UNIVERSTIY OF CALIFORNIA
RIVERSIDE

Metabolomics of Complex Biological Systems to Uncover Molecular Mechanisms in Rice and Other Organisms

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

Doctor of Philosophy

in

Chemistry

by

Gregory Alan Barding, Jr.

June 2013

Dissertation Committee:
Dr. Cynthia K. Larive, Co-Chairperson
Dr. Julia Bailey-Serres, Co-Chairperson
Dr. Yinsheng Wang

The Dissertation of Gregory Alan Barding, Jr. is approved:

_____

_____

Committee Co-Chairperson

_____

Committee Co-Chairperson

University of California, Riverside

Acknowledgements

There are many to acknowledge for their support and guidance on this adventure. First, I would like to thank my wife for her endless support, guidance, and wisdom during graduate school. You kept me motivated and never complained about how occupied school kept me. Next, my family. There are many to list. My parents, of course, whom always encouraged me to excel in everything I attempted and provided me with the tools and discipline to be successful. My grandma and grandpa Barding, for giving me a room unconditionally after my brief venture in Florida. I am not sure I would have been able to attend graduate school without that opportunity. My grandparents Dwyer, whose wisdom kept me going, and for being a short enough drive away for visits but far enough to call a vacation. Those brief breaks were essential to my well-being. My grandma and grandpa Albers for always being supportive. All of my siblings, of course, are awesome. My coworkers also deserve a lot of thanks for putting up with me over the last 5 years. My friends on the other side of the Atlantic: Szabolcs and Orsi, helping with my experiments and being great hosts while I was visiting Budapest. Hans, for putting up with my craziness while I was in Utrecht. Divya, an amazing individual, always smiling and upbeat even while studying in California (Black Friday will never be the same). We look forward to visiting all of them over the next few years. Lastly, I would like to thank my advisors, Prof. Cynthia Larive and Prof. Julia Bailey-Serres. I could not have asked for better guidance and support during this time of my education and personal growth.

Dedication

This dissertation is dedicated to my wife, Kristina Marie, our son, Alexander Robert, and to the members of our family we have not yet had the opportunity to meet.

Acknowledgment of Copyright

ABSTRACT OF THE DISSERTATION

Metabolomics of Complex Biological Systems to Uncover Molecular Mechanisms in Rice and Other Organisms

by

Gregory A Barding, Jr.

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, June 2013
Dr. Cynthia K. Larive, Co-Chairperson
Dr. Julia Bailey-Serres, Co-Chairperson

The goal of this dissertation is to explore the metabolic reconfiguration of the submergence intolerant rice variety *Oryza sativa* ssp. *japonica* cv. M202 and the submergence tolerant cv. M202(*Sub1*) during and after submergence stress. This work uses multiple analytical techniques, including nuclear magnetic resonance (NMR), gas chromatography – mass spectrometry (GC-MS), and reversed-phase ion-pair ultra-high performance liquid chromatography – mass spectrometry (RPIP-UPLC-MS) to explore the metabolite profiles of the tolerant and intolerant rice varieties in response to abiotic stress.

Untargeted metabolomics was used to query the metabolic impact of submergence on the tolerant and intolerant rice varieties. In results obtained by both NMR spectroscopy and GC-MS, the metabolic profiles of the two varieties diverged as early as 1 d of submergence. Using both techniques, the tolerant variety was shown to significantly conserve carbohydrate resources. A comparison of the NMR and GC-MS

results demonstrated that multiple analytical techniques provide the best metabolite coverage. Additionally, a previously unreported NMR resonance was identified as belonging to the methyl protons of the dipeptide alanylglycine (AlaGly). Although the metabolic function of the dipeptide is unclear, AlaGly was the only metabolite to decrease during submergence and not recover during the 1 d recovery period.

RPIP-UPLC-MS analysis provided a more targeted approach to metabolite profiling because of the sample preparation method employed and the selectivity of the RPIP separation for hydrophilic and ionic analytes. Phosphorylated mono- and disaccharides were targeted because of their signaling role in plant metabolism. For example, changes in trehalose-6-phosphate (T6P) levels, a key metabolite responsible for carbon sensing and flowering in plants, were monitored during and after submergence in the tolerant and intolerant rice varieties.

The development of automated and high-throughput chemometric techniques is critical for advances in disease detection, biomarker identification, and toxicological profiling. The coupling of untargeted metabolite profiling by NMR spectroscopy with a new chemometric approach based on *z*-score analyses, Visual Interpretation of Z-Score Ratios (VIZR), was used to determine differences in spectra from human urine and due to dietary supplementation. This approach can also be extended to biomarker identification in plants and other organisms.

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction

Based in part on papers published in the Encyclopedia of Magnetic Resonance, Journal of Proteome Research, and Analytical and Bioanalytical Chemistry

Encycolpedia of Magnetic Resonance, Harris, R. K.; Wasylishen, R. E., Eds. John Wiley: Chichester, 2011

Anal. Bioanal. Chem. 2012, 404, 1165-1179

J. Proteome Res. 2012, 11, 320-330

J. Proteome Res. 2013, 12, 898-909

The primary goal of the research presented in this dissertation is to apply analytical techniques in biological systems to better understand the metabolic impact of abiotic stress on crop species. Specifically, the metabolism of the near isogenic lines *Oryza sativa* ssp. *japonica* cv. M202 (submergence intolerant) and *Oryza sativa* ssp. *japonica* cv. M202(*Sub1*) (submergence tolerant) will be compared during and after submergence stress to better understand the effect a single gene can have on crop survival during extreme flooding events. Because the metabolome is comprised of a diverse array of compounds with a wide variety of sizes, structures, and functional groups, the use of multiple analytical techniques is required to fully characterize metabolic changes. Metabolite profiling experiments can be conducted using targeted and untargeted approaches, and both are used in this work. This dissertation describes the identification and quantitation of a wide variety of metabolites using mass spectrometry (MS) and

nuclear magnetic resonance (NMR) experiments as well as the development of a novel chemometric technique for biomarker identification and comparison of individual samples within a group.

This dissertation builds upon the work of previous group member Dr. Kayla Kaiser to better understand the effects of low-oxygen stress on plant metabolism and applies ^{1}H NMR, gas chromatography - mass spectrometry (GC-MS), and solid phase extraction (SPE) and reverse-phase ion-pair (RPIP) liquid chromatography - mass spectrometry (LC-MS) for the targeted identification and quantitation of anionic compounds, and the development of chemometric techniques for biomarker discovery, toxicological studies, and sample discrimination. The goals of the research presented in this dissertation are carried out through the following objectives:

Objective 1: Compare the metabolic differences of the M202 and M202(*Sub1*) cultivars resulting from both short term and long term submergence stress using ^{1}H NMR spectroscopy (Chapter 2).

Objective 2: Expand the coverage of the metabolomes of the two cultivars during short term submergence stress and control conditions using GC-MS analysis and compare the results with those obtained by NMR spectroscopy (Chapter 3).

Objective 3: Develop an RPIP-LC-MS method for the detection and quantitation of phosphorylated compounds and monitor their changes in rice tissue during submergence and recovery (Chapter 4).

Objective 4: Develop an automated chemometric technique using *z*-scores for the statistical discrimination of individual biosamples and identification of potential biomarkers (Chapter 5).

This introductory chapter first presents background information about metabolomics followed by discussions of the applications of NMR, GC-MS and LC-MS for the identification and quantitation of metabolites. The chapter is concluded with a discussion of chemometric methods for the analysis of biological data sets.

1.1 Metabolomics

Metabolomics focuses on measurements of small molecule metabolites in complex biological samples. Recent advances in analytical platforms, sample preparation protocols, hyphenated techniques, throughput, and statistical analysis methods have enabled metabolic measurements addressing increasingly complex problems in plant biology. These advances are enabling a deeper understanding of the metabolic responses of plants and other organisms to genetic and environmental perturbations.

The study of metabolism can be pursued through a variety of strategies.[1] One strategy is to adopt a "targeted analysis" approach which studies a specific genetic alteration and the resulting metabolic product. "Metabolite profiling" experiments seek to understand the function of an entire pathway or interactions between pathways by

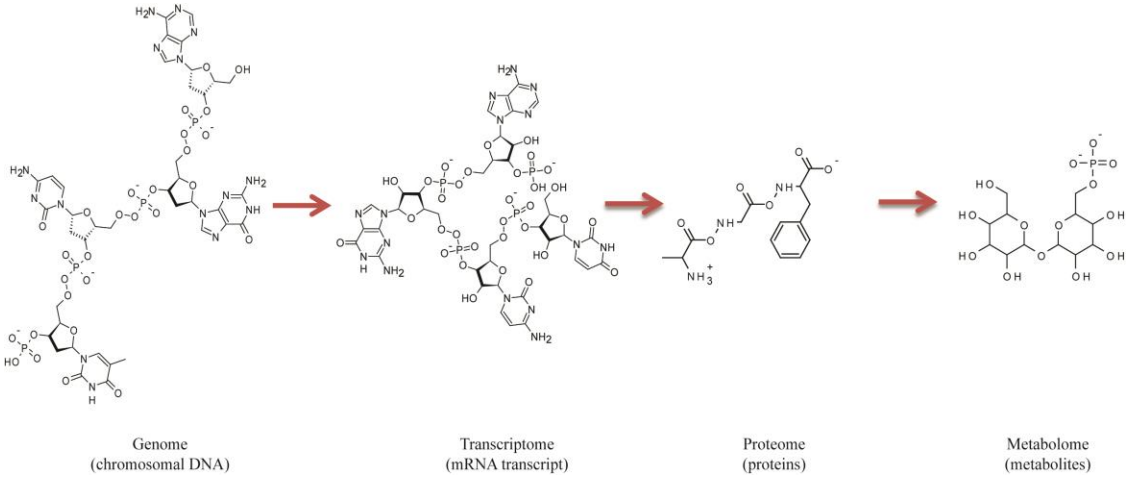


Figure 1.1. A general schematic representing the organization and relationship of the components of systems biology. The production of a metabolite does not necessarily correlate with the response of the preceding steps but can be dependent on translation, protein activation, and other biological signals.

targeting a specific class of molecules for analysis, often using enzymatic assays or labeled compounds. In contrast to targeted approaches, "metabolite fingerprinting" does not necessarily focus on the identification of specific metabolites but rather examines global metabolic changes as a function of a known perturbation such as a genetic mutation, biotic or abiotic stress, etc. Finally, the comprehensive analysis of the metabolites in a biological system is termed "metabolomics" although this term is sometimes mistakenly applied to the above strategies. Regardless of the approach chosen, delineating metabolic responses of an organism to stress conditions, genetic modifications, and environment perturbations is pivotal to understanding plant biology. Figure 1.1 shows the general organization and relationship of the "omics" from DNA to metabolites.

Although the relationships in Figure 1.1 appear to be linear, many factors can influence the production of a metabolite. Metabolites are produced via biochemical pathways operated by specific enzymes. An enzyme is a single protein or a complex of proteins encoded by one or multiple genes in the chromosome. Coordinated production of a gene transcript and the encoded enzyme, as well as other enzymes required for a pathway, is one prerequisite for metabolite production. However, there are multiple levels of regulation following gene transcription, including the synthesis, accumulation and subcellular targeting of the protein. There are many examples where the abundance of mRNA transcript levels (transcriptome) do not necessarily correlate with the production of the encoded protein, such as an enzyme. When pathways require multiple enzymes, the absence of one enzyme can block the accumulation of the metabolite, sometimes

resulting in increased levels of the precursor. Similarly an enzymatic reaction to produce or catabolize a metabolite may be suppressed because the enzyme is not in an active state or because the reaction may be limited by the levels of a required cofactor. Only when all components of the schematic represented in Figure 1.1 are considered together, can a more thorough understanding of systems biology be obtained.

