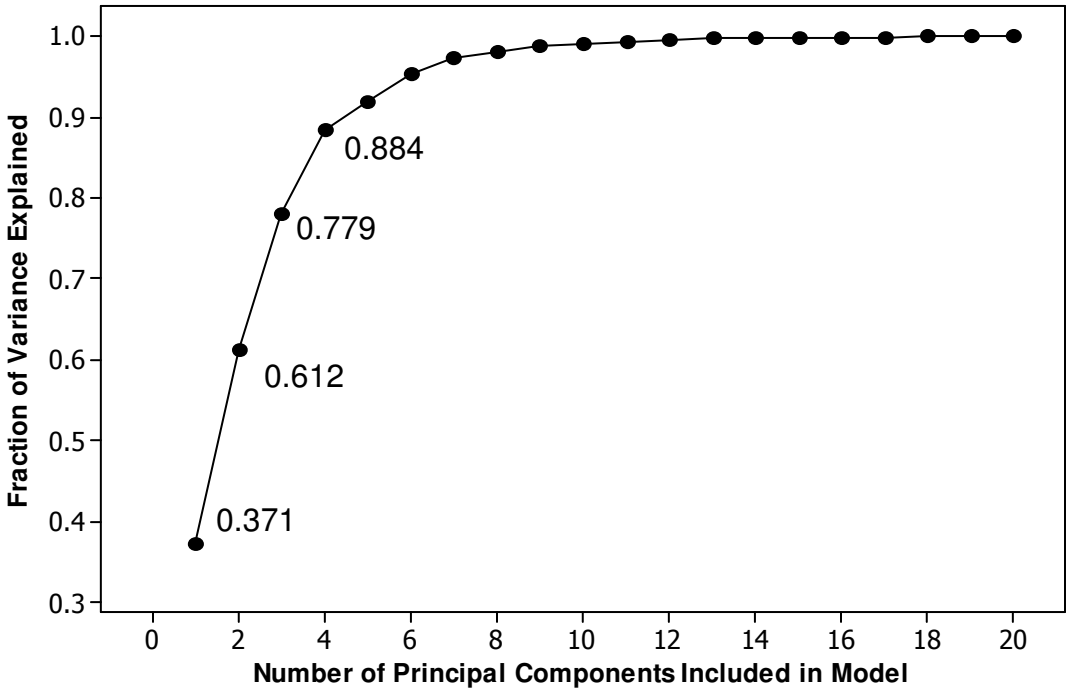


Figure 4.6 Explained variance plot demonstrating that a single component accounts for 37.1% of the variance in the roots only dataset, while the first two components together explain 61.2% of the variance.

Examining a different subset of the original data, shoots can be examined analogously. Figure 4.7 represents each NMR spectrum of shoot tissue extracts as a single point in multidimensional space, here reduced to two dimensions resulting from a linear combination of the original variable space. Control shoots (open triangles) are clearly distinguishable from hypoxic stressed shoots (closed triangles). Instead of clear segregation along one principal component axis, it seems both of the first two

components are necessary to describe the differences between hypoxia stressed shoot extracts and respective control tissue extracts.

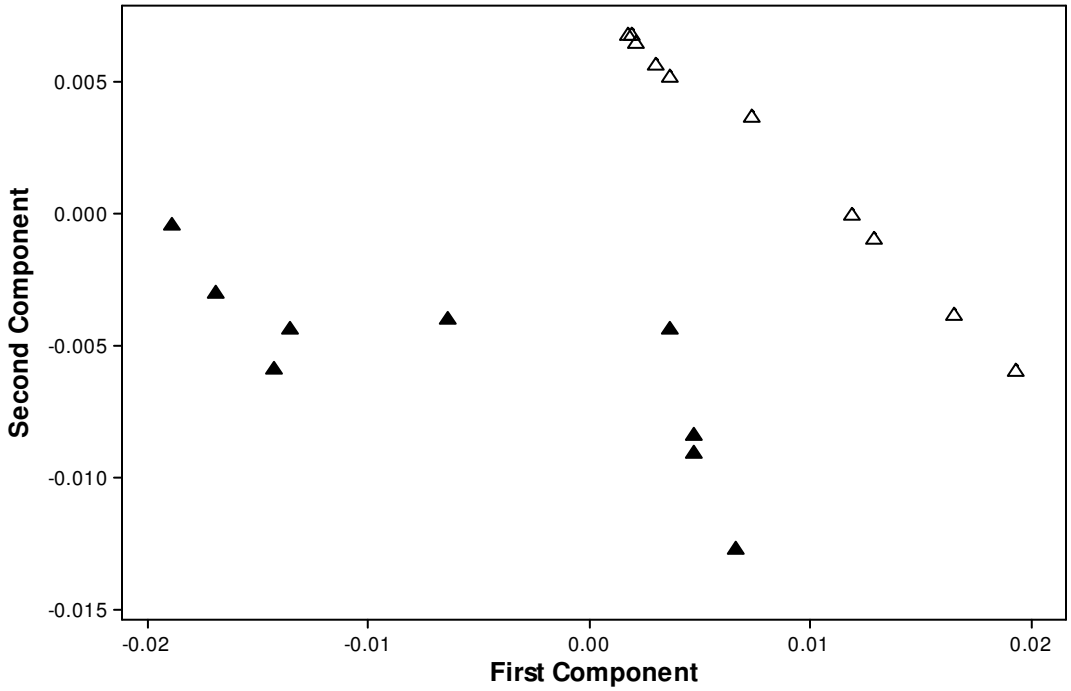


Figure 4.7 Scores plot resulting from PCA using a covariance matrix for shoots under stress and control conditions. Inputs to the model were NMR spectra binned to 0.02 ppm bin widths, excluding regions occupied by solvents and constituents of growth medium. Open symbols are control conditions and closed symbols show the effect of 2 h of Ar treatment (hypoxia stress) on shoots (triangles).

The corresponding loadings plot (Figure 4.8) for shoots is considerably simpler than that obtained in Figure 4.5 for the roots. The Gln β -CH_2- (2.13, 2.15 ppm) and γ -CH_2- resonances (2.45 and 2.47 ppm) extend from the origin towards the lower right quadrant. Other nearby regions contribute in smaller ways (2.17, 2.43 ppm). The β -CH_3 resonance of Ala occupies 1.47, 1.49 ppm, extending from the origin toward the lower left quadrant. Other prominent features can be attributed to the β -CH_3 of lactate (1.33

ppm), covarying with Ala, as evidenced by the direction of its projection in principal component space. Extending in an upwards direction is the spectral region centered around 2.41 ppm. This may be contributed by a molecule that results in a singlet, unidentifiable due to overlap with Gln, since it appears to vary in a unique way.

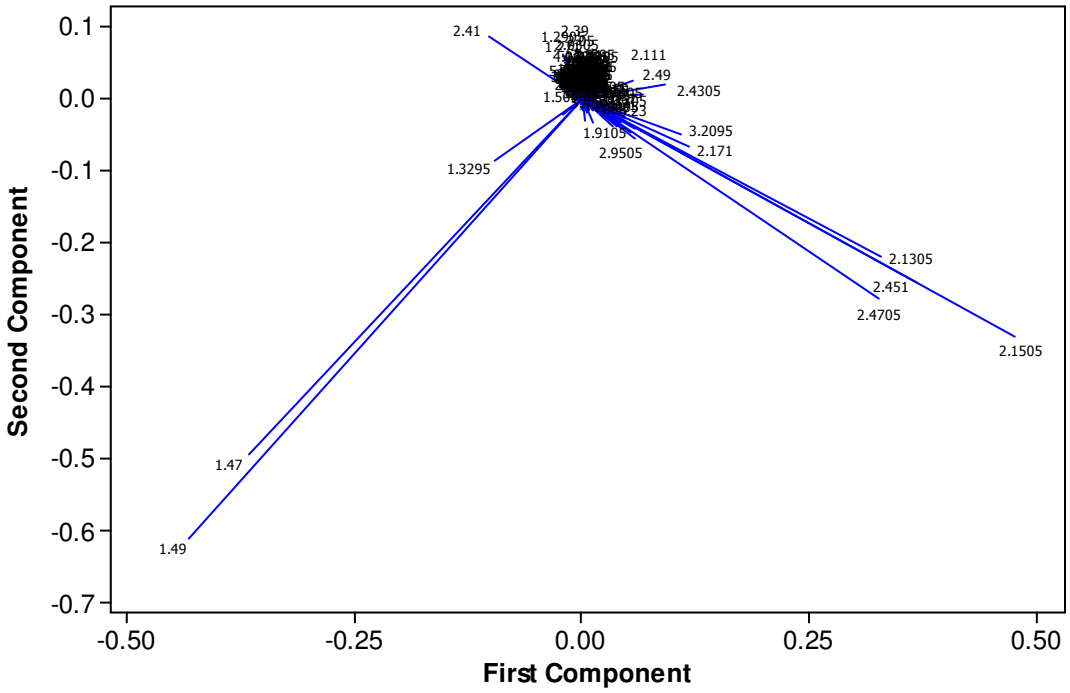


Figure 4.8 Loadings plot indicates variables that contribute to the segregation observed in the scores plot for shoots under hypoxia stress and control conditions. PC1 and 2 together represent the difference between treatments.

The number of components needed to describe the variance in the shoots only dataset is determined in Figure 4.9. Two components are more than satisfactory to adequately represent shoot responses to oxygen deprivation. Taken together, these

results validate our experimental model and suggest we can further explore the data using univariate statistics, where possible, to identify additional contributions to the metabolic response of root and shoot tissue to hypoxic stress.

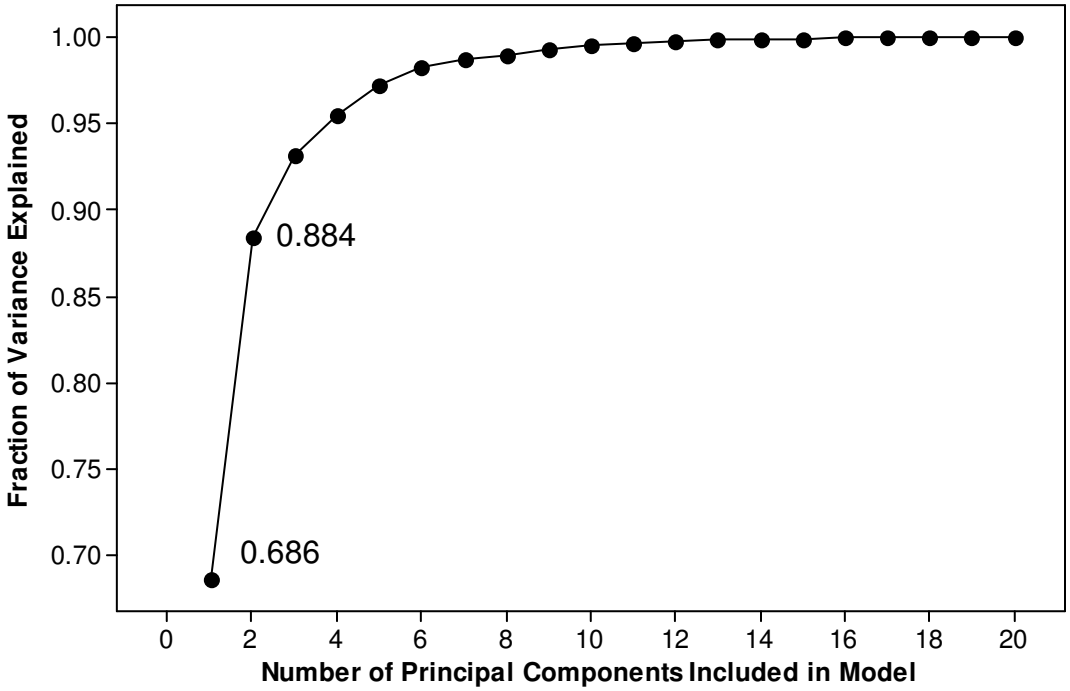


Figure 4.9 Explained variance plot demonstrating that a single component accounts for 68.6% of the variance in the shoots dataset, while the first two components together explain 88.4% of the variance.

4.3.2 Univariate Analysis of NMR-based metabolite profiling data

Non-targeted exploration of the one-dimensional ^{1}H NMR metabolic fingerprints revealed that the distinguishing features in loadings plots for shoots (Figure 4.8) and roots (Figure 4.5) were distinct when seedlings were subjected to 2 h of oxygen deficiency. The identification of integral regions responsible for segregation of samples according to

treatment for each organ encouraged the application of univariate statistics for pairwise comparison of relative metabolite levels between stressed and control conditions. In this way, quantitative fold changes are calculated by integration of specific regions of the NMR spectra corresponding to unobstructed resonances of specific metabolites identified through the acquisition of spectra of pure standards in the neutral buffer used in these experiments. These analyses are summarized in Table 4.1.

Table 4.1 demonstrates that general agreement was observed between the previous (Chapter 3) and this (Chapter 4) study of 2 h oxygen deficiency stress on whole seedlings of *Arabidopsis thaliana*. Clear distinctions can be seen between root and shoot mechanisms for maintaining homeostasis under short-term (2 h) low oxygen stress. What was observed for whole tissue is in many cases representative of the mass-weighted average of root and shoot individual metabolite analyses, Ala is one example, where a fold change between stress and control was 3.3 for whole seedlings and 2.4 or 4.6 for roots or shoots, respectively. Pooling roots and shoots together masked the effect of more dramatic accumulation of Ala in shoots. Conversely, no significant fold change in glutamate (Glu) was observed in whole seedlings, whereas a statistically significant fold change was observed in roots but not in shoots. Only at more severe stress time points is the depletion of Glu, a statistically-significant fold change of 0.7, observed after 9 h oxygen deprivation stress in whole seedlings. Since root tissue is less massive than shoot tissue at this growth stage, the significant change in Glu contributed by the root tissue was masked when whole seedlings were sampled.

Table 4.1 Metabolites quantified by ^{1}H NMR measurements of direct tissue extracts at approximately pH 7. Chemical shifts regions integrated are listed. Numbers reported for each treatment represent the normalized signal intensity of a metabolite under stress divided by the normalized signal intensity of the same metabolite in the corresponding control condition, yielding a fold change for each metabolite. Fold changes report the effect of 2 h stress (2HS/2NS) in roots, shoots, and whole seedlings, as well as 9 h stress (9HS/9NS). Significance is denoted by * at the 95% confidence interval and by ** at the 99% confidence interval. For lactate, the fold change could not be quantified because of low levels in control tissue and spectral overlap, therefore + and +++ denote moderate and strong increases, respectively. The metabolite Asp was not detected (N. D.) in whole seedlings due to resonance overlap at roughly pH 5, as in Chapter 3.

	pH 7 Chapter 4	This work, Chapter 4 (pH 7)		This work, Chapter 3 (pH 5)	
Metabolite	Chemical shift, δ (ppm)	Roots 2HS/2NS	Shoots 2HS/2NS	Whole seedling 2HS/2NS	Whole seedling 9HS/9NS
Ala	1.42	2.4**	4.6**	3.3*	6.5**
Asn	2.90	0.9	0.9	1.0	0.7*
Asp	2.80	0.6**	0.8**	N. D.	N. D.
Choline	3.20	0.9**	0.9	0.8	0.8*
GABA	2.30	1.9**	1.0	1.0	1.3**
Glu	2.35	0.6**	1.0	1.0	0.7*
Gln	2.46	0.9**	0.9**	0.9	0.7**
Lactate	1.34	1.3**	1.6**	+	+ + +
Val	1.04	1.0	1.1	1.1	1.3

The depletion of Gln (insignificant fold change in whole seedlings, statistically significant but small difference, fold change 0.9, in roots and shoots) is masked when there are not enough bioreplicates to achieve statistical power. Only at more severe stress time points is the statistically significant depletion of Gln, fold change 0.7, observed in whole seedlings. This work, with N = 9-10 instead of N = 3-4 (Chapter 3) represents an improvement with respect to sampling.

Rather than reducing the data to a fold change with t-test for significance, the data can be inspected visually by plotting average relative abundances for selected metabolites

Figure 4.10. Standard deviation is displayed as y-error bars for each average reported (N=10) among bioreplicates.

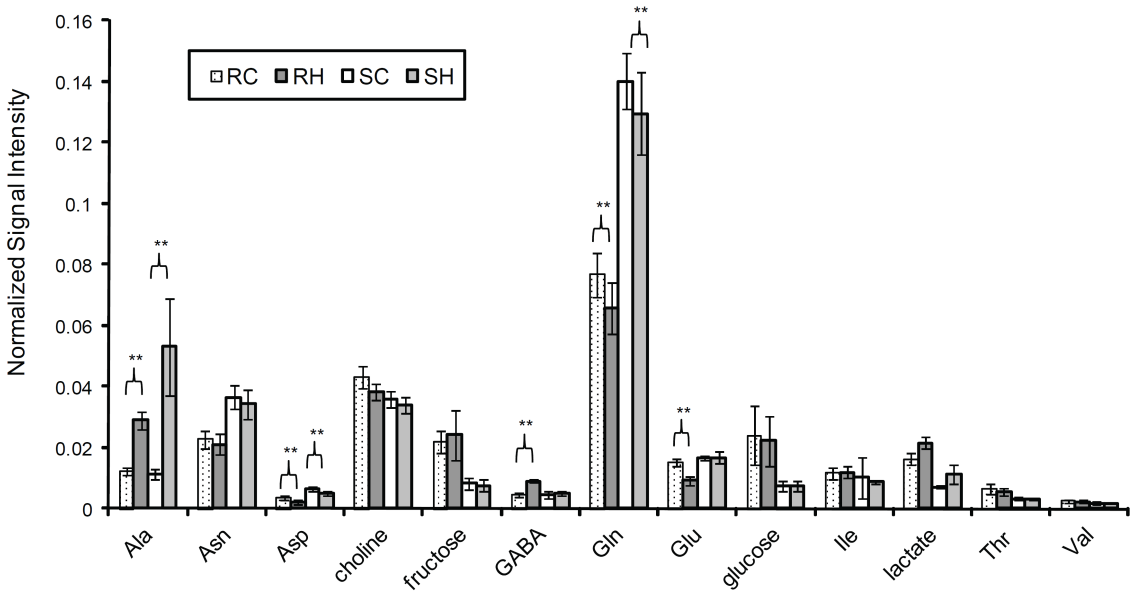


Figure 4.10 Average relative metabolite abundance in extracts representing Arabidopsis seedlings (N=10) for controls (C) and hypoxia treatments (H). Root control and hypoxia stress (RC, RH) are compared with shoot control and hypoxia stress (SC, SH). Standard deviations among bioreplicates are shown as y-error bars. Significant differences between hypoxia and control are clear in Ala, while significant differences between organs are clear in Asn.

Other improvements of this work (Chapter 4) over our previous study of oxygen deprived Arabidopsis seedlings, is the ability to quantify fold changes of asparagine (Asn). The titration of all extracts to neutral pH facilitated peak integration of the Asn β - CH_2 resonance without spectral overlap. The ability to quantify lactate in this study was a result of utilizing the lipid elimination strategies developed in Chapter 2. Quantitative metabolite data can be confirmed by visual inspection of the NMR spectra. Figure 4.11 presents representative spectra for each two hour treatment control (A,C) and low oxygen stress (B,D) when seedlings have been dissected and pooled into a homogenized sample representing shoots (A,B) and roots (C,D).

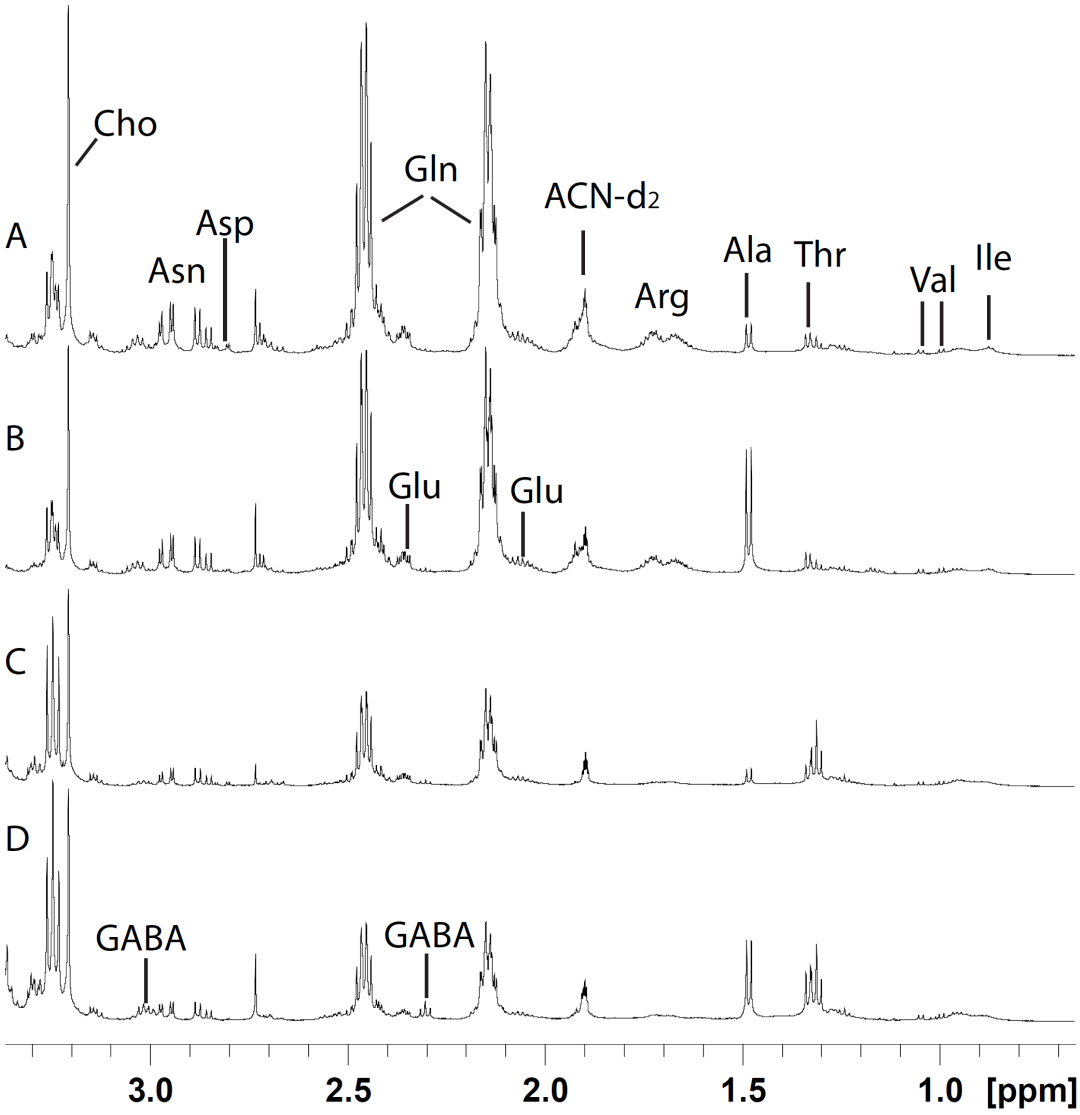


Figure 4.11 NMR spectra of samples representing each treatment. Expanded to facilitate discussion of nitrogen metabolism and amino acid metabolic processes. Shoots under control (A) and low oxygen stress (B) are distinct from roots under control (C) and hypoxic conditions (D).

4.4.3 Insights from GeneChip Measurements

The advantage to applying orthogonal approaches in systems biology is that it can allow the investigator to make inferences about biochemical and genetic networks.

Performing relatively quantitative measurements of actively-translated mRNAs, expressed as fold changes, can inform cellular priorities under energy shortage. Using clustering to mine transcription and translation profiling data proved an effective way to identify processes relevant to the problem under study. Gene ontology (GO) classifications and putative gene assignments in online databases[21, 22] provided clues as to the identity and molecular function of transcripts that were enriched or depleted in polysome complexes.

4.4.4 Pathway Modeling

In a world where all higher order organisms have evolved to utilize molecular oxygen, there exist alternative routes to sustain life during periods of low oxygen. The primary requirements for sustaining metabolism during anaerobic stress are: stored substrate, redox balance and ATP yield. Sugar and starch can be mobilized from storage locations in leaves or roots and enter glycolysis. Redox balance can be maintained by interdependent cycles, which can be facilitated by membrane transport and/or shuttling of carbon skeletons into necessary substrates. Specialized metabolic pathways such as fermentation are likely to be organ-specific due to the accessibility of each organ to macro- and micronutrients and the likelihood of encountering different types of biotic or abiotic stressors.

A pathway diagram is presented in Figure 4.12 highlighting metabolic connections between small molecule metabolites (12 pt font) and enzymes responsible

for their interconversions (18 pt font). Coupling intermediates NAD^+ and ATP are shown in bold.

Oxygen-limited conditions necessitate a switch from aerobic respiration to anaerobic fermentation, with electron acceptors such as pyruvate and acetaldehyde replacing oxygen. These metabolic adaptations help maintain ATP production and NAD^+ regeneration. Anaerobic fermentation is far less efficient in producing energy than aerobic respiration. Carbohydrate availability, either from photosynthesis or from stored reserves, becomes increasingly critical the longer the plant remains submerged. [23] Important contributions to the understanding of these processes are afforded here as they relate to interpretation of the molecular data we obtained on short-term oxygen deprived 7 d old Arabidopsis seedlings. Metabolic pathways discussed include central carbon metabolism (sucrose and starch mobilization, glycolysis, TCA cycle), nitrogen metabolism (nitrogen assimilation and denitrification, GABA shunt, polyamine biosynthesis), and amino acid metabolic processes (catabolism and biosynthesis). In each metabolic discussion the changes noted in root and shoot metabolites, transcript abundance, and localization of expression will be treated individually and quantitatively. These will be discussed in the context of the literature to evaluate whether the finding is novel, contradictory or complementary to previous work. Standing on the existing body

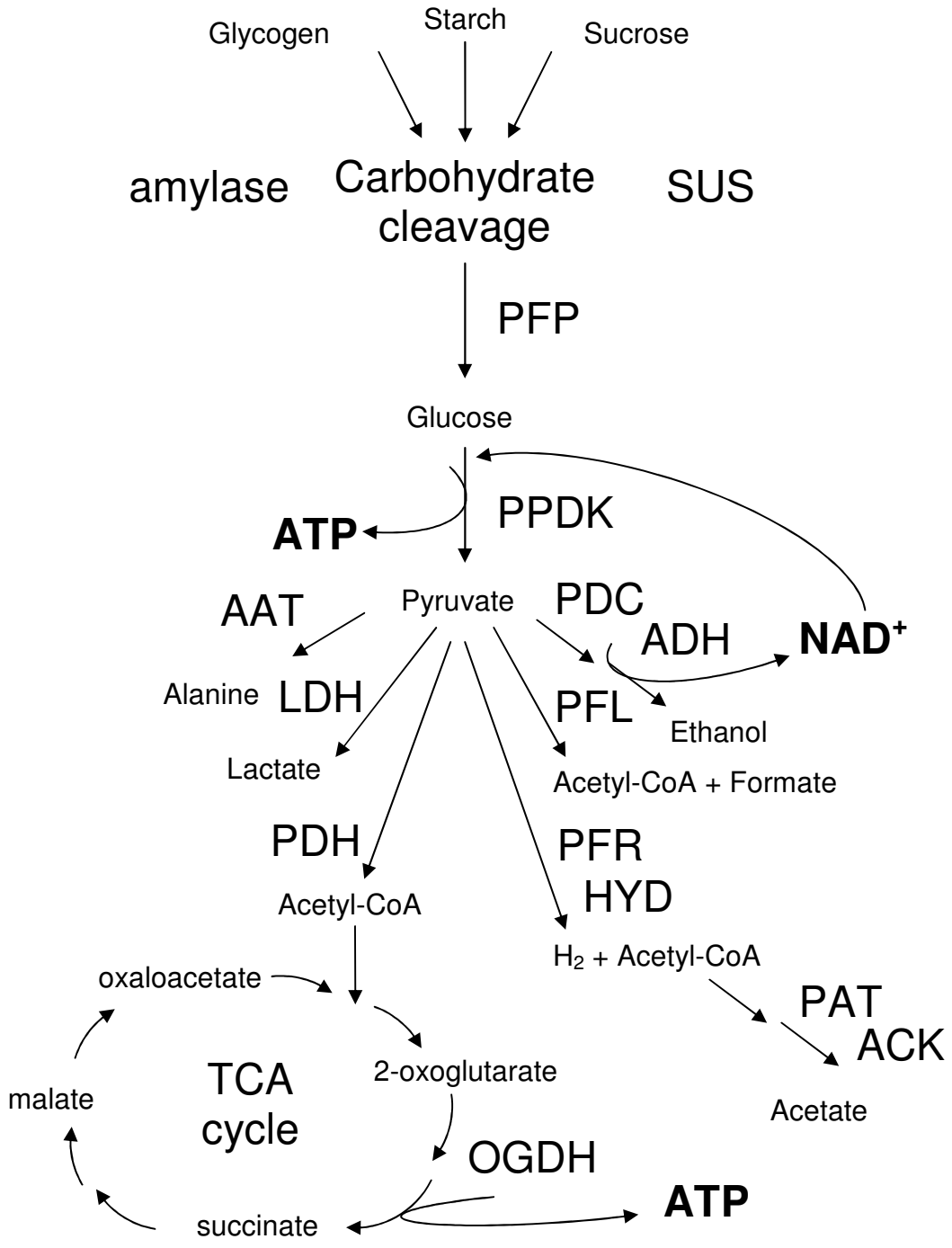


Figure 4.12 Utilization of substrates for ATP generation in aerobic and anaerobic conditions involves carbohydrate flux through glycolysis and generation of substrates for maintaining redox balance for mitochondrial electron transport via the TCA cycle. Fermentative pathways terminating in Ala, lactate, acetate, formate, or ethanol are common responses to oxygen deprivation in eukaryotes. In plants, sucrose consumption is favored; glycogen is not present. The formation of Acetyl CoA + formate and acetate does not occur in plants or animals but is characteristic of the oxygen deprivation response of algae, such as *Chlamydomonas*,

of knowledge of organ-specific adaptations in roots and shoots together with a broad understanding of issues related to metabolite profiling (Chapter 1), methodology (Chapter 2), and practice (Chapter 3), it was proposed that new knowledge in this area could be generated by application of a discovery pipeline to such a problem. Combining commercially-available transcriptomic technology with the rapidly-growing area of metabonomics provided us a nontargeted systems biological approach for investigating organ specialization in terms of cellular contents and regulation under abiotic (low oxygen) stress in seedlings of a model plant.

Keeping in mind the body of knowledge in plant metabolic responses to low oxygen, we applied a supervised approach to combing through the transcripome and translatome. Gene ontology (GO) analysis is a way to cluster transcripts according to the function of the protein encoded; transcripts made from genes in the same family encoding enzymes in the same pathways are grouped together. Some GO classifications contain a handful (2-8 transcripts) and others contain hundreds (200-500 transcripts). Important pathways tend to involve more functional redundancy. Also, one can take a bird's eye view of a pathway or focus on a specific branch. The following section will discuss specific GO classifications that are related to metabolism and for which transcript abundance and polysome loading status were found to change significantly in response to short-term (2 h) oxygen deprivation stress. Moreover, the differences between root and shoot responses will be discussed independently.

4.4.4.1 Central Carbon Metabolism

Sugar and starch measurements revealed that rice (submergence-tolerant) roots and shoots consumed carbohydrate reserves during stress more slowly than wheat (submergence-intolerant).[1] Despite lower fermentation rates in both plant species in the light, carbohydrate contents decreased to almost the same extent as in complete darkness.[2] Sugars decreased during waterlogging in roots, and levels returned to their initial amount during the recovery period.[24] In roots of the submergence-tolerant gray poplar (*Populus x canescens*), starch and sucrose degradation were altered under hypoxia, and concurrently, the availability of carbohydrates was enhanced, concomitant with depletion of sucrose from leaves and elevation of sucrose in the phloem.[12] Induction of sucrose synthase (SUS) versus the invertase route for mobilizing sucrose and starch can lead to improved energy efficiency to avoid energy crisis. Studies of submergence-tolerant (Sub1) rice and a near-isogenic (intolerant) line have revealed a quiescence strategy, mediated by a cascade involving the plant hormones ethylene, abscisic acid (ABA) and gibberellic acid (GA), which demonstrated reduced mobilization of sucrose and starch reserves leads to improved survival of transient low-oxygen stress.[26, 27]

Glycolysis requires two molecules of NAD^+ and produces two molecules of ATP. The rate of glucose uptake and glycolytic flux at 3% oxygen were found to be similar to those under normoxia, but with nearly 90% of glycolytic flux going to fermentation.[28] In the flood-tolerant plant gray poplar, glycolytic flux was stimulated stressing waterlogged roots but not in aerated leaves.[12] Shoot growth was not affected by root

hypoxia, suggesting that the localized up-regulation of glycolysis did not compromise shoot growth.

Transcript data from GeneChips can be painted onto seedling drawings using the electronic fluorescence (eFP) browser. Figure 4.13 provides such data for selected transcripts encoding enzymes involved in glycolytic flux. Key regulatory steps are given special attention. This reveals considerable cell- and region-specific regulation in response to hypoxia. For example, phosphoglycerate kinase (PK) transcript translation appears to become reduced under hypoxia stress but only in shoots, a similar trend being demonstrated by one gene encoding PK (At5g52920).