Plant metabolomics poses a unique challenge compared with mammalian systems due to the complex and sessile nature of plants. Plants are estimated to produce as many as 200,000 metabolites, for many of which structures are unknown.[2, 1] Plant metabolomics has been employed to understand the response of plants to hypoxia stress,[3] improve crop stress-resistance,[4, 5] identify the botanical origins of a product, such as honey,[6] and evaluate wine quality.[7]

A variety of analytical platforms have been used in plant metabolomics studies. These techniques include nuclear magnetic resonance spectroscopy (NMR)[8-10] as well as mass spectrometry (MS) hyphenated with gas chromatography (GC),[11, 12] capillary electrophoresis (CE),[13] and liquid chromatography (LC).[14] Mass spectrometry is thus far the favored platform for plant metabolomics studies because it offers greater sensitivity and simple coupling to chromatographic separations, however a more thorough exploration of the metabolome is best achieved through a combination of both NMR and MS.

1.2 NMR-based metabolomics studies

NMR has been widely used over several decades to explore biochemical pathways of mammals, plants, fish, fungi, and insects as well as in biomarker

identification for disease diagnosis and juice adulteration. [15, 8, 16-20] Although less sensitive than MS-based analyses, NMR is non-destructive, inherently quantitative, and requires no sample derivatization.[21] With these advantages, NMR was introduced for metabolite profiling as early as the 1970's when Brown and coworkers used ^{1}H NMR spectra to profile metabolites in intact human erythrocytes.[16] Additional metabolite profiling experiments were reported in the early 1980's by Moore et al.[22] and Nicholson and coworkers.[23] Recent work using NMR in mammalian systems has led to clinical application of NMR metabolite profiling of intact tissues,[24] pharmacogenomics,[25] and evaluation of the geographical dependence of cardiovascular risks.[19]

Despite the wide variety of organisms studied by NMR spectroscopy, mammalian and plant systems have been the most widely examined due to the relationship of metabolism to human disease and the interest in crop viability for increased food production and sustainability.[26, 27] Due to pronounced biochemical differences between organisms, even within a given family, and tissue types, the methods selected for metabolite extraction, and data analysis, pre-processing, and interpretation must be addressed specifically for the intended study due to their effects on absolute or relative quantitation in addition to the consideration of platform specific experimental parameters.[10, 28]

The goal of a metabolomics experiment is to not only to identify the metabolites present in a sample but to quantify changes in the levels of detected metabolites in response to a biotic or abiotic stress. NMR experiments for either absolute or relative quantitation require careful consideration of the acquisition and processing parameters to

ensure a robust analysis.[29] The resonances of interest must be of adequate signal-to-noise ratio (S/N). For absolute quantitation, a S/N in excess of 100:1 is preferred while for relative quantitation, a S/N in excess of 10:1 can be acceptable.[30, 31] Signal averaging is a simple method for increasing S/N, however, since gains are proportional to the square root of the number of scans coadded, significantly longer analysis times may provide only modest S/N improvements.[32]

1.2.1 Considerations for Quantitative Metabolomics Experiments by NMR

Szántay et al. discussed other ways of increasing S/N, including using higher magnetic fields, cryoprobes, and microcoil probes.[33] NMR sensitivity increases roughly as $B_0^{3/2}$, where B_0 is the strength of the applied magnetic field. Therefore switching from a 9.4 T magnet (400 MHz ^{1}H) to one operating at 18.8 T (800 MHz ^{1}H) will increase sensitivity by a factor of about 2.8.[34] Unfortunately, as the incremental cost for purchasing higher field magnets is also nonlinear, this option may be financially impractical and investigators are typically limited to performing experiments using the instruments available at their site. Despite their expense, higher field magnets are important for the advancement of NMR applications because of increased resolution, relaxation effects, and physical effects (such as TROSY) and the spectrometer should be chosen according the experimental requirements. Cryogenically cooled NMR probes and receivers, also known as cryoprobes or cold probes, are an alternative approach to increasing S/N, providing up to a factor of four increase in sensitivity compared to conventional probes.[35] Cryoprobes increase S/N by decreasing contributions from Johnson noise by cooling the coil and preamplifier to ~25 K. The S/N improvement

actually achieved may be significantly less than four as cryoprobe performance can be adversely affected by high salt concentrations which are common in biological samples. Despite their sensitivity advantage, cryoprobes are expensive to maintain and are often used with a dedicated magnet. For mass-limited samples, microcoil or microstrip probes provide an alternative mechanism for increasing NMR sensitivity.[34, 36] These probes have been reported to detect as low as picomole quantities of analytes, and can enable the acquisition of NMR spectra for samples only available in limited amounts, which would be difficult or impossible to analyze using conventional instrumentation. Additionally, microcoil probes can be easily coupled with online separations such as capillary electrophoresis or capillary isotachophoresis, as reviewed by Jones and Larive.[37]

Regardless of the hardware available to the experimenter, careful selection of acquisition and processing parameters is important in quantitative NMR experiments. Keeping acquisition times short without truncating the free induction decay (FID) reduces the amount of noise collected; an acquisition time that is three times the apparent transverse relaxation time (T_2*) will acquire 95% of the signal. Adequate digital resolution, which depends on the number of data points sampled during the acquisition time and the spectral width, is required if the spectra are to be used for quantitation. Each well-resolved resonance of interest should be represented by at least 5 data points to obtain a reliable integral, and more may be needed to accurately deconvolve overlapped resonances, as discussed in section 1.2.7.[38] The digital resolution of the spectrum can be improved by zero filling, which involves the addition of zeros to the end of the acquired FID. Linear prediction can increase digital resolution by predicting the values of missing

data points, providing a better approximation of the data than is achieved through zero filling. Use of linear prediction can also decrease the degree of apodization required by reducing FID truncation errors. The application of weighting functions can increase the S/N but at a cost in spectral resolution.[39, 38]

The quality of quantitative NMR experiments can be evaluated by determining the accuracy, precision, and error/uncertainty associated with the measurement. Precision is defined as the closeness of agreement between a series of measurements of the same sample. Precision of integration is also dependent on the S/N of the measured signals. If the S/N is greater than 200:1, then a precision greater than 99% can be achieved.[30] To evaluate precision, the repeatability and comparability can be determined. Measurement repeatability measures the same sample multiple times by the same instrument.[31] Comparability can be evaluated through inter-laboratory measurements,[40] where the same sample is measured using different spectrometers operating at different magnetic fields and by different users. If the measurements are performed carefully with the same acquisition parameters, a coefficient of variation less than 1% has been observed for both repeatability and comparability.[30] Additionally, all errors and uncertainties should be evaluated for the whole measurement procedure, including molar ratios, assay determinations, atomic masses, uncertainties in instruments, and the standard reference material.[30, 29]

1.2.2 Absolute and Relative Quantitation by NMR

The longitudinal (T_1) relaxation rate is also an important consideration in a NMR experiment. Absolute quantitation requires the nuclei to be fully relaxed prior to each

scan using a recycle period that is at least 5 times the longest T_1 relaxation time of the species to be measured.[30, 29] Alternatively, it is possible to use a shorter relaxation delay and correct for incomplete relaxation using T_1 relaxation times measured in the sample matrix. For metabolomics experiments, which are mostly based on relative quantitation, repetition rates faster than $5T_1$ increase sample throughput and permit comparison of samples in different classes provided that all experimental parameters, including the sample matrix, are consistent.

1.2.3 Spectral Referencing for NMR Experiments

Spectral referencing (chemical shift referencing) and internal calibration standards are another important consideration for an NMR metabolomics experiment. Several compounds, such as TMSP (sodium-3-trimethylsilylpropionate-d_4) and DSS (sodium-2,2-dimethyl-2-silapentane-5-sulfonate-d_6) are commonly used to reference spectra and can also be used as internal calibration standards.[41, 42] There are several specific requirements that chemical shift reference and internal calibration standards must meet: the reference peaks should not overlap with other signals in the sample, the compound should produce a small number of well-defined signals (such as singlets), it should be chemically stable and not adversely effected by the experimental pH.[41] Several recent articles have described a variety of compounds potentially useful as internal standards and chemical shift reference compounds.[31, 41, 42]

1.2.4 Solvent Suppression

Because many biological samples (for example, urine and serum) are in an aqueous matrix, solvent suppression is an important consideration for ^{1}H NMR

experiments. By allowing use of the maximum receiver gain, suppression of the solvent resonance can increase the S/N of the spectrum allowing better quantitation of lower abundance metabolites. Solvent suppression also reduces radiation damping, a common problem in high Q-factor probes such as cryoprobes. Although a wide range of solvent suppression methods have been introduced, most have the disadvantage that they also suppress (at least partially) analyte resonances close in frequency to the solvent peak, rendering their quantification unreliable.

Presaturation or selective saturation, which uses a long low power pulse at the frequency of the solvent resonance, is the most common water suppression technique and among the simplest to implement.[43] A disadvantage of presaturation is that it can result in the transfer of saturation to exchangeable protons, such as amides, making their quantification impossible. One of the biggest drawbacks of presaturation is that it can cause baseline distortions around the suppressed region, negatively impacting quantitation of analyte signals.

For effective solvent suppression with minor effects on exchangeable protons, gradient-based suppression techniques can be used. Water suppression enhanced through T_1 effects (WET) is a widely used solvent suppression method. WET is a pulsed-field gradient (PFG) based experiment which uses a selective pulse to excite the solvent resonance which is then attenuated using dephasing gradients.[44] WET uses a shorter pulse than presaturation so it is less likely to attenuate exchangeable protons. It avoids baseline distortions and nulls, and can easily be used to simultaneously suppress multiple solvent signals. There is also a version of WET, termed Secure WET, for use on

cryoprobes.[45] Disadvantages of WET solvent suppression include the need to create the shaped pulse and optimize the gradient powers to ensure optimal solvent suppression. If the WET parameters are set improperly, baseline distortions around the solvent signal can occur interfering with quantitation of neighboring resonances.

The WATERGATE (water suppression by gradient-tailored excitation) pulse sequence also employs PFG's to attenuate the water resonance. In this method, a nonselective excitation pulse is followed by a gradient pulse that dephases all the resonances. Then a composite pulse is applied to effectively invert all of the resonances except for the solvent.[46] A second gradient pulse rephases the analyte resonances while further dephasing the solvent resonance. Although WATERGATE is a useful suppression technique for quantitation of exchangeable protons, the delays used in the composite pulse create "nulls" or regions of the spectrum that are not rephased by the second gradient pulse. As a result, a flat excitation profile is not obtained and resonances in the regions around the nulls will be attenuated to varying degrees. An additional disadvantage is that WATERGATE can have an adverse effect on strongly coupled spin systems, creating baseline distortions and hindering quantitation of these peaks. An alternative approach, excitation sculpting, uses selective 180° pulses bracketed by gradients to dephase the solvent resonance while keeping the desired resonances in the detection plane.[47] Excitation sculpting produces spectra with a flat baseline and gives a uniform excitation profile, giving it an advantage over other solvent suppression pulse sequences.

1.2.5 Baseline Correction and Background Subtraction

A flat spectral baseline without phase distortions, rolling, and artifacts increases the reliability of quantitative data. Baseline correction is a more significant problem in spectra measured on older instruments that use analog filters. The digital receivers used in newer spectrometers eliminate the distortions produced by analog filters and also remove the imbalance between quadrature channels that can occur with analog receivers.[48] Although the digital receivers used in newer instruments produce spectra with less baseline distortion, other factors can contribute to baseline problems in both instrument architectures. One source of baseline distortion is the first order phase correction required to compensate for the dead time following the rf pulse and before data acquisition. While ideally data acquisition would begin immediately following the rf pulse, this would distort the intensity of the first points in the FID due to acoustic ringing in probe. In addition, once turned on, the receiver requires some time to produce a linear response. Correction of the first points in the time domain can be achieved through backwards linear prediction to replace the corrupted data points. A common frequency domain correction method involves subtracting a polynomial from the spectrum to give a flattened baseline, but alternative methods such as cubic spline can also be applied. Baseline correction in the frequency domain requires large peak-free noise regions to define the baseline. An alternative approach to baseline correction in the frequency domain described by Xi and Rocke is based on a penalized parametric smoothing model that works especially well with the crowded spectra encountered in metabolomics experiments.[49]

A high spectral background due to signal overlap with the broad resonances of proteins and lipids or the sharp peaks of metabolites, such as glucose, present at high abundance can interfere with the quantitation of analyte resonances. Although several sample preparation methods allow for removal of classes of compounds (i.e., protein precipitation, solid-phase extraction, liquid/liquid extraction, etc.) these can also reduce the precision of the analysis, increase the experiment time and sometimes lead to the analyte loss in the process of removing the targeted interferent. Several spectroscopic methods have been developed as alternatives to address the problem of signal overlap. Spectral subtraction (or difference) is a technique in which a spectrum containing only the background is subtracted from that of the sample. [50] Ideally, the resulting spectrum will show only the resonances of interest; however, spectral subtraction can also result in signal loss, subtraction artifacts and baseline distortions that limit the quantitative information obtained from an NMR experiment. Removal of high concentration analytes while maintaining the quantitative information provided by NMR would greatly increase the capabilities of NMR for biomarker detection. For instance, in normal blood plasma or serum samples or in the urine of diabetic patients, glucose is present at high concentrations complicating the measurement of less abundant metabolites due to resonance overlap.[51] To improve the suppression of the glucose peaks, Ye and coworkers developed a novel approach for compound-specific resonance removal, termed the "add to subtract" method.[52] This method is in essence an NMR-specific version of the general method of standard addition and calls for the addition of glucose to roughly double its concentration in the sample with acquisition of identical spectra before and after spiking.

Using this method, iterative manual fitting within the Bruker Topspin 3.0 software permitted effective subtraction of the glucose resonances decreasing their intensity by 98% while permitting visualization of resonances hidden under glucose. Comparison of the add to subtract method with normal spectral subtraction for urine and protein-precipitated blood serum samples showed the advantages of this approach in producing a clean baseline free of distortions and enabling the quantitation of metabolite resonances otherwise obscured by the more intense glucose peaks.

An alternative approach is the suppression of background resonances by taking advantage of differences in the NMR properties of analytes and those molecules that give rise to the background. The CPMG pulse sequence can be used to allow broad resonances, such as those produced by proteins, to decay through transverse (T_2) relaxation, leaving only the resonances of small molecules with slower T_2 relaxation rates to be detected.[53, 54] There are disadvantages to the CPMG approach for metabolomics measurements including baseline and phase distortions as well as intensity losses due to the T_2 relaxation of the analyte nuclei. Lucas et al. proposed the use of the stimulated echo (STE) pulse sequence as an alternative to the CPMG experiment and demonstrated that STE spectra of blood serum had better phase and baseline character, providing greater reproducibility.[55] Although the STE experiment is normally used for NMR diffusion measurements, the resonances of the slowly diffusing serum macromolecules were not selectively suppressed in the STE spectra through differences in diffusion, but rather by T_2 relaxation during the gradient pulses and associated delays.

1.2.6 Analyte Derivatization

For complex samples, such as those of biofluids or tissue extracts, ^{1}H NMR spectra even at high fields (i.e. > 500 MHz) suffer from problems of resonance overlap. One way of improving spectral resolution is through 2D HSQC experiments that correlate ^{1}H resonances with directly coupled ^{13}C or ^{15}N nuclei. These experiments detect ^{1}H while benefitting from the greater chemical shift dispersion of the heteroatom. To overcome the sensitivity limitations of natural abundance ^{13}C and ^{15}N experiments, chemical derivatization methods have been used to target specific functional groups and enhance the signals from lower abundance molecules not normally detectable by HSQC. Shanaiah and coworkers reported the acetylation of amines with ^{13}C labeled acetic anhydride for the selective detection and identification of amino acids and related metabolites in human urine and serum in the ^{1}H,^{13}C HSQC spectra.[56] Ye et al. expanded on this concept by adding a ^{15}N label to carboxylate groups to detect the labeled metabolites in ^{1}H,^{15}N HSQC spectra, greatly improving spectra resolution as well as the quantitative capacity of the NMR measurements.[57] Using selective ^{13}C and/or ^{15}N labeling, the coverage of targeted metabolic profiling experiments can be increased providing new insights into metabolic processes that might otherwise be undetectable.

1.2.7 Integration, Deconvolution and Binning

The parameters and methods used for post-acquisition data analysis can have an enormous impact on the outcome of a metabolomics NMR experiment.[58] Resonance integration or deconvolution is used for analyte quantitation or to generate the data inputs for metabolomics experiments. Figure 1.2 illustrates the different approaches to

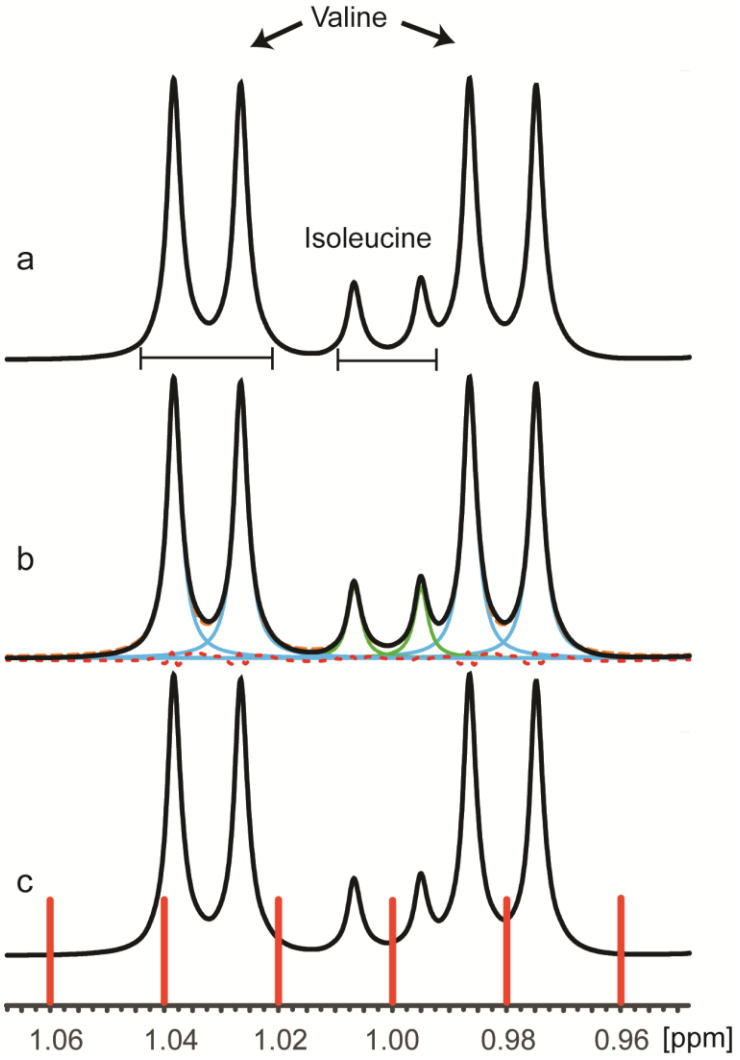


Figure 1.2. Representative 600 MHz ^{1}H NMR spectra showing the methyl resonances of 20 mM valine and 5 mM isoleucine, a) with integral regions manually defined for each resonance, b) after deconvolution with peak fitting, and c) using binned integral regions (shown in red). Fitted data in panel b) is shown in blue and green for valine and isoleucine, respectively, with the residual discrepancy between the calculated and actual spectrum shown in red. Peak fitting was performed using ACDlabs Spectrus Processor.

integration for a region of the 600 MHz ^{1}H NMR spectrum containing the methyl resonances of valine and isoleucine. As discussed in section 1.2.2, for any quantitative NMR measurement, it is important to have a flat baseline. For relative quantitation, a S/N greater than 10 is required, while absolute quantitation requires the S/N to be greater than 100 for the resonances of interest. The ^{13}C satellites of the more abundant compounds can interfere with the analysis of lower abundance compounds due to overlap.[59] Through application of a decoupling sequence on the carbon channel, those satellites can be eliminated and the impurities evaluated quantitatively. Most simply, integration is performed by defining an integral region around the resonances of interest (Figure 1.2a). Integral regions should be chosen to minimize contributions from overlapped resonances, although in complex biological samples this may not be possible.[60]

Due to the Lorentzian character of NMR resonances, integrals can be adversely affected even if neighboring peaks are visually resolved, a problem that is exacerbated when one of the overlapped resonances is of much greater intensity. For example, even in the simple mixture used for measurement of the spectra in Figure 1.2, the partial overlap of the isoleucine and valine doublets in the region between 0.97 and 1.01 ppm artificially increases the isoleucine integral value.

Because of the complexity of a biofluid or tissue extract, resonance overlap can be a significant problem in defining integral regions. A variety of spectral deconvolution approaches can be used to mathematically fit a Lorentzian or Gaussian peak shape to partially overlapped resonances, resulting in a "pure" extracted spectrum that can be used for quantitative analysis.[58, 61, 62] Software for deconvolution is provided by the instrument

vendors and independent commercial platforms, such as ACDlabs Spectrus Processor (Advanced Chemistry Development, Inc. Toronto, CAN) and MestReNova (Santiago de Compostela, ES), are also available. Depending on the question to be addressed, resonances can be selected for deconvolution within a small region or the full spectrum can be analyzed. The integrals of the extracted peaks can be treated as either individual resonances or combined with other extracted peaks from the same spin system. Figure 1.2b demonstrates the ability of deconvolution to resolve the overlapped isoleucine and valine resonances providing more accurate integrals. One of the disadvantages of peak fitting is the time and processing-intensive nature of the fitting algorithm. MestReNova has recently developed a new algorithm, termed Global Spectral Deconvolution (GSD), which quickly fits resonances with a high degree of accuracy. Unlike peak fitting, which typically requires a number of user defined parameters including the peak frequency and width, GSD inputs only require the user to select the desired resolution (for example, if a single resonance shows minor splitting due to coupling, should that resonance be split into two or treated as one), the number of fitting cycles (which influences the speed of the GSD analysis) and a signal cutoff (or threshold) to limit the number of resonances analyzed.

Binning is another approach to integration. While deconvolution seeks to resolve partially overlapped resonances, binning (or bucketing) decreases the spectral resolution by integrating regions of the spectra based on either equidistant bin widths (e.g. 0.02 or 0.04 ppm) or other binning techniques as described below. Binning is not useful for absolute quantitation, however it is widely used in metabolomics studies. Figure 1.2c

shows the binned spectrum of the methyl resonances of valine and isoleucine. Some bins include portions of the resonances of both compounds, resulting in a decrease in resolution.