Glycolysis requires 2 molecules of NAD^+ and produces 2 molecules of ATP. The rate of glucose uptake and glycolytic flux at 3% oxygen were found to be similar to those under normoxia, but with nearly 90% of glycolytic flux going to fermentation.[28]
Transcript data from GeneChips can be painted onto seedling drawings using the electronic fluorescence (eFP) browser. Figure 4.13 provides such data for the transcripts involved in glycolytic flux, key regulatory steps are given special attention.
In the flood-tolerant plant gray poplar, glycolytic flux was stimulated stressing waterlogged roots but not in aerated leaves.[12] Shoot growth was not affected by root hypoxia, suggesting that the localized up-regulation of glycolysis did not compromise shoot growth. In this work, phosphoglycerate kinase appears to become depleted under

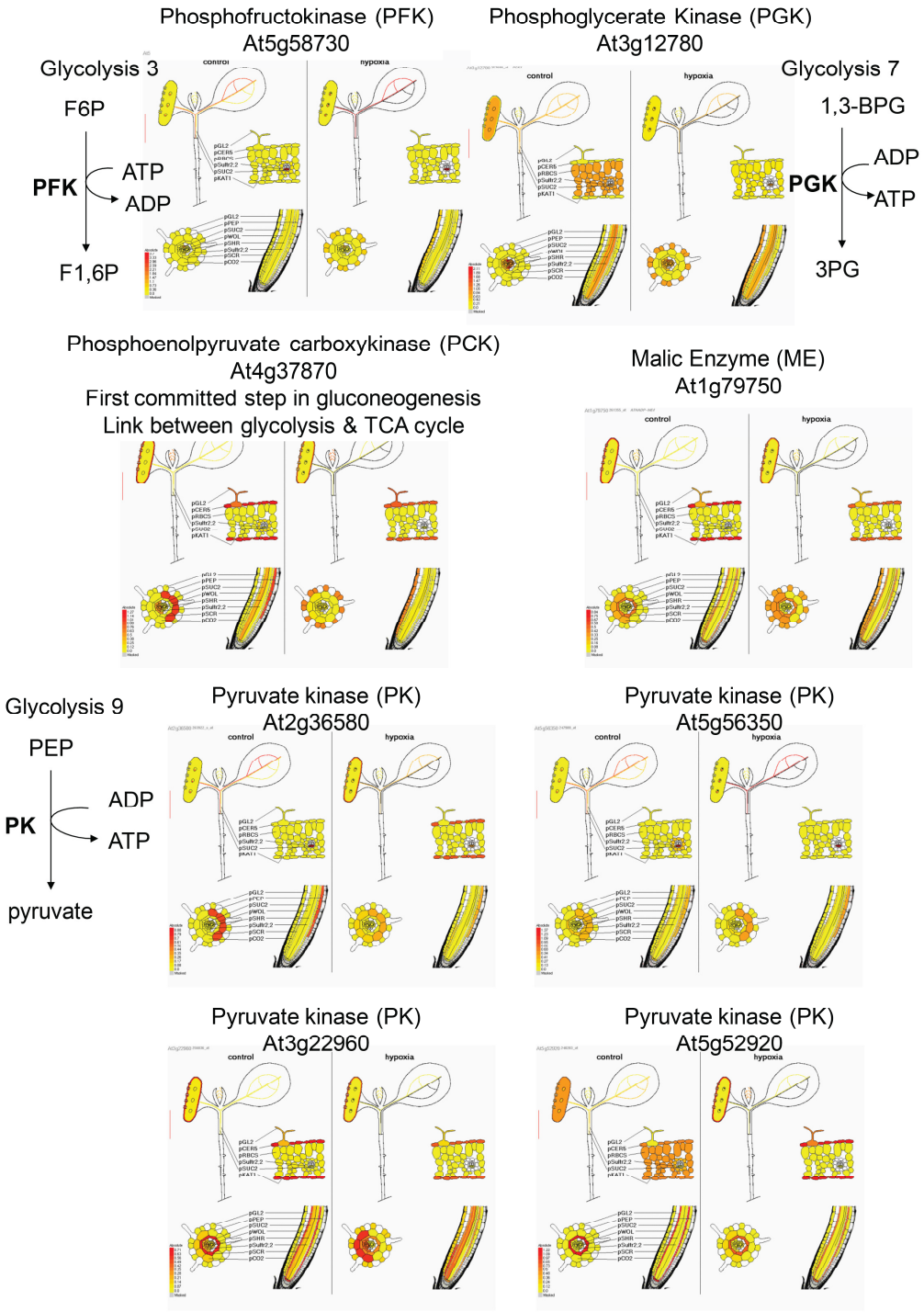


Figure 4.13 Arabidopsis eFP browser mappings of immunopurified mRNA abundance onto seedling cell-types. Reactions catalyzed by the reactions in glycolysis are shown where an adenylate is a necessary cofactor. In these heatmaps, yellow indicates that the transcript is not present (or present at low or nondetectable levels) and red indicates transcript is present. For each transcript, the control condition is shown in the left panel and the 2 h hypoxia-stressed condition is shown at right.

hypoxia stress but only in shoots, a similar trend being demonstrated by one isoform of pyruvate kinase.

Selected transcripts encoding enzymes involved in the classical TCA cycle were also mapped onto the eFP template, shown in Figure 4.14. The depletion of succinate dehydrogenase under low oxygen stress is observed in roots and shoots, however the depletion of malate dehydrogenase (MDH) is more dramatic in the shoot.

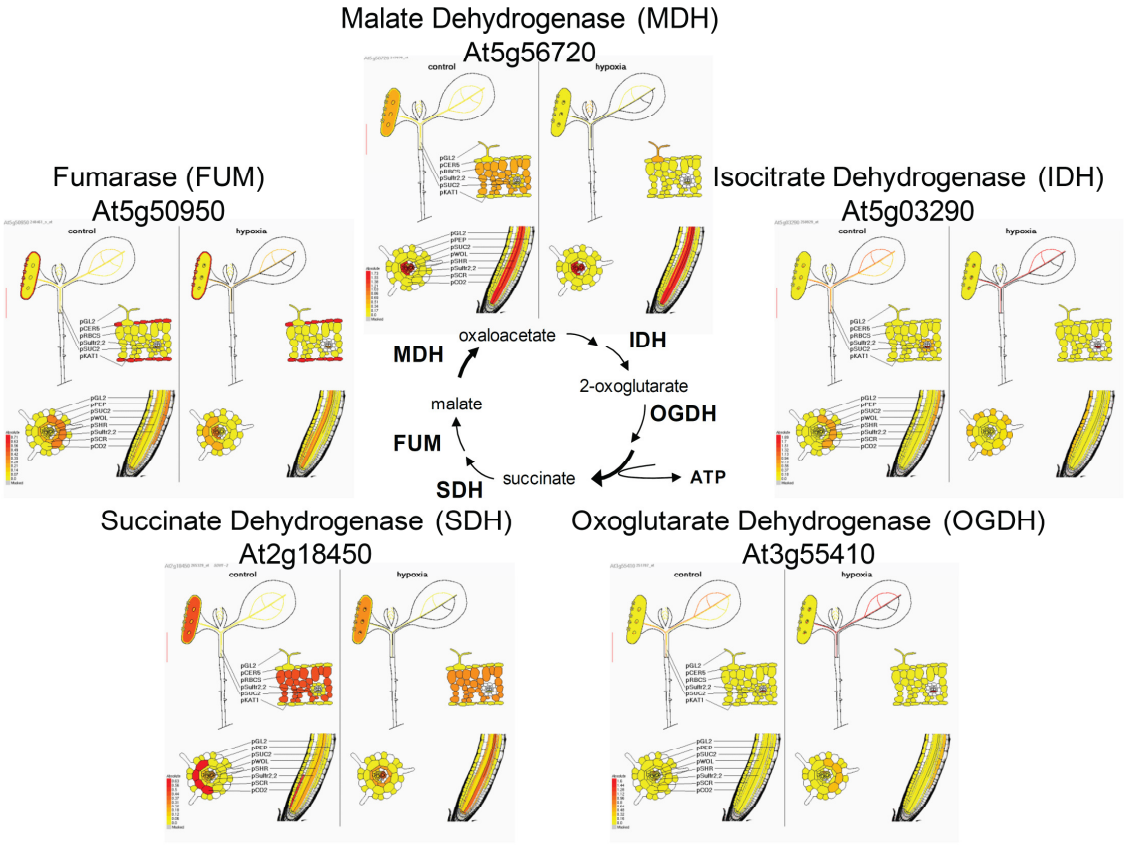


Figure 4.14 Arabidopsis eFP browser mappings of immunopurified mRNA abundance onto seedling cell-types. Reactions catalyzed by the reactions in the TCA cycle are shown, additionally steps where an adenylate is a necessary cofactor are indicated. Yellow indicates that the transcript is not present (or present at low or nondetectable levels) and red indicates transcript is present. For each transcript, the control condition is shown in the left panel and the 2 h hypoxia-stressed condition is shown at right.

One GO upregulated in shoots was GO:0005985, related to sucrose metabolism. This set of genes encodes an isoform of SUS (At3g43190) and starch synthase (At1g32900), indicating that stored carbohydrate reserves are being mobilized for energy production. In roots, a sucrose transport protein (At1g22710) is depleted in polysomes whereas a glucose transporter (At1g11260) is enriched. This may be indicative of a preference for monosaccharides, which can directly enter glycolysis, rather than disaccharides, which may be utilized under prolonged stress.

Carbohydrate metabolic process (GO:0044262) overall includes every pathway covered in this subsection, comprising 379 distinct genes. Disaccharide metabolic process (GO:0005984) is more specific to sucrose and triose mobilization, consisting of 82 loci. These processes were all induced in roots under hypoxia. Fold changes for selected genes under these ontologies are presented in Table 4.2. Metabolite fold changes, where correlation to gene function can be made, are presented in italics.

A general trend in translational status can be observed. mRNAs encoding the ATP-consumptive regulatory steps (e.g., PFK, phosphorylates F6P forming F1,6P) are downregulated at the translational level. PFK is one of several regulatory enzymes, responsible for control of glycolytic flux. By contrast, mRNAs encoding enzymes catalyzing ATP producing reactions (e.g., PK, dephosphorylates PEP forming pyruvate) show enhanced translational status under oxygen deprivation.

Translational status for TCA cycle enzymes is statistically insignificantly different between 2 h stress and control populations. Interestingly, the transcript for malic enzyme (ME), responsible for conversion of malate to pyruvate, generating

NADPH and carbon dioxide is significantly reduced in polysomes. Some have suggested malic enzyme is responsible for regeneration of this key biosynthetic intermediate (NADPH) to allow an alternate pathway for carbon to reenter glycolysis under low oxygen stress, but this study suggests that is not due to upregulation of ME synthesis. Other possible control mechanisms may be pH changes or posttranslational protein modifications acting directly on the enzyme.

Table 4.2 Carbohydrate Metabolic Process (GO:0005975) transcript and metabolite levels as measured in immunopurified mRNA populations isolated from roots (R) and shoots (S) where N=3. Values reported result from fold changes (FC) between hypoxia-stressed tissue and respective control tissue. Metabolite fold changes (*italicized*) are reported for roots and shoots (N = 10). Fold changes (relative abundance in stressed tissue extracts divided by relative abundance of the same variable in control extracts) in response to two hours of low oxygen are given. **designates confidence at the 99% interval or greater, *designates confidence between 95-99% that the relative abundances are different between hypoxia stress treatment and control.

Pathway / Enzyme	**Abbrev.**	**Gene Location** (OrgXsomePos)	**R**	**S**
Glycolysis and TCA Cycle GO:0006096, GO:0006099, GO:0006097				
phosphofructokinase	PFK	At5g58730	0.80	0.76
phosphoglycerate kinase	PGK	At3g12780	0.87	1.3
pyruvate kinase	PK	At2g36580, At5g56350	1.6**	1.4*
pyruvate kinase	PK	At3g22960, At5g52920	0.83	0.93
phosphoenolpyruvate carboxylase	PEPC	At3g14940	0.6	0.8
pyruvate dehydrogenase	PDH	At1g01090	0.93	1.0
pyruvate dehydrogenase	PDH	At2g34590	0.77	0.88
acetyl-CoA carboxylase	ACC	At1g36050	0.55	0.71
isocitrate dehydrogenase	IDH	At5g03290	0.98	0.97
oxoglutarate dehydrogenase	OGDH	At3g55410	0.86	0.92
succinate dehydrogenase	SDH	At2g18450	1.1	1.1
fumarase	FUM	At5g50950	0.75	0.90
malate dehydrogenase	MDH	At5g56720	1.5	1.2
malic enzyme	ME	At1g79750	0.79	0.85
Response to hypoxia GO:0001666				
alcohol dehydrogenase	ADH	At1g77120	120**	14**
pyruvate decarboxylase	PDC	At4g33070, At5g54960, At5g01320	280**	41**
lactate dehydrogenase	LDH	At4g17260	6.8**	2.4**
lactate			*1.3***	*1.6***

One process strongly induced involves ethanolic fermentation via ADH (At1g77120) and PDC (At4g33070), strongly enriched in polysome complexes tens to hundreds of times greater in abundance in 2 h stressed tissues. The enhancement of lactate dehydrogenase (LDH) mRNA in polysomes was not as dramatic, however it is statistically significant and is observed with a corresponding accumulation of the enzymatic reaction's product – lactate. The transcripts encoding fermentative enzymes are more strongly loaded in polysomes in roots than shoots. Possibly this is because under submergence, roots can extrude lactate and other organic and amino acids into waterlogged soil, whereas shoots can more easily release volatile byproducts such as ethanol and acetaldehyde.

Plants possess several unusual features of central metabolism, specifically in sucrose degradation, that may decrease oxygen consumption. The breakdown of one molecule of sucrose by invertase and hexokinase requires two molecules of ATP, whereas its breakdown by SUS and UDP-glucose pyrophosphorylase requires only one molecule of inorganic pyrophosphate (PPi).[40, 41] Unfortunately, due to contamination by sucrose-containing growth medium, our attempts to quantify endogenous sucrose were unsuccessful (Figure 4.15).

Figure 4.15 shows the region of the NMR spectrum occupied by central carbon metabolites. Shoots under control (A) and low oxygen stress (B) are distinct from roots under control (D) and hypoxic conditions (E). An NMR spectrum of a sample of growth medium is presented for comparison in (C). Sugar anomeric protons, clearly available for integration, were ignored due to media contamination of particularly root tissue.

Our examination also considered fatty acid metabolism. The regulation of synthesis of Acetyl-CoA carboxylase (ACC), a key enzyme in fatty acid biosynthesis, is altered under low oxygen. *ACC* transcripts decrease in abundance among polysome complexes. The mitochondrion maintains redox balances via 3-hydroxybutyrate dehydrogenase, fatty acid chain elongation, and most importantly reversal of the TCA cycle with succinate serving as a carbon sink. There are two routes to succinate: Asp, malate and fumarate as substrates will flow via fumarate reductase catalyzed reactions while Glu or α-ketoglutarate/2-oxogluatrate will flow via succinate CoA ligase (SCS) catalysis. In addition to contributing to redox balance, succinate accumulation proceeds with the production of molar equivalents of ATP, influencing also the energy yield of anaerobic metabolism.[38] Our attempts to quantify TCA cycle intermediates by NMR were unsuccessful due to severe spectral overlap in regions where small molecule resonances would have been observed (Figure 4.16). Adjusting sample pH did not alleviate this overlap.

There is interdependence between ethanol and lactate fermentation pathways. In the absence of ethanol fermentation, lactate fermentation is favored. Both pathways result in increases in lactate production or PDC enzyme activity in tomato. Cellular lactate can be converted back to pyruvate producing NADH upon reoxygenation.

The primary requirements for anaerobic metabolism are: stored substrate, redox balance and ATP yield. Stored energy in the form of starch or sucrose has been more

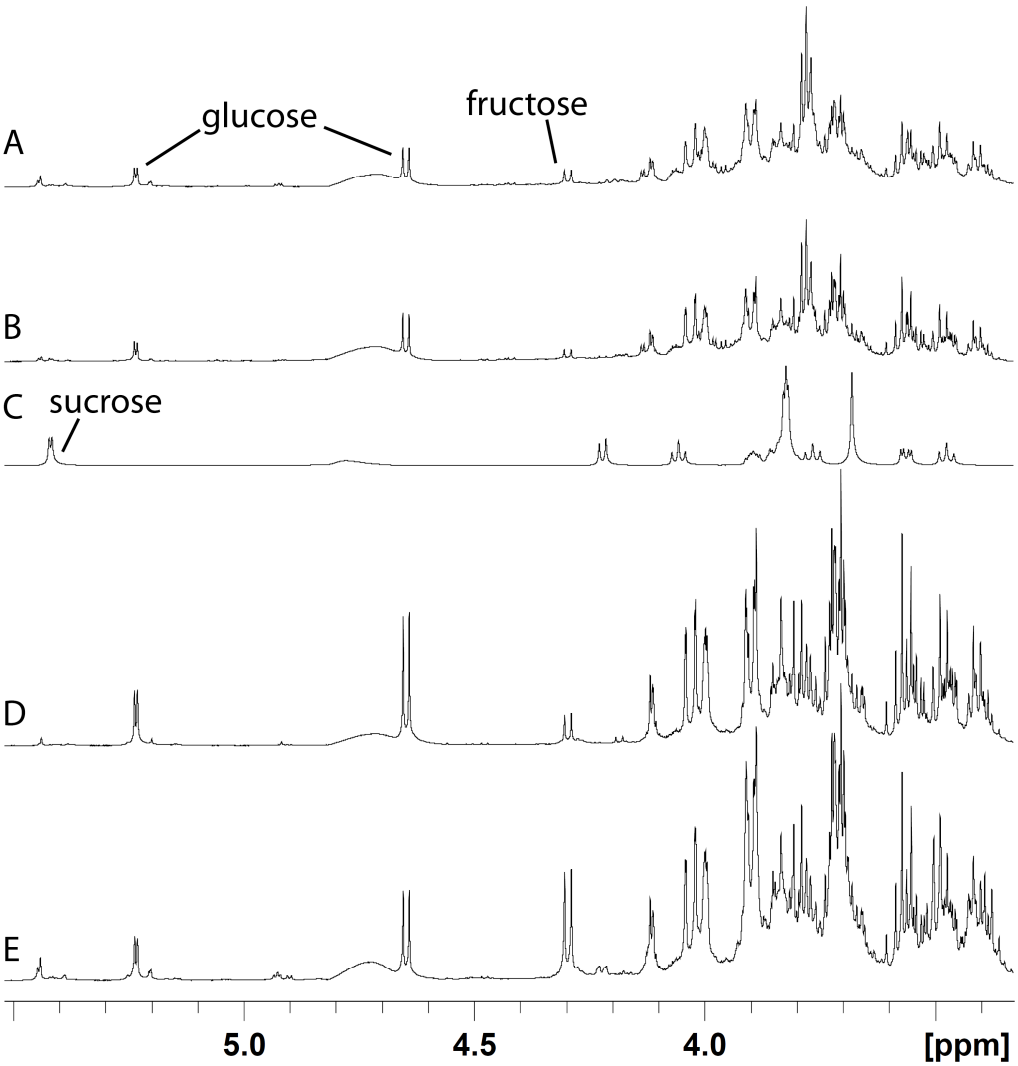


Figure 4.15 NMR spectra of extracts (intensity x 1) representing seedlings (A, B, D, E) and growth medium (C). This expansion provides a basis for discussion of carbohydrate metabolic processes. Shoots under control (A) and low oxygen stress (B) are distinct from roots under control (D) and hypoxic conditions (E). The predominant sugar in tissue extracts is glucose, followed by fructose whereas the medium contains a large excess of sucrose and trace amounts of other sugars.

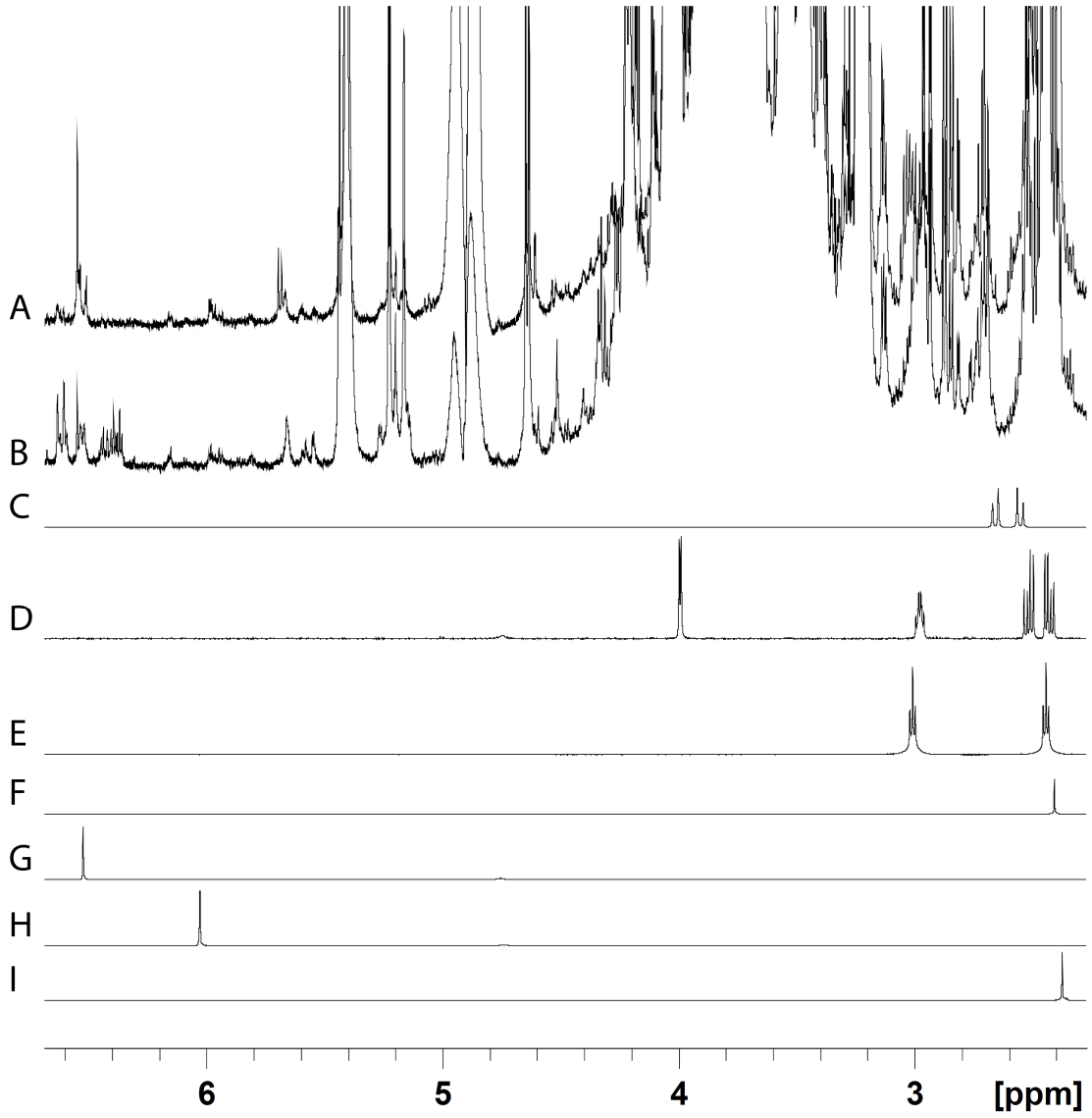


Figure 4.16 NMR spectra of samples representing each biological treatment (intensity x 5) expanded to facilitate discussion of tricarboxylic acid cycle (TCA) intermediates. Shoot (A) and root (B) control tissue extracts provide scant information on TCA cycle due to severe overlap in the region occupied by methylene protons. Citrate (C), isocitrate (D), 2-oxoglutarate (E), succinate (F), and pyruvate (I) occupy this crowded region. Resonances of fumarate (G) and malate (H) are also not observable in Arabidopsis tissue extracts.

compete for the substrate pyruvate. Lactate fermentation may be required to start ethanol fermentation. As lactate accumulation increases, ethanol fermentation may be favored. This may be pH-driven or due to substrate interactions. Overexpression of LDH did not

carefully examined in Section 4.2.1. Redox balance and ATP yield are inherently coupled and will be further discussed here without treating each separately.

In order to sustain anoxia, each oxidative step in fermentative metabolism must be balanced by a reductive step. There are two important aspects to redox regulation that must be considered at the outset: the oxidation state of the cytosol and mitochondria must be maintained; and the oxidation state of coenzymes (NAD^+/NADH, FAD^+/FADH) must be maintained.[38] The standard mechanism of redox regulation in the cytosol is the coupled reaction shown in Equation 4.1

$$\text{pyruvate} + \text{NADH} + H^+ \leftrightarrow \text{lactate} + NAD^+ \qquad (4.1)$$

Auxiliary mechanisms in the cytosol may involve MDH and glutamate dehydrogenase. Under low oxygen conditions in plants, ADH is important for NAD+ production. Overall these reactions oxidize glycolytically formed NADH in the cytosol.[38]

Another critical mechanism for balancing redox substrates and moving them where necessary for the action of the electron transport system is the malate-Asp shuttle. This mechanism involves cytosolic and mitochondrial forms of the enzymes aspartate aminotransferase and MDH and the substrates/products malate, 2-oxoglutarate, Glu and Asp.

4.4.4.2 Nitrogen Metabolism

In higher plants, only root mitochondria but not leaf mitochondria reduce nitrate to nitric oxide at the expense of NADH.[44] Nitrate and nitrite may serve as alternative electron acceptors and may partly replace fermentation for NADH reoxidation during

glycolysis, even though electron flow through NR was comparatively low. A "hidden" nitrite/NO cycle, similar to hemoglobin, could oxidize NO to nitrate, which would maintain a much higher electron flow for the maintenance of minimum energy metabolism.[44] Nitric oxide has been suggested to signal oxygen concentrations, regulating the citric acid cycle (by inhibiting aconitase) and respiration (by inhibiting cytochrome *c* oxidase and activating the mitochondrial alternative oxidase (AOX)). Figure 4.17 presents a metabolic pathway map with enzymes designated for processes that will be discussed in this section. Pathways where metabolite information was relevant are in black. Accumulating substrates are in large font compared, whereas those which were depleted by 2 h of hypoxia are shown in a smaller font. The contribution of GABA and nitrate reductase (NR) was noted only in roots but not shoots. GABA transaminase (GABA-T) appears to be responsible for only a portion of the Ala

A key intermediate between carbon and nitrogen metabolism is Ala, a zwitterion between pH 3.3 and 8.8 making it a neutral substrate that can be remobilized rather than lost via evaporation or efflux. Metabolites such as lactate, ethanol, acetaldehyde, succiate and malate are transported across plant cell membranes by passive or active mechanisms, and may be lost to the plant entirely via excretion. The operation of the GABA shunt contributes to Ala accumulation in the roots of *Arabidopsis* but not shoots. GABA-T appears to be responsible for only a portion of Ala synthesized in roots during hypoxic conditions, and other mechanism(s) are probably responsible for rapid Ala accumulation in roots during the initial phase of hypoxia and for Ala accumulation in shoots. There is evidence of translational regulation or a difference in protein stability of GABA-T

between roots and shoots; GABA-T transcript levels were higher in shoots than roots although enzymatic activity is similar in both roots and shoots.

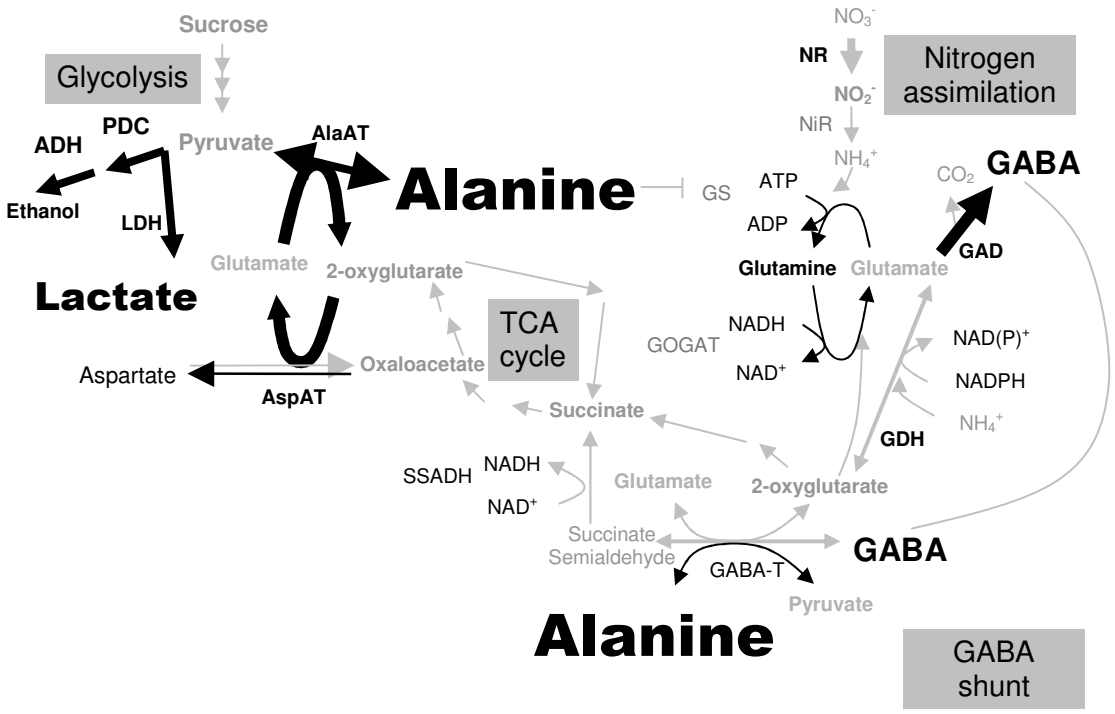


Figure 4.17 Metabolic pathways involved in nitrogen metabolism. Pathways of nitrogen assimilation and the GABA shunt are highlighted as they relate to the TCA cycle and glycolysis. Our metabolite observations together with transcription profiling indicated that the fermentation pathways, transamination and dehydrogenases. Succinate semialdehyde (SSADH), 2-oxyglutarate (GDH), and glutamate (GAD) dehydrogenases are all involved in the generation of the metabolic coupling intermediate NADH and the energy storage and transport molecule ATP.

A defect in the GABA shunt pathway (loss of either GAD or GABA-T) affected Ala accumulation only in roots but not shoots of Arabidopsis. It appears hypoxia-induced Ala accumulation can be divided into two phases: rapid build-up (first 4 h) and

subsequent slow increase (beyond 4 h). This work suggests that the GABA shunt is responsible for the slow phase of Ala accumulation in roots.

A hypothesis for the physiological role of the GABA shunt is based on an experiment involving *gaba-t1* mutants grown on GABA-supplemented media, where it was demonstrated that wild type plants can utilize GABA as an alternative N source but not without GABA-T enzyme. Chemical inhibition of GABA-T in an *ssadh* mutant of Arabidopsis demonstrated that reactive oxygen intermediates and γ-hydroxybutyrate (GHB) buildup are coupled with this pathway.

Our results for the transcript and metabolite comparison according to functional annotation by gene ontology (GO) classification are presented together with metabolite data in Table 4.3. We interpret these quantitative and statistically robust data in the framework of the pathway diagram to discern processes critical for survival of hypoxia.

Our observations are well supported by the literature. Mutating GAD1 in Arabidopsis resulted in changes in metabolite levels in roots. Expressing GAD lacking calmodulin-binding domain in tobacco and rice resulted in extremely high levels of GABA accumulation. GABA was also found to accumulated in roots and shoots of anoxia-stressed maize after several hours.[6] It was not found to correlate with cytoplasmic acidosis, perhaps this transition is triggered by the rise in cytosolic calcium ions during anoxia.

In the water-adapted plant *Lotus japonicus,* under complete submergence the total amount of free amino acids was unaffected in leaves, but reduced by half in roots.[24] In general, amino acids derived from precursors from the glycolytic pathway

Table 4.3 Nitrogen Compound Metabolic Process (GO:0006807) transcript and metabolite levels as measured in immunopurified mRNA populations isolated from roots (R) and shoots (S) where N=3. Values reported result from fold changes (FC) between hypoxia-stressed tissue and respective control tissue. Metabolite fold changes (*italicized*) in roots and shoots (N = 10). Fold changes (relative abundance in stressed tissue extracts divided by relative abundance of the same variable in control extracts) in response to two hours of low oxygen are given. **designates confidence at the 99% interval or greater, *designates confidence between 95-99% that the relative abundances are different between hypoxia stress treatment and control.

Pathway / Enzyme	**Abbrev.**	**Gene Location** (OrgXsomePos)	**R**	**S**
Nitrogen Metabolism GO:0071941, GO:0008940, GO:0006809				
nitrate reductase	NR	At1g77760, At1g37130	4.6**	1.1
nitrite reductase	NIR	At2g15620	1.1	1.0
nitrogen fixation aminotransferase	NFS/CDS	At2g30970, At1g08490	0.86	1.0
nitric oxide synthase	NOA/NOS	At3g47450		
glutamate synthetase	GOGAT/GLT	At5g04140, At5g53460, At2g41220	0.91	0.86
glutamate-ammonia ligase aka glutamine synthetase	GS/GLN	At5g37600, At5g35630, At5g16570, At3g53180, At1g66200, At3g17820, At1g48470	1.0	1.0
glutamate	*Glu*		*0.6***	*1.0*
glutamine	*Gln*		*0.9***	*0.9**
GABA Shunt GO:0006540				
glutamate decarboxylase	GAD	At3g17720, At5g17330, At1g65960, At2g02010	1.0	1.0
gamma-aminobutyrate transaminase	GABA-T	At3g22200	0.6*	1.0
gamma-aminobutyrate	*GABA*		*1.9***	*1.0*
succinic semialdehyde dehydrogenase	SSADH	At1g79440	0.9	1.0

increased during waterlogging, whereas levels declined for most amino acids linked to the TCA cycle, with the exception of Ala, GABA, and Glu which increased during waterlogging.

Accumulation of succinate, lactate, GABA and Ala has been reported in shoots of rice, wheat, corn, barley, rye and *Echinochloa sp.* (barnyard grass) following waterlogging or submergence. Plants that were more anoxia tolerant (i.e. rice) had a high

succinate to lactate ratio and cell sap became alkaline. Plants that were nonrestant produced more lactate than succinate and cell sap became acidic, and moderately resistant plants produced equal amounts of lactate and succinate and showed little variation in cell sap pH. The authors suggest GABA is an important means of consumption of H^+, but it alone cannot explain the alkaline shift.[45] Ala accumulation cannot explain the effect because the extent of Ala production was similar across plants tested. There is evidence that decarboxylation of arginine to putrescine may be responsible for alkalinization in rice. Succinate does not accumulate in roots, as it does in shoots.[45] Accumulation of GABA during hypoxia has been explained by activation of Glu decarboxylase by the lowered cytosolic pH, but this is unlikely because other enzymes in the GABA shunt perform optimally at pH 8-10.

Having explored the functional classification GO processes involving central carbon metabolism and nitrogen metabolism, we will now consider the effects of low oxygen as it impacts the pathways coupled to central carbon and nitrogen metabolism. Figure 3.1 affords a scheme for the biosynthesis of the amino acids from branch points in central carbon metabolism.

Ala is a source of Thr, Ile, Met, Lys via Asp, which can alternately become Asn. Branched-chain amino acids Leu and Val, as well as Ser, Cys and Gly are derived from pyruvate, competing with the production of lactate and ethanol under hypoxia. The aromatic amino acids (Trp, Phe, Tyr) are made via the shikimate pathway from phosphoenolpyruvate (PEP). Amino acids grouped with nitrogen metabolism include Glu, Pro, Gln, His, and Arg, which are connected to the TCA cycle at 2-oxoglutarate.

Significantly altered processes will be considered as they were observed in our translatomes and metabolome.

Revisiting the idea that root mitochondria utilize nitrate and nitrite as alternative electron acceptors, partly replacing fermentation for NADH reoxidation during glycolysis, we can examine these processes using eFP browser for these transcripts. Figure 4.18 shows that many dynamic changes take place in the roots and shoots even after only two hours of ephemeral stress. Specifically, NR in roots seems to completely change its expression pattern from being predominantly in the root cortex along the maturation zone to the root epidermis.

4.4.4.3 Cellular Amino Acid Metabolic Processes

Besides carbon metabolism, nitrogen metabolism in gray poplar was severely affected in waterlogged roots, as seen from numerous changes in the transcriptome and the metabolome related to nitrogen uptake, nitrogen assimilation, and amino acid metabolism.[12] With decreased demand for amino acid incorporation into proteins, because protein synthesis is an ATP-consuming process, regulation of amino acid biosynthesis by feedback inhibition or other means is likely to occur. Additionally, with amino acids available as substrates under energy crisis, we propose that amino acid catabolism for entry into glycolysis and subsequent shuttling of those products into ATP-producing branches of the TCA cycle were the main pathways induced by 2 h low oxygen stress.