A commonly encountered problem using biofluids or tissue extracts is that variations in sample ionic strength (i.e., salt content) and pH can complicate metabolite analysis by NMR. Buffers are often used to minimize these effects, however, deviations in chemical shifts can still be observed. Binning can accommodate subtle differences in chemical shifts, but sometimes the end of a bin might fall on the top of a peak, and if there are pH-induced frequency shifts, changes in the binned integrals could provide misleading information about metabolite changes. Adaptive binning, a technique using a wavelet transform of the data, provides a means of spectral alignment to negate the effects of pH or ionic strength differences among the spectra being compared.[63] This method creates a variable bin width dependent on the detected peaks to account for chemical shift changes within a set of resonances. A disadvantage of this method is that it requires user-defined variable inputs and creates a reference spectrum for wavelet optimization, resulting in a loss of information. Adaptive intelligent binning (AI-Binning) builds upon the adaptive binning technique in that it requires less user input while still providing variable bin length to account for chemical shift differences between spectra.[64] This method divides the spectra into successive bins based on the intensity values of the bin edge, essentially defining the bin widths based on local minima.

1.2.8 Data Normalization

Data normalization is an integral part of studies employing NMR quantitation of biological samples, accounting for differences in dilution (a common problem with urine samples), extraction efficiency, pipetting errors, receiver gain, and water content. Where the extraction efficiency can be reliably determined or sample dilution is not a problem, integrals can be normalized to the tissue dry weight or biofluid volume, respectively. In other cases, a method such as sum-normalization, which accounts for concentration differences by summing the area of the full spectrum, may be useful. In sum- or bucket-normalization the NMR spectrum is divided into bins (removing solvent and contaminant signals), the bins are summed, and each individual integral is divided by the summed value.[65] However, sum-normalization has been criticized because it assumes that the relative abundance of compounds, and therefore the resonances they produce, is similar across sample groups. For example, glucose levels in diabetic patients can vary significantly and in some samples may be present at high levels, causing glucose to weight the sum-normalization factor in a highly variable fashion. To compensate for this limitation, other normalization techniques have been developed. In probabilistic quotient normalization (PQN), individual spectra are compared to a reference spectrum (normally a control spectrum or average of several spectra) by taking the quotient of the bins.[66] The median of the quotients is then chosen as the normalization factor, and all variables of the spectra are divided by that quotient. PQN is an effective method for normalization shown to be robust in a variety of studies.

1.3 Mass Spectrometry for Metabolomics Experiments

Despite the advantages of NMR for metabolomics experiments, the greater sensitivity of hyphenated-MS techniques can allow the quantitation of less abundant metabolites.[67] Because of the complexity of metabolite extracts, a separation is typically incorporated into the analysis prior to MS detection. Popular chromatographic approaches prior to MS detection include gas chromatography (GC) and liquid chromatography (LC); however, the analytical platform chosen is largely dependent on the goal of the experiment.

1.3.1 Gas Chromatography-Mass Spectrometry for Metabolomics

GC-MS using electron ionization (GC-EI-MS) is a popular analytical platform for metabolomics studies due to its sensitivity and reproducibility as methods are readily transferable between labs and instruments. The application of GC-MS for the identification of metabolites has been reported as early as 1967 however, its popularity and use for general metabolite profiling has increased significantly in the last 15 years.[68, 69] For example, GC-MS metabolite profiling has been used in urine analysis,[70] the analysis of mice brain tissue,[71] blood serum,[72] and plant tissue.[12] The robustness of GC-MS measurements has largely been facilitated by the adoption of uniform MS parameters affecting fragmentation and through the use of retention indices (RI) that minimize discrepancies in the chromatography.[73, 74] The inclusion of metabolite fragmentation patterns and retention index information into publicly available libraries has increased the utility of GC-MS for metabolomics studies of complex biological organisms.[75, 76] GC-MS has been successfully used to query plants, fungi, mammals, fish, and amphibians for

their metabolic response to a variety of biotic and abiotic stressors.[11] GC-MS is particularly effective in the analysis of primary metabolites, specifically those involved in central carbon metabolism. To circumvent the low volatility of most biological compounds, prior to GC-MS analysis molecules are silylated and sugar cyclization is reduced through methoximation of ketones and aldehydes.[11] Because of extensive prior analytical method development, GC-MS is a popular and reliable analytical platform for the sensitive detection of a wide variety of metabolite classes. However not all molecules are amenable to derivatization and, once derivatized, some analytes are not sufficiently volatile or stable for GC separations.

Despite the inclusion of a separation prior to detection, peak overlap still occurs and can complicate chromatographic interpretation and metabolite identification. There are several available programs used for deconvolution, the most widely of which is the Automated Mass Spectral Deconvolution and Identification System (AMDIS).[77] The AMDIS program is freely available and can be coupled to a variety of libraries for peak identification. In a two-step process, AMDIS first deconvolves the chromatogram of EI-MS data and then performs a library search of each deconvoluted peak. If RI markers are used, the AMDIS program also calculates the RI for the extracted peak, enabling identification not only by fragmentation patterns but retention indices as well.

1.3.2 High Performance Liquid Chromatography - Mass Spectrometry for Metabolomics

High Performance Liquid chromatography-mass spectrometry (HPLC-MS) is more commonly used for targeted metabolite profiling of secondary metabolism, such as phenolic compounds[78] and sugar phosphates.[79] It is often the preferred tool for

pharmacological studies,[80] and for the identification of biomarkers to aid in disease detection because of the wide selection of stationary phases available with selectivities for different classes of compounds.[81] Additionally, HPLC-MS can be hyphenated with NMR to assist with component identification, although such instruments are not widely available.[82] Electrospray ionization (ESI) is typically used for HPLC-MS metabolite analysis. The advantage of ESI ionization is that it is a soft ionization source that results in minimal fragmentation, allowing the analyte of interest to be observed as a molecular ion peak. In some cases, however, fragmentation can occur even under optimized conditions and salt adducts are not uncommon, complicating spectral interpretation.

One reason that HPLC-MS is popular is that it is a very versatile method for bioanalysis. Different types of LC separations can be used to resolve a wide range of compound classes, providing that the mobile phase is MS compatible. Unlike GC-MS, HPLC-MS can be used to analyze both hydrophilic and hydrophobic compounds without derivatization by changing the mobile or stationary phases. HPLC separations are performed at lower temperatures than are required in GC, avoiding decomposition of thermally unstable compounds.[82] HPLC-MS, however, is more susceptible to matrix effects compared with GC-MS, such as sample pH and salt content, contributing to variation in observed retention times for given analytes between chromatographic analyses. Matrix effect can be mitigated by increasing sample preparation methods, such as liquid-liquid extraction, however the potential for sample loss increases with each additional preparation step.[83]

Hydrophilic interaction chromatography (HILIC) has been used to separate polar small molecules in metabolite extracts.[84, 85] The HILIC separation mechanism is primarily due to partitioning of the analyte between a water layer adsorbed on the stationary phase surface and the organic mobile phase.[86] As the aqueous composition of the mobile phase increases, the analyte will partition from the adsorbed layer of water into the mobile phase.[87] Aspects of this mechanism are still not well-understood, and additional experimental evidence will be required to fully explain the HILIC separation mechanism.[84] The advantage of HILIC is that it can separate highly charged hydrophilic analytes and is MS-friendly.[87] The drawbacks of HILIC are its susceptibility to matrix effects due to ionic interactions with the stationary phase and the analytes, the low flow rate which reduces sample throughput, and long equilibration period required prior to each injection.[85]

Reverse-phase high pressure liquid chromatography (RP-HPLC), the most widely applied separation method for LC-MS, resolves analytes based on differences in their interactions with a hydrophobic stationary phase. RP-HPLC is a popular method because the separations are reproducible and predictable. The nature of the interactions of analytes with RP stationary phases have been discussed extensively and elution order can often be anticipated, with analyte retention generally increasing with hydrophobicity.[88] Because RP-HPLC retains non-polar compounds, polar compounds elute early in the separation or are not retained at all, preventing their resolution. Glauser and coworkers demonstrated the utility of RP-HPLC-MS, combined with NMR, to separate and identify stress biomarkers in *Arabidopsis thaliana*.[89] They showed the transferability and

reproducibility of RP-HPLC separation from an analytical separation to a semi-preparative separation, enabling efficient component resolution simultaneously with MS detection.

Reverse-phase ion-pair high pressure liquid chromatography (RPIP-HPLC) is a separation approach that uses standard reverse-phase columns with a mobile phase containing an ion pairing reagent (IPR) in the separation buffer. The IPR, such as tributylamine, contains both lipophilic and charged moieties. Due to the ion-pairing interaction, charged analytes (independent of hydrophilicity) can be retained on a reversed-phase column. An alternative method to RPIP-HPLC is strong-anion exchange (SAX) HPLC, in which charged analytes are bound by the oppositely charged stationary phase and a salt gradient is used for elution. The high salt concentration required to elute analytes from the SAX column is not compatible with MS detection. In contrast to SAX-HPLC, RPIP-HPLC uses volatile MS-compatible ion-pairing reagents in the mobile phase.

Figure 1.3 shows the two currently accepted mechanisms of RPIP-HPLC separations. As illustrated in Figure 1.3a, adsorption of the IPR onto the stationary phase separates analytes through a pseudo-ion exchange mechanism.[90] In the mechanism shown in Figure 1.3b, the IPR and the charged analyte form an ion pair in solution that partitions onto the hydrophobic stationary phase.[90, 91] It is likely that both mechanisms occur in most RPIP-HPLC separations and that the balance between these competing processes is influenced by experimental conditions.[92] As a result, RPIP separations methods tend to be less robust than RP-HPLC separations. Extensive method

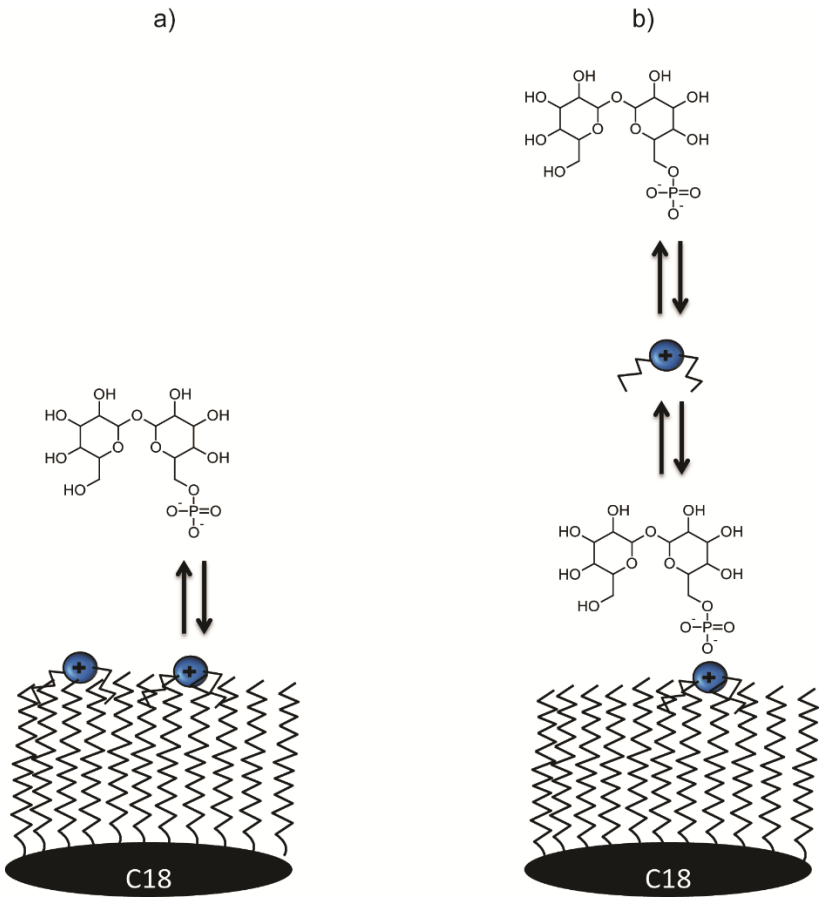


Figure 1.3. The RPIP separation mechanisms illustrated with trehalose-6-phosphate as the analyte and dibutylamine as the IPR. (a) The dynamic exchange model for RPIP in which the IPR coats the stationary phase and acts by an anion-exchange mechanism to retain oppositely charged ions.[90] (b) The formation of an ion-pair in solution allows analyte retention through the interaction of the lipophilic side chains of the IPR with the stationary phase.[91]

optimization is often required for each new analyte targeted, and may even be necessary when extending the separation to a new sample matrix. Although RPIP-HPLC has been less commonly reported in metabolomics experiments, there is growing interest in its application for the separation of phosphorylated and other anionic compounds.[93, 94]

1.3.3 Considerations for Mass Spectrometry-based Metabolomics Experiments

As with NMR, there are several factors that require consideration when performing MS-based metabolomics experiments. One factor is the type of mass analyzer used as each has different attributes in terms of mass accuracy, sensitivity, reliability for quantitation, cost and ease of use. Data analysis (including data alignment, integration, and quantitation) is another consideration. There are a wide variety of software tools to assist in data mining and the different approaches of each program can impact the analysis results.

1.3.3.1 Mass Analyzers

The choice of mass analyzer is an integral part of an MS-based metabolomics experiment. There are a variety of mass analyzers that can be used for the detection, identification, and quantitation of metabolites each with their own advantages and disadvantages. Two of the main challenges in metabolomics experiments are metabolite identification and achieving sufficient measurement sensitivity to reliably quantify compound present at low abundance. Many metabolites found in trace amounts hold important biological functions[95] and not all mass spectrometers are equally sensitive. Although other types of mass spectrometers have been employed for metabolite

identification and quantitation, the most widely used mass analyzers are the quadrupole mass filter and the time-of-flight mass spectrometer (TOF-MS).

Quadrupole instruments are well-suited for the detection and quantitation of trace quantities since they are able to detect sub-picomolar amounts of a compound.[96] Quadrupole instruments also have a dynamic range of approximately 6 orders of magnitude, comparable to that of NMR, allowing a wide range of concentrations to be analyzed without risk of detector saturation.[96] Additionally, several quadrupoles can be operated in tandem, allowing mass-selective isolation, fragmentation, and detection to facilitate compound identification. A disadvantage is that the mass accuracy of the quadrupole mass analyzer is inherently low, typically only allowing unit resolution thereby limiting compound identification supported by the m/z ratio.[97]

In contrast, the TOF instrument can obtain a high mass accuracy, down to 1 ppm (or better, depending on the instrument) allowing a more accurate assignment of empirical formula to the detected m/z. Additionally, TOF instruments can be linked with a quadrupole (Q-TOF), allowing MS/MS in addition to high mass accuracy measurements to further facilitate compound identification. However, TOF instruments are limited to a dynamic range of 10^4 at best and are not as sensitive as their quadrupole counterparts. Because of the differences between instrumentation, metabolomics experiments must be designed to account for the merits inherent in each analyzer. However, as in the discussion of high field NMR, practical limitations may dictate that the experimenter matches his/her analyses to the locally available instrumentation and a choice of mass analyzer based solely on figures of merit may not be realistic.

1.3.3.2 Quantitation with Mass Spectrometry

As with NMR, quantitation in MS-based metabolomics experiments is either performed relatively or to determine absolute quantities. For relative quantitation, the peak areas in the extracted ion chromatograms are compared across samples to reveal differences as a function of treatment. Normalization for relative quantitation can be carried out by summing the areas of identified peaks corresponding to each metabolite and dividing this value by that sum of the areas for all metabolites (similar to sum normalization described in section 1.2.8). Normalization can reduce the contribution of differences due to dilution, however like for NMR, changes in the most abundant signals can bias the results. Another common problem with MS measurements is the potential for changes in sensitivity over the course of several injections. An internal surrogate, an analyte that is otherwise not found in the sample analyzed, can be added to compensate for changes in ionization efficiency. The effect of changes in ionization during metabolomics experiments can also be mitigated through randomization of the sample order so that samples from different replicates are not measured sequentially but intermixed throughout the run.

Absolute quantitation in mass spectrometry is more complicated to carry out effectively, especially in metabolomics experiments due to their goal of simultaneously determining the levels of many compounds. Because ionization is compound-specific for all ionization sources, quantitation is accomplished through creation of a calibration curve, using the method of standard additions, or by addition of an isotopically labeled

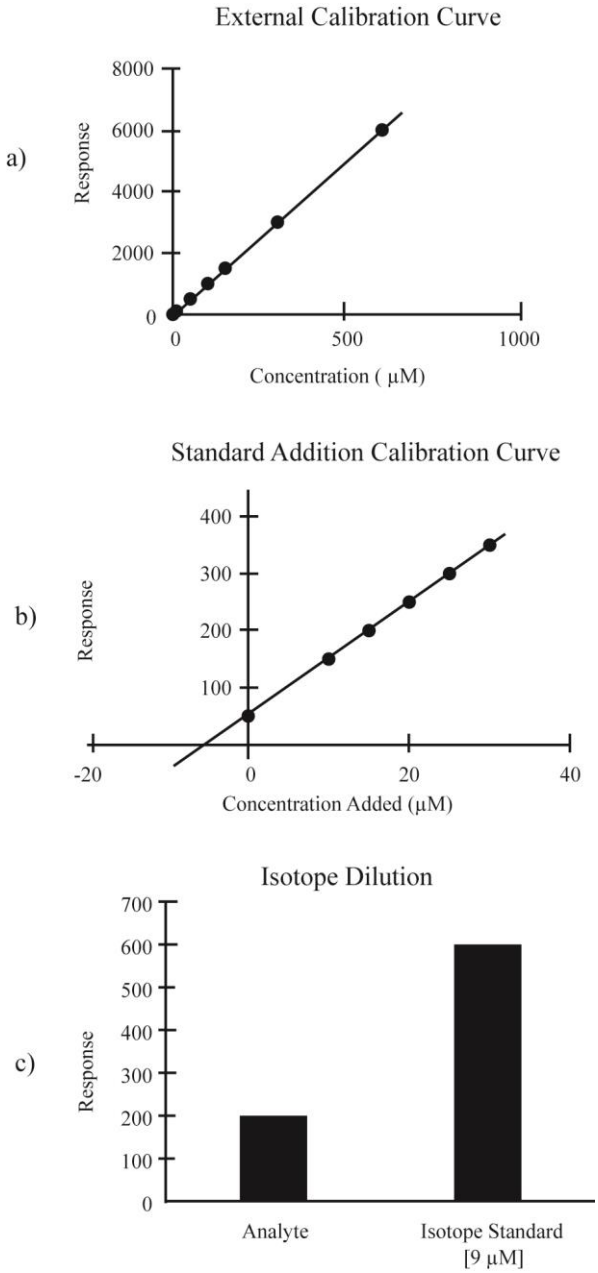


Figure 1.4. (a) Analyte quantitation by an external calibration curve. The equation of the best fit line can be used to calculate the concentration of the analyte using the response measured for the sample. (b) A standard addition calibration curve generated by adding known amounts of the analyte to the sample. The concentration of the analyte in the sample is determined by the x-axis intercept. (c) For quantitation by isotope dilution, a known amount of an isotopically labeled analogue of the analyte of interest is added to the sample and the analyte and standard peaks are resolved due to differences in m/z.

analogue of the targeted analyte.[97] Figure 1.4 illustrates three different calibration methods for absolute quantitation.

External standardization is carried out using a calibration plot generated using a standard solution of the targeted analyte (Figure 1.4a). The equation generated by a best fit function of the calibration plot is used to determine the concentration of analyte in a sample from the detected response. Although the calibration plot may be best fit by a nonlinear function, often only the linear region of the calibration curve is used for quantitation.[97] MS calibration using external standards works well with relatively simple samples; however, matrix effects from complex biological samples can dramatically affect ionization efficiency reducing the accuracy of the determination. Another problem with external standardization is that changes in ionization efficiency over the course of the experiment can mean that samples analyzed at the end of the experiment may give a lower response than those determined at the start of the run. This limitation is commonly addressed by interleaving a standard with the sample queue, for example by re-measuring a standard after every 9 sample injections to compensate for instrument drift.

The method of standard additions overcomes the limitation of external standardization by performing the calibration in the sample matrix (Figure 1.4b).[98] In this method, the sample is analyzed and then successive aliquots of an analyte standard at known concentration are added to the sample with the measurement repeated following each addition. The instrument response is plotted vs. the concentrations of the added standard and the resulting plot fit by linear regression. The x-intercept of the calibration plot indicates the concentration of the analyte in the original sample. However, because

standard addition is used to compensate for matrix effects, a plot must be generated for each sample separately for each analyte of interest making this technique impractical for many metabolomics experiments.

Isotope dilution is a favored alternative for MS quantitation in which a known amount of an isotopically labeled analog of the compound of interest is added to the sample (Figure 1.4c).[99] Isotope dilution is a special case of the method of standard addition that avoids the need to perform multiple analyses of the sample because the analyte and standard can be discriminated based on differences in *m/z*. For example, glucose uniformly enriched in ^{13}C ([U-^{13}C]-glucose) and having an observed *m/z* six Daltons higher than the naturally abundant [U-^{12}C]-glucose might be used for MS quantification of glucose. This mass difference is sufficient to permit resolution and integration of both glucose analogs in the resulting mass spectrum allowing quantification of the glucose concentration based on the ratio of peak areas. In an LC- or GC-MS experiment, because both the analyte of interest and the isotopically labeled analog are chemically equivalent, they will have the same chromatographic retention time and will experience identical ionization conditions. Isotopically labeled versions of many metabolically interesting compounds are commercially available but are not available for all metabolites and a labeled analog of each analyte of interest must be added to every sample making this approach costly.

1.3.3.3 Data Extraction for Integration

Compared with NMR spectra, data extraction in a GC- or LC-MS experiment is less complex. Although chromatograms may themselves be complex with significant

peak overlap, it is the mass spectral data, not the total ion chromatogram (TIC) that is used for data extraction. Most metabolites produce a molecular ion with unique mass-to-charge ratio (*m/z*), although there are prominent exceptions including sugars and isomeric phenolic compounds. For ESI, a soft ionization source commonly used in LC-MS, the resulting mass spectra ideally contain only one m/z ratio for the molecular ion of each analyte, however, in-source fragmentation and adduct formation commonly occur complicating the analyses. Spectra obtained with other ionization methods can be even more complex. For EI, a hard ionization source commonly used in GC-MS, the mass spectral data is typically a series of *m/z* values representative of characteristic fragments and the molecular ion peak may be of very low intensity, or even be undetectable above the spectral noise. In such spectra, quantitation may be carried out using the base peak (the most intense mass spectral peak) or using several characteristic fragments.