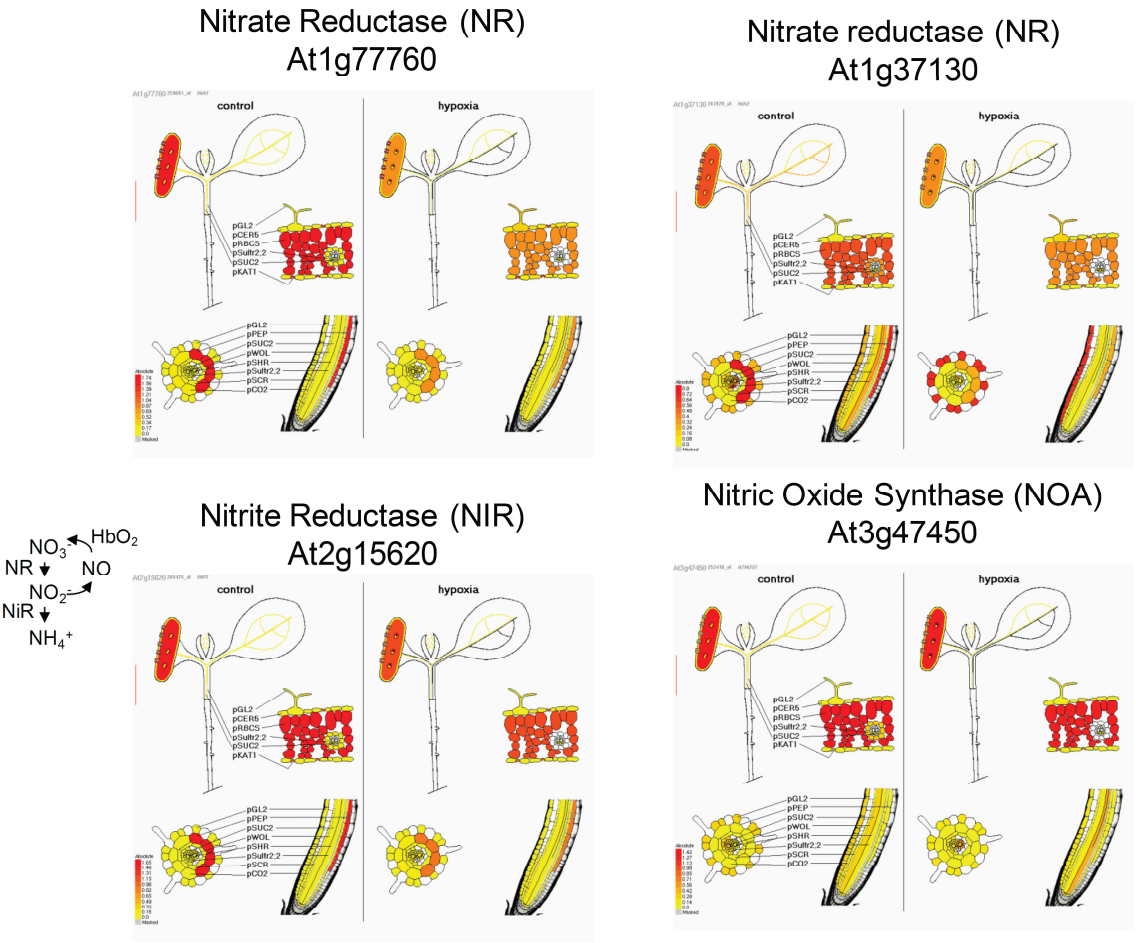


Figure 4.18 Arabidopsis eFP browser mappings of immunopurified mRNA abundance onto seedling cell-types. Reactions catalyzed by the reactions in the nitrogen assimilation pathway are shown. Yellow indicates that the transcript is not present (or present at low or nondetectable levels) and red indicates transcript is present. For each transcript, the control condition is shown in the left panel and the hypoxia-stressed condition is shown at right.

Gene ontology analysis revealed induction of NR (GO:0008940) and nitric oxide biosynthesis (GO:0006809) in roots. Shoots instead induced nitrogen compound metabolic process (GO:0051171) which is populated with many regulatory enzymes. A summary of fold changes in mRNA in polysome complexes and corresponding metabolite concentrations is provided by Table 4.3. Not surprisingly shoots do not alter

polysome enrichment of NR, since aerial tissues are not the predominant site of nitrogen assimilation.

Once inorganic nitrogen is absorbed by the roots as nitrate, it is reduced to nitrite and subsequently converted to Gln in an ATP-consuming step catalyzed by glutamine synthase (GS). Although this transcript is not altered in abundance, the endproduct Gln is significantly decreased after 2 hours of oxygen-limited conditions. Catalysis via glutamate synthase (GOGAT) converts Gln to Glu which can then participate in the GABA shunt involving glutamate GAD and GABA-T. The abundance of *GABA-T* mRNA in polysomes is decreased under low oxygen compared with control conditions, which may be a contributing factor to the observed accumulation of the reaction substrate, GABA.

Other significant changes under low oxygen include a depletion of transcripts encoding proline and ammonium transporters (At3g55740 and At1g64780, respectively) in polysomes of roots. Shoots loaded glutamate decarboxylase (GAD) mRNAs into polysomes but only in some cell types (data not shown).

Similarly to our previous study of low oxygen on whole Arabidopsis seedlings (Chapter 3), many transcripts for amino acid catabolism were loaded in polysomes. The enzymes discussed in the previous chapter were the transaminases (Section 3.3.4.3, Table 3.5) and again we see AlaAT (At1g17290) and AspAT (At5g19550) strongly accumulating, and this time more dramatically in roots than shoots (Table 4.4). Another transferase GGAT (At1g70580) seems to accumulate more in shoots, possibly responsible for the observed consumption of Glu.

A glance at Table 4.4, which summarizes cellular amino acid metabolic processes (GO:0006520), reveals a general upregulation of processes leading to amino acid catabolism (GO:0009063) and a downregulation of processes leading to amino acid biosynthesis (GO:0008652). Pathways that remain unaffected by 2 hours of low oxygen appear to include branched-chain amino acid biosynthesis, since we observe no significant alteration in loading status of *BCKD* gene mRNAs and a corresponding end product, Val. Conversely, we observe large decreases in the polysome loading of mRNAs encoding enzymes such as asparaginase (At5g08100), which catalyzes the conversion of Asn to Asp and corresponding increases in asparagine synthase (At3g47340) indicating that production of the nitrogen transporter Asn is favored. This may serve to divert carbon away from the Asp family amino acid biosynthetic processes. Asn may also play an important role in maintenance of redox balance or denitrification. Most likely Asp may becomes depleted as a metabolite because it is a cosubstrate for the reaction catalyzed by AspAT that is coupled with alanine production from pyruvate.

Notably, *AlaAT1/2* and *AspAT2* mRNAs showed greater elevation in the translatomes of roots than in shoots, whereas Ala accumulation and Asp depletion were more evident in shoots than roots (Tables 4.1 and 4.4; Figure 4.10). The discrepancy between gene regulation and metabolite accumulation could be due to pathway regulation based on substrate rather than enzyme availability. However, an alternative possibility is that Ala (pyruvate or other intermediates) produced in the root is transported to the shoot.

The observations presented here are interpreted in the context of the current literature to determine whether these data are representative. From these connections, we

can put stock in drawing hypotheses based on other unexpected changes that can be identified by either unsupervised multivariate statistics, or clustering and cutting large matrices of biologically-informative data until a useful global picture or subset is culled.

What follows is an interpretation of our observations in the context of the literature in the area of low oxygen responses in plants. Nitric oxide has been proposed as a signaling molecule. The evidence from this study suggests that under low oxygen, pathways leading to the production of nitric oxide are stimulated in roots within 2 hours of hypoxia onset. GABA is formed *in vivo* via a metabolic pathway called the GABA shunt. The initial step in this pathway utilizes 2-oxoglutarate formed from glucose metabolism via the TCA cycle. 2-oxoglutarate is then transaminated by α-oxoglutarate transaminase (GABA-T) to form Glu, the immediate precursor of GABA. Finally, Glu is decarboxylated to form GABA by the enzyme(s) glutamic acid decarboxylase (GAD).[41]

During episodes of heat stress, GABA accumulates to levels six- to 10-fold higher than in unstressed plants. The synthesis of GABA from Glu consumes protons and thus the GAD-catalyzed reaction is one of several reactions involved in the regulation of cytosolic pH. GABA is then transported into the mitochondrion where it is converted to succinic semialdehyde (SSA) in a coupled reaction consuming pyruvate and producing Ala via GABA-T [EC 2.6.1.19]. SSA is converted to succinate by SSA-DH [EC 1.2.1.16 or 1.2.1.24] regenerating NADPH. Arabidopsis mutants of SSA-DH demonstrate increased hydrogen peroxide production during episodes of heat stress and show increased programmed cell death (PCD). Thus, the function of the GABA metabolic

Table 4.4 Cellular Amino Acid Metabolic Process (GO:0006520) transcript and metabolite levels as measured in this study. Enzymes whose transcript abundance was significantly altered among immunopurified mRNA populations isolated from roots (R) and shoots (S) where N=3. Values reported result from fold changes (FC) between hypoxia-stressed tissue and respective control tissue. Metabolite fold changes (*italicized*) in roots and shoots (N = 10). **designates confidence at the 99% interval or greater, *designates confidence between 95-99%.

Pathway / Enzyme	**Abbrev.**	**Gene Location** (OrgXsomePos)	**R**	**S**
Amino Acid Catabolism GO:0009063				
hydroxyphenylpyruvate dioxygenase	HPD	At1g06570	2.5*	1.8*
gamma-hydroxybutyrate dehydrogenase	GHBD	At3g25530	1.1	1.3
glutamate dehydrogenase	GDH	At5g18170, At5g07440	1.7	1.0
proline oxidase		At3g30775	2.5**	1.8
alanine aminotransferase	AlaAT	At1g17290	8.6**	4.4**
alanine	*Ala*		*2.4***	*4.6***
alanine:2-oxoglu. aminotransferase	AOAT	At1g23310	1.0	1.1
asparaginase		At3g16150	3.7**	0.86
glutamate:glyoxylate aminotransferase	GGAT	At1g70580	1.1	1.5
broad specificity (Asp, Tyr, Phe) aminotransferase	ACS	At5g51690	1.1	1.0
broad specificity (Asp, Tyr, Phe) aminotransferase	ACS	At1g62960	0.7	1.0
branched-chain alpha-keto acid dehydrogenase	BCKD	At3g13450, At3g06850	1.2	1.2
valine	*Val*		*1.0*	*1.1*
aspartate aminotransferase	AspAT 2	At5g19550	5.2**	4.2**
aspartate aminotransferase	AspAT 1,3,5	At5g11520, At4g31990, At2g30970	1.1	1.2
aspartate aminotransferase	AspAT putative	At5g51690, At1g80360, At1g62960, At1g62800, At2g22250	1.1	1.2
asparagine synthase	AS	At3g47340	1.5	1.5
Amino Acid Biosynthesis GO:0008652				
asparaginase	ASN	At5g08100	0.78	0.76
S-adenosylmethionine decarboxylase		At3g02470, At3g25570, At5g15950	0.75	0.82
arginine decarboxylase		At2g16500, At4g34710	0.48*	0.66
glycine decarboxylase		At2g35120	0.95	0.88
isopropylmalate dehydrogenase	IMDH	At1g31180	0.51*	1.1
aspartate kinase	ASP putative	At3g02020	0.28**	0.70
aspartate	*Asp*		*0.6***	*0.8***
asparagine	*Asn*		*0.9*	*0.9*

pathway appears to modulate the oxidative status of the cell during heat stress, and may be involved in the production of a cell signal, effectively modulating PCD.[41]

Regulation of cytoplasmic pH by glutamate decarboxylase (GAD, EC 4.1.1.15) has been proposed since GABA synthesis by GAD consumes a proton, and its activity is stimulated in acidic conditions (pH ~5.8). The role of Ala and GABA in maintenance of osmotic potential in stressed tissues may be important to counteract the rapid fall of regular carbohydrate levels. Glu is converted to GABA, which is converted to SSA by GABA-T, simultaneously producing Ala from pyruvate. Succinate is produced from succinic semialdehyde under normoxia, however under hypoxia SSA can also be reduced to GHB.[46]

Comparison of *AlaAT* knockout mutants and wild-type plants for Ala concentrations over hypoxic stress and subsequent recovery indicated that AlaAT1 is primarily involved in Ala breakdown during recovery from low oxygen stress. Ala biosynthesis under hypoxia does not require AlaAT1 activity.[47]

After a thorough look at primary metabolic pathways, we aimed to ascertain whether the aromatic region of the NMR spectrum would yield information about Phe, Trp, and Tyr. The S/N of these metabolites was not sufficient to quantitatively evaluate fold-changes (Figure 4.19) although it is clear that shoots are distinct from roots and hypoxic conditions in roots are distinct from controls (7.6 and 7.8 ppm).

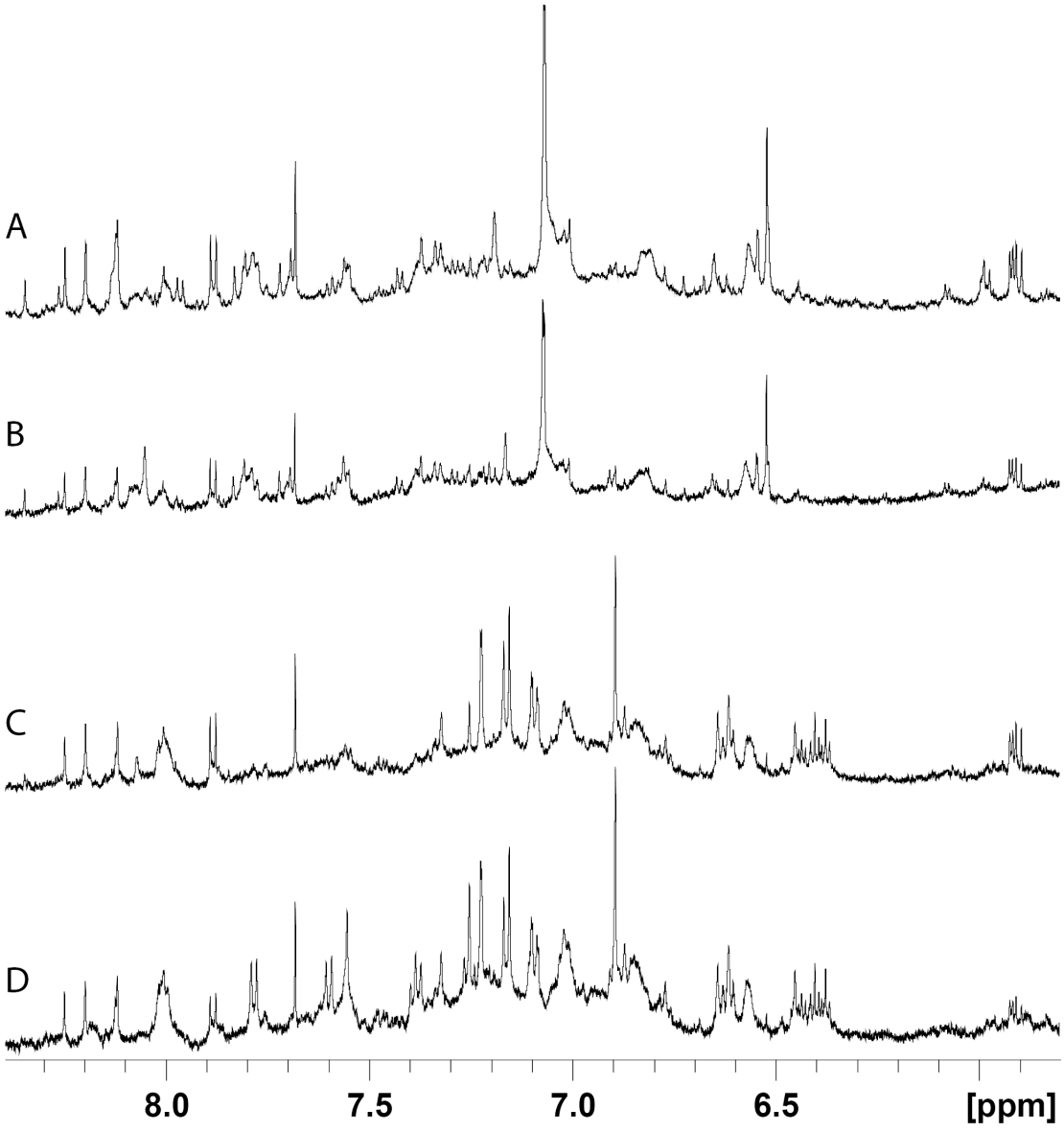


Figure 4.19 NMR spectra of samples representing each treatment. Expanded to facilitate discussion of aromatic amino acids and secondary metabolites. Shoots under control (A) and low oxygen stress (B) are distinct from roots under control (C) and hypoxic conditions (D).

4.5 Conclusions

Overall this study enforced our confidence in ribosome-immunopurification for the study of translational regulation and metabolic network reconfiguration under low

oxygen. We hypothesized that a different stress response would be observed in aerial shoots and terrestrial roots, and as expected, there were aspects of adaptive mechanisms that were distinct in each organ. The reasons for this specialization are likely because each organ has access to different detoxification routes and a different capacity for stored substrate. The research also raises the possibility that transport of metabolites may occur between the root and the shoot during low oxygen stress. Typically when a plant encounters a situation of oxygen limitation, it will either conserve its resources (e.g., submergence tolerance by quiescence) or expend them in an effort to escape the stressful situation (e.g., submergence survival by escape).[48] Understanding these processes at the molecular level may address agricultural challenges we currently face, and those changes which may result upon further progress of global climate change.

In summary, ATP-generating processes are stimulated while those consuming ATP are inhibited in part through translational regulation. Specifically PK and SCS-mediated substrate level ATP production may attempt to maintain a supply of energy sufficient to survive until reoxygenation. The reliance on fermentation of pyruvate to lactae, ethanol, alanine or GABA to maintain this production may be distinct in the root versus the shoot. Moreover, transport of intermediates from the root to the shoot, such as pyruvate, Ala may also occur. It will be important to assess whether the observed distinctions in gene regulation and metabolism contribute to energy management in the two organ systems.

From the analytical perspective, improvements over our previous study arise from modifications to the experimental design (more bioreplicates), sampling

(microdissection), sample preparation (lipid removal and solvent homogeneity), data acquisition (improved signal-to-noise), resonance assignment (use of databases at same pH), and data interpretation (pmn.aracyc). A non-targeted systems biological approach allowed us to examine pathways in which no preexisting hypotheses existed. In conclusion, this work illustrates the utility of a model organism for integrated studies at multiple levels of biological regulation and its application to a problem of agricultural relevance.

4.6 References

1. Mustroph, A.; Boamfa, E. I.; Laarhoven, L. J. J.; Harren, F. J. M.; Albrecht, G.; Grimm, B., Organ-specific analysis of the anaerobic primary metabolism in rice and wheat seedlings, I: Dark ethanol production is dominated by the shoots. *Planta* **2006,** 225, 103-114.

2. Mustroph, A.; Boamfa, E. I.; Laarhoven, L. J. J.; Harren, F. J. M.; Pors, Y.; Grimm, B., Organ specific analysis of the anaerobic primary metabolism in rice and wheat seedlings II: Light exposure reduces needs for fermentation and extends survival during anaerobiosis. *Planta* **2006,** 225, 139-152.

3. Stitt, M.; Bulpin, P. V.; ap Rees, T., Pathway of starch breakdown in photosynthetic tissues of *Pisum sativum*. *Biochem. Biophys. Acta* **1978,** 544, 200-214.

4. Waters, I.; Morell, S.; Greenway, H.; Colmer, T. D., Effects of anoxia on wheat seedlings II Influence of O_2 supply prior to anoxia on tolerance to anoxia, alcoholic fermentation and sugar levels. *J. Exp. Bot.* **1991,** 42, 1437-1447.

5. Limami, A. M.; Glevarec, G.; Ricoult, C.; Cliquet, J. B.; Planchet, E., Concerted modulation of alanine and glutamate metabolism in young Medicago truncatula seedlings under hypoxic stress. *J.Exp. Bot.* **2008,** 59, (9), 2325-2335.

6. Drew, M. C., Oxygen deficiency and root metabolism: injury and acclimation under hypoxia and anoxia. *Ann. Rev.Plant Phys.* **1997,** 48, 223-250.

7. Branco-Price, C.; Kaiser, K. A.; Jang, C. J. H.; Larive, C. K.; Bailey-Serres, J., Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana. *Plant J.* **2008,** 56, (5), 743-755.

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

9. Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006,** 2, (4), 171-196.

10. Steinfath, M.; Groth, D.; Lisec, J.; Selbig, J., Metabolite profile analysis: from raw data to regression and classification. *Physiol. Plantarum* **2008,** 132, 150-161.

11. Ellis, M. H.; Dennis, E. S.; Peacock, W. J., Arabidopsis roots and shoots have different mechanisms for hypoxic stress tolerance. *Plant Physiol.* **1999,** 119, 57-64.

12. Kreuzwieser, J.; Hauberg, J.; Howell, K. A.; Carroll, A.; Rennenberg, H.; Millar, A. H.; Whelan, J., Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia. *Plant Physiol.* **2009,** 149, (1), 461-473.

13. Werner, T.; Holst, K.; Pors, Y.; Guivarc'h, A.; Mustroph, A.; Chriqui, D.; Grimm, B.; Schmulling, T., Cytokinin deficiency causes distinct changes of sink and source parameters in tobacco shoots and roots. *J. Exp. Bot.* **2008,** 59, (10), 2659-2672.

14. Govindaraju, V.; Young, K.; Maudsley, A. A., Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **2000,** 13, 129-153.

15. Ratcliffe, R. G., *In vivo* NMR studies of the metabolic response of plant tissues to anoxia. *Ann. Bot.-London* **1997,** 79, (Supplement A), 39-48.

16. Wiklund, S.; Nilsson, D.; Eriksson, L.; Sjostrom, M.; Wold, S.; Faber, K., A randomization test for PLS component selection. *J. Chemometr.* **2007,** 21, 427-439.

17. Schripsema, J., Application of NMR in plant metabolomics: techniques, problems and prospects. *Phytochem. Anal.* **2010,** 21, 14-21.

18. Keun, H. C.; Ebbels, T. M. D.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Analytical reproducibility in ^{1}H NMR-based metabonomic urinalysis. *Chem. Res. Toxicol.* **2002,** 15, 1380-1386.

19. Dai, H.; Xiao, C.; Lui, H.; Tang, H., Combined NMR and LC-MS analysis reveals the metabonomic changes in *Salvia miltiorrhiza* Bunge induced by water depletion. *J. Proteome Res.* **2010,** 9, (3), 1460-1475.

20. Wiklund, S. Spectroscopic Data and Multivariate Analysis ~Tools to Study Genetic Perturbations in Poplar Trees~. Umea University, Umea, Sweden, 2007.

21. The Arabidopsis Information Resource (TAIR). http://www.arabidopsis.org (accessed Mar 10, 2007)

22. Bais, P.; Moon, S. M.; He, K.; Leitao, R.; Dreher, K.; Walk, T.; Sucaet, Y.; Barkan, L.; Wohlgemuth, G.; Roth, M. R.; Wurtele, E. S.; Dixon, P.; Fiehn, O.; Lange, B. M.; Schulaev, V.; Sumner, L. W.; Welti, R.; Nikolau, B. J.; Rhee, S. Y.; Dickerson, J. A., PlantMetabolomics.org: a web portal for plant metabolomics experiments. *Plant Physiol.* **2010,** 152, 1807-1816.

23. Hinz, M.; Wilson, I. W.; Yang, J.; Buerstenbinder, K.; Llewellyn, D.; Dennis, E. S.; Sauter, M.; Dolferus, R., Arabidopsis RAP2.2: an ethylene response transcription factor that is important for hypoxia survival. *Plant Physiol.* **2010,** 153, (2), 757-772.

24. Rocha, M.; Licausi, F.; Araujo, W. L.; Nunes-Nesi, A.; Sodek, L.; Fernie, A. R.; van Dongen, J. T., Glycolysis and the tricarboxylic acid cycle are linked by alanine aminotransferase during hypoxia induced by waterlogging of *Lotus japonicus*. *Plant Physiol.* **2010,** 152, (3), 1501-1513.

25. Vashisht, D.; Hesselink, A.; Pierik, R.; Ammerlaan, J. M. H.; Bailey-Serres, J.; Visser, E. J. W.; Pedersen, O.; Zanten, M. V.; Vreugdenhil, D.; Jamar, D. C. L.; Voesenek, L. A. C. J.; Sasidharan, R., Natural variation of submergence tolerance among *Arabidopsis thaliana* accessions. *New Phytol.* **2011**, 190, 299-310.

26. Fukao, T.; Bailey-Serres, J., Plant responses to hypoxia - is survival a balancing act? *Trends Plant Sci.* **2004,** 9, (9), 449-456.

27. Bailey-Serres, J.; Voesenek, L., Life in the balance: a signaling network controlling survival of flooding. *Curr. Opin. Plant Biol.* **2010,** 13, (5), 489-494.

28. Alonso, A. P.; Raymond, P.; Rolin, D.; Dieuaide-Noubhani, M., Substrate cycles in the central metabolism of maize root tips under hypoxia. *Phytochemistry* **2007,** 68, 2222-2231.

29. Grossman, A. R.; Catalanotti, C.; Yang, W. Q.; Dubini, A.; Magneschi, L.; Subramanian, V.; Posewitz, M. C.; Seibert, M., Multiple facets of anoxic metabolism and hydrogen production in the unicellular green alga *Chlamydomonas reinhardtii*. *New Phytol.* **2011,** 190, (2), 279-288.

30. Sweetlove, L. J.; Beard, K. F. M.; Nunes-Nesi, A.; Fernie, A. R.; Ratcliffe, R. G., Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci.* **2010,** 15, 462-470.

31. Vanlerberghe, G. C.; Feil, R.; Turpin, D. H., Anaerobic metabolism in the N-limited green alga *Selenastrum minutum* I. regulation of carbon metabolism and succinate as a fermentation product. *Plant Physiol.* **1990,** 94, 1116-1123.

32. Grafahrend-Belau, E.; Schreiber, F.; Koschutzki, D.; Junker, B. H., Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. *Plant Physiol.* **2009,** 149, 585-598.

33. Blokhina, O.; Lackman, P.; Maaheimo, H.; Rischer, H.; Kallio, P. T.; Haggman, H.; Fagerstedt, K., Metabolic profiling of *Arabidopsis thaliana* expressing Vitreoscilla haemoglobin (VHb) under oxygen deprivation. Presented at the 10th conference of the *International Society for Plant Anaerobiosis (ISPA)*, Volterra, Italy, 2010.

34. Zabalza, A.; Van Dongen, J. T.; Froehlich, A.; Oliver, S. N.; Faix, B.; Gupta, K. J.; Schmalzlin, E.; Igal, M.; Orcaray, L.; Royuela, M.; Geigenberger, P., Regulation of respiration and fermentation to control the plant internal oxygen concentration. *Plant Physiol.* **2009,** 149, (2), 1087-1098.

35. Dolferus, R.; Wolansky, M.; Carroll, R.; Miyashita, Y.; Ismond, K.; Good, A., Functional analysis of lactate dehydrogenase during hypoxic stress in Arabidopsis. *Funct. Plant Biol.* **2008,** 35, 131-140.

36. Huang, S.; Ishizawa, K.; Greenway, H.; Colmer, T. D., Manipulation of ethanol production in anoxic rice coleoptiles by exogenous glucose determines rates of ion fluxes and provides estimates of energy requirements for cell maintenance during anoxia. *J.Exp. Bot.* **2005,** 56, (419), 2453-2463.

37. Storey, K. B., *Functional metabolism: regulation and adaptation*. John Wiley & Sons: Hoboken, NJ, 2004; p 594.

38. Hochachka, P. W., *Living without oxygen: closed and open systems in hypoxia tolerance*. Harvard University Press: Cambridge, MA, 1980; p 181.

39. Oliver, S. N.; Lunn, J. E.; Urbanczyk-Wochniak, E.; Lytovchenko, A.; Dongen, J. T. v.; Faix, B.; Schmalzlin, E.; Fernie, A. R.; Geigenberger, P., Decreased expression of a cytosolic pyruvate kinase in potato tubers leads to a decline in pyruvate resulting in an *in vivo* repression of the alternative oxidase. *Plant Physiol.* **2008,** 148, 1640-1654.

40. Geigenberger, P., Response of plant metabolism to too little oxygen. *Curr. Opin. Plant Biol.* **2003,** 6, (3), 247-256.

41. Taiz, L.; Zeiger, E., *Plant Physiology*. 4th ed.; Sinauer Associates: Sunderland, MA, 2006.

42. Gupta, K. J.; Zabalza, A.; van Dongen, J. T., Regulation of respiration when the oxygen availability changes. *Physiol. Plantarum* **2009,** 137, (4), 383-391.

43. Storey, K. B., *Functional Metabolism: regulation and adaptation*. John Wiley & Sons: Hoboken, NJ, 2004; p 417.

44. Gupta, K. J.; Stoimenova, M.; Kaiser, W. M., In higher plants, only root mitochondria, but not leaf mitochondria reduce nitrite to NO, *in vitro* and *in situ*. *J.Exp. Bot.* **2005,** 56, (420), 2601-2609.

45. Menegus, F.; Cattaruzza, L.; Chersi, A.; Fronza, G., Differences in the anaerobic lactate-cuccinate production and in the changes of cell sap pH for plants with high and low resistance to anoxia. *Plant Physiol.* **1989,** 90, 29-32.

46. Miyashita, Y.; Good, A. G., Contribution of the GABA shunt to hypoxia-induced alanine accumulation in roots of *Arabidopsis thaliana*. *Plant Cell Physiol.* **2008,** 49, (1), 92-102.

47. Miyashita, Y.; Dolferus, R.; Ismond, K.; Good, A. G., Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in *Arabidopsis thaliana*. *Plant J.* **2007,** 49, 1108-1121.

48. Voesenek, L. A. C. J. and Bailey-Serres, J., Life in the balance: a signaling network controlling survival of flooding. *Curr. Op. Plant Bio*. **2010,** 13, 489-494.

Chapter Five

LC-NMR and LC-MS Structural Elucidation

of Arabidopsis Secondary Metabolites

John Limtiaco facilitated LC-NMR based studies for establishing these metabolite identifications. Szabolcs Beni recognized the NMR signature of flavonol aromatic proton chemical shifts and couplings, giving us a structural foothold from which to work forward in creating an in house *Arabidopsis thaliana* secondary metabolite database. Vishwa Shah assisted in preparation of the plant material for this work and in populating the cheminformatic database with secondary metabolite structural information.

In Chapter 4, comparison of root and shoot tissue extracts after brief low oxygen stress suggested changes in metabolite abundance in the aromatic region of the NMR spectra, where secondary metabolite resonances are found (Figure 4.19). Our focus in this chapter was to isolate chromatographically and determine the structure of a set of Arabidopsis secondary metabolites. Mass information alone was not sufficient for structural assignments, even when collision induced dissociation (CID) tandem MS was applied. To elucidate structures of Arabidopsis secondary metabolites, LC was directly coupled to NMR. UV absorbance was recorded online at 254 nm, initiating automated loop trapping to store single components and binary mixtures isolated by LC. After tentative structural assignments were made, pure reference material for one compound was obtained and interrogated by identical methods. Assignments for eight secondary metabolites of *Arabidopsis thaliana* are presented here. These are supported by a database populated with structures and calculated values for mass with proposed fragments, distribution coefficient, proton and carbon chemical shifts and coupling constants. This chapter establishes the foundation to address a biological question: hypoxia, which will be the focus of Chapter 6.

5.1 Introduction to Plant Secondary Metabolism

Secondary metabolism is defined as the sum of chemical reactions which are not absolutely required for the survival of the organism. Primary metabolic reactions involve biosynthesis of amino acids, lipids, carbohydrates and nucleotides; providing building blocks for respiration and other basic cellular functions (transport, reproduction, energy conservation). Secondary metabolic reactions are responsible for synthesis of molecules such as pigments and antimicrobial compounds, which mediate interactions with other species but are not directly responsible for basic cellular functions.

5.1.1 Combinatorial, Plasticity, Dynamic Range

Metabolite biosynthesis in plants follows a combinatorial model where a limited number of scaffold molecules are decorated with a limited number of substituents in any given plant.[1, 2] Analytically this results in a large number of constitutional isomers, presenting a challenge for structural assignment by UV or MS only, even utilizing exact mass information.[3]

Plasticity in plant secondary metabolomes results from ongoing evolution. In this so-called "arms race," secondary metabolites play defensive or offensive roles in the plant life cycle, naturally resulting in molecules with bioactivity. The plant metabolome is chemically diverse in terms of polarity, molecular mass and amount present. As of the year 2005, already 200,000 secondary metabolite structures had been elucidated from plants and microbes.[4] Metabolites observed *in planta* span seven orders of magnitude in concentration.[5] Secondary metabolites derive their bioactivity in nonconspecific

organisms from their unusual chemical nature, circumventing deactivation and/or transport which would normally result in detoxification.[6]

Between 15-20% of Arabidopsis genes are predicted to encode enzymes of secondary metabolism, far more than necessary to produce the number of metabolites identified so far. Because a significant portion of its genome is devoted to secondary metabolism, some claim there exist undiscovered molecules yet to be isolated and characterized.[7] An alternative hypothesis is that gene duplication is responsible for the redundancy observed in secondary metabolism. Recent work has suggested gene duplication for functional compensation is the reason large numbers of genes seem to encode tens or hundreds of combinatorially diversified secondary metabolites, underscoring the current interest in developing robust methods to profile the complement of secondary metabolic pathways in model systems such as Arabidopsis.[8]

5.1.2 Metabolomics and Metabolic Profiling

The study of secondary metabolites is an interdisciplinary field comprising pharmaceutical natural product discovery, plant behavioral studies documenting molecules responsible for communication among and between organisms, and taxonomic classification of plant species by their secondary metabolite profiles.[9] Careful studies of plants for medicinal compounds were carried out as early as 1791 in Paris.[10] The complexity of isolating single components from among a mixture containing potentially toxic substances required marrying interests of purveyors of medicine (human physiology) and chemistry (molecular reactivity).

NMR, despite its limitations in terms of sensitivity, is a powerful tool in dereplication.[11-14] The application of LC-NMR to profiling of human urine in 1992 demonstrated its suitability for metabolite analysis.[15] The technique became useful to researchers in plant natural product screening throughout the late 90s and early 00s.[16] The development of WET solvent suppression in the mid 90s significantly improved spectral quality and was a contributing factor to the popularization of the technique.[17] Commercial availability of LC-NMR flow probes and interfaces led to wide-spread adoption of LC-NMR.[12, 16, 18, 19] An engaging review summarizing challenges and applications of LC-NMR in plant studies is afforded by one of the leading groups in the field.[20]

5.1.3 Challenges in Chemical Space Exploration

The plasticity and combinatorially-biosynthesized diversity in secondary metabolism presents both a challenge and an advantage in studies aimed toward dereplication.[21] Taxonomically-related plants produce similar types of molecules, thereby reducing the number of potential identifications given only mass information.[22, 23]

Pure standards for primary metabolites are readily available, whereas for secondary metabolites, their low abundance and diversity make standards expensive (often chromatographically isolated from biological material) or in some cases commercially unavailable (when plant material is limited and benchtop synthesis is not an option).

We took an approach of limiting the region of chemical space under study. Three strategies accomplished this (1) enrichment via solid-phase extraction capture, (2) a liquid chromatographic method focused on the hydrophilic metabolites eluting between 2 and 20% of the organic mobile phase discarding even moderately hydrophobic molecules, and (3) monitoring UV absorbance for peak trapping at 254 nm. Secondary metabolites in Arabidopsis fall into several classes; phenylpropanoids, glucosinolates, terpenoids and phytoalexins.[24] Plant phenolics represent a large group of defensive compounds having a phenol (hydroxybenzene) moiety, which absorb UV light at 254 nm.[25]

Phenolics range in complexity from one to three rings, and more if hydrolysable tannins and condensed tannins are considered. Phenolics that are colored serve roles in pollinator attractants; those that have bitter taste are herbivore deterrants.[26] The planar ring systems can mimic ATP and other redox coenzymes including NADPH, NADH, $FMNH_2$ and $FADH_2$. Anti-inflammatory antioxidant[26, 27] properties arise through covalent interaction with free radicals, notably ROS such as superoxide (O_2^-). Many phenolics have antimicrobial (antibacterial or antifungal) properties.[28]

It is estimated that about 2% of all carbon synthesized by plants (1×10^9 t per annum) is converted into flavonoids, one of the largest groups of naturally occurring phenols.[29] A generalized structure is provided in Figure 5.1. Flavonoids are virtually ubiquitous in green plants, therefore will be encountered in nearly any plant extract. Structurally, the flavonoid scaffold contains fifteen carbon atoms arranged in C_6-C_3-C_6 configuration. Two aromatic rings are linked by a three carbon unit which may or may

not form a third ring. Further modification of the flavonoid scaffold may result in additional (or reduced) hydroxylation, methylation of hydroxyl groups or flavonoid nucleus; isoprenylation of hydroxyl groups or flavonoid nucleus; and most importantly glycosylation of hydroxyl groups (producing *O*-glycosides) or the flavonoid nucleus (producing C-glycosides). The effect of glycosylation is to make the flavonoid less reactive and more water (sap) soluble. Glycosides occasionally have one (or more) of their sugar hydroxyls derivatized with an acid such as acetic or ferulic. Among the plant kingdom, flavonoids are restricted to Bryophyta or higher classes suggesting evolutionary development as recently as 460 million years ago.

Flowering plants containing large amounts of flavonoid glycosides are summarized in the following table (Table 5.1). It can be seen that a variety of flavonoid skeletons are available as well as decorative groups. The focus of this work was on phenylpropanoids, based on our preliminary structural evidence by 1D ^{1}H NMR of low-oxygen stressed Arabidopsis.

Phenylpropanoids, such as resveratrol found in red wine, have generated public interest having shown potential as therapeutic agents in cancer treatment and prevention,[30, 31] protection against neurodegenerative disease,[32] and management of obesity and related cardiovascular problems.[33] In plants they are thought to play roles in regulating polar auxin transport, a plant hormone which regulates plant growth and development, which would be of critical relevance in young seedlings.[24, 34] Phenylpropanoids are also credited with providing protection from UV light and defense against herbivores and pathogens.

Table 5.1 Sources of plant-based drugs of the flavonoid class. Scaffolds quercetin (Q), luteolin (L), kaempferol (K), isorhamnetin (I), apigenin (A), peonidin (P) and eriodictyol (E) are decorated with glucose (G), galactose (gal), and rhamnose (R) sugar substitutions at the C3, C7 and less commonly at C8 positions. Q3GR is commonly called rutin. The substituent RG is commonly called neohesperidose.

Plant source	Common name	Flavonoid constituents
Arnica montana	Arnica	Q3G, Q3Ggal, L7G, K3G
Calendula officinalis	Marigold	I3G, I3RG, I3RGR, Q3G, Q3GR
Chamaemelum nobile	Chamomile	A7G, L7G
Chamomila recutita	German Chamomile	A7G, L7G, Q7G, P7G
Cereus grandiflorus	Night-blooming Cereus	I3gal, I3galRG, I3RG
Crategus monogyna	Hawthorne	Q3gal, Q3RG, Q3Rgal, A8G
Robinia pseudoacacia	Acacia	K3Rgal7R, A7RG
Tussilago farfara	Farfarae	Q3G, Q3gal, Q3RG
Aurantii pericarpium	Seville Orange	E7RG, Q7RG, N7GR

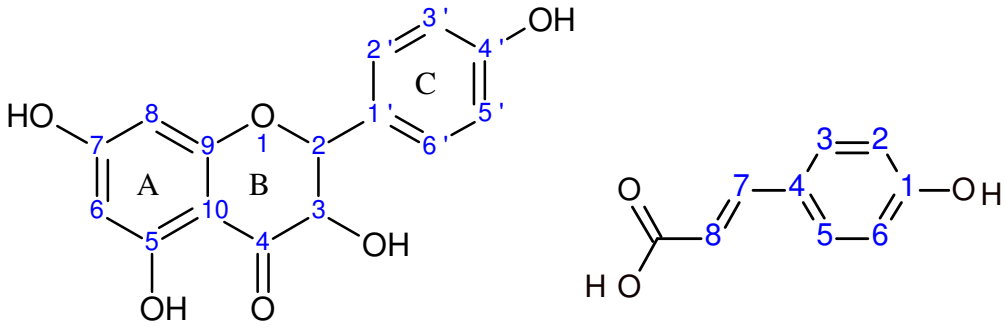


Figure 5.1 General skeleton of a flavonol, which generates resonances by magnetically equivalent first-order coupled protons at the 2',6' and 3',5' positions. The C8 and C6 positions also contribute resonances to the aromatic reigon, but are more upfield. The ring enclosed by C5-10 is called the A ring. The heterocyclic ring enclosed by C2-4 is called the C ring. The ring enclosed by C1'-6' is called the B ring. Numbering for cinnamic acid skeletons is shown also.

Phenylpropanoids are naturally occurring phenolic compounds which have an aromatic ring to which a three-carbon side chain is attached. They are derived biosynthetically from the amino acid Phe and contain one or more C_6-C_3 residues.

Hydroxycinnamic acids form the building blocks of lignin and regulate growth and provide disease resistance. Nearly ubiquitous in plants are ferulic, sinapic, caffeic and p-coumaric acid. Other rarer forms include 3-hydroxy-4-methoxycinnamic, o-coumaric, and p-methoxycinnamic acids.[35, 36] These are obtained in best yield by mild alkaline hydrolysis since hot acid will cause decarboxylation to the corresponding styrene.[37]

In Arabidopsis the end products of the phenylpropanoid pathway include sinapoyl malate and sinapoyl choline.[38] Sinapoyl glucose is the substrate used for biosynthesis of sinapoyl choline.[39] The motivation to pursue phenylpropanoids was threefold: (1) their potentially interesting roles in regulating plant responses to low oxygen stress given what is already proposed regarding their functions *in planta,* (2) we had observed induction of aromatic compounds in the NMR spectrum that correlated with hypoxia treatment, although we did not know which resonances were bioactive or even which corresponded to phenylpropenoids, and (3) evidence of large fold changes in the transcript abundance and polysome loading for messages encoding genes involved in secondary metabolite biosynthetic pathways had been observed in several studies.[40, 41]

5.2 Experimental Approach

The abundance of secondary compounds was selectively enriched via capture and release using a hydrophilic-lipophilic balanced (HLB) solid-phase extraction (SPE) cartridge.[42] Areas of chemical space captured by HLB SPE were found to be similar to that which adsorbs to C18, when identical solvation, conditioning, loading, washing and elution conditions were applied (data not shown).

Attempts to interrogate the intact mixture representing secondary metabolite-enriched Arabidopsis tissue extracts were unsuccessful due to excessive resonance overlap in the aromatic region of the spectrum even when two dimensional techniques were applied (Chapter 1, Figures 1.7-1.9).

To alleviate crowding, a chromatographic separation was developed to target a small portion of the secondary metabolome, resolving approximately fifteen metabolites. Ultra-high pressure liquid chromatography (UPLC) shortened the time necessary for method development, as well as providing greater sensitivity and conserving solvent.[43] Quadrupole time-of-flight (QTOF) mass spectrometry (MS) was used to detect metabolites, which lowers the detection limit of our method approximately 3 orders of magnitude. Incorporating SPE in the sample preparation served to protect the integrity of the MS instrument, and improved data quality by reducing levels of interfering ions and lipids/waxes.

The separation was performed in reverse phase, and only early-eluting compounds (2 – 20% organic) were analyzed by MS. Tandem mass spectrometry, by means of collision-induced dissociation (CID), was employed to gain structural information. It soon became evident that mass information alone, even with fragmentation, could not establish molecular identity.[44] Plant secondary compounds are often not available as pure standards for traditional annotation by spiking.[45] For this reason, we returned to NMR. In particular, LC-NMR was useful in automated sample handling and one-dimensional proton spectra were acquired in sufficient signal-to-noise ratio (S/N) to establish molecular identity for several key metabolites.

This chapter outlines our approach to annotation and dereplication of a small portion of the secondary metabolites made by seedlings of *Arabidopsis thaliana*. Levels of secondary metabolites such as terpenoids, alkaloids, glucosonolates and flavonoids are often higher in young tissue,[46] we expected to find these compounds in abundance in our Arabidopsis seedlings. Because of the low sensitivity of NMR and the low concentration *in planta* of these molecules, many seedlings had to be pooled to constitute a sample of sufficient concentration for NMR. In a single concentrated extract resulting from nearly 5,000 pooled seedlings, it was undertaken to focus on dereplication of the most hydrophilic secondary metabolites found in Arabidopsis seedling tissue extracts.

5.2.1 Plant Growth

Arabidopsis thaliana (ecotype Columbia-0) seedlings were prepared as described in Chapter 2, section 2.2.2.2 with the following timing specifications. Briefly, seeds were surface-sterilized and vernalized for 3 days prior to being transferred to 1x MS solid media containing 1% sucrose (w/v). Plates were placed vertically in a growth chamber at 2:00 pm (7 h into the light cycle) on day 0 (September 1, 2009). Seedlings were harvested 7 d later between 12:30 and 3:30 pm into liquid nitrogen and directly homogenized into a single pool of tissue, which was lyophilized overnight in 6 x 50 mL Falcon tubes. The total tissue pool (1.7745g DW) resulted from 46 plates containing approximately 100 seeds per plate.

5.2.2 Tissue Extraction

Extraction was performed at low pH (pH meter reading 2.89) using 50/50 (v/v) acetonitrile-d_3/water-d_2 as the primary extraction buffer (EB). Approximately 20 mg DW of tissue was combined with 900 µL of EB, which was homogenized with a micropestle (Eppendorf) for 4 min manually in 1.5 mL microcentrifuge tubes (Eppendorf). Primary extracts were clarified by centrifugation at 13,200 x *g*. Two of these supernatants were combined per 2 mL microcentrifuge tube for installation into speedvacuum (to remove EB). Time to dryness was 3-5 h. Dried extracts were stored under refrigeration (-4 ^{o}C) overnight.

5.2.3 Enrichment of Secondary Metabolites by SPE Capture

A buffered solution (pH meter reading 2.92) containing 100 µM acetate-d_4 in Burdick and Jackson (B & J) Water (ultrapure, UPLC-grade) was prepared for a reconstitution buffer (RB). Dried extracts were combined with 200 µL RB, vortexed 60 s, sonicated 10 min, and vortexed 60 s to resuspend water-soluble metabolites. Centrifugation at 13,200 x *g* precipitated insoluble material.

Waters Oasis HLB SPE cartridges containing 60 mg sorbent with a liquid volume capacity of 3 mL were solvated with 1 mL methanol (solvent contacted sorbent for a minimum of 40 s prior to evacuation). Cartridges were prepared for loading by introduction of 1 mL of RB. Extracts were loaded onto cartridges at approximately the loading of 20 mg DW tissue per cartridge with a flow rate equivalent to 0.2 mL / min, suspended in RB. After loading, cartridges were washed (to remove loosely bound and

abundant primary metabolites) using 800 µL of 50/50 (v/v) MeOH/H_2O. Secondary metabolites were eluted in 150 µL MeOH. Eluates were dried by speedvacuum overnight.

5.2.4 Chromatographic Separation

Waters (Part No. 186002353) Acquity UPLC BEH C18 (particle size 1.7 µm) column was used for method development, operating at a flow rate of 0.5 mL / min at 40^{o}C. The analytical column (2.1 x 150 mm) was protected with a Waters VanGuard column of identical diameter and packing material (Part No. 186003975). The work of Jean-Luc Wolfender at the University of Geneva Laboratory of Pharmacognosy is considered to be at the forefront of UPLC-MS based plant metabolomics, and therefore served as a template for early experiments.[47] Detailed studies on flavonoid metabolism in Lamiales (which include popular ornamental flowers and dietary herbs such as snapdragon, basil and spearmint) utilized exclusively LC-MS.[48] Many classes of secondary metabolites in strawberry flowers were differentiated by organ specialization.[49] Information from UV/Visible absorption spectra, MS/MS fragmentation patterns, and comparison between calculated octanol-water partition coefficients (log P) and retention time were used to establish putative assignments.

Due to the low abundance of compounds in a plant extract and our desire for a more targeted approach, amenable to integration for quantitation and loop trapping by LC-NMR for structural elucidation, i.e. fully resolved chromatogram, chromatographic conditions were adjusted from the method of Wolfender (1% per min gradient) until less

than twenty unique signals were detected by UPLC-MS. By UPLC a short ramp gave way to a long shallow ramp (0.82% change in solvent composition per min). When this method was migrated to HPLC, the short ramp was eliminated, as clarified in Table 5.2.

Table 5.2 Comparison between UPLC and HPLC for gradient elution by aqueous (0.1% formic acid in water) and organic mobile phases (B, 0.1% formic acid in acetonitrile).

UPLC Method	Time (min)	%B	%B	Time (min)	HPLC Method
wolfender_082809d	0	1	5	0	kayla_UPLCmigration
	2	1			
	3	6	5	10	
	20	20	20	68	
	25	98	98	86	
	28	98	98	95	
	28.1	1	5	96.5	
	30	1	5	100	

A comparison of chromatograms from these two separations is shown in a previous chapter (Figure 1.11), demonstrating how the efficiency of UPLC (over 3.3 times faster) and the loading capacity of HPLC (greater than 8 times the material per peak) are complementary and both are useful for exploration of natural product space. An alternative approach to overloading an HPLC column on a single injection is to apply post-column SPE trapping by a process called "cumulative loading" wherein multiple chromatographic injections can be stored on a single set of SPE cartridges.[50] Drawbacks to this method are it does not work well for hydrophilic molecules because they are not trapped efficiently in the SPE step. An additional advantage of LC-SPE-NMR is that deuterated solvents can be used for SPE elution and regular (non-deuterated) solvents can be used elsewhere (throughout chromatography and UV detection).

The HPLC separation was performed using Waters Xbridge C18 column (3.5 μm diameter packing material) operating at 1.1 mL / min and 25 ^{o}C (Part No: 186003943). The column dimensions (4.6 x 250 mm) all around exceeded the UPLC column by a factor of 4 in terms of surface area and 8 in terms of volume. A Sentry Xbridge guard column (length 20 mm) of identical packing material was fitted to the inlet of the analytical column (Part No: 186003061). Constraints on the method were imposed by system pressure in the LC-NMR loop storage and transfer lines. Deuterated mobile phases were prepared (Cambridge Isotopes) and utilized throughout the NMR portion of these experiments. To mimic the conditions utilized for UPLC-MS, mobile phase solvents containing 0.1% formate-d in ACN-d_3 (pump C) and D_2O (pump D) were introduced to the column according to Table 5.2.

5.2.5 Loop Storage Triggered via UV Absorbance and On-line LC-NMR

For LC-NMR three modes are possible: on-flow, direct stopped flow, and loop storage. In on-flow mode, NMR spectra are continuously acquired while peaks as the peaks elute. The disadvantage of this mode is that signal averaging can only be applied to improve S/N of transients acquired while the component is in the active volume of the flow probe. In direct stopped flow mode, timing is calibrated such that a detector detects peaks eluting from the column. When a peak is detected, the flow continues until the peak arrives in the NMR cell. Then, chromatography is halted allowing NMR experiments requiring signal averaging to be performed. Once the NMR experiments are completed, the separation resumes until the next peak is found. This process can be

repeated several times within one chromatogram. This method suffers from diffusion of analyte during chromatographic stop-and-go, which can cause peak broadening in the chromatographic dimension. Despite these drawbacks, LC-NMR has been very useful for plant metabolite profiling.[13, 14]

Loop collection and transfer, sometimes referred to as "peak parking," is an established technique in LC-NMR. Loop storage involves fluidic diversion of up to 20 μL of detected peaks without interrupting the separation. Typical peak widths in HPLC exceeded 200 μL, so some sample loss occurs under automation. When the separation is completed, an HPLC pump (Agilent 1100) is used to transfer material from the loops into the NMR probe. The Bruker peak sampling unit (BPSU-36) stores material from up to 36 peaks in a single experiment. Collection can be triggered manually or in an automated mode by UV absorbance. Deuterated solvents are used to push material from each loop into the NMR flow probe, one at a time. Nondestructive measurements can be recorded (1 and 2D NMR), and peaks can be collected after each loop transfer for further interrogation.

The NMR probe used in these experiments was a 3 mm flow cell Bruker LCSSEI (^{1}H-^{13}C) with Z gradients (H8400/0207). The probe was tuned, matched and shimmed using a 5% *tert*-butanol solution in D_2O. Timing was calibrated using a 100 mg/mL solution of acetaminophen.

5.2.6 NMR Acquisition

NMR spectra were recorded using water suppression through enhanced T_1 effects (WET), which can accommodate excitation sculpting for the purpose of suppressing multiple solvent resonances.[17] In this procedure a 90° pulse-acquire experiment is performed to center the transmitter frequency (O1 = 2819.80 Hz) on the residual water resonance around 4.7 ppm. A frequency list is defined including the resonance of acetontrile-d_2 (O2 = 1178.85 Hz) around 2 ppm on a spectrometer operating at 599.84 MHz. A spectral width of 6613.7 Hz was excited (11 ppm) for adequate coverage of ^{1}H resonances of plant secondary metabolites. Another simple 90° pulse-acquire experiment is performed to write these values to the disk. A new experiment employing WET solvent suppression is opened and O1 is set precisely to match the parameters of the 90° pulse-acquire experiment. A sinc function is created using the shape tool and modulated to match O1 precisely again, calling in the defined frequency list. The shaped pulse length should be 10,000 μs and must be saved and called in to the acquisition parameters of the *wet* experiment and recalled for the gradient pulses (SPNAM 7…10). Suppression of residual acetonitrile and water resonances can be optimized in real time by adjusting the power levels of shaped pulses (*sp10* and *sp7*).

5.2.7 ESI-qTOF Mass Measurements

Mass spectra were obtained using an ESI source on a quadrupole time-of-flight instrument (Waters Corporation, Millford, MA). A calibration was performed on both the quadrupole (q) and the time-of-flight (TOF) independently using sodium formate

solution. Mass resolution between 5000-6000 and mass accuracy between 5-600 ppm was obtained throughout the course of plant metabolomic experiments, checked against the peptide leucine enkephalin in both positive and negative modes. Even after careful calibration before and between runs, mass accuracy of no less than 200 ppm was typical. Mass accuracy of <5 ppm is routine when a lockspray device is employed. This mode of operation was not employed in this dataset to optimize discovery capability. Implementation of lockspray reduces effective sampling rate, generating fewer mass spectra per chromatographic time unit. Running a blank every 3 runs was helpful. Isotopic envelopes were used to distinguish between plant-derived molecules and system peaks / noise.

Peaks collected after LC-UV-BPSU-36-NMR were dried in a speedvacuum. Collected fractions were reconstituted in 50/50 MeOH/H_2O for direct infusion electrospray ionization (ESI) time-of-flight (TOF) mass spectrometry (MS). Mass spectra were recorded for each collected peak in both positive and negative modes.

5.2.8 MS/MS Fragmentation

Intact mixtures were concentrated and fractionated by UPLC for tandem mass spectrometry (MS/MS). Experiments were performed using Ar as a collision gas accelerated to a range of energies: 10, 15 and 30 eV. Fragmentations observed were used to support structural hypothesis anchored by 1D ^{1}H NMR data. The source temperature was maintained at 120 ^{o}C. Selected tuning paramaters are listed as follows: capillary 3000 V, sample cone 25 V, extractor cone 1 V.

5.2.9 Dereplication and Databasing

Major secondary metabolites, such as phenylpropanoids and glucosinolates in Arabidopsis, are readily detected in extracts.[51-53] Identity is established by comparison with libraries of reference compounds, or by two-dimensional NMR. Proton NMR chemical shift (δ) information was difficult to obtain directly from LC-BPSU peak trapping-NMR due to a lack of chemical shift reference compound in the samples. Referencing to solvents and residual formic acid (non-deuterated) gave approximate δ (ppm) values. Unlike metabolomic analyses, a major effort is needed for unequivocal identification of metabolites.[54] Pattern recognition based on an experienced natural product spectroscopist's eye (Szabolcs Beni) was ultimately what provided the structural foothold necessary to recognize the flavonoid ring signature in the aromatic region of the spectrum. The hydrophilicity of the molecules and a set of resonances in the sugar region of the spectrum together pointed toward glycosylation.

Restricting database searching by mass was not particularly helpful in establishing molecular identities because of the diversity of structural isomers generated by plant secondary metabolism. MS/MS fragmentation patterns were somewhat more informative since hexose (-164 Da) and deoxyhexose (or methyl-pentose, -147 Da) losses could be clearly observed. Also sinapic acid daughter ions (223 Da) were observed.

ACD/Labs software suite including ChemSketch, HNMR Predictor, CNMR Predictor, Processor and DB were utilized in the management and cross-referencing of data from these experiments. Molecular structures isolated from Arabidopsis and other angiosperms reported in the literature were drawn in ChemSketch. Corresponding

references were entered into a database along with descriptors such as class (flavone, flavonol, isoflavone, hydroxycinnamic acid, glucosinolate, etc.), common name of the molecule (if such exists), localization *in planta*, and the tissue type in which the metabolite was found (if reported). Not entered into the database were: retention time, UV-Visible absorption spectra, chromatograms, mass spectra, tandem mass spectra. ACD/Labs software was found to be particularly well suited to NMR-based information. Filtering of the database was permitted by keyword, category, or a range of chemical shift or mass values which were useful in sorting and filtering information throughout the dereplication process.

5.2.10 Simulation and Standards

ACD/Labs was used to predict ^{1}H chemical shift information (δ, ppm), coupling constants (J, Hz) and log P for all molecules entered into the database. Exact masses ($[M-H]^-$, Da) of molecular ions were calculated. Calculated and predicted values, together with support from the literature, were used to make preliminary assignments. One compound (kaempferol-3,7-*O*-dirhamnoside, K37R, common name Kaempferitrin) was available as a pure standard. It was ordered from ChromaDex (Irvine, CA). Agreement was observed between the standard and the endogenous metabolite we observed in extracts of *Arabidopsis thaliana* in terms of ESI-MS spectral features, ^{1}H NMR predictions and measurements, retention time on our UPLC column and ratio between dual channel UV absorbance (254 and 280 nm). Carbon chemical shift information and number of attached protons was obtained through a multiplicity-edited

HSQC experiment (me-HSQC), wherein phase of 2D cross peaks indicates methyl/methyne (positive phase) or methylene (negative phase) protons.

Once this assignment had been confirmed against a high-purity standard, the assignments of the remaining signals were established in a similar manner, although without the benefit of pure standards for confirmation of structure.

5.3 Results

Enrichment by SPE conferred specificity towards a particular type of molecule. Resonances attributed to secondary metabolites rise above the noise as a greater amount of biological material is resuspended in a constant small volume (400 μL, accommodated into our standard 5 mm NMR probe via Shigemi tube). These one-dimensional spectra were acquired on 7-d-old seedling extracts. It was determined that spectral overlap in this sample was prohibitive for facilitating structure elucidation of mixture components. Two-dimensional homonuclear experiments (COSY, TOCSY) were performed on this sample, as well as diffusion-ordered spectroscopy (DOSY), presented previously in Chapter 1 Figures 1.7-1.9.

Expansions of the 2D NMR spectra reveal the difficulties in making assignments for aromatic compounds. The COSY and TOCSY spectra are not useful. Even HSQC provides limited information. The best way to make connections from the ring to the rest of the molecule is HMBC and that is also complicated in a complex mixture, motivating the need to separate the components for identification. A less complex spectrum is

generated by HSQC, wherein assignment can be facilitated by an authentic standard, as attempted in Figure 5.2.

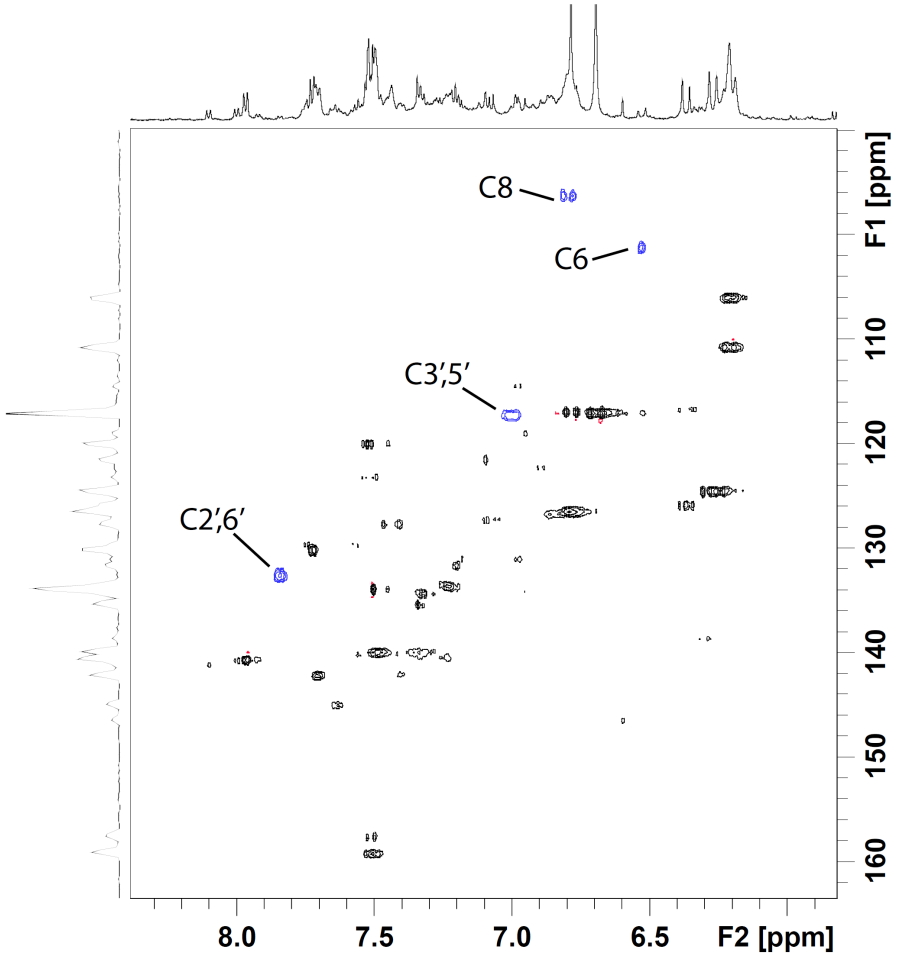


Figure 5.2 HSQC (hsqcetgpsi2) spectrum of SPE-enriched ~1.8 g DW *Arabidopsis thaliana* (Col-0) seedling (7 d-old) tissue shown in black/red. The HSQC of K37R is overlaid in blue with its assigned resonances labeled. The HSQC spectra were acquired with a 1000 µs dephasing gradient to spoil first-order coherence and a J[H-C] is set to 158 Hz.

The signals in the HSQC are difficult to divide into individual molecules, however HMBC can facilitate the grouping of resonances into connected spin systems. Unfortunately, while connectivity within sugar rings can be observed, it is still not possible to observe crosspeaks between the methyl protons of the rhamnose sugar substituents and any of the protons on the flavonoid skeleton. The HMBC spectrum

(Figure 5.3) is overlaid with the HSQC of the same sample (seedling extract enriched by SPE) which is also overlaid with the HSQC of a standard of K37R in methanol-d_4.

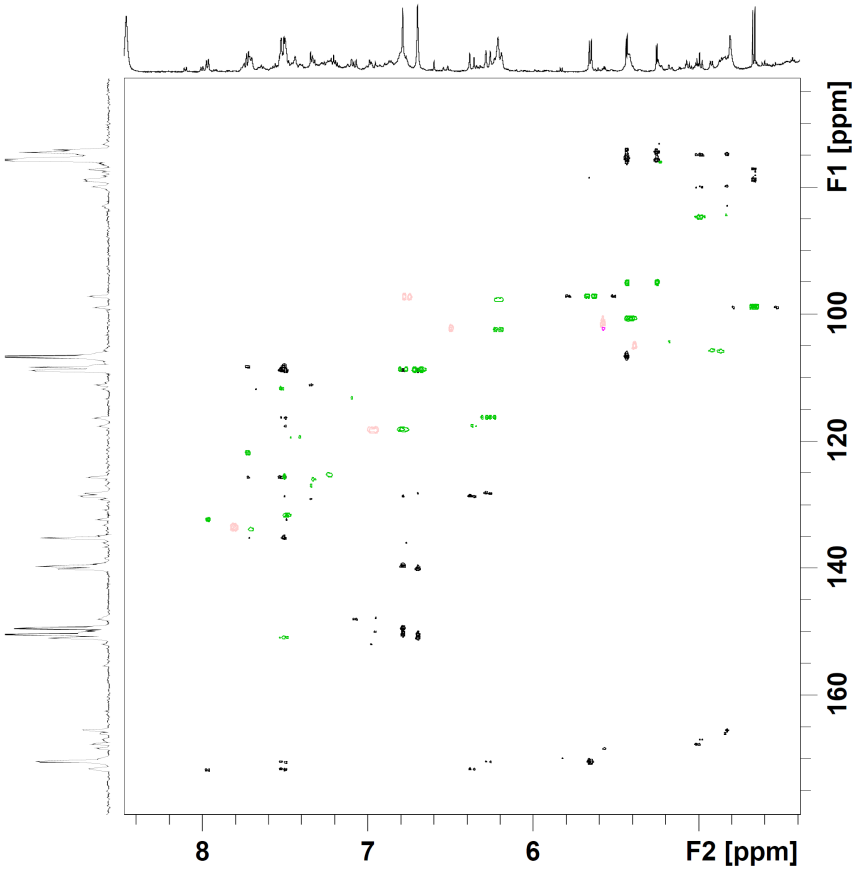


Figure 5.3 HMBC (hmbcgplpndqf) spectrum shown in black and HSQC spectrum shown in green yield complementary information. The HMBC experiment was optimized for visualizing long-range coupling while suppressing one-bond correlations. The HSQC spectrum of K37R standard is overlaid for comparison in pink.

When options for intact mixture analysis were exhausted, a reverse phase chromatographic method development was undertaken by UPLC (see Chapter 1, Figure 1.14). The method was migrated to HPLC using Waters Xbridge technology, specifically designed for UPLC-HPLC method portability. Following a solvent delay for elution of the void, UV absorbance triggered diversion of 120 µL aliquots of eluting peaks into

storage loops. Figure 5.4 presents the entire chromatographic separation monitored by absorbance at 254 nm resulting from injection of 2x diluted superextract (20 μL total injection volume). Our goal was for targeted peaks to be loaded on to the column such that the maximum UV absorbance for the peaks of interest was around 1000 units, but shown here the most abundant peak had a height of only half that, to prevent loss of chromatographic resolution by overloading the column. Each peak is labeled with the loop number in which it was stored and the retention time in min.

On-flow LC-NMR suffered from limitations in dynamic range despite our attempts to suppress solvent resonances (Figure 5.5). Across the x-axis is a one-dimensional ^{1}H NMR spectrum. Along the y-axis is plotted time, with zero at the bottom of the figure and 80 min close to the top. Between 30 and 50 min, some information seems to be present, but the S/N is too poor to support structural conclusions. Even summing transients did not sufficiently improve spectral quality to allow this data to be informative.

In separate experiments, spectra were recorded on collected loop material using a minimum of 512 transients. WET parameters were calibrated for each component on a

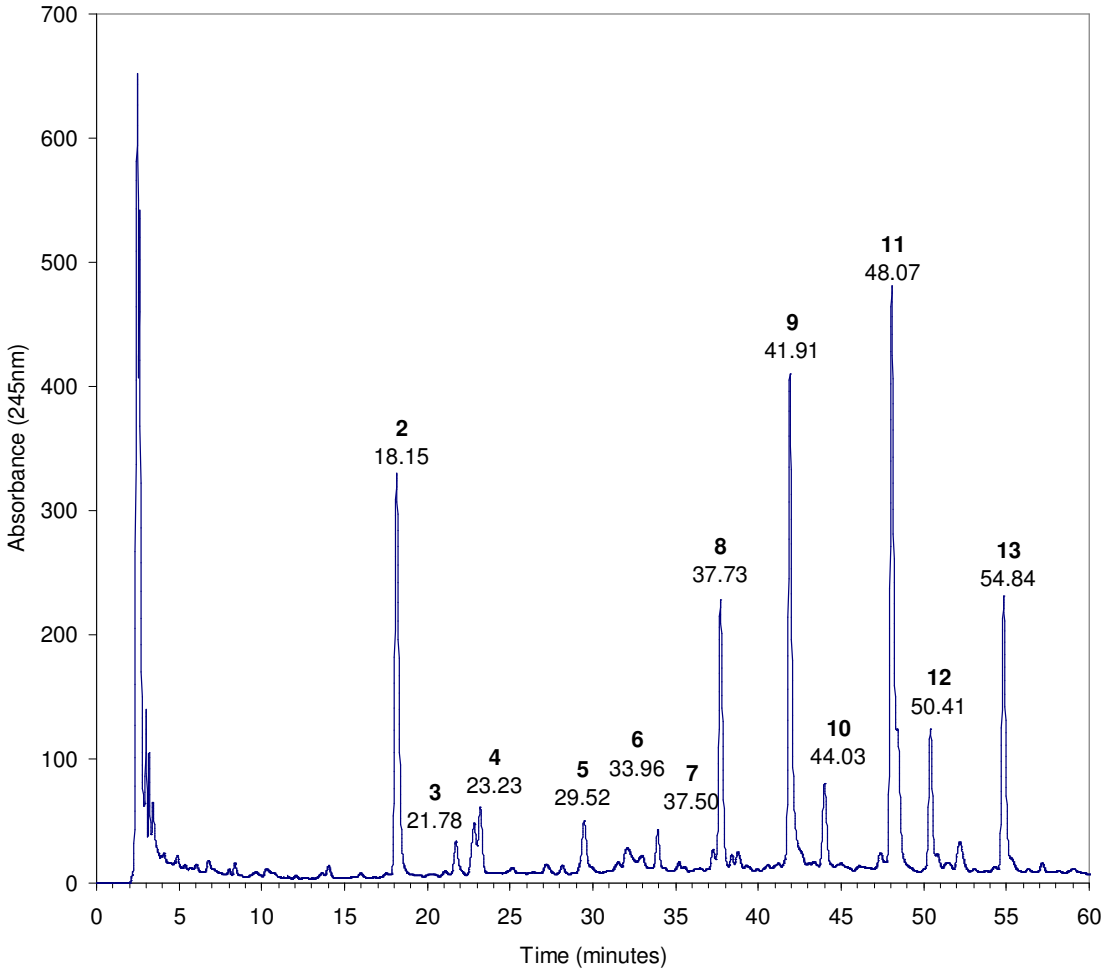


Figure 5.4 LC chromatogram of superextract wherein UV absorbance at 254 nm triggered automated peak trapping of plant natural products. Numbers of the sampled peaks are assigned here with retention time in min listed below each peak number.