Regardless of the ionization source employed, data extraction in MS-based metabolomics experiments is performed by plotting each m/z value present in the spectra as a function of time creating individual extracted ion chromatograms (XIC). Because of the chemical complexity of the metabolome, there may be several chromatographic peaks for each m/z value selected, each representing a distinct metabolite. Each peak is assigned a chromatographic retention time, and is subsequently identified as a mass-retention time pair to better facilitate identification. The peaks are integrated and an area is assigned to the mass-retention time pair. Although this process can be performed manually, for large data sets several computer programs are available that automatically extract individual components including XCMS,[100] MarkerLynx (Waters Corporation,

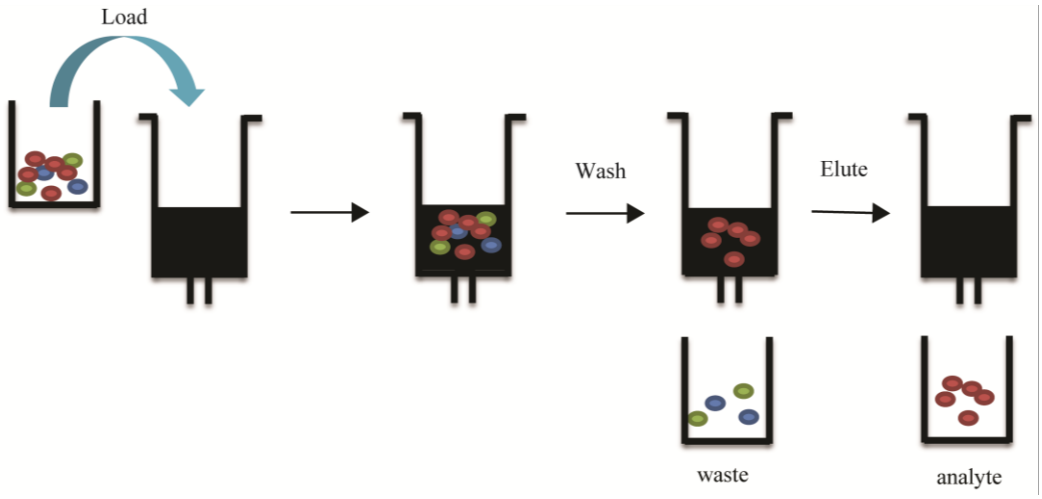


Figure 1.5. A schematic showing the isolation of an analyte by solid-phase extraction (SPE). Following cartridge-specific washing steps, the sample containing the analyte of interest (red circles) and other components (green and blue circles) is loaded into the cartridge. A washing step removes poorly retained matrix components as waste while the analyte is retained on the cartridge. The analyte is subsequently eluted from the cartridge and collected for analysis.

Milford, MA), Mass Hunter (Agilent Technologies, Santa Clara, CA) and MetaboAnalyst.[101] Additionally, each of these software packages automatically baseline correct and align peaks between data sets to simplify data processing.

1.4 Solid-phase Extraction for Sample Preparation

Solid-phase extraction (SPE) is a useful technique for sample clean-up or isolation of a desired class of molecules with similar structures or properties.[102] Figure 1.5 shows a general workflow utilizing SPE for the isolation of a targeted analyte.[103] A cartridge with a specific affinity for a class of compounds or type of structures is chosen. Prior to loading the sample the cartridge is prepared by washing to remove process impurities and stabilizers and to leave it in a form that will best retain the analytes of interest. As shown in Figure 1.5, the sample is first loaded onto the cartridge and components that are not retained by the solid phase are removed in the wash step. The compounds retaining on the cartridge are then eluted. Typically the eluent contains the analytes of interest as well as additional matrix components having similar physico-chemical characteristics.

The availability of several different SPE stationary phases including weak/strong anion exchange (WAX/SAX), weak/strong cation exchange (WCX/SCX), reverse-phase C18, and hydrophilic-lipophilic balanced (HLB) materials provide the possibility of isolating and separating diverse classes of compounds. For example, Beretta and coworkers applied reverse-phase SPE-NMR for the identification of the botanical origin of honey.[6] Application of SPE allowed the isolation of hydrophobic markers that could be used for the identification of the type of honey and any contaminants that might be

present. Another recent example of SPE-NMR includes the enrichment of phenolic compounds from grape extracts during various grape berry development stages.[104] Rolcik and coworkers used C18-SPE prior to GC-MS analysis to purify indole-3-acetic acid, a common plant metabolite to increase the sensitivity and selectivity of GC-MS analysis for phytohormones.[105] Bruce and coworkers used HLB cartridges and LC-MS analysis to develop a high throughput quantitative method for determining vitamin D metabolites in serum.[106] Similarly, Delatte and coworkers used a WAX SPE cartridge built on an HLB packing material and subsequent LC-MS analysis to target the isolation of phosphorylated disaccharides in plant materials, including trehalose-6-phosphate.[79] These examples demonstrate that SPE cleanup/isolation has a wide range of applications, can be used with a variety of sample matrices and is compatible with all major analytical platforms used for metabolomics measurements.

Analyte recovery after SPE is not always 100%.[103, 107] Analyte recovery can be reduced by incomplete retention, analyte loss during the washing step, or by failure to completely elute the analyte from the SPE cartridge. Because the sample matrix can influence recovery, samples that have been spiked with a known amount of an analyte standard are typically used in the determination of recoveries. This matrix almost always contains background levels of the analyte of interest therefore addition of an isotopically labeled analog allows the determination of recovery using LC- or GC-MS, as described in section 1.4. As the labeled analog and the target analyte are chemically equivalent, they will be affected similarly during SPE. In addition to matrix effects, the flow rate during the loading, washing, and elution steps can also influence analyte recovery. A

sufficiently slow flow rate should be used to allow the analyte enough time to interact with the cartridge stationary phase during loading and likewise, interact with the mobile phase during elution.

1.5 Multi-Platform Analyses for Compound Identification and Characterization

Online HPLC-UV/PDA-SPE-NMR-MS (or -MS-NMR) allows for the identification and detection of metabolites using multiple detection platforms in a single experiment. Hyphenation of NMR and MS can be valuable in evaluating a plant's metabolome due to the complementary nature of the data produced by each technique. The sensitivity of MS analysis enables the detection of low concentration analytes and provides some structural information about the identity. NMR can be especially informative in the identification of metabolites that are otherwise unknown or only putatively assigned. For example, Exarchou and coworkers demonstrated that online HPLC-UV-SPE-NMR-MS is capable of identifying flavonoids in Greek oregano.[108] This was possible by first performing a separation using HPLC with protonated solvents and then splitting 5% of the peak for MS analysis with 95% directed to UV-SPE-NMR. Prior to NMR analysis, peaks were trapped on SPE cartridges, dried, and eluted with deuterated solvents. SPE provided a method for online concentration of the sample for NMR detection along with exchange of the protonated solvent used in the separation with the more expensive deuterated solvent for ^{1}H NMR spectral acquisition. Peaks were trapped multiple times on the SPE cartridge prior to elution to increase the concentration of the analyte prior to NMR analysis. To ensure a higher recovery rate, the analytes were automatically diluted with water after elution from the HPLC column to decrease the

organic mobile phase composition and encourage efficient SPE trapping. Agnolet and coworkers compared online HPLC-PDA-MS-SPE-NMR with traditional ^{1}H-NMR metabolite profiling for the evaluation of commercial *Gingko biloba* extracts.[109] The purpose of these experiments was to evaluate ^{1}H NMR as a replacement for MS for the analysis of glycosides and terpene trilactones (TTLs) in *Ginkgo biloba*. Using HPLC-PDA-MS-SPE-NMR, the authors were able to identify and characterize the composition of several unique glycosides and TTLs in commercial extracts and validate ^{1}H NMR for global profiling of extracts for quality control purposes. Additionally, these techniques allowed the identification of potentially harmful components of the commercial extracts. The disadvantage of online methods is that the differences in analyte concentration required for MS and NMR analysis may be difficult to control. A fully hyphenated method requires that the NMR spectrum of the analyte of interest be acquired online, limiting the number of scans and subsequently, the quality of the spectrum obtained.[110] HPLC-MS-NMR analysis can also be completed offline, enabling the isolation of selected analytes followed by a concentration step for NMR analysis. Rezzi and coworkers demonstrated an automated SPE-HPLC fractionation method for biofluids followed by NMR analysis.[111] Although MS detection was not employed, it can be implemented afterwards for offline sample analysis, taking advantage of the inherently non-destructive nature of NMR in combination with the sensitivity of MS.

Multi-platform approaches should be considered in the design of plant metabolomics experiments as they have shown to be advantageous for sample enrichment, component identification, and method verification. The combination of

HPLC, MS, SPE, and NMR allowed researchers to simultaneously reduce sample complexity, identify contaminants, target specific classes of compounds and evaluate biological diversity quantitatively using a single analysis. Because of the complexity of plant systems and the large number of predicted metabolites, further development and application of hyphenated HPLC-MS-NMR techniques will be beneficial for future plant metabolomics studies.

1.6 Statistical and Chemometric Analysis of Metabolomics Datasets

Statistical analysis is important to ensure appropriate representation of quantitative data especially when relative quantitation is performed for a large dataset. Several different statistical methods are available for the evaluation of metabolomics data. For comparison of the integrated areas of individual peaks (univariate analysis), common statistical analysis methods can be applied, such as the Student's t-test and p-value calculations, analysis of variance (ANOVA), and regression analysis. These techniques are described in a variety of textbooks and articles and will not be discussed here.[112, 98, 113, 114] The focus of this section is on multivariate data analysis for evaluating global changes in a set of NMR or MS spectra. Multivariate data analysis is widely used in metabolomics studies and for evaluating the results of method development experiments.[115, 10] There are numerous types of multivariate analysis methods that can be applied to quantitative datasets, however only a few of the more relevant techniques will be discussed in this section. The review by Broadhurst and coworkers provides an in-depth analysis and description of many statistical approaches including their advantages and disadvantages.[116]

1.6.1 Principal Components Analysis for Metabolomics

Principal components analysis (PCA) is a multivariate statistical method that has the advantage of being an unsupervised technique, meaning the analysis is completed independent of variable identification.[117] PCA identifies statistical patterns within a dataset, independent of the experimental conditions (such as genotype, stress, drug dose, subject weight, diet, gender, etc.), making correlations that can later be attributed by the researcher to specific experimental conditions or sample classes. For PCA of NMR data, spectra are compiled into a data table containing intensity (either as integral bins or individual data points) as a function of chemical shift for each sample. For PCA of MS data, the XICs are compiled into a data table containing the area of each mass-retention time pair. In PCA of GC-MS analysis, only the mass-retention time pairs of known metabolites are used to avoid inclusion of derivatization side products.