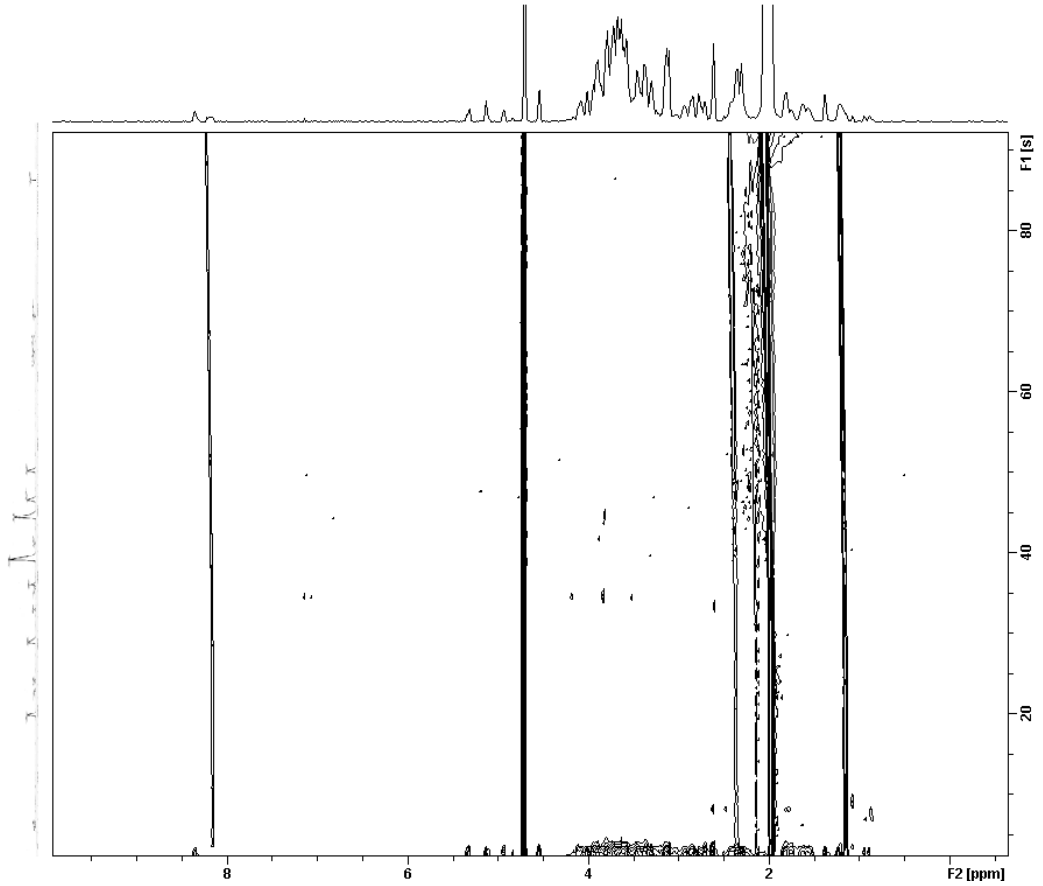


Figure 5.5 On-flow pseudo-2D spectrum in (*lc2wetdc*) using WET solvent suppression of HOD and ACN-d_2 resonances via shaped pulses with ^{13}C decoupling.

case-by-case basis. In natural product dereplication studies involving a single taxa, genus, phyla, or species, it is often most useful to begin elucidation of the most complicated component in the mixture. Other components likely to be built on similar scaffolds sharing common decorations. The most highly retained molecule in the Arabidopsis 7-d-old seedling HLB-enriched extract was stored in loop 13. Its NMR spectrum is shown in Figure 5.6.

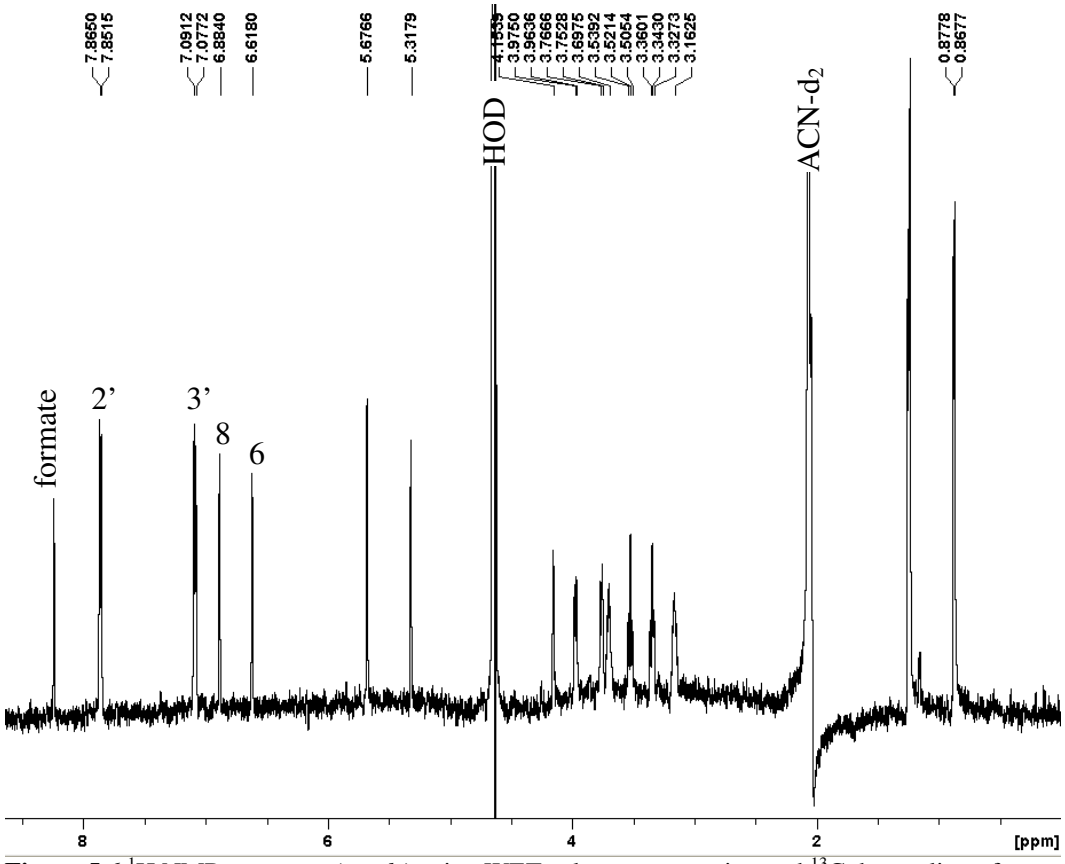


Figure 5.6 ^{1}H NMR spectrum (*wetdc*) using WET solvent suppression and ^{13}C decoupling of a natural product isolated from 7-d-old seedlings of *Arabidopsis thaliana*. Chemical shifts listed above resonances have been reported referenced to acetonitrile-d_2 in water. This NMR spectrum corresponds to peak 13 in the chromatogram shown in Figure 5.4 which elutes at 54.84 min. Assignments correspond with numbering scheme in Figure 5.1.

The most deshielded protons in this molecule can be found on the C ring. Resonances contributed by protons on C2' and C6' are found at the most downfield region near 8 ppm. The next most deshielded protons, bound to C3' and C5', resonate closer to 7 ppm and are first-order coupled to C2' and C6', respectively. This results in two doublets of equivalent intensity, integrating to two protons each. The protons on the A ring at positions C8 and C6 contribute singlets between 6-7 ppm since they represent isolated spin systems. The proton on C8 is more deshielded than the proton on C6.[38, 55, 56]

An important point of diversification of the flavonol scaffold is at the kaempferol (K) C3'position: hydroxylation yields quercetin (Q) while hydroxymethylation yields isorhamnetin (I). Common glycan substituents of flavonols include: glucose, rhamnose, galactose and xylose. These are found both as monosaccharide substituents and as larger oligosaccharides bound at the 3-*O* or 7-*O* positions. Hydroxycinnamic acids substituents can occur atthe 3'-*O* and 4'-*O* positions of the kaempferol core. Malonate, isoprene, and hydroxycinnamic acids can decorate the hydroxyl groups of attached sugars, adding complexity and diversity. Flavonol 3,7-di-*O*-glycosides more readily lose a glycan at the 3 position than the 7 position in positive mode MS,[57] and vice versa in negative mode, so MS/MS is particularly useful when the glycan substituents are of different masses.[58] Proton NMR chemical shifts are also structurally informative.[59]

Combining a structural foothold on a flavonoid scaffold together with UPLC-MS data, comparisons were made with what had been reported in the literature according to age of tissue, localization of the sample, and specification defined by the taxonomical constraints. These were compared with UPLC-MS-based metabolic profiles of extracts of different organs and developmental stages, assisting in logical assignments for compounds isolated from peaks 8, 9, 11 and 13 of Figure 5.2. We propose that the identity peak 13 is kaempferol-3,7-*O*-dirhamnoside, K37R, which has been reported in Arabidopsis. Numbering of glycan substituents and nomenclature for fragmentation is shown in Figure 5.7.

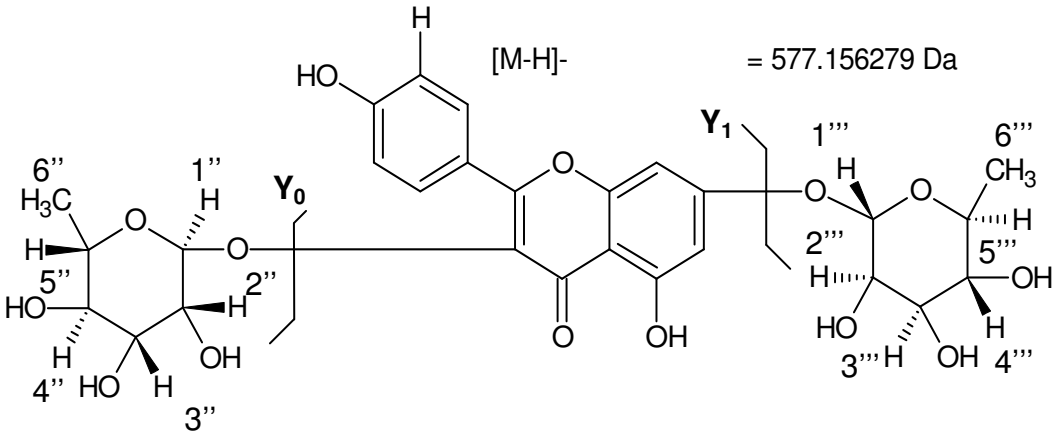


Figure 5.7 Structure of kaempferol-3,7-*O*-dirhamnoside, K37R, common name Kaempferitrin and exact mass of molecular ion predicted by ACD/Labs. Fragmentation patterns are labeled. Initial loss of the glycone at the C3 position leaving the charge on the singly substituted flavonol glycoside (Y_0) is followed by a loss of the glycone at the C7 position, leaving the charge on the aglycone (Y_1). Loss of a terminal rhamnose is recognizable by mass difference of 147 Da. Numbering of the glycones is shown, extending the numbering scheme to include double and triple prime sites. The protons on C6'' and C6''' are characteristic of rhamnose substituents.

The proposed identification of peak 13 was supported by collision-induced dissociation (CID) tandem mass spectrometry, shown in Figure 5.8. In source fragmentation is sufficient to observe loss of the C3 *O*-linked rhamnose (Y_0). As greater collision energy is applied, the molecular ion ($[M-H]^-$) is no longer the dominant peak. The aglycone (Y_1) is observed via loss of the C7 *O*-linked rhamnose. Loss of 147 Da is observed indicating terminal rhamnose groups.

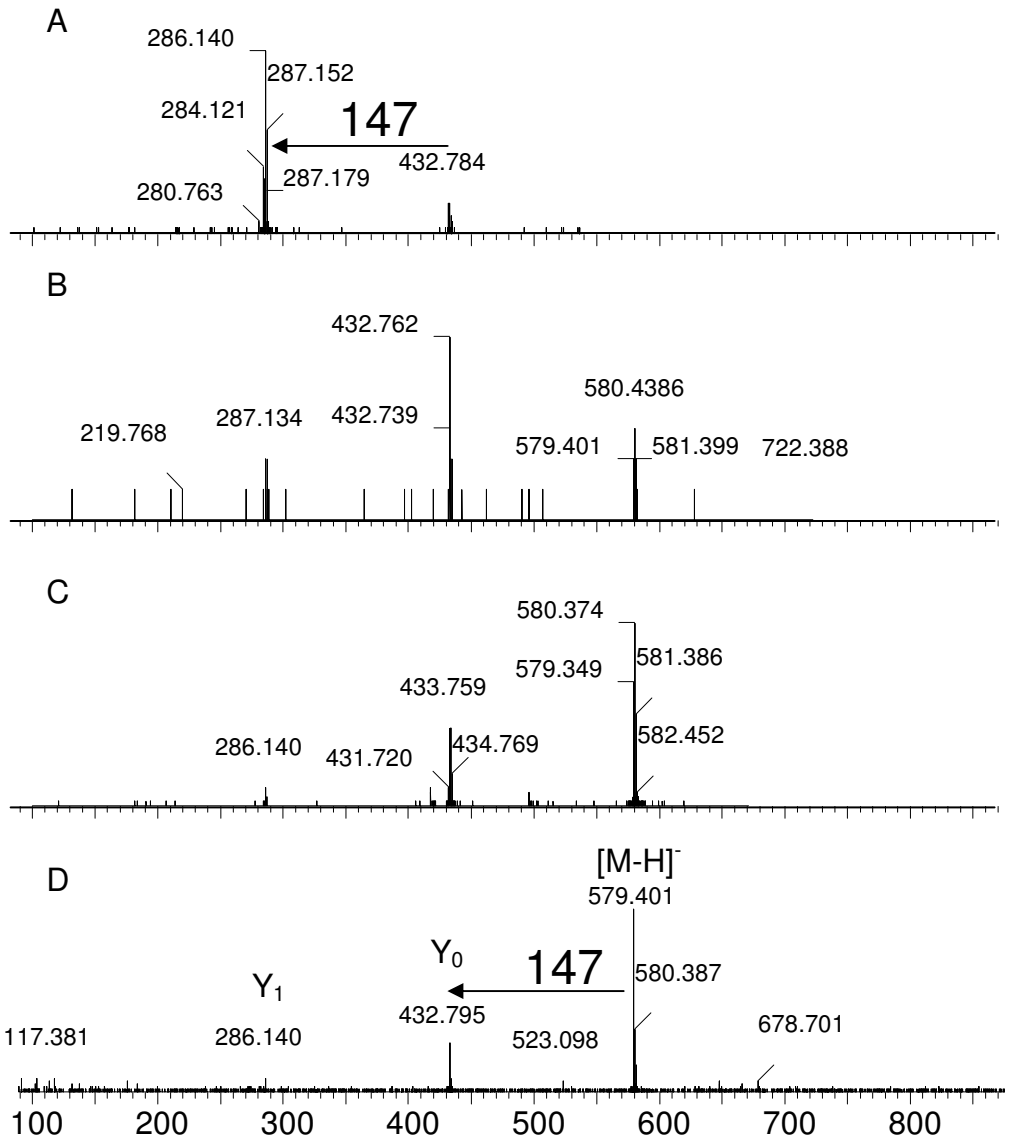


Figure 5.8 MS/MS Fragmentation pattern of UPLC fraction corresponding to peak 13, putatively assigned via HPLC-NMR to the identity K37R, under zero (D), ten (C), fifteen (B) and thirty (A) electron-Volts (eV) of acceleration voltage. In source fragmentation is sufficient to observe loss of the C3 *O*-linked rhamnose (Y_0). As greater collision energy is applied, the molecular ion ($[M-H]^-$) is no longer the dominant peak. The aglycone (Y_1) is observed via loss of the C7 *O*-linked rhamnose. Loss of 147 Da is observed indicating terminal rhamnose groups.

After structural hypotheses were proposed based on literature, preliminary MS, MS/MS and 1D ^{1}H NMR spectra of a pure standard of kaempferol-3,7-*O*-dirhamnoside, K37R, common name Kaempferitrin were obtained in methanol-d_4. When the standard arrived, it was checked to be sure its retention time matched that observed in plant extracts and its purity was evaluated chromatographically. It was injected into the UPLC in 50/50 MeOH/H_2O and spiked into an extract to eliminate chromatographic or spectrometric differences contributed by matrix effects in its retention and ionization/fragmentation behavior.

The NMR sample was prepared in the vial in which it arrived. The me-HSQC experiment was applied to yield chemical shift information in both proton and carbon dimensions. This information was compared with what has been previously reported for this molecule (by NMR and MS) and what ACD/Labs predicted for these values based on a structural hypothesis. A comparison of this type is presented in Table 5.3.

Table 5.3 Comparison between chemical shifts measured in mixed D_2O/ACN-d_3 (LC-NMR) from a plant extract, chemical shifts measured in pure MeOH from a pure analytical standard. Multiplicity, *m*, and coupling constant, *J*, for proton-proton interactions were used for assignment of all resonances observed by HSQC, providing carbon chemical shifts. Predictions by ACD/Labs are also compared with values reported in the literature.

Atom No.	^{1}H chemical shift, δ (ppm)				*m*	J (H,H) (Hz)	^{13}C chemical shift, δ (ppm)		
	LC-NMR	Std in MeOH	Predicted	Literature			Std in MeOH	Predicted	Literature
2' 6'	7.85	7.84	7.96	7.80	d	8.8	132.55	131.96	131.96
3' 5'	7.08	7.00	6.86	7.01	d	8.8	117.3	116.59	116.59
8	6.87	6.8	6.63	6.79	d	2.1	96.33	95.68	95.68
6	6.62	6.53	6.47	6.52	d	2.1	101.29	100.54	100.54
1"	5.32	5.42	5.34	5.32	d	1.9	104.03	103.38	103.38
2"	4.61	4.27	3.86	4.24	dd	1.9/3.5	72.48	71.79	71.79
3"	3.76	3.76	3.78	3.74	dd	3.5/9.5	72.78	72.02	72.02
4"	3.34	3.38	3.52	3.33	t	9.5	74.27	73.11	72.11
5"	3.15	3.36	3.46	3.28	dq	5.9/9.5	73.98	72.09	72.09
6"	0.87	0.97	1.21	0.89	dd	5.9	18.37	17.85	17.85
1'''	5.68	5.61	5.90	5.60	d	1.9	100.32	99.88	99.88
2'''	4.15	4.07	3.83	4.07	dd	1.9/3.5	72.28	73.33	71.57
3'''	3.96	3.88	3.82	3.88	dd	3.5/9.5	72.68	72.03	72.02
4'''	3.35	3.53	3.67	3.51	t	9.5	74.27	73.78	73.52
5'''	3.69	3.64	3.7	3.61	dq	5.9/9.5	71.85	70.78	71.26
6'''	1.24	1.30	1.31	1.24	dd	5.9	18.83	17.45	18.24

Progress in dreplication was assisted by a combined approach involving measurement, prediction, comparison, and validation (when an authentic standard was available). At the time this work was carried out, only Kaempfeterin was available for spiking to establish 100% confidence in assignment of metabolite signal to a precise molecule. Putative assignments based on structural evidence and comparison with literature [2, 25, 34, 47, 49, 53, 56-58, 60-78] yielded Table 5.4. Full spectral information for each peak is presented in Figure 5.9.

Distinguishing features for kaempferol skeletons included a doublet for H2' whereas quercetin (hydroxylated at C3') shows a singlet for H2' around 7.6 ppm. Singlets are observed for H6 and H8 for both flavonoids. The scaffold isorhamnetin (*O*-methylated at C3') was distinguished by its unique mass in combination with the appearance of an *O*-methyl resonance around 4.0 ppm in the NMR spectrum. Likewise, the presence of a sinapate scaffold (a hydroxycinnamic acid with *O*-methylation at C2 and C6) was confirmed by the appearance of *O*-methyl resonances. Other distinguishing features of sinapate are the H7 and H8 doublets between 7.5-8.0 ppm and 6.0-6.7 ppm, respectively. The sinapate resonances are also more dispersed than the doublets of the flavonoids. Choline is distinguished from malate and glucose substituents on sinapate by mass, and also by the appearance of *N*-methyl resonances around 3.7 ppm.

The spectra measured for peaks 4 and 10 did not contain sufficient material for ^{1}H NMR analysis even though they gave rise to an intense peak in the UV. The spectra measured for peaks 3, 6, and 7 suggested that a mixture of components were overlapped in these chromatographic peaks and the complexity of the resulting spectra precluded a confident assignment of the individual components. Although peak 12 consisted of a mixture, the integrated intensities of the resonances allowed us to confidently assign two structures (Table 5.4). Full NMR and MS spectra are provided for each peak in Appendix B.

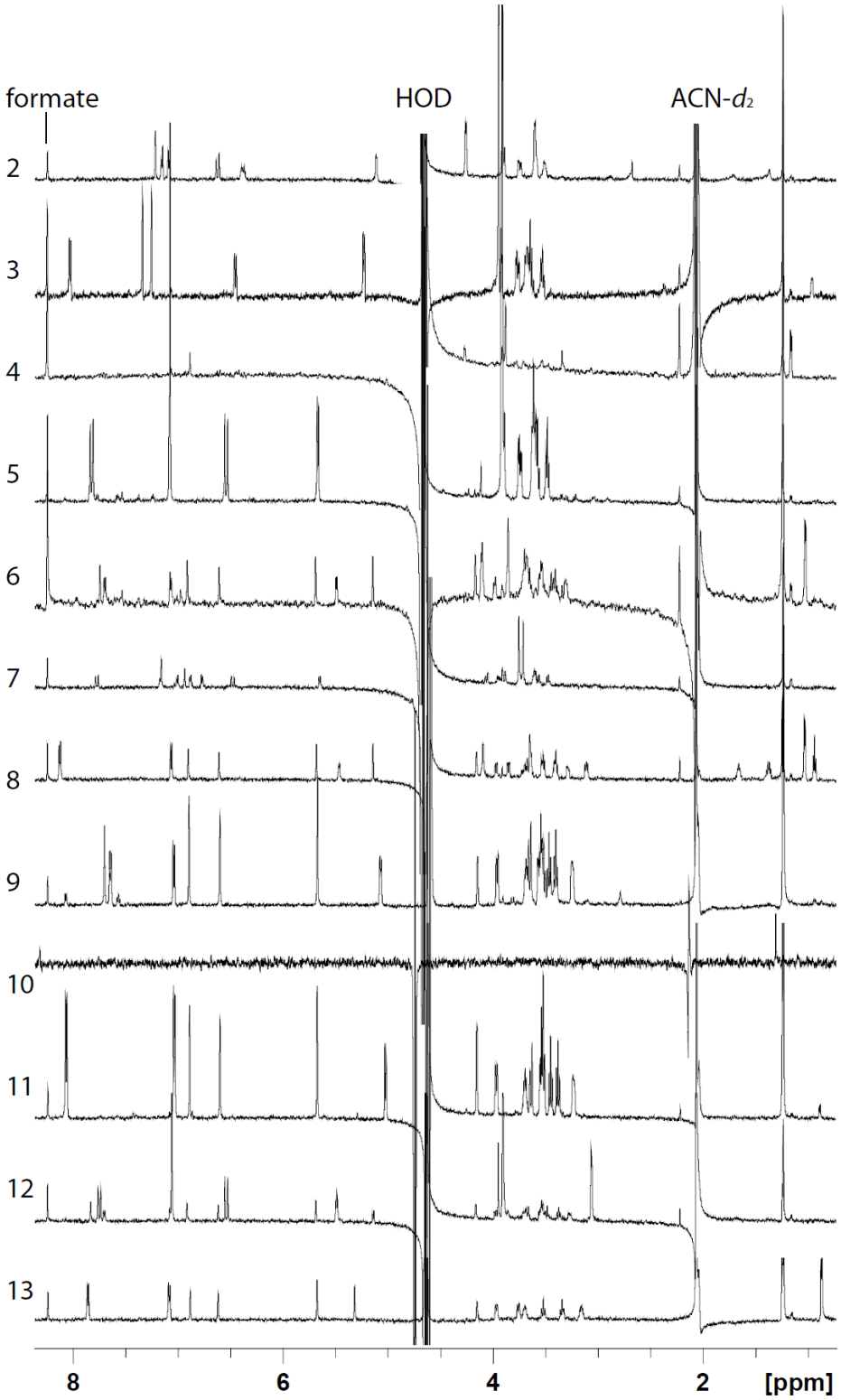


Figure 5.9 ^{1}H NMR spectra of metabolites recorded using automated LC-NMR peak selection using absorbance at 254 nm.

Table 5.4 Putative metabolite identifications according to elution order with corresponding references. An appendix with full spectral and spectrometric data is provided at the end of this dissertation, Appendix B. Retention time (t_R) is in min for the UPLC separation in Figure 5.2. Ion mass-to-charge (*m/z*) ratios are rounded to four significant digits to account for the observed 600 ppm difference in mass in the course of a set of experiments. Abbreviations: Q = quercetin, R = rhamnose, G = glucose, K = kaempferol, I = isorhamnetin, S = sinapoyl, M = malate.

Metab	t_R (min)	*m/z*	LC-NMR	References
Q3RG7R	10.5	755.2	Peak 6	77
K3RG7R	11.5	739.7	Peak 8	77, 79, 80
Q3G7R	13.0	609.6	Peak 9	61, 62, 79
K3G7R	15.0	593.3	Peak 11	77
I3G7R	15.7	623.0	Peak 12	57
K37R	17.2	577.3	Peak 13	59, 79
SG	8-9	385.1	Peak 2	57, 76
SM	15.7	339.1	Peak 12	81

5.4 Discussion

In Chapters 3 and 4, the metabonomic approach was employed successfully using 1D ^{1}H NMR spectroscopy of intact mixtures. Limited sample preparation allowed relative quantitation of metabolites affected by hypoxia stress in seedling extracts of *Arabidopsis thaliana*. The advantages of ^{1}H NMR are clear; it is nondestructive, quantitative, and nonselective. However sensitivity and mixture complexity were limiting factors for interrogation of secondary metabolites generally present in lower abundance in plant tissues.

5.4.1 Coverage of Chemical Space by SPE Capture

Changing SPE sorbent may be an effective way to diversify coverage of the secondary metabolome, as shown in Figure 5.10. Corresponding structures for diverse

sorbents are shown in Figure 5.11. This approach is analgous to the work of Carlson et al.[82] who designed a suite of sorbent materials for capturing regions of chemical space in mammalian-based metabolic profiling. The HLB stationary phase is more desirable because it does not have the wetting requirements which accompany C18 phases. This gives the operator more freedom in experimental design for maximum recovery of target analytes. In a follow-up study (data not shown) the secondary metabolite profiles of *Xiphidium caeruleum* root exudates recorded using LC-DAD were indistinguishable regardless of whether C18 or HLB was used, when following manufacturers recommended guidelines for washing and elution of the C18 phase. With our study, we took a widely-targeted approach, meaning that we hoped to capture an array of secondary metabolites while excluding interfering ions and lipds/waxes which tend to cause a loss of instrument performance and complicate the analysis and a uniform approach for washing and elution were applied to all stationary phases in Figures 5.10 and 5.11.

Although not firmly established, we believe our current method has the potential to detect glucosinolates.[83] Since not all glucosinolates contain a moiety which absorbs at 254 nm, many of these molecules although observable in our ESI-MS studies were not automatically trapped by the BPSU due to a lack of UV absorbance signal at that wavelength. Therefore of the secondary metabolites synthesized by Arabidopsis, we have covered all classes excluding alkaloids, which are more hydrophobic and if present in this organism we have biased our methods against their detection by focusing on hydrophilic metabolites only.

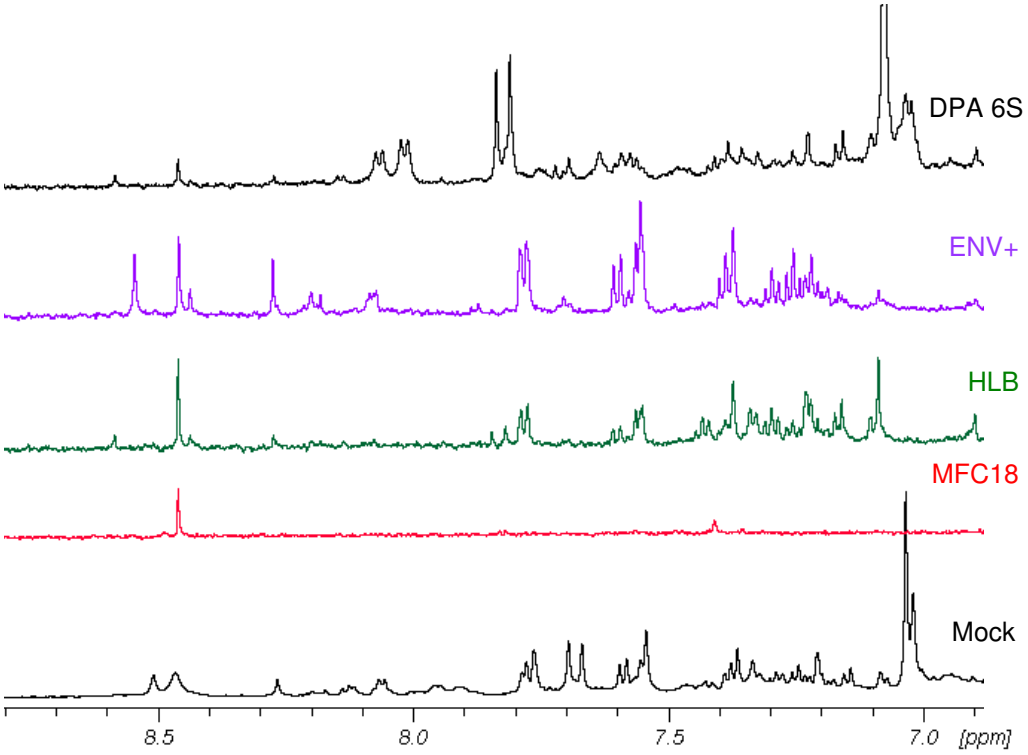


Figure 5.10 Aromatic region of ^{1}H NMR spectra recorded on *Arabidopsis* extracts enriched by solid-phase extraction by discovery polyamide (DPA 6S), crosslinked polystyrene (ENV+), hydrophilic-lipophilic balanced copolymer (HLB), and monofunctional C18 (MFC18) showing selectivity for different components based on differences in structure. Sorbent structures are presented in Figure 5.11 for comparison. A non-SPE treated sample is presented for comparison (Mock).

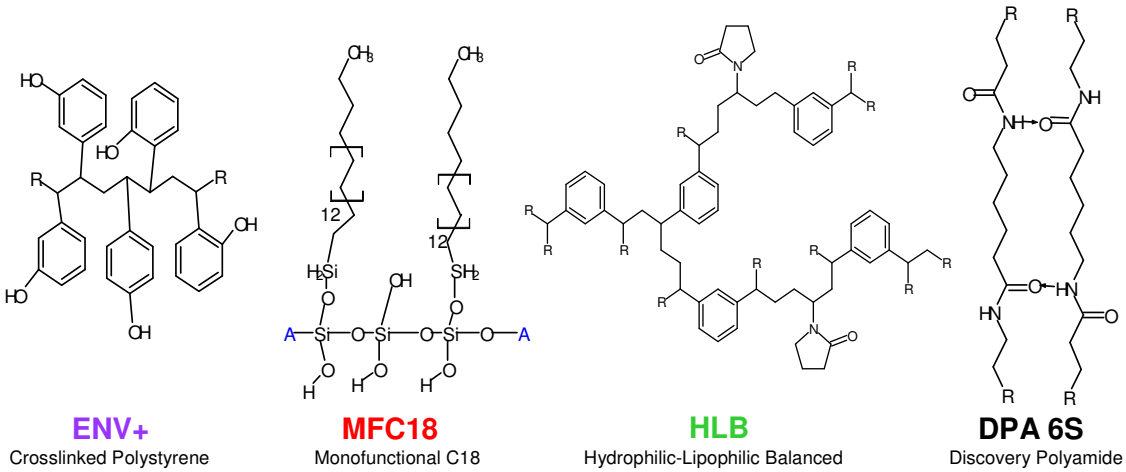


Figure 5.11 Solid-phase extraction (SPE) sorbent polymeric structures. A portion of each is shown for comparison of structural similarities and differences.

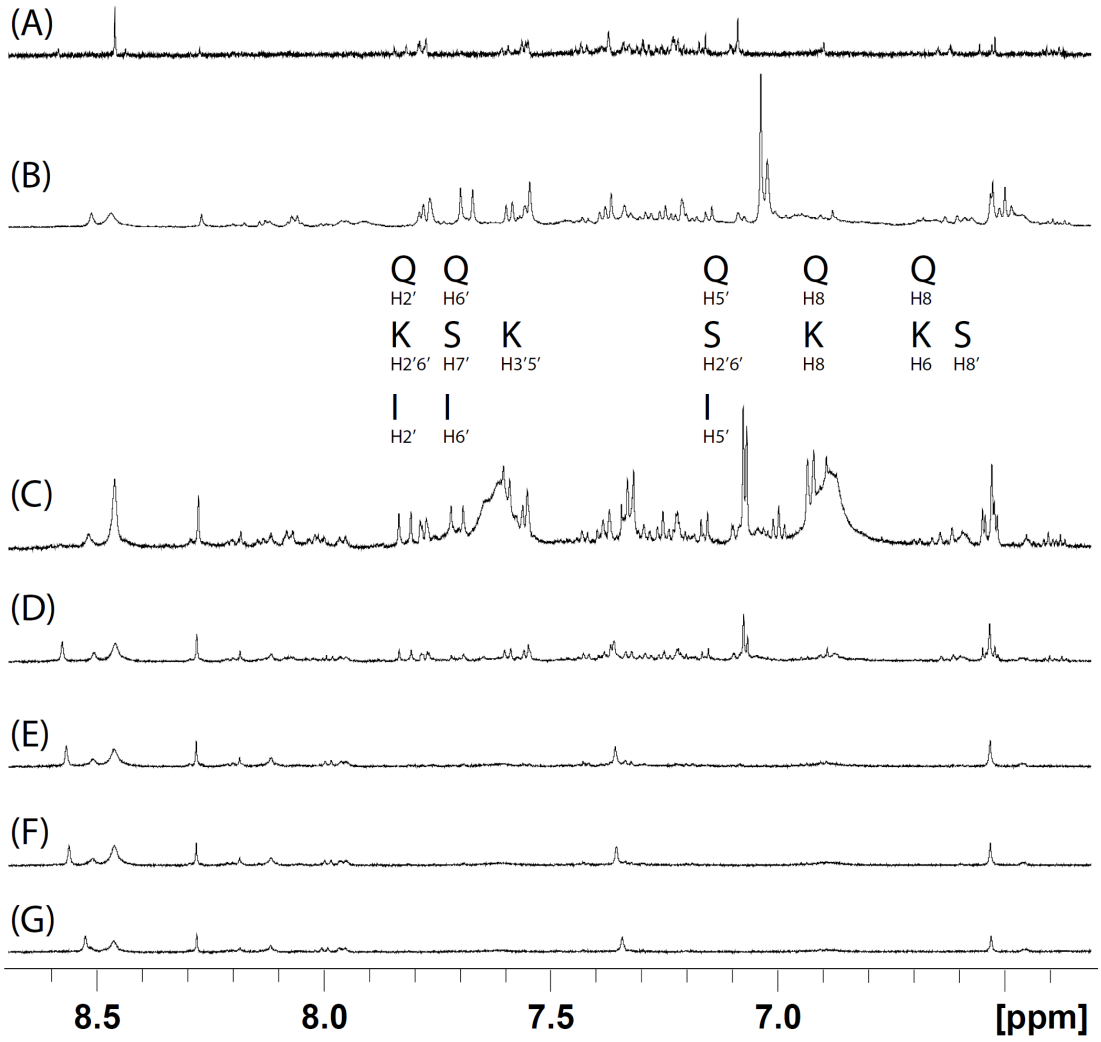


Figure 5.12 Arabidopsis seedling extract (A) without SPE-enrichment is compared with the 1D ^{1}H NMR spectral quality obtained when using HLB for secondary metabolite enrichment (B). Optimization of elution conditions: 100% MeOH (C), 80% MeOH (D), 50% MeOH (E), 20% MeOH (F), and 0% MeOH in H_2O (G) indicated that a strong solvent is necessary to elute the metabolites of interest from the cartridge. Regions occupied by resonances of Arabidopsis secondary metabolites kaempferol (K), quercetin (Q), isorhamnetin (I), and sinapate (S) are indicated, although chemical shifts will vary slightly depending on substitution patterns.

5.4.2 Method Portability using Waters BEH C18 Chromatography Columns

Good agreement was observed between retention behavior of molecules introduced into the C18 column for HPLC or UPLC. The use of deuterated mobile phase

solvents did not appreciably affect retention behavior or elution order. By correlating peak height in the UV-detected chromatograms using absorbance measured at 254 nm, it was facile to relate peaks from HPLC back to the UPLC-MS separation. Because LC-NMR is non-destructive, reclaimed peaks could be introduced to the direct infusion port of the ESI-MS for confirmation of mass-to-charge ratios recorded for peaks in a UPLC separation, which was useful in these experiments. This also ensured that compound degradation did not occur during loop storage in the lengthy LC-NMR analyses.

5.4.3 UV Absorbance

In this study, UV-Vis was not particularly useful except for triggering peak trapping. Had it been available on the chromatographic systems used for these experiments, diode-array detection would have offered an advantage for this study. Diode array detection is particularly useful in plant metabolomics for identification of groups of chemically/structurally related molecules in a complex mixture.[18] The wavelength 254 nm was selected in this work for its common use in industrial settings, considered a "catch all" for phenolics and organic molecules containing conjugated ring systems. For UPLC-MS, a dual channel wavelength detector was available, therefore ratios of absorbance could be calculated, but this feature was underutilized for dereplication.

5.4.4 Suitability of LC-NMR for Exploratory Analysis

LC-NMR has great potential to facilitate plant metabolomic studies.[12] In our lab, it was an intergral part of the pipeline that allowed us to secure a sturdy foothold with which to tackle the literature and catalog searches that ultimately led us to confident assignment of one secondary metabolite, and to establish putative assignments for 7 more. In this work only one-dimensional proton spectra were acquired on each loop, which was time-consuming in itself, due to the small amount of material present even from a pooled superextract. Sensitivity being a clear issue in natural product studies by NMR, typically a cryoprobe is used, but this was not available for this work. Two-dimensional (multiplicity-edited HSQC and HMBC) spectra of loop material would have been useful to obtain proton and carbon chemical shift information and long-range couplings (via heteronuclear multiple bond correlation, HMBC) that are used to establish sugar-scaffold linkages and pin down de novo structural assignments in the absence of authentic standards. An alternative approach is proposed by Wolfender et al. using microcoil NMR spectroscopy for this purpose.[68]

5.4.5 Role of Mass Spectrometry in Structural Analysis

Mass spectrometry for structural analysis in plant secondary metabolism can be misleading, given the combinatorial nature of these molecules. Fragmentation patterns can only establish the building blocks that were used to assemble the molecule, not necessarily the way in which the parts were connected, which is often the most critical aspect for biological activity. The mass-to-charge ratios of molecular ions in negative

mode were used to search against open-source repositories of information such as the METLIN database and the Madison Metabolomics Database Consortium. Literature accessed through Web of Science provided known molecules and there seem to be an abundance of publications using ESI-MS for flavonoid analysis.[84]

Tandem mass spectrometry (MS/MS) was useful to confirm structural hypotheses, but without a foothold gained by NMR, the information was difficult to interpret directly.[84] Losses of 146 and 163 Da indicated the presence of non-terminal rhamnose and glucose, respectively. Terminal sugars were visualized as mass losses 1 Da heavier. Observation of a fragment at 223 Da indicated the presence of sinapic acid. The mass-to-charge ratio observed for the aglycone upon application of 30 eV collision energy to the parent ion pointed toward identities of flavonoid scaffolds, but precise positioning of hydroxyl and methyl substitution patterns were not observed via cross ring collisions, as proposed by others.[58]

5.4.6 On-line and Off-line Measurements: Automating Analyses

Performing a metabolomic exploration on-line is a smart approach since secondary metabolite abundances appear to be tightly regulated *in planta*. Many of these molecules become unstable outside the vacuolar compartments in which they are maintained while part of the living system. Modifications can occur as a result of exposure to solvent (acetylation is common, methylation less so, acyl migration has been reported for acylation modifications along a sugar substituent in MeOH), UV-visible light, oxygen in the atmosphere, etc. Maintaining the molecules inside the on-line

analytical system ensures that what one is measuring is not an artifact but is biologically relevant and identical to the molecule as it exists *in planta*.

Another advantage to automation is reproducibility. The SPE protocol described here is labor-intensive and time-consuming. The use of a BioMek robot may have improved these parts of the method. A dispersive SPE would have perhaps been easier to automate than SPE cartridges. The recovery obtained using the SPE in the manner described here for K37R standard was 17% $\pm$ 2%. The maximum recovery observed for K37R standard in a buffer using the HLB system were 88% $\pm$ 5% using 5 replicates, optimized iteratively by varying volume and identity of solvents used for reconstitution, dilution, conditioning, washing, and elution.

5.4.7 Data Management for Dereplication

Managing data from a variety of instrument platforms was a real challenge. Software packages to support databasing were either picture- or number-oriented, either MS- or NMR-based, either too stripped-down or too cumbersome to accommodate the functionality we were seeking to facilitate the dreplication process. In the end, data management and integration was accomplished using Powerpoint, which could display pictures, text and tables. Text searching was an option, although searching by a range of values was not possible. Final assignments were made at our data processing workbench, which supports ACD/Labs, Waters Agilent MassLynx, Bruker Topspin, and Microsoft Office, with access to the internet/databases/literature and 4 large screens. Dynamic switching and realtime browsing of raw data, references, and simulated data were critical

for putative assignments to be established. Often in the absence of pure standards for verification of chemical identity, chemical synthesis provides structural insight.[71]

5.5 Conclusions

By combining liquid chromatography with UV detection and mass spectrometry, a large amount of preliminary information can be obtained about the structures contained in an extract. Molecular mass, number of methoxyl and hydroxyl groups, number of sugars, their sequence and certain elementary information about the substitution pattern on specific scaffold molecules can be established with little sample required.[85] Other spectroscopies such as IR or NMR can be added to the pipeline once molecules have been targeted for isolation. Compilation of a spectral library is recommended for data management and to facilitate the pattern recognition that often is the crux of natural product dereplication and chemical space explorations.

So-called "top down" systems biological approaches begin by chunking a system into complex pieces and further subdividing the pieces into smaller parts. For proteomics, this means an intact protein is studied in the MS. In "forward" genetics a large-scale mutagenesis (on several million seeds) is carried out, or in the case of chemical genetics a large library of molecules (200,000 – 600,000 compounds) are screened, in the search for genes or gene products that have a particular function. Map-based cloning or target identification, respectively, are used to restrict the problem, to ultimately arrive at knowledge of the function of one gene.

"Bottom up" approaches start with basic elements first, following a seed model where complexity is built up upon a foundation of small known elements. For proteomics, this means proteins are digested into peptides, which are characterized and reassembled to provide a model of the original protein. Reverse genetics searches for an observable (measurable) phenotype. Directed mutagenesis or gene silencing is applied to a specific gene, and a corresponding phenotype is observed to facilitate determination of gene function. Thus, starting from one or a few known smaller elements, larger systems can be understood.

Analogously to genomics or proteomics, metabolomics can be approached either forward/reverse or top-down/bottom-up. If nothing is known about the system *a priori* and comparative metabonomics is applied, the forward/"top down" approach is emulated, as was the case for Chapters 3. If targeted metabolite profiling examines in detail one pathway or related family of metabolites, this is more analogous to the reverse/"bottom up" approach, which was more like Chapter 4 (by NMR) and will be the thrust of Chapter 6 (by LC-MS). Working entirely in MS would not have been possible without support by these structural assignments made using hyphenated LC-NMR and LC-MS/MS. In this chapter, broad coverage of secondary metabolic biosynthetic pathways in *Arabidopsis* was mapped.

This work was extended to more fully elucidate the nature of low oxygen stress in a model plant *Arabidopsis thaliana,* a workhorse for the new frontiers in systems biology. Quantitative interrogation of secondary metabolic pathways became the focus. This material is the basis for Chapter 6.

5.6 References

1. Wang, Y., Needs for new plant-derived pharmaceuticals in the post-genome era: an industrial view in drug research and development. *Phytochem. Rev.* **2008,** 7, 395-406.

2. Dixon, R. A.; Achnine, L.; Deavours, B. E.; Naoumkina, M., Metabolomics and gene identification in plant natural product pathways. In *Plant Metabolomics*, Saito, K.; Dixon, R. A.; Willmitzer, L., Eds. Springer-Verlag: New York, NY, 2006; Vol. 57, pp 243-259.

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

4. Hartmann, T.; Kutchan, T. M.; Strack, D., Evolution of metabolic diversity. *Phytochemistry* **2005,** 66, 1198-1199.

5. Fiehn, O.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Lee, D. Y.; Lu, Y.; Moon, S.; Nikolau, B., Quality control for plant metabolomics: reporting MSI-compliant studies. *The Plant Journal* **2008,** 53, 691-704.

6. Berenbaum, M. R., Chemical Ecology. In *Chemical Ecology*, Meinwald, E. a., Ed. National Academy Press: Washington, DC, 1995; pp 1-16.

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

8. Hananda, K.; Sawada, Y.; Kuromori, T.; Klausnitzer, R.; Saito, K.; Toyoda, T.; Shinozaki, K.; Li, W.-H.; Hirai, M. Y., Functional Compensation of Primary and Secondary Metabolites by Duplicate Genes in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **2011,** 28, (1), 377-382.

9. Dixon, R. A.; Achnine, L.; Deavours, B. E.; Naoumkina, M., Metabolomics and Gene Identification in Plant Natural Product Pathways. In *Biotechnology in Agriculture and Forestry*, Saito, K.; Dixon, R. A.; Willmitzer, L., Eds. Springer-Verlag: Berlin, Germany, 2006; Vol. 57, pp 243-258.

10. Sneader, W., *Drug Discovery: the evolution of modern medicines*. John Wiley & Sons: New York, NY, 1985; p 435.

11. Seger, C.; Sturm, S., Analytical aspects of plant metabolite profiling platforms: current standings and future aims. *J. Proteome Res.* **2007,** 6, 480-497.

12. Corcoran, O.; Spraul, M., LC-NMR-MS in drug discovery. *Drug Discovery Today* **2003,** 8, (14), 624-631.

13. Iwasa, K.; Cui, W.; Sugiura, M.; Takeuchi, A.; Moriyasu, M.; Takeda, K., Structural analyses of metabolites of phenolic 1-Benzyltetrahydroisoquinolines in plant cell cultures by LC/NMR, LC/MS, and LC/CD. *J. Nat. Prod.* **2005,** 68, 992-1000.

14. Rijke, E. d.; Kanter, F. d.; Ariese, F.; Brinkman, U. A. T.; Gooijer, C., Liquid chromatography coupled to nuclear magnetic resonance spectroscopy for the identification of isoflavone glucoside malonates in *T. pratense* L. leaves. *J. Sep. Sci.* **2004,** 27, 1061-1070.

15. Spraul, M.; Hofmann, M.; Dvortsak, P.; Nicholson, J. K.; Wilson, I. D., Liquid chromatography coupled with high-field proton NMR for profiling human urine for endogenous compounds and drug metabolites. *J. Pharm. Biomed.* **1992,** 10, (8), 601-605.

16. Exarchou, V.; Godejohann, M.; Beek, T. A. v.; Gerothanassis, I. P.; Vervoort, J., LC-UV-solid-phase extraction-NMR-MS combined with a cryogenic flow probe and its application to the identification of compounds present in Greek oregano. *Anal. Chem.* **2003,** 75, 6288-6294.

17. Smallcombe, S. H.; Patt, S. L.; Keifer, P. A., WET solvent suppression and its applications to LC NMR and high-resolution NMR spectroscopy. *J. Magn. Reson.* **1995,** 117, 295-303.

18. Wolfender, J.-L.; Queiroz, E. F.; Hostettmann, K., Phytochemistry in the microgram domain - a LC-NMR Perspective. *Magn. Reson. Chem.* **2005,** 43, 697-709.

19. Albert, K., *On-line LC-NMR and Related Techniques*. John Wiley & Sons, Inc.: West Sussex, England, 2002; p p 45-178.

20. Wolfender, J.-L.; Marti, G.; Queiroz, E. F., Advances in techniques for profiling crude extracts and for the rapid identification of natural products: dereplication, quality control and metabolomics. *Curr. Org. Chem.* **2010,** 14, 1808-1832.

21. Austin, M. B.; O'Maille, P. E.; Noel, J. P., Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nature Chemical Biology* **2008,** 4, 217-222.

22. Reynolds, T., The evolution of chemosystematics. *Phytochemistry* **2007,** 68, 2887-2897.

23. Hadacek, F., Secondary metabolites as plant traits: current assessment and future perspectives. *Crit. Rev. Plant Sci.* **2002,** 21, (4), 273-322.

24. Kliebenstein, D. J., Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant Cell Environ.* **2004,** 27, 675-684.

25. Stalikas, C. D., Extraction, separation, and detection methods for phenolic acids and flavonoids. *J. Sep. Sci.* **2007,** 30, 3268-3295.

26. Cooper-Driver, G. A.; Bhattacharya, M., Role of phenolics in plant evolution. *Phytochemistry* **1998,** 49, (5), 1165-1174.

27. Rao, L. J. M.; Yada, H.; Ono, H.; Yoshida, M., Acylated and non-acylated flavonol monoglycosides from the Indian minor spice nagkesar (*Mammea longifolia*). *J. Agric. Food Chem.* **2002,** 50, 3143-3146.

28. Polya, G., *Biochemical Targets of Plant Bioactive Compounds: A pharmacological reference guide to sites of action and biological effects*. CRC Press: New York, NY, 2003.

29. Markham, K. R., *Techniques of Flavonoid Identification*. Academic Press: San Diego, CA, 1982; p 113.

30. Rabi, T.; Bishayee, A., Terpenoids and breast cancer chemoprevention. *Breast Cancer Res. Treat.* **2009,** 115, 223-239.

31. Rajamanickam, S.; Agarwal, R., Natural products and colon cancer: current status and future prospects. *Drug Develop. Res.* **2008,** 69, 460-471.

32. Sun, A. Y.; Wang, Q.; Simonyi, A.; Sun, G. Y., Botanical phenolics and brain health. *Neuromolecular Med.* **2008,** 10, 259-274.

33. Hsu, C.-L.; Yen, G.-C., Phenolic compounds: evidence for inhibitory effects against obesity and their underlying molecular signaling mechanisms. *Mol. Nutr. Food Res.* **2008,** 52, 53-61.

34. D'Auria, J. C.; Gershenzon, J., The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr. Opin. Plant Biol.* **2005,** 8, 308-316.

35. Vermerris, W.; Nicholson, R., Families of phenolic compounds and means of classification. In *Phenolic Compound Biochemistry*, Springer: New York, 2006.

36. Harborne, J. B., *Phytochemical Methods: A Guide to Modern Techniques of Plant Analysis*. Chapman and Hall: London, 1973.

37. Harborne, J. B., *Phytochemical Methods*. John Wiley & Sons: New York, NY, 1973; p 278.

38. Hagemeier, J.; Schneider, B.; Oldham, N. J.; Hahlbrock, K., Accumulation of soluble and wall-bound indolic metabolites in *Arabidopsis thaliana* leaves infected with virulent or avirulent Pseudomonas syringae pathovar tomato strains. *Proc. Nat. Acad. Sci* **2001,** 98, (2), 753-758.

39. Shirley, A. M.; Chapple, C., Biochemical characterization of sinapoylglucose:choline sinapoyltransferase, a serine carboxypeptidase-like protein that functions as an acyltransferase in plant secondary metabolism. *J. Biol. Chem.* **2003,** 278, (22), 19870-19877.

40. Branco-Price, C.; Kaiser, K. A.; Jang, C. J. H.; Larive, C. K.; Bailey-Serres, J., Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* **2008,** 56, (5), 743-755.

41. Mustroph, A.; Zanetti, M. E.; Jang, C. J. H.; Holtan, H. E.; Repetti, P. P.; Galbraith, D. W.; Girke, T.; Bailey-Serres, J., Immunopurified mRNA-ribosome complexes expose cell-type specific plasticity in response to an ephemeral stress in Arabidopsis. *Proc. Nat. Acad. Sci.* **2009,** 106, (44), 18843-18848.

42. Noteborn, H. P. J. M.; Lommen, A.; Jagt, R. C. v. d.; Weseman, J. M., Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J. Biotechnol.* **2000,** 77, 103-114.

43. Spacil, Z.; Novakova, L.; Solich, P., Analysis of phenolic compounds by high performance liquid chromatography and ultra performance liquid chromatography. *Talanta* **2008,** 76, 189-199.

44. Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *J. Proteome Res.* **2007,** 6, (2), 443-458.

45. Rochfort, S.; Caridi, D.; Stinton, M.; Trenerry, V. C.; Jones, R., The isolation and purification of glucoraphanin from broccoli seeds by solid phase extraction and preparative high performance liquid chromatography. *J. Chromatogr. A* **2006,** 1120, 205-210.

46. Opitz, S.; Schneider, B., Organ-specific analysis of phenylphenalenone-related compounds in Xiphidium caeruleum. *Phytochemistry* **2002,** 61, (7), 819-825.

47. Grata, E.; Boccard, J.; Glauser, G.; Carrupt, P.-A.; Farmer, E. E.; Wolfender, J.-L.; Rudaz, S., Development of a two-step screening ESI-TOF-MS method for rapid determination of significant stress-induced metabolome modifications in plant leaf extracts: The wound response in *Arabidopsis thaliana* as a case study. *J. Sep. Sci.* **2007,** 30, 2268-2278.

48. Noguchi, A.; Horikawa, M.; Fukui, Y.; Fukuchi-Mizutani, M.; Iuchi-Okada, A.; Ishiguro, M.; Kiso, Y.; Nakayama, T.; Ono, E., Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in Lamiales. *Plant Cell* **2009,** 21, 1556-1572.

49. Hanhineva, K.; Rogachev, I.; Kokko, H.; Mintz-Oron, S.; Venger, I.; Karenlampi, S.; Aharoni, A., Non-targeted analysis of spatial metabolite composition in strawberry (Fragaria x ananassa) flowers. *Phytochemistry* **2008,** 69, 2463-2481.

50. Sprogøe, K.; Stærk, D.; Ziegler, H. L.; Jensen, T. H.; Holm-Møller, S. B.; Jaroszewski, J. W., Combining HPLC-PDA-MS-SPE-NMR with circular dichroism for complete natural product characterization in crude extracts: levorotatory gossypol in *Thespesia danis*. *J. Nat. Prod.* 71, 516-519.

51. Colquhoun, I. J., Use of NMR for metabolic profiling in plant systems. *J. Pestic. Sci.* **2007,** 32, 200-212.

52. Reichelt, M.; Brown, P. D.; Schneider, B.; Oldham, N. J.; Stauber, E.; Tokuhisa, J.; Kliebenstein, D. J.; Mitchell-Olds, T.; Gershenzon, J., Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochemistry* **2002,** 99, 663-671.

53. Pedras, M. S. C.; Zheng, Q. A., Metabolic responses of Thellungiella halophila/salsuginea to biotic and abiotic stresses: Metabolite profiles and quantitative analyses. *Phytochemistry* **2010,** 71, (5-6), 581-589.

54. Moco, S.; Bino, R. J.; Vos, R. C. H. D.; Vervoort, J., Metabolomics technologies and metabolite identification. *Trends Anal. Chem.* **2007,** 26, 855-866.

55. Tan, J.; Bednarek, P.; Liu, J.; Schneider, B.; Svatos, A.; Hahlbrock, K., Universally occurring phenylpropanoid and species-specific indolic metabolites in infected and uninfected *Arabidopsis thaliana* roots and leaves. *Phytochemistry* **2004,** 65, 691-699.

56. Bednarek, P.; Schneider, B.; Svatos, A.; Oldham, N. J.; Hahlbrock, K., Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism in Arabidopsis roots. *Plant Phys.* **2005,** 138, 1058-1070.

57. Stobiecki, M.; Skirycz, A.; Kerhoas, L.; Kachlicki, P.; Muth, D.; Einhorn, J.; Mueller-Roeber, B., Profiling of phenolic glycosidic conjugates in leaves of *Arabidopsis thaliana* using LC/MS. *Metabolomics* **2006,** 2, (4), 197-219.

58. Cuyckens, F.; Claeys, M., Mass spectrometry in the structural analysis of flavonoids. *J. Mass Spectrom.* **2004,** 39, 1-15.

59. Pauli, G. F., Higher Order and Substituent Chemical Shift Effects in the Proton NMR of Glycosides. *J. Nat. Prod.* **2000,** 63, 834-838.

60. Stracke, R.; Jahns, O.; Keck, M.; Tohge, T.; Niehaus, K.; Fernie, A. R.; Weisshaar, B., Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12-and MYB111-independent flavonol glycoside accumulation. *New Phytol.* **2010,** 188, (4), 985-1000.

61. Kerhoas, L.; Aouak, D.; Cingoz, A.; Routaboul, J. M.; Lepiniec, L.; Einhorn, J.; Birlirakis, N., Structural characterization of the major flavonoid glycosides from *Arabidopsis thaliana* seeds. *J. Agr. Food Chem.* **2006,** 54, (18), 6603-6612.

62. Nakabayashi, R.; Kusano, M.; Kobayashi, M.; Tohge, T.; Yonekura-Sakakibara, K.; Kogure, N.; Yamazaki, M.; Kitajima, M.; Saito, K.; Takayama, H., Metabolomics-oriented isolation and structure elucidation of 37 compounds including two anthocyanins from *Arabidopsis thaliana*. *Phytochemistry* **2009,** 70, (8), 1017-1029.

63. Last, R. L.; Jones, A. D.; Shachar-Hill, Y., Towards the plant metabolome and beyond. *Nat. Rev. Mol. Cell Biol.* **2007,** 8, 167-174.

64. Zheng, X.; Shi, P.; Cheng, Y.; Qu, H., Rapid analysis of a Chinese herbal prescription by liquid chromatography-time-of-flight tandem mass spectrometry. *J. Chromatogr. A* **2008,** 1206, 140-146.

65. Hanhineva, K.; Kokko, H.; Siljanen, H.; Rogachev, I.; Aharoni, A.; Karenlampi, S. O., Stilbene synthase gene transfer caused alterations in the phenylpropanoid metabolism of transgenic strawberry (Fragaria x ananassa). *J. Exp. Bot.* **2009,** 60, (7), 2093-2106.

66. Grata, E.; Boccard, J.; Guillarme, D.; Glauser, G.; Carrupt, P.-A.; Farmer, E. E.; Wolfender, J.-L.; Rudaz, S., UPLC-TOF-MS for plant metabolomcs: A sequential approach for wound marker analysis in *Arabidopsis thaliana*. *J. Chromatogr. B* **2008,** 871, (2), 261-270.

67. Boccard, J.; Grata, E.; Thiocone, A.; Gauvrit, J.-Y.; Lanteri, P.; Carrupt, P.-A.; Wolfender, J.-L.; Rudaz, S., Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding. *Chemom. Intell. Lab. Syst.* **2007,** 86, 189-197.

68. Glauser, G.; Guillarme, D.; Grata, E.; Boccard, J.; Thiocone, A.; Carrupt, P.-A.; Veuthey, J.-L.; Rudaz, S.; Wolfender, J.-L., Optimized liquid chromatography-mass spectrometry approach for the isolation of minor stress biomarkers in plant extracts and their identification by capillary nuclear magnetic resonance. *J. Chromatogr. A* **2008,** 1180, 90-98.

69. Bottcher, C.; Roepenack-Lahaye, E. v.; Schmidt, J.; Schmotz, C.; Neumann, S.; Schell, D.; Clemens, S., Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant Phys.* **2008,** 147, 2107-2120.

70. Bottcher, C.; Centero, D.; Freitag, J.; Hofgen, R.; Kohl, K.; Kopka, J.; Kroymann, J.; Matros, A.; Mock, H.-P.; Neumann, S.; Pfalz, M.; Von Roepenack-Lahaye, E.; Schauer, N.; Trenkamp, S.; Zurbriggen, M.; Fernie, A. R., Teaching (and learning from) metabolomics: The 2006 PlantMetaNet ETNA metabolomics research school. *Physiol. Plantarum* **2008,** 132, 136-149.

71. Bottcher, C.; Roepenack-Lahaye, E. v.; Schmidt, J.; Clemens, S.; Scheel, D., Analysis of phenolic choline esters from seeds of *Arabidopsis thaliana* and *Brassica napus* by capillary liquid chromatography/electrospray-tandem mass spectrometry. *J. Mass Spectrom.* **2008,** 44, 466-476.

72. Yonekura-Sakakibara, K.; Tohge, T.; Matsuda, F.; Nakabayashi, R.; Takayama, H.; Niida, R.; Wantanabe-Takahashi, A.; Inoue, N.; Saito, K., Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in Arabidopsis. *Plant Cell* **2008,** 20, 2160-2176.

73. Jaroszynska, J., Isolation of free phenolic compounds from arboreal leaves by use of the Florisil/C18 system. *Anal. Bioanal. Chem.* **2003,** 377, 702-708.

74. Robbins, R. J., Phenolic acids in foods: an overview of analytical methodology. *J. Agric. Food Chem.* **2003,** 51, 2866-2887.

75. Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M., HPLC-MS-based methods for the study of metabonomics. *J. Chromatogr. B* **2005,** 817, 67-76.

76. Hendrawati, O.; Yao, Q.; Kim, H. K.; Linthorst, H. J. M.; Erkelens, C.; Lefeber, A. W. M.; Choi, Y. H.; Verpoorte, R., Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Sci.* **2006,** 170, 1118-1124.

77. Gall, G. L.; Metzdorff, S. B.; Pedersen, J.; Bennett, R. N.; Colquhoun, I. J., Metabolite profiling of *Arabidopsis thaliana* (L.) plants transformed with an antisense chalcone synthase gene. *Metabolomics* **2005,** 1, (2), 181-198.

78. Downey, M. O.; Mazza, M.; Krstic, M. P., Development of a stable extract for anthocyanins and flavonols from grape skin. *Am. J. Enol. Vitic* **2007,** 58, (3), 358-364.

79. Mulinacci, N.; Vincieri, F. F.; Baldi, A.; Bambagiotti-Alberti, M.; Wagner, H., Flavonol glycosides from *sedum telephium* subspecies *Maximum* leaves. *Phytochemistry* **1995,** 38, (2), 531-533.

80. Bloor, S. J.; Abrahams, S., The structure of the major anthocyanin in *Arabidopsis thaliana*. *Phytochemistry* **2002,** 59, 343-346.

81. Widarto, H. T.; Meijden, E. V. D.; Lefeber, A. W. M.; Erkelens, C., Metabolomic differentiation of *Brassica rapa* following herbivory by different insect instars using two-dimensional nuclear magnetic resonance spectroscopy. *J. Chem. Ecol.* **2006,** 32, 2417-2428.

82. Carlson, E. E.; Cravatt, B. F., Chemoselective probes for metabolite enrichment and profiling. *Nat. Methods* **2007,** 4, (5), 429-435.

83. Hirai, M. Y.; Sugiyama, K.; Sawada, Y.; Tohge, T.; Obayashi, T.; Suzuki, A.; Araki, R.; Sakurai, N.; Suzuki, H.; Aoki, K.; Goda, H.; Nishizawa, O. I.; Shibata, D.; Saito, K., Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *P. Natl. Acad. Sci. USA* **2007,** 104, (15), 6478-6483.

84. Zhou, D.-y.; Xu, Q.; Xue, X.-y.; Zhang, F.-f.; Liang, X.-m., Identification of O-diglycosyl flavanones in Fructus aurantii by liquid chromatography with electrospray ionization and collision-induced dissociation mass spectrometry. *J. Pharm. Biomed. Anal.* **2006,** 42, 441-448.

85. Hostettmann, K.; Marston, A.; Wolfender, J.-L., Strategy in the search for new biologically active plant constituents. In *Phytochemistry of Plants Used in Traditional Medicine*, Hostettmann, K.; Marston, A.; Maillard, M.; Hamburger, M., Eds. Clarendon Press: Oxford, 1995; Vol. 37, pp 17-45.

Chapter Six

LC-MS study of hypoxic Arabidopsis

Sean Cutler and Andrew Defries facilitated delivery of selected chemicals from CePCeB's Microsource Spectrum library. These compounds were used for initial dereplication and attempts at method standardization. AD was particularly helpful in discussions regarding optimization of chromatographic conditions. Christopher Jones assisted with UPLC-MS method development and operation of the instrument for the data presented in this chapter. Greg Barding applied multivariable data analysis to search for correlations within the dataset. Julia Bailey-Serres and Cynthia Larive provided insight and direction throughout in the iterative process that led to these experiments.

6.1 Introduction

In Chapter 5, HPLC-NMR was employed in parallel with UPLC-MS to establish the molecular identities of several classes of secondary metabolites isolated by HLB-SPE from tissue extracts of the model plant *Arabidopsis thaliana*. Eight molecules were assigned to peaks in UPLC-MS chromatograms, six flavonol glycosides, and two phenylpropanoids. This chapter builds on these structural assignments and extends the work to quantitatively monitor secondary metabolites in Arabidopsis extracts using UPLC-MS. Exploratory analysis of secondary metabolite profiles using multivariate data analysis (MVDA) was applied to reveal changes in metabolite abundance that are not apparent in raw data.

Metabolite measurements by ^{1}H NMR of whole seedling or specific organ extracts (highlighted in Chapters 3 and 4) were not applicable to secondary metabolism due to limitations in sensitivity and resolution. We sought to determine how hypoxia stress affects secondary metabolite pathways using SPE enrichment of several classes of secondary metabolites followed by UPLC-MS. By enrichment of the extract on a solid

phase cartridge, sensitivity can be improved. Through chromatographic separation, resolution challenges were addressed. This work represents the first study to focus on secondary metabolic adaptations of molecules in the functional classes phenylpropanoids, flavonoid glycosides, and glucosinolates under hypoxia stress in a plant.

Secondary metabolism (as articulated in Chapter 5) comprises all biosynthetic pathways *in planta* that lead to the formation of molecules inessential for the growth and development of the plant.[1] These secondary metabolites are not without function; chemical ecology is the study of secondary metabolites that mediate communication, perhaps to manipulate partners, competitors, and ecosystems.[2, 3] In response to biotic interactions, such as herbivory, molecules are synthesized to promote symbiotic relationships or deter harmful organisms from interacting with the plant.[4, 5] Abiotic (e.g.,UV) stress can alter pigmentation, or conversely a mutation can cause sensitivity to UV light,[6] due to re-coordination of biosynthetic pathways leading to colored secondary metabolites for purposes of "sunscreen", most likely by protecting against reactive oxygen species that damage DNA and membranes.

Primary and secondary metabolic pathways are interrelated. Figure 6.1 shows phosphoenolpyruvate (PEP), a key coupling intermediate (Chapter 1) for glycolysis, which can be converted to shikimate rather than pyruvate by a different series of enzymatic reactions. The common precursor to aromatic amino acids and many secondary metabolites is chorismate, which itself is a precursor to vitamins B9, K1, and the signaling molecule salicylate. The aromatic amino acids Tyr and Phe can be biosynthesized through a common intermediate, prephrenate (not shown), in three steps,

or directly via two distinct intermediates, which are acted upon by aromatic amino acid transaminase (AAAT). Both of these amino acids provide a source of phenylpropanoids, the class of secondary metabolites, which became the focus of this work.

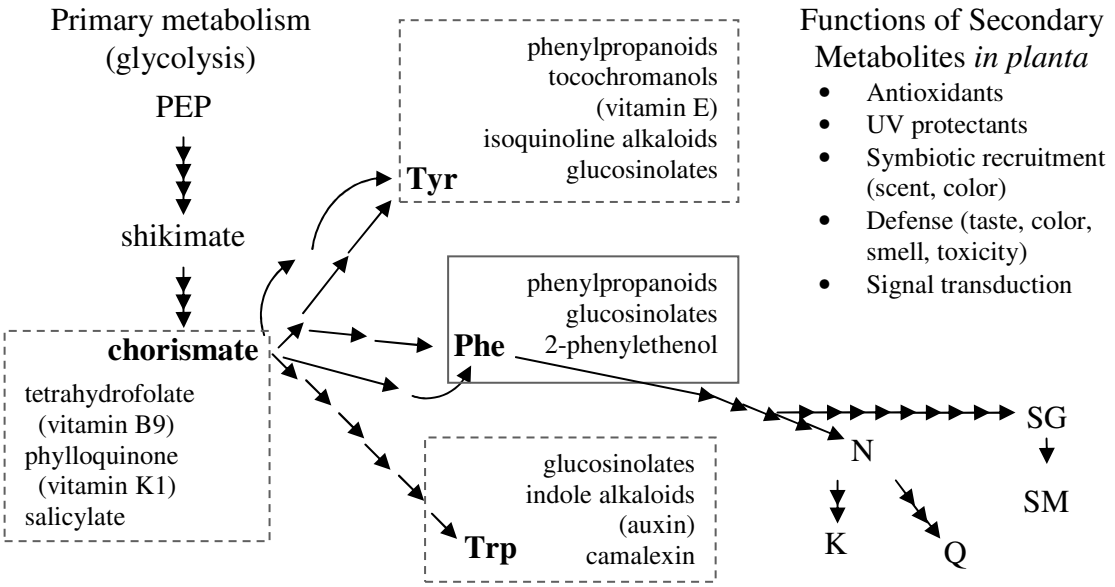


Figure 6.1 Metabolic pathways in higher plants leading from primary to secondary metabolism. Key intermediates phosphoenolpyruvate (PEP), shikimate, and chorismate are precursors to the aromatic amino acids Tyr, Phe and Trp. Aromatic amino acids are further diversified into secondary metabolites of the structural and/or functional classifications summarized in boxes around each intermediate. Phe gives rise to the flavonoids naringenin (N), kaempferol (K), and quercetin (Q), as well as sinapoyl-glucose (SG) and sinapoyl malate (SM). Flavonoids are commonly decorated with glucose (G) and rhamnose (R) sugars.

The roles of secondary compounds are summarized in Figure 6.1, in no particular order. Tocochromanols, such as vitamin E, function as antioxidants. Phenylpropanoids in Arabidopsis have been attributed to have a UV protecting function[7, 8] as well as having been found in root exudates,[4, 9] suggesting a role in plant-environment communication. Response of plants including changes in the levels of phenylpropanoids

to oomycete and pathogen attack has also been reported.[10] Flavonoids produced in roots provide protection against fungi and nematodes and can exhibit allelopathic effects. Ferulic and coumaric acids in soil may affect root water absorption.[11]

Historically, the development of chromatography was facilitated by plant pigments. Genetic screens seek observable phenotypes, and a change in color is easy to spot by eye. For these reasons, flavonoid biosynthetic pathways have been elucidated using a series of transparent testa (*tt*) mutants, which lacked pigmentation in the testa (seed coat) of Arabidopsis.[12, 13] Thin-layer chromatography (TLC) has been the workhorse for natural product studies, fluorescence of flavonoids is enhanced by diphenylboronic acid (DPBA), which can be used for *in vivo* staining to obtain information about flavonoid localization as well as quantitation of flavonoids in plant extracts resolved by TLC or HPLC.

Until this work, scant evidence of metabolite level changes in secondary pathways in response to oxygen deprivation was reported in the literature. A nontargeted UPLC/ESI-QTOF-MS method facilitated detection of >1,000 of mass signals in roots and leaves with S/N > 5. Nontargeted detection of alterations in metabolic profiles between wild-type and mutant plants combined with multivariate analysis facilitated phenylpropanoid and flavonoid pathway analysis, unraveling metabolic connections.[14] For these reasons, we included secondary metabolite profiling in this study, believing that observed changes in transcript abundance may be reflected in metabolite levels.