The multidimensional NMR or MS data matrix is reduced by PCA to a lower dimensional space based on 2 or more orthogonal principal components, which depend on the sample variance. Although statistical significance cannot be deduced based on the relationship between sample groups in a PCA plot (if indeed the different sample classes, treatments, conditions, etc. group differently than the controls), such groupings do provide insight into the statistical differences between samples (Figure 1.6). If there are large differences between sample groups in the PCA scores plot, several of the variables (or bins) likely contain resonances that vary significantly; therefore, sample classes that demonstrate clearly resolved groupings will likely also be distinguished by univariate comparisons of selected peak integrals. Figure 1.6 shows an example of PCA scores

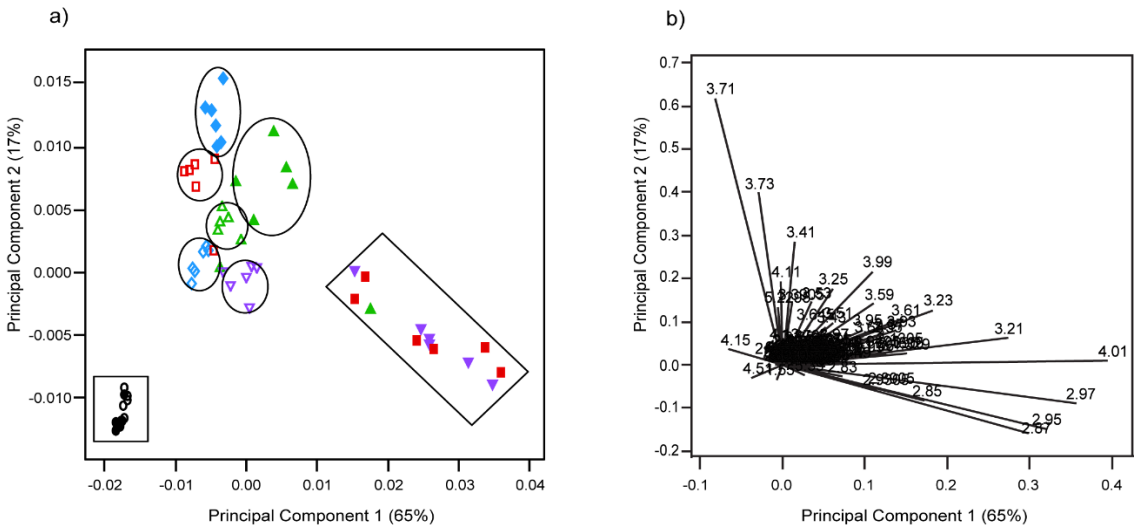


Figure 1.6. An example of PCA performed on an NMR metabolomics dataset showing the (a) scores and (b) loadings plots. The various symbols in the scores plot (a) represent the 10 individual treatments, with five samples from each treatment analyzed. Consistent sample groupings within a treatment are highlighted by ellipses. The samples that do not group separately are enclosed by a rectangle. The numbers in the loadings plot (b) are the different NMR bins that are responsible for the separation of the different treatments in (a).

(1.6a) and loadings (1.6b) plots for an NMR metabolomics dataset. The symbols in the scores plot represent replicate analyses of samples from 10 different treatments, distinguished using a different symbol for each treatment. Distinct sample groupings are highlighted by ellipses, show clear separation of samples from several of the treatments along principal component one (PC1) and principal component two (PC2). However, not all of the treatments produced results that group separately and these are highlighted with rectangles. The loadings plot (Figure 1.6b) shows which variables (in this case data bins) contribute to the separation along PC1 or PC2, identifying the metabolites that contribute most to the sample groupings in the scores plot (Figure 1.6a). In the example data set shown in Figure 1.6, PC1 and PC2 describe 65 % and 17% of the variance, respectively. If the first two principal components do not explain a majority of the variance (for example, if PC1, PC2, and PC3 account for 20%, 18%, and 16%, respectively), then other principal components will need to evaluated.

1.6.2 Other Statistical Approaches for the Analysis of Metabolomics Data

Orthogonal projections to latent structures discriminate analysis (O-PLS-DA), also known as partial least squares, is another multivariate analysis technique used in the analysis of biological datasets.[118, 113, 119] This is a supervised classification model that differs from PCA by addition of grouping variables that indicate in which class the samples belong. As in PCA, if the data originated from NMR analysis, the data input is a table of binned integrals (or the intensity of individual data points) as a function of chemical shift. If MS data is to be analyzed, the input is mass-retention time pairs and the corresponding areas. PLS reveals the relationship between the original data matrix

and the representation generated by the analysis, providing information on how the different sample sets can be described based on the grouping information provided. As with PCA, definitive statistical correlations between the sample classes can only be loosely interpreted from the PLS results. Problems with over fitting, where too many variables are used for too few samples (as often the case in metabolomics), additionally complicate PLS interpretation.[112, 114, 120] To overcome this limitation, cross-validation is becoming more common for PLS measurements.[120] For a more in-depth description, several reviews and articles discussing advantages and disadvantages of PLS, describe different variations of PLS, and the types of datasets for which PLS is most applicable.[112, 114, 120]

Statistical total correlation spectroscopy (STOCSY), a pseudo-2D NMR plot, compares the intensities of resonances to generate correlations between these resonances.[121] The main advantage of STOCSY is the generation of a pseudo 2D NMR spectrum based on the statistical correlation of a specific resonance, allowing facile identification of resonances that differ between sample classes. Not only does STOCSY provide a measure of the statistical differences between resonances it also indicates the magnitude of the change, revealing either negative or positive correlations corresponding to an increase or decrease in the relative levels of a particular metabolite. Because the STOCSY plot serves as a visual aid for identifying resonances that differ between sample classes, an additional level of correlation between resonances can be elucidated. Resonances demonstrating similar statistical differences can be interpreted as arising from the same molecule or from molecules of related pathways. STOCSY has been

applied to a variety of organisms and has been adapted to include additional chemometrics.[122-124]

1.7 Isotope Enrichment and Metabolic Flux

One of the advantages of NMR spectroscopy is the ability to detect a variety of biologically important nuclei, such as ^{1}H, ^{13}C ^{15}N, and ^{31}P. Similarly, MS analysis provides isotope-specific information enabling the evaluation of a variety of nuclei through their isotopic distribution. Stable isotope labeling of molecules (such as drugs, sugars, metabolic precursors, etc.) with ^{13}C or ^{15}N can be a useful tool for elucidating specific biochemical reactions to stresses, genetic differences, toxicants and pharmaceuticals. Quantitative evaluation of compounds before, during, and after metabolism is of growing interest. Twaddle and coworkers used deuterated bisphenol A in rats to track and quantify its various metabolites by LC-MS in tissues, fluids, and excreta, enabling researchers to evaluate concentration affects.[125] Fluorine (^{19}F) NMR is also often utilized for pharmacokinetic studies due to its high sensitivity and low natural background, making fluorine-labeled pharmaceuticals easier to track throughout their adsorption and eventual catabolism. Recently, Do and coworkers reported the use of ^{19}F for reaction monitoring and yield determinations for pharmaceutical development, taking advantage of the non-destructive, and inherently quantitative nature of NMR.[126] Other studies using fluorine analysis include *in vivo* magnetic resonance studies of psychiatric drugs and high throughput NMR for identification of enzyme inhibitors.[127, 128]

Metabolic flux analysis uses compounds with radioactive or stable isotope labels to determine the flow of metabolites through biochemical pathways.[129] Plant and cell-

based studies have demonstrated the utility of specifically labeling nutrients for metabolic flux analysis. Using NMR for isotopomer analysis, the metabolism of labeled materials provides insights into the biochemical regulation by following the extent of incorporation the stable isotope labels as well as their position within the metabolites. Lane and coworkers demonstrated the utility of isotopomer analysis in cell culture through introduction of [U-^{13}C]-glucose and tracing the positional enrichment of ^{13}C isotopes in glycerophospholipids.[130] Similarly, isotopomer analysis was conducted by GC-MS on metabolites resulting from usnic acid, a suspected hepatotoxin found in lichen-derived food supplements used for weight loss.[131] NMR has also been used to discern the assimilation preferences for CO_2 in two plant species in high and low CO_2 atmospheres. Biochemical pathways were identified through labeled $^{13}CO_2$ using solid state NMR and steady state metabolic flux experiments.[132]

1.7.1 Increasing the Sensitivity of NMR

For most compounds NMR suffers from poorer sensitivity than can be obtained with MS detection, limiting its application in metabolic flux experiments. Although it requires specialized instrumentation that is not yet universally available, dynamic nuclear polarization (DNP), greatly enhances NMR sensitivity making this method especially valuable for metabolic flux experiments. DNP increases the sensitivity of NMR by as much as six orders of magnitude by transferring electron spin polarization to nuclear spin polarization.[133] Once the compound is transferred from the DNP magnet to the NMR instrument, the analyte nuclei quickly relax back to their equilibrium populations. Because of the limited polarization lifetimes (seconds to minutes), *in vivo* metabolic flux

experiments using DNP must be performed quickly so that substrate uptake and conversion occur before the polarization is lost. These limitations suggest DNP to be useful primarily for relative quantitation, however recent reports comparing LC-MS analysis and DNP for drug metabolism indicate DNP can be absolutely quantitative.[134] Meier et al. reported a DNP experiment using uniformly labeled ^{13}C and ^{2}H glucose in a yeast cell culture.[135] These experiments showed central carbon metabolism regulation as a result of nutritional changes, chemical treatment, and a comparison between *E. coli* and yeast metabolic profiles.

1.8 Application of Metabolomics for Understanding Crop Stress Responses

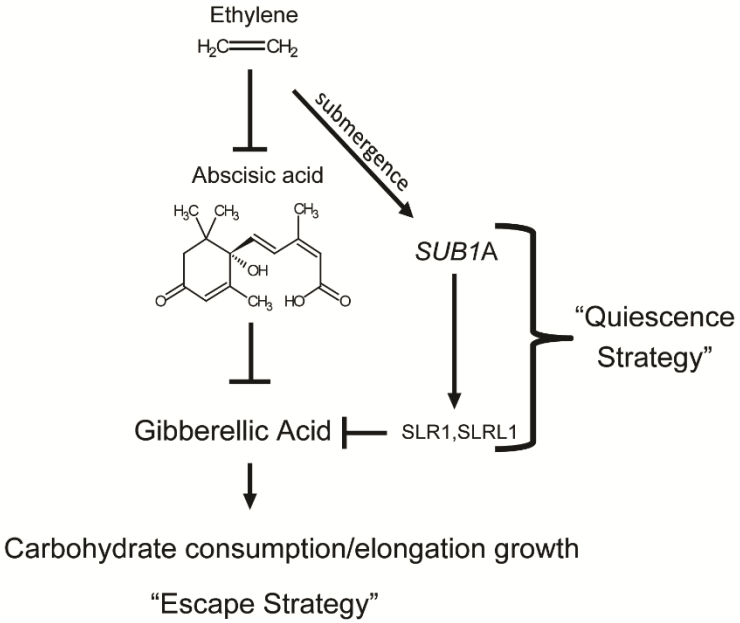
The genetic manipulation of crop plants to improve agronomic traits by use of breeding or transformation technology is routine and large amounts of seed can be easily generated for storage and distribution from accessible seed banks.[136, 137] The genetic variation underlying trait differences can be interrogated to provide information specific to biochemical pathways and the combination of this information with the results of metabolomics experiments can together give a deeper understanding of systems biology. Arbona and Steinfath have recently demonstrated the utility of metabolomics for phenotyping plant materials of closely related genotypes[138] and genotypes grown under different agriculturally relevant conditions.[139] The results of these studies indicate the sensitivity of the plant metabolome to differences in environment, subtle genotypic differences, and agricultural stress.