In preliminary work to support the quantitation of secondary metabolites, a small library of secondary metabolites was surveyed (Spectrum collection, MicroSource

Discovery Systems, Gaylordsville, CT), to provide retention times for evaluation for their use as a surrogate and internal standard. These standards were arrayed in a 386 well plate in the amount of 2 μL in DMSO. The molarity range was 10-6.5 mM. These compounds were diluted in 200 μL mobile phase and transferred to LC-MS vials. Under the same chromatographic conditions, Table 6.1 reports retention times and predicted partition coefficients (log P) for molecules explored in preparation and throughout this study.

The molecules listed in Table 6.1 are representative of secondary metabolites found in Angiosperms. This small database was populated by text-based searching of chemical literature by ISI Web of Science. Search terms included "SPE" and "plant secondary metabolism." Structural similarity of secondary metabolite scaffolds against the Center for Plant Cell Biology (CePCeB) at UCR's collection of screening libraries was performed in ChemMine.[15] Compounds for which a sample was available through the CePCeB core facility, with a similarity to plant secondary metabolites, were all found in the Microsource Spectrum library. Other compounds for which vendors could be located through similarity searching in SciFinder scholar were ordered (kaempferol-3-R-7-R and quercetin-3-R from Chromadex, Irvine, CA and naringenin-7-G from Idofine, Hillsborough, NJ). Aglycones hesperetin and naringenin were available from Sigma-Aldrich (St. Louis, MO). Chemical abstracts service (CAS) issues unique numerical identifiers to every chemical described in the open scientific literature. The numerical designation of chemical substances is convenient for database searches and transcends language, nomenclature, conventions, and governments, facilitating international accessibility to chemical information.

Table 6.1 Secondary metabolites evaluated. N. A. for a CAS number means that no such designation has been assigned to this molecule. In cases where solubility was a problem, e.g., quercetin aglycone, no retention time and observed mass-to-charge ratio (*m/z)* information is recorded although a standard has been purchased. R = rhamnose, G = glucose.

Secondary metabolite	Obs. t_R (min)	Predicted log P	Molecular formula	Observed *m/z*	Mol. wt. (g / mol)	CAS no.
gallic acid	2.3	0.91 ± 0.33	$C_7H_6O_5$	169	170.12	149-91-7
glucoibervirin	4.4	1.31 ± 0.77	$C_{11}H_{21}NO_9S_3$	421.7	407.48	26888-03-9
gentisic acid	5.6	1.56 ± 0.26	$C_7H_6O_4$	153	154.12	490-79-9
glucobrassicin	5.7	1.35 ± 1.37	$C_{16}H_{19}N_2O_9S_2^-$	448.8	447.46	4356-52-9
epigallocatechin	6.1	-0.10 ± 0.38	$C_{15}H_{14}O_7$	305	306.27	1617-55-6
cianidanol	6.6	0.49 ± 0.38	$C_{15}H_{14}O_6$	289	290.26	7295-85-4
caffeic acid	7.2	1.42 ± 0.36	$C_9H_8O_4$	179	180.16	331-39-5
glucohirsutin	7.7	1.27 ± 0.77	$C_{16}H_{31}NO_{10}S_3$	494	493.62	21973-60-4
epicatechin	9.1	0.49 ± 0.38	$C_{15}H_{14}O_6$	289	290.26	7295-85-4
sinapoyl glucose	9.4	0.45 ± 0.54	$C_{17}H_{22}O_{10}$	386.6	386.35	29881-39-8
quercetin-3-R-G-7-R	10.4	0.69 ± 1.50	$C_{33}H_{40}O_{20}$	755.2	756.66	162062-89-7
kampferol-3-R-G-7-R	11.6	0.89 ± 1.47	$C_{33}H_{40}O_{15}$	742	740.65	162062-89-7
ferulic acid	11.9	1.64 ± 0.36	$C_{10}H_{10}O_4$	193	194.18	537-98-4
quercetin-3-G-7-R	13.1	1.03 ± 1.48	$C_{27}H_{30}O_{16}$	611.5	610.51	18016-58-5
ellagic acid	14.2	0.52 ± 1.48	$C_{14}H_6O_8$	301	302.2	476-66-4
3-hydroxycoumarin	15.2	1.60 ± 0.75	$C_9H_6O_3$	161	162.14	939-19-5
kampferol-3-R-7-G	15.3	0.20 ± 1.43	$C_{27}H_{30}O_{15}$	595.4	594.52	N. A.
isorhamnetin-3-G-7-R	15.7	1.00 ± 1.45	$C_{28}H_{32}O_{16}$	625.6	624.54	N. A.
kaempferol-3-R-7-R	17.2	1.66 ± 1.45	$C_{27}H_{30}O_{14}$	579.4	578.52	482-38-2
?	17.4	-	-	496	-	-
fisetin	17.9	2.52 ± 0.62	$C_{15}H_{10}O_6$	287	286.23	528-48-3
quercetin-3-R	18	2.17 ± 1.46	$C_{21}H_{19}O_{11}$	448.4	447.3	522-12-3
naringenin-7-G	18.5	0.82 ± 0.83	$C_{21}H_{22}O_{10}$	433	434.4	529-55-5
?	18.8	-	-	463.9	-	-
daidzein	21.5	2.78 ± 1.33	$C_{15}H_{10}O_4$	257	254.23	486-66-8
hieracin	21.7	1.86 ± 0.68	$C_{15}H_{10}O_7$	301	302.23	1621-84-7
neoglucobrassicin	22	1.07 ± 1.38	$C_{17}H_{22}N_2O_{10}S_2$	478	478.5	5187-84-8
4-methoxyglucobrassicin	22	1.07 ± 1.38	$C_{17}H_{22}N_2O_{10}S_2$	478	478.5	83327-21-3
genistein	22.2	2.96 ± 1.40	$C_{15}H_{10}O_5$	269	270.24	446-72-0
apigenin	22.3	2.10 ± 0.56	$C_{15}H_{10}O_5$	332	270.24	520-36-5
naringenin	22.3	3.19 ± 0.38	$C_{15}H_{12}O_5$	271	272.26	480-41-1 67604-48-2
mucronulatol	22.7	2.27 ± 0.38	$C_{17}H_{18}O_5$	303		N. A.
rhamnetin	22.9	2.58 ± 0.62	$C_{16}H_{12}O_7$	315	316.26	90-19-7
?	23.2	-	-	306.3	-	-
hesperetin	23.3	2.90 ± 0.39	$C_{16}H_{14}O_6$	285	302.27	520-33-2 41001-90-5
quercetin	-	2.07 ± 0.72	$C_{15}H_{10}O_7$	-	302.24	117-39-5
glucoiberin	-	-0.81 ± 0.77	$C_{11}H_{21}NO_{11}S_3$	-	439.48	554-88-1
melatonin	-	0.96 ± 0.44	$C_{13}H_{16}N_2O_2$	-	232.28	73-31-4

6.2 Experimental Approach

6.2.1 Preparation of Plant Material

Approximately 100 seeds of *Arabidopsis thaliana* ecotype Columbia-0 were sown on sterile Murashige and Skoog basal salt mixture (MS medium) containing 1.0% w/v sucrose and 0.4% agar (Phytagel, Sigma). Seeds were arranged on plates in a single horizontal line 2/3 up from the bottom of the plate. Prior to sowing, seeds were sterilized as described in Chapter 2, section 2.2.2.2. After sowing the seeds, plates were stored horizontally in a positive flow hood for 1 h before excess water was removed by sterile Pasteur pipet. Plates were transferred to the growth chamber, and positioned vertically in wooden holders for 7 days. A single Scotch tape hinge was placed on the top of each plate to secure the lid and facilitate handling.

Seedlings were prepared in the controlled facility maintained in the Genomics Building, University of California, Riverside, Bailey-Serres laboratory. A single genotype was grown on all plates, which were randomized according to proximity to lights. Light period: 16 h day at ~80 µmol photons m^{-2} s^{-1}/ 8 h darkness at 23 °C (Percival Scientific, model CU36L5C8). No water or additional nutrition was supplied to the aseptic closed plates for the duration of 7 day germination and growth phase.

On the 7^{th} day after plant establishment, plates were transferred to the stress room. Plates were placed vertically into Lucite chambers, 16-21 plates per chamber vertically, into which 99.998 % (v/v) argon gas was pumped and ambient air was allowed to exit under positive pressure. Dim light was provided to simulate light levels during submergence under murky water (5 to 7 µmol photons m^{-2} s^{-1}) at room temperature (23 to

25 °C). Supplemental light provided by Verilux HappyLite MiniPlus Sunshine Supplement System (26 Watt Spiral Fluorescent Light Bulb, P/N CFS26GU24VLX) and Good Earth Portable Luminaire (05/06 E180723) was filtered to the appropriate dimness using a ChemWipe. Measurement of light levels was carried out by Licor Quantum Sensor (S/N Q24051) and Light Meter (S/N LM2-1144). Timing of day, night, and application of low oxygen stress were consistent with Chapters 3-5.

Controls were placed in identical chambers open to laboratory air. Oxygen deprivation was applied similarly to the work presented in Chapter 3. Briefly, 9 h of argon treatment, equivalent to 1.5 h gradual stress followed by 7.5 h anoxia, was applied to seedlings placed vertically in stress chambers. The argon was bubbled through water to minimize dessication stress. Three groups were treated simultaneously in each bioreplicate: 9 h stress (H), 9 h control (C), and 9 h stress followed by 1 h recovery (R). Seedlings were stressed at growth stage 1.0, which corresponds to an average of 6 days from the date of sowing, where cotyledons, or the first light-harvesting organs to emerge from the seed, are fully opened.[16] This experiment was designed to be similar to that used for NMR metabonomic and translatome analysis in Chapter 3 and therefore the results should be comparable.

Immersion in liquid nitrogen (LN_2) within 5 min of release from stress treatment was the metabolism quenching method. Petri plates were opened and the whole frozen seedlings were transferred via sterile tweezers to a mortar and pestle containing liquid nitrogen (LN_2). Plant material was homogenized until it reached the consistency of a fine powder. Suspended in LN_2, powdered plant material was transferred from the mortar to

three individual 50 mL Falcon tubes, which were transported to Chemical Sciences in a Dewar immersed in LN_2. A slurry of dry ice and acetone (-78 ^oC) was prepared to boil off excess LN_2 (-196^oC) prior to lyophilization. After all moisture was removed from the plant material, it was stored at -20^oC for ~1 year prior to analysis.

6.2.2 Tissue Extraction

Tissue dry weight was obtained prior to the addition of an extraction buffer (EB) containing a surrogate (naringenin-7-*O*-glucoside, N7G, 20 µM, Idofine, Hillsborough, NJ). EB was prepared by mixing acetic acid with Burdick and Jackson (B & J) ultrapure water (Optima grade, Fisher Scientific), adjusting to pH 7.0 using ammonium hydroxide, resulting in a 100 µM buffer. To this, acetonitrile (Optima grade) was added in a 1:1 volume ratio. Surrogate (N7G) suspended in 100% MeOH was added to give 20 µM. This buffer was mixed by vortexing and many inversions. Approximately 20 mg of tissue constituted each sample where possible (some pools of tissue were limited to 5 mg) to which 900 µL EB was added. Manual homogenization with a micropestle for 4 min was used for metabolite extraction. After centrifugation for 10 min at 13,200 x *g*, supernatants were transferred to a new tube for concentration to dryness under speedvacuum (Savant) for approximately 3 h.

6.2.3 SPE Enrichment

Solid-phase extraction with Waters Oasis HLB cartridges was employed for the enrichment of target analytes and elimination of interfering compounds. Each extract

was reconstituted into 200 µL of 50/50 MeOH/H_2O, vortexed 60 s, sonciated 10 min, and vortexed 60 s more then centrifuged for 4 min at 5,000 x *g* to remove insoluble material. This supernatant was diluted to a volume of 5 mL using B & J water in a volumetric flask. SPE cartridges were conditioned with 2 separate treatments (40 s each) of 3 mL MeOH, which was evacuated for 10 min by vacuum (20 kPa) prior to equilibration. This was found to be an important step in maximizing recovery of target metabolites. Cartridges were equilibrated using 3-5 treatments of 3 mL B & J water. During this time, flow rates were equalized among all cartridges attached to the SPE vacuum manifold.

Each reconstituted extract was loaded onto an individual cartridge by 1000 µL Eppendorf pipet, being careful to avoid cross contamination via pipet tips. Flow rates of approximately 0.2 mL / min were used to maximize sample loading driven by gravity. Occasionally an air bubble would decrease the flow rate of a cartridge, in these cases positive pressure driven by a pipet bulb was applied. Water was used to wash loosely bound and noninteracting compounds from the cartridge, 2 mL was added and removed under gentle pressure. Cartridges were dried under 20 kPa vacuum for 10 min. Metabolite elution was achieved in 1 mL of acetonitrile (ACN). This strong solvent was found to be necessary for adequate recovery of metabolites (~20% recovery under MeOH elution, ~88% recovery under ACN elution, 1 mL each, data not shown). A Teflon needle guided eluent into microcentrifuge tubes directly below each cartridge enclosed in an adjustable collection rack system, inside the manifold. Strong solvent (ACN) had to be removed prior to analysis by LC-MS, therefore eluates were dried for 3 h by speedvacuum. Samples were stored overnight at -20^{o}C.

6.2.4 Chromatography

A 10% MeOH reconstitution buffer (RB) prepared in B & J water containing 0.5 μM quercetin-3-*O*-rhamnoside (Q3R, Chromadex, Irvine, CA) was used for internal standardization. Each extract was reconstituted in 100 μL of RB, vortexed for 60 s, sonicated for 10 min, and vortexed for 60 s again. Centrifugation for 4 min at 13,200 x *g* ensured that no particulates were introduced to the chromatographic instrument. Supernatants were transferred to conical-bottom MS vials with plastic screw cap lids and recycled septa.

Although it is common practice to reconstitute samples for LC-MS analysis in the mobile phase used for separation, preliminary studies revealed sample instabilities when 0.1% formic acid was present in RB over the duration samples remained in the autosampler (0 - 48 h). Therefore we used an RB which contained no buffering agent, particularly no acid. All procedures were carried out at room temperature (23-25^{o}C).

Blanks (B) were run at the beginning and end of each batch. Standards (STD) were also run at the beginning and end of each batch, consisting of the surrogate (N7G) and the internal standard (Q3R) added at 50 μM in 50/50 MeOH/H_2O. A quality control (QC) was prepared for each bioreplicate by pooling 25 μL aliquots from each treatment (75 μL total volume of QC).[17]

Run order is shown in Figure 6.2. The plant tissues resulting from the first bioreplicate assayed, consisting of three treatments (C = 9 NS, S= 9 HS, R = 9 + 1R), which were sampled once (1 analytical replicate) per tissue pool. Refined extracts

reconstituted in RB were sampled 3 times (machine replicates 1, 2, 3). A batch is considered what can be accomplished either during one workday or one overnight run. Batch supervision is on the front end (calibration, resolution check) and back end, to ensure that each batch was analyzed successfully and no degradation of instrument performance occurred during the batch. Figure 6.3 shows the scheme for sampling of technical replicates and corresponding filenames of UPLC-MS chromatograms.

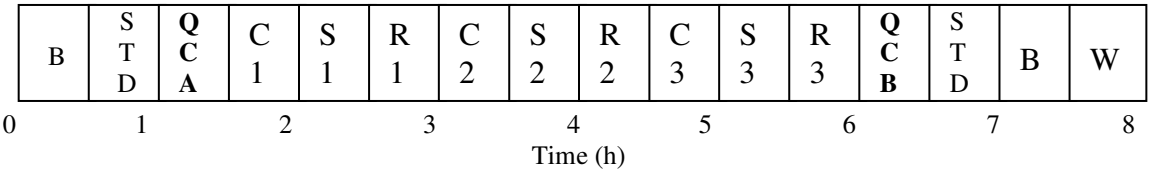


Figure 6.2 Run order for a batch. The first injection of any day is a blank (B) to prepare the chromatographic system and ensure that the column is clean. Then a sample containing the surrogate and internal standard (STD) is injected. This is to record a baseline response for mass and retention times in the absence of a matrix. Then a pooled sample representing all three biological treatments (QC) is injected to record mass and retention times for surrogate, standard, and endogenous metabolites in the matrix. Each treatment (C, S, R) is sampled three times. QC is run again to determine mass and retention time drifts. The STD is run again as well as a blank and the column is washed if no more batches are to be run immediately. For this method, 30 min per injection is the runtime, so one batch can be easily accommodated into a standard workday.

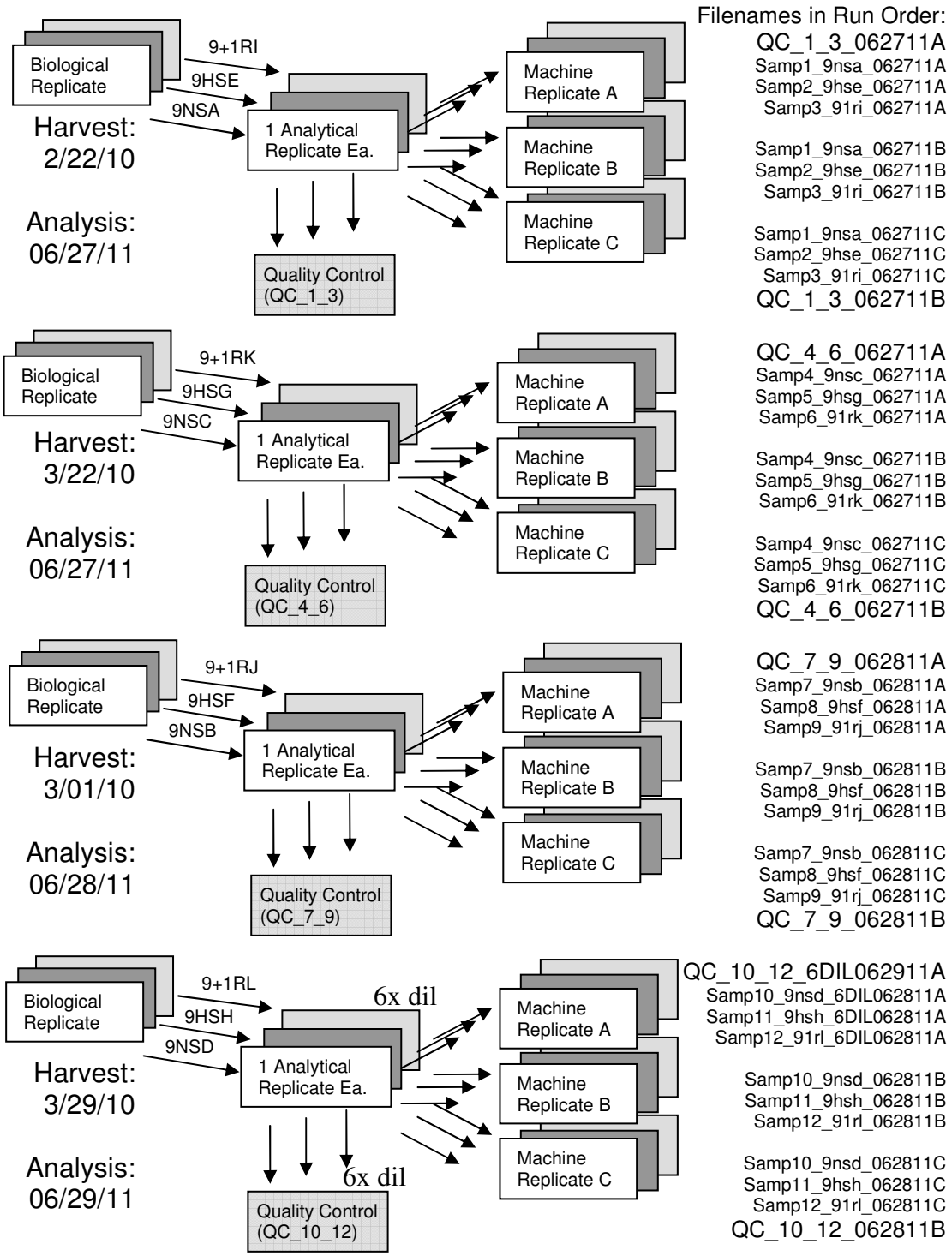


Figure 6.3 Run order and filenames corresponding to LC-MS metabolic profiling experiment carried out on 7-d-old seedlings of *Arabidopsis thaliana*. Each biological replicate involved 3 treatments. Quality controls for each batch were prepared by pooling an aliquot representing each treatment to account for all metabolites. The final bioreplicate was diluted 6-fold, as indicated in filenames for machine replicate A.

Samples were injected within 9 h of reconstitution using a sequence template shown in Figure 6.4. A Waters Acquity UPLC was used for these experiments, controlled using MassLynx software. The chromatographic separation was carried out in a UPLC BEH C18 column (2.1 μm i.d., 1.7 μm diameter particles, 150 mm length, fitted with a VanGuard guard column of identical diameter and packing material. The flow rate used was 0.4 mL / min and the column was maintained at 40°C. Gradient elution in 0.1% formic acid was carried out according to Table 6.2. Briefly, an isocratic 5% ACN stage was maintained for 2 min, most of which was sent to waste using a solvent delay to prevent early eluting hydrophilic ions from entering and contaminating the mass spectrometer. A gradient between 2 and 20 min increased the organic content to 20%, wherein the separation of glycosylated secondary metabolites was achieved by a relatively shallow ramp of approximately 0.83% B. After 20 min, the eluent was sent to waste using a solvent delay and the column was cleaned using 98% ACN for 3 min.

Table 6.2 Separation parameters for non-targeted profiling of Arabidopsis secondary metabolites. A UPLC BEH C18 column (2.1 μm i.d., 1.7 μm particles, 150 mm length) under gradient elution, flowing at 0.4 mL / min at 40°C. To promote ionization, 0.1% formic acid was added to the aqueous (H_2O) and organic (ACN) mobile phases.

Time (min)	**% B (ACN)**	**Action / post-column activity**
0	5	Eluent to waste (SD1), Isocratic
2	5	Eluent to ESI-MS Begin Gradient (G1)
20	20	End G1, Eluent to waste (SD2)
25	98	Gradient (G2)
28	98	Column cleaned
28.1	5	Column reequilibration
30	5	Ready for next injection

6.2.5 UV Absorbance

Dual channel measurements of UV absorbance were collected throughout all chromatography at 254 nm and 365 nm. These wavelengths were chosen to correspond to the flavonoid A ring and flavonoid C rings, respectively. Detection at 280 nm is most generally used for simultaneous separation of mixtures of phenolic acids although for dual monitoring, (254 and 280) or (280 and 320) can be ideal wavelengths.[18]

6.2.6 MS (TOF) Mass Measurements

A Waters QTOF Micromass Mass Spectrometer operated by Chris Jones was used to acquire these data. Electrospray ionization and ion guiding was operated in negative mode. The capillary was maintained at 3 kV bias. The source was maintained at 120 ^{o}C. A desolvation gas (N_2) was used flowing at 650 L / h and 250 ^{o}C cooperatively with a cone gas flowing at 10 L / h. The sample cone maintained at 1 V guided droplets and ions toward the extraction cone, which was also maintained at 1 V.

After a solvent delay (0-1.5 min) the mass range of 75-1500 *m/*z was interrogated using a scan time of 0.5 sec. Resolution between 5000-6000 and mass accuracy between 5-600 ppm (0.004-0.428 Da) across the dataset was observed. Another solvent delay was utilized during column washing (20-30 min) and re-equilibration.

Metabolite identifications were made using retention time window, mass spectral similarity, and coordinated information from literature and Table 6.1.

6.2.7 Data Pre-processing

To ensure unbiased analysis, peaks (unique *m/z*) were automatically detected in the UPLC-MS TIC using MassLynx software using a peak-to-peak baseline noise threshold above 1000 counts. The observed mass of our surrogate was 433.2329 Da, while the exact mass calculated by ACD/Labs for [M-H$^-$] was 433.1140 Da, leaving a 274.7 ppm systematic bias throughout this dataset.

6.2.8 MarkerLynx Multivariate Data Analysis

MarkerLynx was used for spectral deconvolution and peak integration with the following parameters: a chromatographic window of 2-20 min, spectral window of 200-1000 *m/z,* 0.20 Da mass accuracy, peak width at 5% height of 9 sec, peak-to-peak baseline noise of 1, marker intensity threshold of 30 counts, mass window of 0.5 Da for peak alignment, noise elimination level of 3.0. Isotopes were treated individually.

Unique retention time and *m/z* pairs were identified automatically and each considered as an independent variable. Signal intensity of each retention time and *m/z* pair were automatically calculated and exported in a table. For some metabolites, the signal was not above a signal-to-noise (S/N) cutoff threshold. These appear as zero in the exported data. To avoid bias in the univariate statistics, these cells were changed from zero (would have been included in calculations) to empty (excluded from calculations). Comparative metabolomics in LC-MS-based metabolomics involves calculating normalized averages representative of treatment and control, then a ratio of these yields a fold change for each retention time and *m/z* pair. Both biological (N = 4) and technical

(N = 3) replicates were recorded across three treatments (S, C, R) as defined in Figure 6.2. Significant differences in normalized metabolite signal intensity between S and C were identified by calculating a p-value using a Student's *t* test.

Retention time and *m/z* pair signal intensities (variables) were normalized to the signal intensity of the surrogate (N7G) in the column representing each chromatogram. Variables with a p-value less than 0.05 were transported to Minitab for multivariable data analysis (MVDA). Principal component analysis (PCA) validated the p-value cutoff to consider significant for consideration as biomarkers of low oxygen stress. Including variables with p-value up to 0.10 was performed for cross-validation (data not shown). When a metabolite was not observed above a S/N cutoff, it was considered not present.

6.3 Results

A key contribution of this work was that it established a pipeline for exploration of chemical space in plant secondary metabolomes. Without analytical strategies for metabolite identification and quantification, as well as data management strategies for extracting knowledge out of huge amounts of data, we cannot make progress with discoveries in this area. Our goal was to explore the spectrum of secondary metabolites of 7 d seedlings of *Arabidopsis thaliana* grown under control growth conditions (C), deprived of oxygen for 9 h (H) or reoxygenated for 1 h following 9 h of low oxygen stress (R) by use of LC-MS metabolic profiling. The HPLC-NMR and other structural clues (Chapter 5) led to the identification of several secondary metabolites, but in this

chapter, UPLC-MS data is treated in an unsupervised way to identify signals contributed by unknown molecules or adducts.

Constraining the data-fitting routines provided by MarkerLynx to a S/N cutoff reduced the dimensionality of the dataset to 3800 retention time and *m/z* pairs (variables). Univariate comparisons between hypoxic stressed (H) and control treated (C) samples were performed for each of the 3800 ions observed above the S/N cutoff. Our calculation of p-value assumed that the distributions of each population was two-tailed with unequal distributions.

Table 6.3 provides the most significantly different retention time and *m/z* pairs (variables). Some of these variables showed a fold change greater than 1.0, indicating these molecules accumulated in response to low oxygen stress. However, for the majority of the significantly altered variables in the extracts, we saw fold changes less than 1.0, indicating depletion of the metabolite in response to the stress. To determine the compounds with these masses, the METLIN database was searched resulting in one or more distinct structural possibilities for each of the observed chemical signals.[19] Proposed empirical and structural information is provided in Table 6.3.

For visualization of the normalized signal intensity and standard deviation for each of the proposed flavonoids in Table 6.3, a plot was generated (Figure 6.4). The trend appears to be that these molecules are less abundant after 9 h of low oxygen and a slight recovery for some compounds is observed after 1 h of reoxygenation.

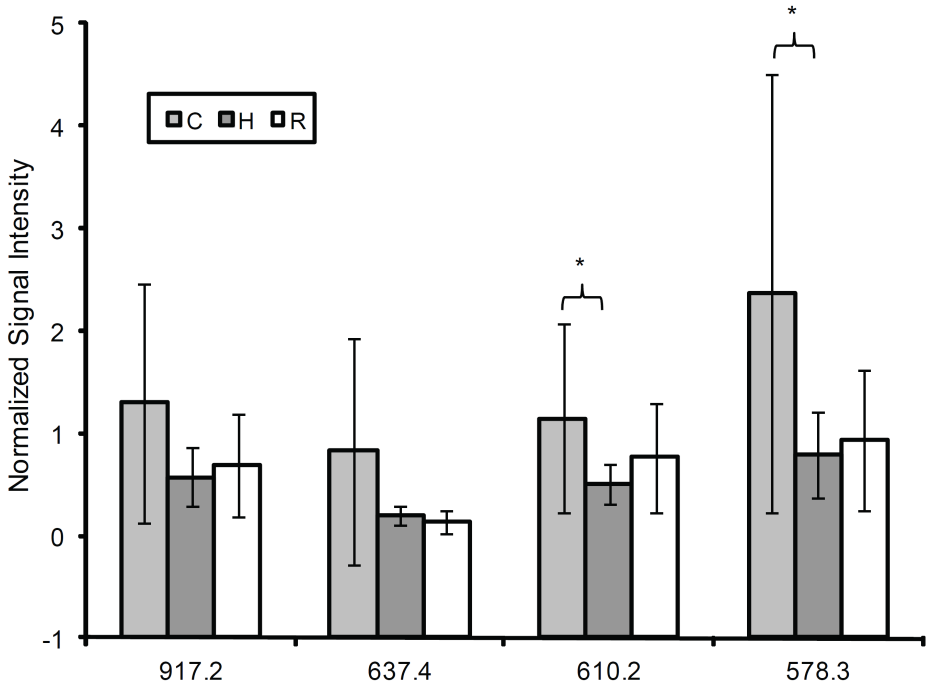


Figure 6.4 Average relative metabolite abundance in extracts representing Arabidopsis seedlings (N=9) for control (C), hypoxia (H), and recovery (R) treatments. Error bars indicate standard deviations.

To continue exploring these molecular signals and changes in their accumulation in response to hypoxia stress, we constructed a PCA model including a small data matrix comprised of only the variables, which had a cutoff of p-value less than 0.05. The scores plot is shown in Figure 6.5. Here each LC-MS chromatogram is represented by a single point. Hypoxia-treated (△) seedling extracts are clearly distinguishable from controls (□). The separation according to biological treatment appears mostly along principal component (PC) 1 with biological or analytical variance contributing to the second component. The corresponding loadings plot shows the *m/z* values responsible for the trends. Comparison with Table 6.3 aids identification of coeluting ions, which may be part of the same molecule.

Table 6.3 Biomarkers of hypoxia stress identified by UPLC-MS combined with MVDA. Retention time (tR) and mass-to-charge ratio (*m/z*) are shown together with the fold change in abundance of the biomarker when the normalized signal intensity in the stress treatment is divided by the normalized signal intensity in the control treated tissue. Significance is quantified by p-value. All ions reported exhibited a statistically significant change above/below the control, with >95% confidence from a dataset consisting of 4 bioreplicates x 3 machine replicates for each treatment. For fold changes designated with NA, the fold change cannot be calculated because the function becomes undefined if the amount of the metabolite in control tissue is below the method detection limits.

t_R (min)	*m/z*	Fold change 9HS/9NS	p-value for 9HS/9NS	Proposed molecular features based on METLIN *m/z* search	Proposed molecular formulae	Present calls out of 36
7.1	460.2	1.97**	0.0012			26
13.4	777.3					14
13.4	721.4					17
13.4	578.3					18
13.4	358.2	6.65**	0.00056			25
13.4	307.1	1.78**	0.00010			18
5.0	553.1	0.40*	0.015			36
5.3	917.2	0.44	0.058	flavonoid, three sugars flavonoid, phenylpropanoid	$C_{39}H_{50}O_{25}$ $C_{42}H_{46}O_{23}$	36
5.4	388.1	0.48*	0.049			36
6.2	683.4	0.53*	0.033		$C_{30}H_{36}O_{18}$	35
8.1	637.4	0.24	0.075	flavonoid, two sugars flavonoid, one sugar and a phenylpropanoid	$C_{29}H_{34}O_{16}$ $C_{32}H_{30}O_{14}$	23
8.1	457.0	0.28*	0.017		$C_{23}H_{22}O_{10}$	36
9.4	469.2	0.46*	0.034			35
9.4	427.2	0.42*	0.015			31
9.8	294.2	0.30*	0.042			27
13.1	610.2	0.44*	0.037	flavonoid, two sugars flavonoid, one sugar, phenylpropanoid	$C_{27}H_{31}O_{16}$ $C_{30}H_{27}O_{14}$	36
16.2	578.3	0.33*	0.027	flavonoid, two sugars flavonoid, sugar, phenylproponoid flavonoid,sugar combined with *o*-methylation and malonation	$C_{27}H_{31}O_{14}$ $C_{30}H_{27}O_{12}$ $C_{26}H_{27}O_{15}$	36
18.3	917.4	0.47*	0.043	flavonoid, 4 sugars flavonoid, 3 sugars, phenylpropanoid	$C_{39}H_{50}O_{25}$ $C_{42}H_{46}O_{23}$	36
18.3	445.1	0.32*	0.025		$C_{22}H_{22}O_{10}$ $C_{21}H_{18}O_{11}$	36
18.6	469.2	0.51*	0.036			36

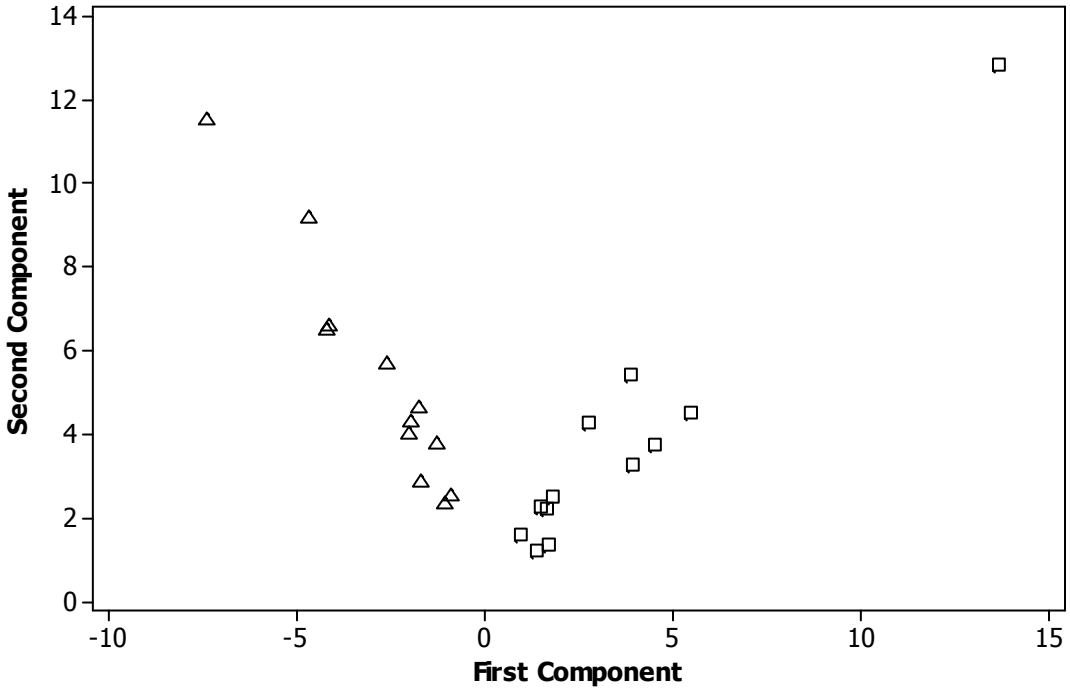


Figure 6.5 PCA scores plot showing UPLC-ESI-MS-based metabolic profiling of 9 h hypoxia stress (H) and 9 h control (C) SPE-treated seedling extracts. The model was constructed using all variables listed in Table 6.3 with the exception of *m/z* 917.4, t_R 18.3, which dominated the loadings plot. C (□) and H (△) seedling extracts are distinguishable as distinct groups along the first component axis (i.e. PC1).

Examining the loadings plot (Figure 6.6) in the context of Table 6.3 grouping of ions can be observed according to whether the metabolite accumulates in response to stress or is depleted. Interestingly one ion varies along PC2, currently unidentified *m/z* 427.2. These two PCs are sufficient to explain 91% of the variance in the selected dataset containing only significantly different ions between C and S. Figure 6.7 demonstrates this graphically in an explained variance plot.

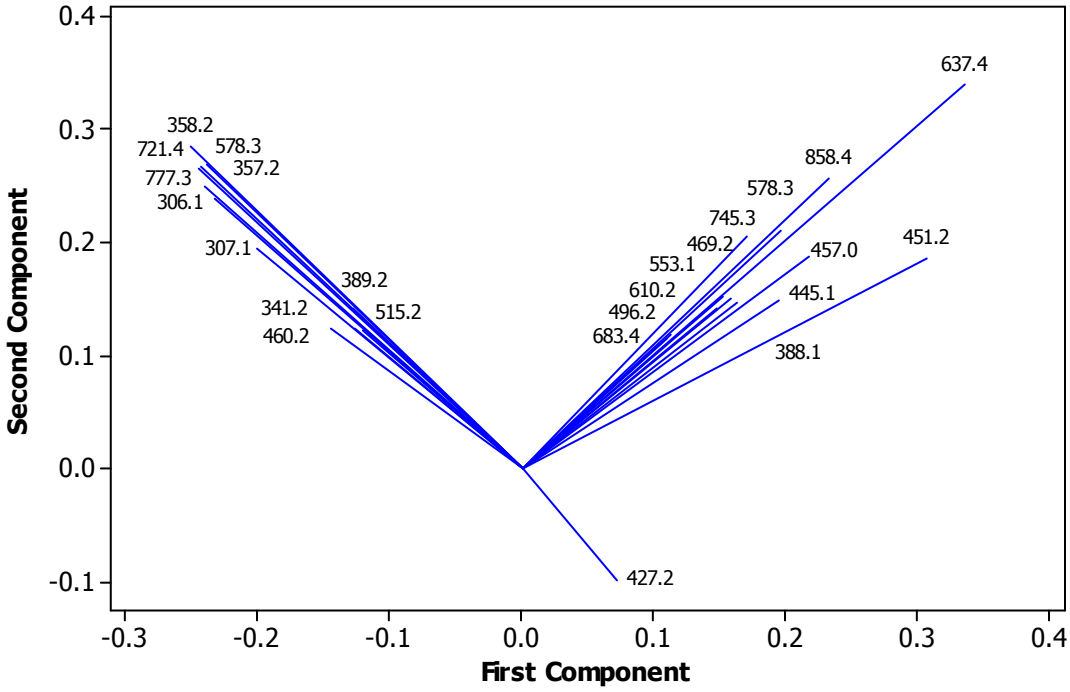


Figure 6.6 Loadings plot generated by PCA demonstrating the variables responsible for the separation between control and hypoxia-stressed 7 d Arabidopsis seedling SPE-enriched extracts. Biomarkers 1-3, 8, and 14-15 (Table 6.3) vary along the second component axis, and are accumulated under hypoxia stress, relative to control. Alternately, biomarkers 4-7, 9-13, 16-19 vary along the first component axis, and are depleted in the HS, relative to NS samples. Interestingly biomarkers 9 and 9b initially appeared to be fragments of the same molecule (same t_R) but they do not covary, indicating they are independent molecules.

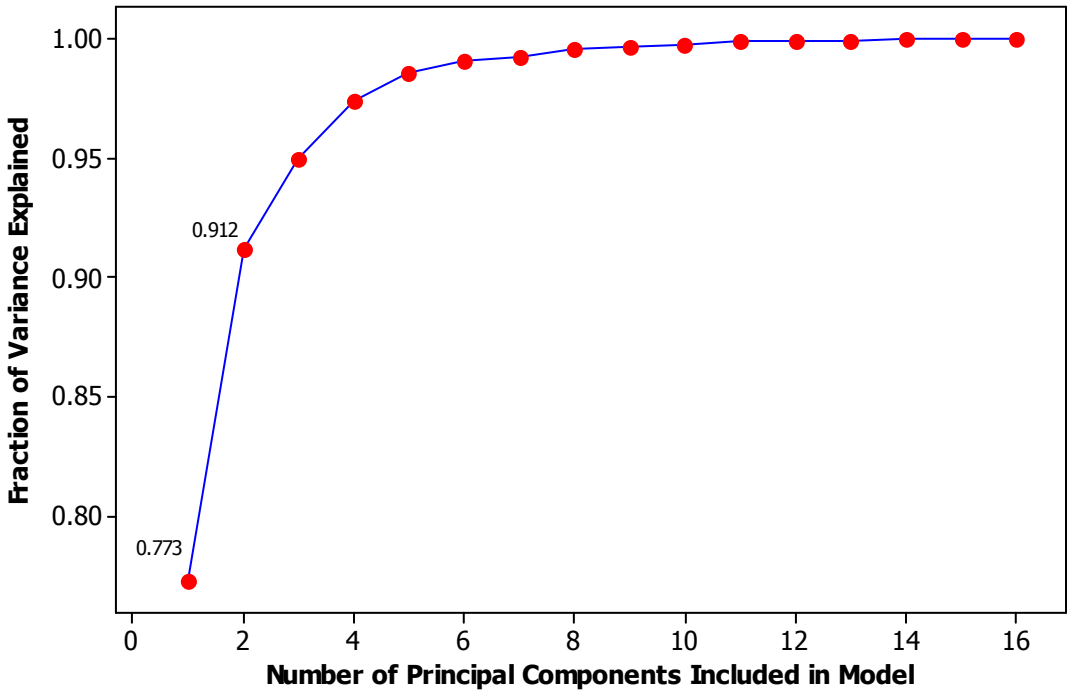


Figure 6.7 Plotting eigenvalue on the y-axis and eigenvector on the x-axis gives an explained variance plot. The scores and loadings plots (Figures 6.5, 6.6) account for 91% of the variance in Table 6.3.

Next, we sought to query the dataset comprising all 3800 ions to determine if the molecules whose identity was established via HPLC-NMR in Chapter 5 were altered in response to extended low oxygen stress. Table 6.4 summarizes these findings. In short, no significant changes in these metabolites were observed.

6.4 Discussion

For LC-MS based studies that aim at even relative quantitation, many aspects must be demonstrated to establish method reliability and robustness. Mass signal detection and quantification must be carried out manually if the exploration of the data by nontargeted (global) analysis is to be trusted. Böttcher et al.[14] report that "features" must

Table 6.4 Changes in levels of identified metabolites in response to oxygen deprivation. None of the relative abundances of the metabolites identified respond (become accumulated or depleted) significantly in response to 9 h low oxygen stress in 7 d Arabidopsis seedlings. Due to their abundance, all of the metabolites reported in this table were present in all samples.

Metabolite identity	t_R (min)	Obs. *m/z*	p-value 9HS/9NS	FC 9HS/9NS
sinapoyl-glucose	9.4	386.658	1.0	1.0
K-3-R-G-7-R	11.6	742.389	0.52	0.79
	11.6	741.338	0.64	0.86
Q-3-G-7-R	13.1	609.208	0.84	0.94
K-3-R-7-G	15.1	594.211	0.56	0.87
	15.2	595.250	0.46	0.69
I-3-G-7R	15.6	624.251	0.61	0.88
K-3-R-7-R	17.2	578.233	0.28	0.73
Q-3-R	18.3	447.109	0.25	0.61

initially be selected manually (11 knowns plus three unknowns over a range of several orders of magnitude in intensity); features were defined as having a unique retention time and *m/z* pair. We applied this approach to our hypoxia and recovery study of Arabidopsis extracts. Böttcher et al. reported retention time shift observed across replicates was 32 $\pm$ 13 s (n = 16 for all statistics). Average mass accuracy of 4.8 $\pm$ 1.9 mD (0.15 ppm) was reached using qTOF MS. Mean relative standard deviation (RSD) values of percentile-normalized feature intensities were 8.0 $\pm$ 3.5% for machine replicates and 9.1 $\pm$ 4.0% between analytical replicates.[20] For a clarification of this terminology see Chapter 1, Figure 1.3. We observed retention time drift on the order of 3-4 s and mass accuracy of 500 ppm across the dataset. Despite the use of a surrogate to correct for sample loss and instrument drift, we did not achieve high accuracy or precision in these metrics.

The most severe drawbacks to LC-MS metabolite profiling is (1) dynamic range of the TOF detector especially when coelution occurs, which cannot be accounted for by performing a dilution and (2) retention time drift, which has been reported in the literature to be compensated for in post-processing using XCMS software.[21] In our hands, this was not possible. MarkerLynx provided a sufficient means for feature selection and data dimensionality reduction via automated peak integration.

In some cases, semi-quantification has been reported by normalizing the internal standard response to 1 in all LC-MS TICs.[14] The abundance of kaempferol glycosides in Arabidopsis leaves was in many cases one order of magnitude higher than quantities observed for quercetin, isorhamnetin and cyanidine, indicating that all samples could be run twice, once at full strength and again at a 10x diultion.[22] Where pure standards are available for the secondary metabolites of interest, concentration can be determined using a standard curve.[23] Fold changes between control and mutant plants have been reported, using biochanin A as an internal standard, as well as distinct differences between root (21 signals) and shoot (24 signals) secondary metabolomes.[24]

A key advantage of multivariate data analysis is that we can identify significantly changing variables that are not apparent by visual inspection of LC-MS chromatograms. The MarkerLynx software was used to deconvolute coeluting ions by reducing the data to retention time and *m/z* pairs prior to integration.

Working in a model system like Arabidopsis gives a researcher the option to order corresponding mutants to test whether the presence of the gene encoding the biosynthetic capability of the plant to make such a molecule is compromised and whether that affects

the plant's ability to survive low oxygen stress. The following literature references point to potential future hypotheses to be tested. In Arabidopsis roots, several genes that encode enzymes in cell-wall, lipid and flavonoid biosynthesis and in defense responses are repressed under low oxygen conditions.[25] Anthocyanin production has been observed under hypoxia and contributes to survivability.[26] In intact plants, hypoxia induces a decrease in chlorogenic acid (less phenolics).[27] Flavonol glycosides are thought to mediate ROS,[28, 29] may be produced under hypoxia in preparation for oxidative burst upon reoxygenation. A coupled reaction quenches ROS by reducing ascorbate to dehydroascorbate (DHA) while the flavonol glycoside is oxidized to a semiquinone.[30, 31] Quantitative real-time PCR was used to evaluate gene expression of messages encoding enzymes in the flavonoid biosynthesis pathway in common buckwheat (*Fagopyrum esculentum*) in conjunction with HPLC for quantitation of flavonoid products in different organs, sprouts, seeds, flowers and leaves. An outcome of this was the finding that the content of rutin was 60% greater at 12 days after sowing when plants were grown in the light instead of the dark. Gallic acid was only found in sprouts grown in light.[23]

There are two possible explanations for why the abundance of these flavonoids are altered in response to low oxygen stress. First, if the molecules are short lived, then they may be turned over and not replenished under NADPH- and ATP-limited conditions. Second, if these flavonoids help to fight ROS, then ROS formed during the transition from normoxia to hypoxia could reduce their abundance and again they would likely not be replenished. There remain open questions in this work with regard to exploration of markers of recovery. Figure 6.8 shows the mRNA data from Chapter 3

with a focus on secondary metabolism. Genes indicated by arrows are identified in the legend, which may provide a foundation for follow-up experiments together with this metabolite data.

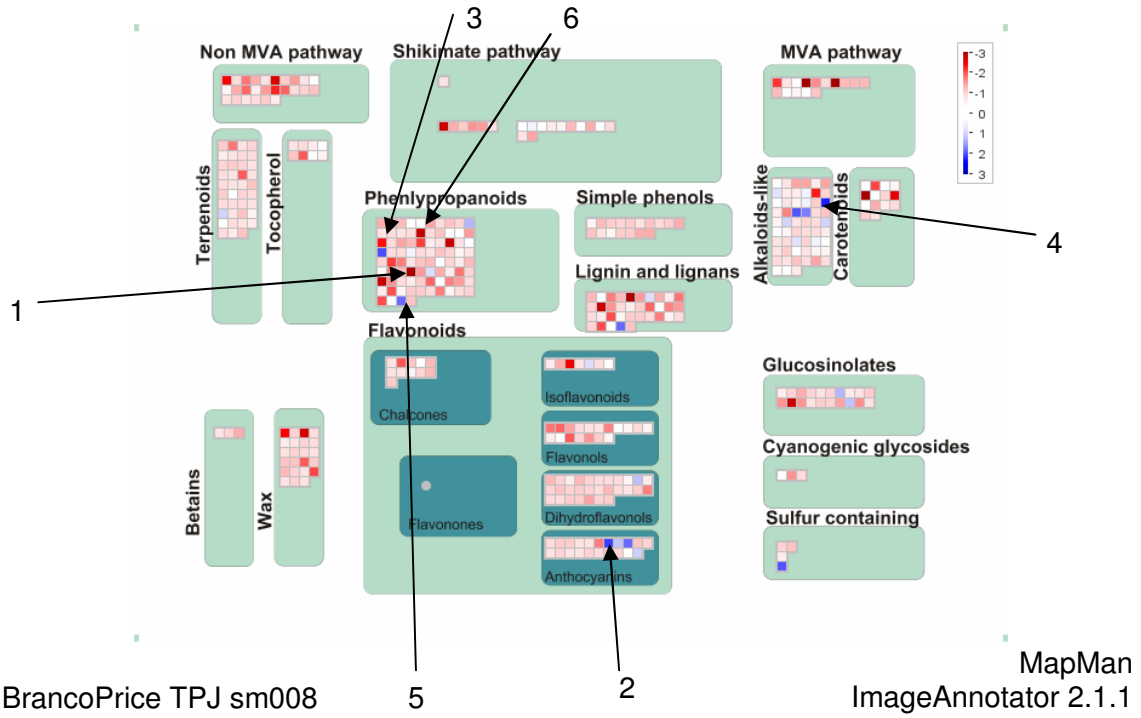


Figure 6.8 MapMan outputs with normalized relative transcript abundance shown as signal $\log_2$ ratio (SLR). Induction is shaded in blue while downregulation is in red. The abundance of mRNAs involved in secondary metabolism are significantly altered in response to the perturbation. (1) At3g10340 Phenylalanine ammonia lyase (PAL) SLR = -3.0 (2) At5g39050 Anthocyanin 5-aromatic acyltransferase (anthocyanin-5-*O*-glucoside-6-*O*-malonyltransferase) SLR = 2.1 (3) At3g50270 anthranilate N-hydroxycinnamoyl/benzoyltransferase SLR = -2.3 and At2g33590 cinnamoyl-CoA reductase SLR = 2.0 (4) At4g23600 tyrosine aminotransferase / tyrosine transaminase SLR = 2.3 (5) At1g72680 cinnamyl-alcohol dehydrogenase (putative) SLR = 1.9 (6) At5g23220 isochorismate hydrolase (superoxide inducible protein) SLR = -2.8.

6.5 Conclusion

In conclusion, the use of software such as MarkerLynx for nontargeted exploration of secondary metabolomic information can facilitate identification of

molecular signatures corresponding to substances which change in abundance, either accumulating or becoming depleted, under low oxygen stress *in planta*. Due to the high biological variability (40-100%) no significant differences were observed between hypoxia-stressed and control seedling extracts in metabolites identified in Chapter 5. Hypotheses regarding the identity and function of biomarkers listed in Table 6.3 may lead to the better understanding of secondary metabolic reconfiguration to survive anoxia.

The biggest roadblocks to turning data into information in this study were two-fold (1) retention time windows were large, 3-4 min, and (2) baseline distortions may have influenced reliability of integrated peak intensities. Digital resolution was sufficient at 1 scan per second. Reproducible injections are critical. This is the reason for an internal standard, and normalizing to N7G was found to be effective for this purpose.

Determining which signals to include in the model was subjective, based on limit of quantification (LOQ), S/N 10 was a good starting point, but there may be a reason to set the threshold higher to exclude variables and generate a dataset which is manageable if noise seems to dominate the scores and loadings plots. This study also demonstrated that both hypothesis generating and targeted analyses can be performed at the data analysis stage, even on the same dataset. As we saw with NMR in Chapters 3 and 4, application of both targeted and unsupervised statistical methods were able to extract more information from the dataset. In this chapter, samples were interrogated by LC-MS only. This study also demonstrated that both hypothesis-generating and targeted analysis can be performed at the data analysis stage, even on the same dataset.

6.6 References

1. American Society of Plant Biologists, *Biochemistry & Molecular Biology of Plants* John Wiley & Sons: Somerset, NJ, 2000.
2. Berenbaum, M. R., The chemistry of defense: theory and practice. In *Chemical Ecology: The Chemistry of Biotic Interaction*, Eisner, T.; Meinwald, J., Eds. National Academy Press: Washington, DC, 1995; pp 1-16.

3. Pichersky, E.; Gang, D. R., Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* **2000,** 5, (10), 439-445.

4. Walker, T. S.; Bais, H. P.; Halligan, K. M.; Stermitz, F. R.; Vivanco, J. M., Metabolic profiling of root exudates of *Arabidopsis thaliana*. *J. Agric. Food Chem.* **2003,** 51, 2548-2554.

5. Boccard, J.; Grata, E.; Thiocone, A.; Gauvrit, J.-Y.; Lanteri, P.; Carrupt, P.-A.; Wolfender, J.-L.; Rudaz, S., Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding. *Chemom. Intell. Lab. Syst.* **2007,** 86, 189-197.

6. Schauer, N.; Fernie, A. R., Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci.* **2006,** 11, (10), 508-516.

7. Kliebenstein, D. J., Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant Cell Environ.* **2004,** 27, 675-684.

8. D'Auria, J. C.; Gershenzon, J., The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr. Opin. Plant Biol.* **2005,** 8, 308-316.

9. Bednarek, P.; Schneider, B.; Svatos, A.; Oldham, N. J.; Hahlbrock, K., Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism in Arabidopsis roots. *Plant Phys.* **2005,** 138, 1058-1070.

10. Tan, J.; Bednarek, P.; Liu, J.; Schneider, B.; Svatos, A.; Hahlbrock, K., Universally occurring phenylpropanoid and species-specific indolic metabolites in infected and uninfected *Arabidopsis thaliana* roots and leaves. *Phytochemistry* **2004,** 65, 691-699.

11. Kramer, P. J.; Boyer, J. S., Chapter 5: Roots and Root Systems. In *Water Relations of Plants and Soils*, Academic Press: San Diego, CA, 1995; p 495.

12. Shirley, B. W.; Kubasek, W. L.; Storz, G.; Bruggemann, E.; Koornneef, M.; Ausubel, F. M.; Goodman, H. M., Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *Plant J.* **1995,** 8, (5), 659-671.

13. Buer, C. S.; Imin, N.; Djordjevic, M. A., Flavonoids: new roles for old molecules. *Journal of Integrative Plant Biology* **2010,** 52, (1), 98-111.

14. Böttcher, C.; Roepenack-Lahaye, E. v.; Schmidt, J.; Clemens, S.; Scheel, D., Analysis of phenolic choline esters from seeds of *Arabidopsis thaliana* and *Brassica napus* by capillary liquid chromatography/electrospray-tandem mass spectrometry. *J. Mass Spectrom.* **2008,** 44, 466-476.

15. Cao, Y.; Jiang, T.; Girke, T., A maximum common substructure-based algorithm for searching and predicting drug-like compounds. *Bioinformatics* **2008,** 24, 13.

16. Boyes, D. C.; Zayed, A. M.; Ascenzi, R.; McCaskill, A. J.; Hoffman, N. E.; Davis, K. R.; Gorlach, J., Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* **2001,** 13, 1499-1510.

17. Greef, J. v. d.; Martin, S.; Johasz, P.; Adourian, A.; Plasterer, T.; Verheij, E. R.; McBurney, R. N., The art and practice of systems biology in medicine: mapping patterns of relationships. *J. Proteome Res.* **2007,** 6, 1540-1559.

18. Stalikas, C. D., Extraction, separation, and detection methods for phenolic acids and flavonoids. *J. Sep. Sci.* **2007,** 30, 3268-3295.

19. Abagyan, R. METLIN metabolite database. http://metlin.scripps.edu/ (accessed Jan 2, 2012)

20. Bottcher, C.; Roepenack-Lahaye, E. v.; Schmidt, J.; Schmotz, C.; Neumann, S.; Schell, D.; Clemens, S., Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant Phys.* **2008,** 147, 2107-2120.

21. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006,** 78, 779-787.

22. Stobiecki, M.; Skirycz, A.; Kerhoas, L.; Kachlicki, P.; Muth, D.; Einhorn, J.; Mueller-Roeber, B., Profiling of phenolic glycosidic conjugates in leaves of *Arabidopsis thaliana* using LC/MS. *Metabolomics* **2006,** 2, (4), 197-219.

23. Li, X.; Park, N. I.; Xu, H.; Woo, S.-H.; Park, C. H.; Park, S. U., Differential expression of flavonoid biosynthesis genes and accumulation of phenolic compounds in common buckwheat *(Fagopyrum esculentum)*. *J. Agric. Food Chem.* **2010,** 58, 12176-12181.

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

25. Geigenberger, P., Response of plant metabolism to too little oxygen. *Curr. Op. Plant Bio.* **2003,** 6, (3), 247-256.

26. Vashisht, D.; Hesselink, A.; Pierik, R.; Ammerlaan, J. M. H.; Bailey-Serres, J.; Visser, E. J. W.; Pedersen, O.; Zanten, M. V.; Vreugdenhil, D.; Jamar, D. C. L.; Voesenek, L. A. C. J.; Sasidharan, R., Natural variation of submergence tolerance among *Arabidopsis thaliana* accessions. *New Phytol.* **2011,** 190, 299-310.

27. Riewe, D.; Grosman, L.; Zauber, H.; Wucke, C.; Fernie, A. R.; Geigenberger, P., Metabolic and Developmental Adaptations of Growing Potato Tubers in Response to Specific Manipulations of the Adenylate Energy Status. *Plant Phys.* **2008,** 146, 1579-1598.

28. Bailey-Serres, J.; Mittler, R., The roles of reactive oxygen species in plant cells. *Plant Physiol.* **2006,** 141, (2), 311-311.

29. Bailey-Serres, J.; Voesenek, L. A. C. J., Flooding Stress: Acclimations and Genetic Diversity. *Annu. Rev. Plant Biol.* **2008,** 59, 313-339.

30. Hernandez, I.; Alegre, L.; Breusegem, F. V.; Munne-Bosch, S., How relevant are flavonoids as antioxidants in plants? *Trends Plant Sci.* **2009,** 14, 125-132.

31. Jaleel, C. A.; Riadh, K.; Gopi, R.; Manivannan, P.; Ines, J.; Al-Jiburi, H. J.; Chang-Xing, Z.; Hong-Bo, S.; Panneerselvam, R., Antioxidant defense responses: physiological plasticity in higher plants under abiotic constraints. *Acta Physiol. Plant* **2009,** 31, 427-436.

Chapter Seven

Conclusions and Future Directions

7.1 Conclusions

This dissertation explores the development of analytical methods suitable for recording global (nontargeted) metabolomic information as well as affording quantitation for a handful of abundant primary metabolites (targeted) in Angiosperms (flowering plants). Analogous measurements were applied to less-abundant secondary metabolites, requiring dereplication in the absence of authentic standards. The methods used included 1D ^{1}H NMR, 2D ^{1}H-^{1}H NMR, 2D ^{1}H-^{13}C NMR, LC-NMR, HPLC, UPLC-MS.

1D ^{1}H NMR proved useful for metabolite identification and relative quantitation in complex mixtures. In Chapter 2, 1D ^{1}H NMR was used to record the effect of sample preparation on the metabolites extracted from wet or dry mature plant leaves. Quality was assessed with respect to reproducibility, where 1D ^{1}H NMR provided excellent performance by contributing immeasurable instrument variability and providing a direct readout of quantity of all mixture components. In Chapters 3 and 4, 1D ^{1}H NMR was applied to investigate a problem of biological interest: reconfiguration of primary metabolism in response to low oxygen stress (hypoxia/anoxia) in the model angiosperm *Arabidopsis thaliana.* Advances in experimental design lead to quantification of higher numbers of metabolites, and higher resolution of organ-specific metabolic profiles in seedlings.

1D ^{1}H NMR was also used for identifying secondary metabolites enriched by solid-phase extraction (SPE) using a hydrophilic-lipophilic balanced (HLB) stationary phase for capture and purification. In Chapter 5, HPLC coupled to 1D ^{1}H NMR provided enough structural information to establish tentative assignments to eight distinct but structurally related phenolic compounds.

LC-MS was also vital in this work for providing complementary structural information, particularly because plant extracts contain thousands of molecules diversified by combinatorial biosynthesis for which standards were not available. In Chapter 5, UPLC-MS provided efficient handling of limited quantities of sample and solvents throughout chromatographic method development. Gradient elution developed by LC-MS established conditions for peak trapping in LC-NMR, where structure assignment in 1D ^{1}H NMR is more straightforward when complete resolution of components is achieved chromatographically. In Chapter 6, UPLC-MS was applied to investigate the secondary metabolism of seedlings under prolonged low oxygen (anoxia) stress and recovery.

7.2 Future work

In the area of hypoxia, open questions remain regarding light availability, nutrient availability, temperature and combination of stresses and the impact on primary metabolism and energy utilization. The interplay between carbon and nitrogen metabolism is of interest.[1, 2] Efforts to improve plant survival of stress may be mediated by exogenous application of nutrients, if signaling pathways and responses are known.

The effects of hypoxia stress pretreatment and its enhancement of survivability have not been fully investigated in Arabidopsis,[3, 4] although such work has increased the performance of heart transplants.[5] From this work it follows that interpretation of the recovery data acquired in Chapters 3 and 6 has potential to inform future hypotheses in this area.

With an understanding of molecular signaling under hypoxia, survivability and biomass or nutrient accumulation of crops could be increased. Chemical simulation of hypoxia pretreatment could be induced by exogenous application of some molecule, analogously to herbicide safeteners, allowing plants to better survive actual environmental crises.[6] Alternately, bioengineered plants with inducible promoters driving the expression of survivability-related genes could be planted and treated with an inducing agent in advance of environmental stress such as a flood or ice storm. The use of transgenic plants to determine which pathways are critical for survival may shed light on areas of current controversy, such as whether the TCA cycle is operating in a reversed mode or two distinct branches.[1, 7] Recent studies suggest metabonomic studies of abiotic stress in plants are ongoing and open questions remain in the field.[8] Environmental stresses caused by drought, elevated temperature, salinity and rising CO_2, or a combination of two or more of these stresses, pose a growing threat to agriculture.[9]

The role of osmolytes has not been thoroughly investigated, nor has the role of reactive oxygen species (ROS) and nitric oxide in signaling changes in oxygen availability. Experiments comparing phenotypic changes of roots and shoots under hypoxia in the light and dark[4] reporting changes in volatile metabolites, adenylates and

sugars, has been completed but a more detailed look at the molecular mechanisms utilized by Arabidopsis to survive hypoxia could be achieved using metabolic profiling. We aimed to determine how those mechanisms may be distinct in different organs. We hypothesized that photosynthetic capability of shoots as well as the specialized mechanisms for nutrient assimilation in roots would be distinctive. This is of relevance to agriculture because some crops are predominantly shoots (broccoli, spinach, lettuce) and some are roots (potato, ginger, carrot).

Analytically, the dedication of a pipeline for sample fractionation and characterization, data management and publication can significantly improve the throughput and quality of plant-based metabolomic studies. Expansion of our database infrastructure became necessary, and further efforts to assimilate computer science, chemometrics, and bioinformatics approaches to managing chemical data will enrich the outcomes. International collaborations become of the utmost importance, through literature connections, travel to scientific meetings or training and outreach within one's own campus. Although some believe that there is one best statistical test for each set of experimental data,[10] others argue that expanding the battery of statistical tests applied to multivariate data may serve the purpose of cross-validation of discovered biomarkers and trends within large datasets.[11]

Advances in computing power have enhanced the tracking and modeling of higher numbers of variables. Systems biology and biologistics represent a paradigm shift in biological research, providing the application of these new modalities of thought and experimental design were applied to a problem (oxygen deprivation) which has evolved

over 439 million years, since the appearance of land plants.[12] One could consider metabolites like passengers on railways, where metabolic pathways are analogous to train lines/routes, if one train is canceled, this eventually causes upstream and downstream effects, analogous to the network disruption we observe in response to low oxygen. The serious application of biologistics to understanding pathways is still in its infancy[13] and attempts at bioengineering have had limited success.[14-16] Further merging of these fields may prove fruitful, and the methods developed in this dissertation could be applied to studies such as these.

Metabolic profiling of primary and secondary pathways in plants will enhance studies of chemical ecology, biodiversity, conservation biology, and systematics. Biological questions regarding cell-to-cell or plant-to-plant communication, networks, signaling, coordination, symbiosis, defense, growth, and adaptation may be addressed using careful sample preparation and dereplication or proposal of a novel structure if necessary. Since estimates of chemical diversity in metabolomes exceed current entries in databases, there exist molecules yet to be described which may have biological activities of interest in isolation or in a concerted mechanism with other metabolites. New solid phases for chromatography and/or sample preparation may be critical for exploration of novel chemical space within secondary metabolomes.

Plant metabolite profiling has the potential to play a key role in the standardization of herbal medicines for quality control of feedstocks, to facilitate metabolic engineering or selective breeding, and for bioactive natural product discovery.[17] A majority of the over 7 billion people in the world (80%) utilize plants to

meet their primary health care needs.[18] From local, native plants, pharmacognosy and ethnobotany are born. Studies of cultural and holistic uses of plants can be treated scientifically, using a plant metabolomic approach. Western medicine, in particular absorption, distribution, metabolism and excretion (ADME) studies investigating the potential of combination therapies, can also benefit from the analytical technologies and computational approaches to data analysis developed to meet the challenges of working in complex plant extracts. Analogously to ADME studies of western medicines (small-molecule drugs), bioavailability of vitamins and secondary compounds is of further interest.

It follows naturally that plant metabolomics and metabolic profiling has the potential to improve agriculture. Non-targeted genome-wide functional genomic studies incorporating metabolic profiling may reveal genes that can improve crop production (yield) through resistance to biotic and/or abiotic stress. Perhaps also the nutritional value (quality) of foods can be improved. In these ways, metabolomics has the potential to aid in the alleviation of poverty and hunger.[19] By improving these factors, the security of all people is improved. In areas where traditional farming employs fertilizers and pesticides against pests and pathogens, issues of pollution and sustainability may be addressed in an ecological (metabolic) context. With globalization becoming the norm, trade among countries in produce and goods is common. Methodology for standardization in terms of quality and source authenticity of plant materials is already of interest. Besides the use of plants for food or pharmaceuticals, industrial uses such as soap, biofuel, even plastics can also continue to be explored.

As a final word on future avenues, it seems we will want biotracking in more than one dimension. Ideally we would track the location and transport, incorporation or modification of a substrate, even when the organism incorporating it is in symbiosis with another organism. Heterorganismal metabolome-metabolome interactomes will perhaps be the next facet of systems biology. This type of experiment would provide broad-spectrum bioactivity by providing a complement of molecules as inputs from organism A to organism B. One can even imagine modeling entire ecosystems.

There exists an exciting future for micrometabolomics as well (i.e. of specific cells, cell types or tissues). Interesting avenues for studies include; localization *in planta*, studies relating structure, amount, regulation, localization and biological activity, cross-talk between secondary metabolite biosynthetic pathways, understanding the ecological importance of each secondary metabolite, and studies correlating natural variation in secondary metabolism with plant fitness or interactions with the environment.[20]

7.3 References

1. Fait, A.; Fromm, H.; Walter, D.; Galili, G.; Fernie, A. R., Highway or byway: the metabolic role of the GABA shunt in plants. *Trends Plant Sci.* **2007,** 13, (1), 14-19.

2. Lam, H.-M.; Coschigano, K. T.; Oliveira, I. C.; Melo-Oliveira, R.; Coruzzi, G. M., The molecular-genetics of nitrogen assimilation into amino acids in higher plants. *Ann. Rev. Plant Phys.* **1996,** 47, 569-593.

3. Mustroph, A.; Albrecht, G., Fermentation metabolism in roots of wheat seedlings after hypoxic pre-treatment in different anoxic incubation systems. *J. Plant Physiol.* **2007,** 164, (4), 394-407.

4. Ellis, M. H.; Dennis, E. S.; Peacock, W. J., Arabidopsis roots and shoots have different mechanisms for hypoxic stress tolerance. *Plant Physiol.* **1999,** 119, 57-64.

5. Oppegard, S. C.; Nam, K. H.; Carr, J. R.; Skaalure, S. C.; Eddington, D. T., Modulating temporal and spatial oxygenation over adherent cellular cultures. *Plos One* **2009,** 4, (9), 8.

6. Davies, J.; Caseley, J. C., Herbicide safeners: a review. *Pestic. Sci.* **1999,** 55, 1043-1058.

7. Sweetlove, L. J.; Beard, K. F. M.; Nunes-Nesi, A.; Fernie, A. R.; Ratcliffe, R. G., Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci.* **2010,** 15, 462-470.

8. Zhang, J.; Zhang, Y.; Du, Y.; Chen, S.; Tang, H., Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome. Res.* **2011,** 10, (4), 1904-1914.

9. Ahuja, I.; Vos, R. C. H. d.; Bones, A. M.; Hall, R. D., Plant molecular stress responses face climate change. *Trends Plant Sci.* **2011,** 15, (12), 664-674.

10. Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006,** 2, (4), 171-196.

11. Harlow, L. L., *The Essence of Multivariate Thinking*. Lawrence Erlbaum Associates: Mahwah, NJ, 2005; p 264.

12. Brown, B., Abiogenesis. In *Evolution: A Historical Perspective*, Baigrie, B., Ed. Greenwood Press: Westport, CT, 2007; pp 129-148

13. Helbing, D.; Deutsch, A.; Diez, S.; Peters, K.; Kalaidzidis, Y.; Padberg-Gehle, K.; Lammer, S.; Johansson, A.; Breier, G.; Schulze, F.; Zerial, M., Biologistics and the struggle for efficiency: concepts and perspectives. *Adv. Complex Syst.* **2009,** 12, (6), 533-548.

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

15. Sriram, G.; Fulton, D. B.; Shanks, J. V., Flux quantification in central carbon metabolism of Catharanthus roseus hairy roots by ^{13}C labeling and comprehensive bondomer balancing. *Phytochemistry* **2007,** 68, 2243-2257.

16. Kruger, N. J.; Ratcliffe, R. G., Dynamic metabolic networks: Going with the flow. *Phytochemistry* **2007,** 68, 2136-2138.

17. Shyur, L.-F.; Yang, N.-S., Metabolomics for phytomedicine research and drug development. *Curr. Opin. Chem. Biol.* **2008,** 12, 66-71.

18. Farnsworth, N. R., Medicinal Plants in Therapy. *Bull. W.H.O.* **1985,** 63, 965-981.

19. Love, B. E.; Spaner, D. *Improvement of the agricultural sustainability and livelihoods of poor farmers through biotechnology: reality or speculation?*; International Development Research Centre (IDRC): 2006.

20. Kliebenstein, D. J., Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant Cell Environ.* **2004,** 27, 675-684.