Flooding, a form of abiotic stress, is a common natural disaster that has a significant impact on crop production, survival, and quality.[140] Rice, an important staple

food crop, is frequently affected by short-term deep flooding especially in regions in South and Southeast Asia, reducing crop yields, increasing the financial burden of farmers and decreasing the food supply. The identification of the *SUB1*A gene of the ca. 125 kB multigenic *SUBMERGENCE 1* (*SUB1*) locus on chromosome 9 as a determinant of enhanced submergence tolerance in rice has led to breeding of popular varieties capable of enduring complete submergence for two weeks or even longer.[141] Such breeding efforts produced the variety, *Oryza sativa* ssp. *japonica* cv. M202(*Sub1*), with a higher tolerance to deep floods compared to *Oryza sativa* ssp. *japonica* cv. M202, a variety grown commercially.[141] The increased submergence tolerance of Sub1 rice lines allows these plants to maintain high grain yields even after a two week flood, easing the economic and agricultural implications of flash floods and providing a more reliable rice crop in rain-fed regions of weather instability.[142]

Under submergence conditions, the plant hormone ethylene is entrapped and accumulates in plant tissues. Figure 1.7 shows the hormone cascade implicated in the opposing escape and growth strategies observed in response to flooding in rice. During flooding, submergence-intolerant rice varieties exhibit an escape strategy in which ethylene increases gibberellic acid (GA) responsiveness resulting in stimulation of shoot elongation (Figure 1.7).[143] In shallow or progressive floods, this allows the plant to grow by elongation to maintain photosynthetic tissue above the water level and return to normal photosynthesis and respiration. If the flood is deep, this escape strategy can result in a depletion of carbohydrate stores leaving the plant unable to recover once the floodwaters recede. Submergence tolerance is conferred by the presence of the *SUB1A-1*

a)



b)

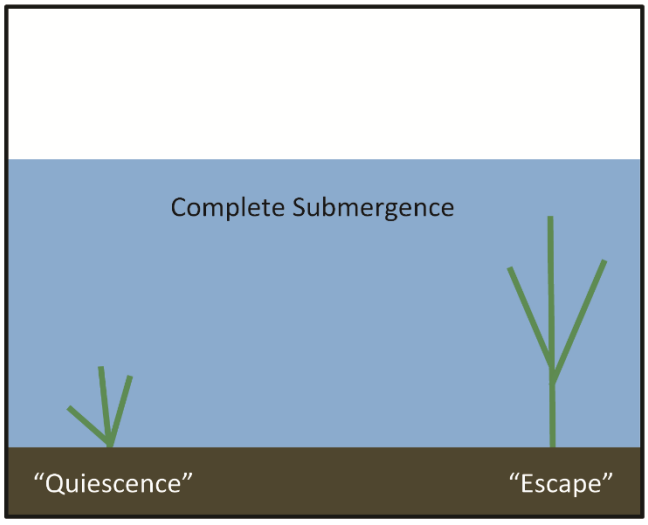


Figure 1.7. (a) The hormone cascade involved in rice submergence response. Upon complete submergence, ethylene is entrapped resulting in decreased bioactivity of abscisic acid and an increase in gibberellic acid activity, resulting in shoot elongation. In the presence of *SUB1*A, however, the activity of gibberellic activity is abrogated, resulting in reduced growth and carbohydrate conservation. (b) An illustration highlighting the difference in the growth of plants exhibiting either the quiescence or escape strategy.

allele of the *SUB1*A gene, encoding a group VII ethylene responsive transcription factor.[144, 141] Ethylene accumulation during submergence induces transcription of *SUB1*A, which reduces responsiveness to GA via the GA-signaling repressors SLENDER RICE-1 (SLR1) and SLR-LIKE-1 (SLRL1), with the effect of abrogating GA driven carbohydrate consumption and elongation, characterized as a quiescence strategy (Figure 1.7).[145] *SUB1*A has also been linked to ethanolic fermentation, with *SUB1*A containing plants showing greater accumulation of mRNAs and enzymatic activity for pyruvate decarboxylase (PDC) and alcohol dehydrogenase (ADH) during submergence compared to *SUB1*A-deficient plants.[144] These differences in transcripts and enzymatic activity suggested that *SUB1*A might promote fermentative metabolism; however, further analysis revealed that M202(*Sub1*) had a reduced ethanol accumulation compared to the intolerant variety. These observations support the hypothesis that *SUB1*A limits carbohydrate consumption under submergence stress. However, they also indicate that there can be inconsistencies between transcript levels and enzymatic activities, emphasizing the benefit of metabolite analysis in conjunction with gene expression or enzyme assays. Because of the complex relationships in systems biology, the connection of transcriptomics experiments with downstream metabolic activity will provide a more detailed understanding of submergence. Figure 1.8 highlights the various metabolites related to carbohydrate consumption/glycolysis including the TCA cycle, nitrogen assimilation/storage, carbon allocation, and amino acid metabolism that might be affected by the presence or absence of the *SUB1A* gene. Although the difference of the M202 and M202(*Sub1*) varieties is a single gene, the metabolic affects can be wide ranging.

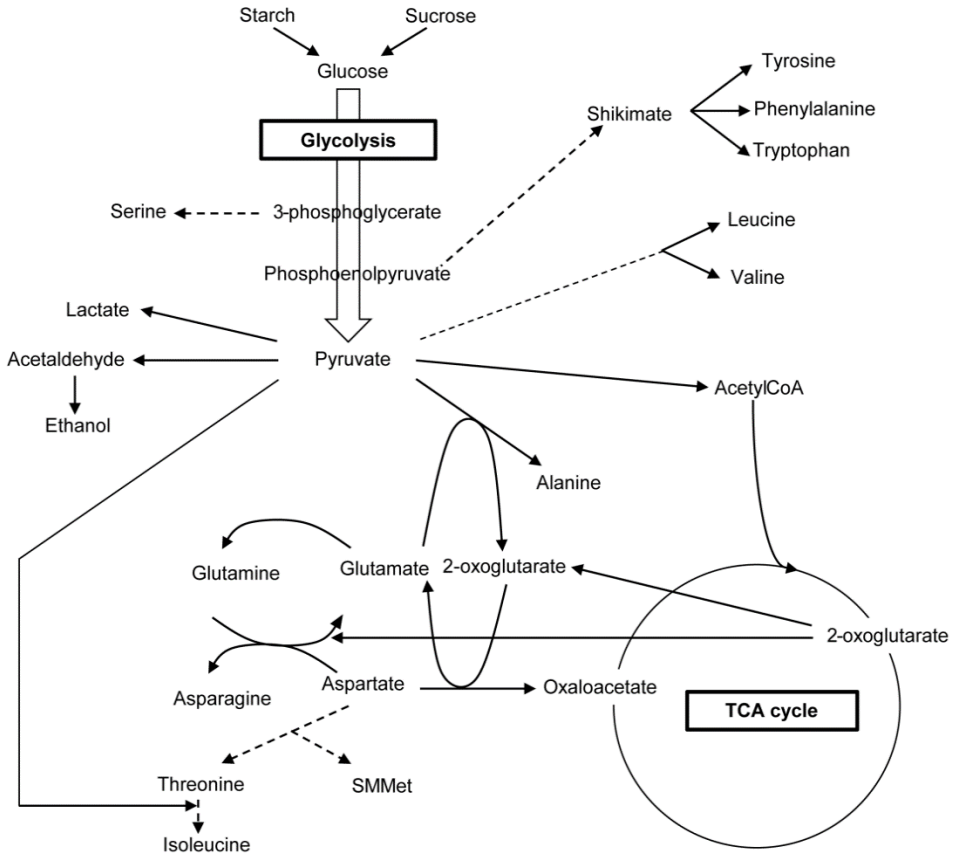


Figure 1.8. The biochemical pathways involved with glycolysis and ethanolic fermentation and some of the primary metabolites that can be affected by pathway perturbations, including the nitrogen metabolism, amino acid production, and TCA cycle metabolites.

In addition to increased submergence tolerance, *SUB1*A containing rice exhibits enhanced recovery from drought stress by decreasing leaf water loss and increasing the expression of genes affiliated with drought adaptation.[146] Interestingly, upon de-submergence rice plants undergo both re-oxygenation and dehydration stresses, however, the effects of *SUB1*A on the metabolite profile during de-submergence have not been determined.[144, 146]

1.9 Summary

The goals of this dissertation are to determine the metabolic changes that occur in rice during and after submergence and improve upon existing chemometric techniques for the automated characterization of metabolomics samples. In Chapter 2, the effects of submergence and recovery on metabolites of shoot tissue of both M202 (submergence intolerant) and M202(*Sub1*) (submergence tolerant) are elucidated using ^{1}H NMR spectroscopy. Further examination of the rice metabolome in response to submergence and recovery stress is carried out in Chapter 3 where GC-MS is used to detect and quantify primary metabolites, and the results compared with those obtained by NMR. Class-specific detection and quantitation is explored in Chapter 4 by incorporating SPE and LC-MS for the analysis of trehalose-6-phosphate and other phosphorylated/anionic components in response to submergence and recovery stress. In Chapter 5 a new chemometric approach, VIZR, for the automated analysis of ^{1}H NMR spectra is reported and its application for identification of unique components in a set of urine spectra is described. Lastly, in Chapter 6 the general conclusions and future directions of the research in this dissertation are discussed.

1.10 References

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

(2) Dixon, R. A. Natural products and plant disease resistance. *Nature* **2001,** *411*, (6839), 843-847.

(3) 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.

(4) Charlton, A. J.; Donarski, J. A.; Harrison, M.; Jones, S. A.; Godward, J.; Oehlschlager, S.; Arques, J. L.; Ambrose, M.; Chinoy, C.; Mullineaux, P. M.; Domoney, C. Responses of the pea (Pisum sativum L.) leaf metabolome to drought stress assessed by nuclear magnetic resonance spectroscopy. *Metabolomics* **2008,** *4*, (4), 312-327.

(5) Narsai, R.; Howell, K. A.; Carroll, A.; Ivanova, A.; Millar, A. H.; Whelan, J. Defining core metabolic and transcriptomic responses to oxygen availability in rice embryos and young seedlings. *Plant Physiol.* **2009,** *151*, (1), 306-322.

(6) Beretta, G.; Caneva, E.; Regazzoni, L.; Bakhtyari, N. G.; Facino, R. M. A solid-phase extraction procedure coupled to H-1 NMR, with chemometric analysis, to seek reliable markers of the botanical origin of honey. *Anal. Chim. Acta* **2008,** *620*, (1-2), 176-182.

(7) Cuadros-Inostroza, A.; Giavalisco, P.; Hummel, J.; Eckardt, A.; Willmitzer, L.; Pena-Cortes, H. Discrimination of wine attributes by metabolome analysis. *Anal. Chem.* **2010,** *82*, (9), 3573-3580.

(8) Barding, G. A.; Fukao, T.; Beni, S.; Bailey-Serres, J.; Larive, C. K. Differential metabolic regulation governed by the rice *SUB1*A gene during submergence stress and identification of alanylglycine by ^{1}H NMR spectroscopy. *J. Proteome Res.* **2012,** *11*, 320-330.

(9) Charlton, A. J.; Robb, P.; Donarski, J. A.; Godward, J. Non-targeted detection of chemical contamination in carbonated soft drinks using NMR spectroscopy, variable selection and chemometrics. *Anal. Chim. Acta* **2008,** *618*, (2), 196-203.

(10) Kaiser, K. A.; Barding, G. A.; Larive, C. K. A comparison of metabolite extraction strategies for H-1-NMR-based metabolic profiling using mature leaf tissue from the model plant *Arabidopsis thaliana*. *Magn. Reson. Chem.* **2009,** *47*, S147-S156.

(11) Fiehn, O.; Kopka, J.; Trethewey, R.; Willmitzer, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadruple mass spectrometry. *Anal. Chem.* **2000,** *72*, (15), 3573-3580.

(12) Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K. Comparison of GC-MS and NMR for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12*, (2), 898-909.

(13) Britz-McKibbin, P.; Terabe, S. On-line preconcentration strategies for trace analysis of metabolites by capillary electrophoresis. *J. Chromatogr. A* **2003,** *1000*, (1-2), 917-934.

(14) t'Kindt, R.; De Veylder, L.; Storme, M.; Deforce, D.; Van Bocxlaer, J. LC-MS metabolic profiling of *Arabidopsis thaliana* plant leaves and cell cultures: Optimization of pre-LC-MS procedure parameters. *J. Chromatogr. B* **2008,** *871*, (1), 37-43.

(15) Kim, J. D.; Kaiser, K.; Larive, C. K.; Borkovich, K. A. Use of H-1 nuclear magnetic resonance to measure intracellular metabolite levels during growth and asexual sporulation in *Neurospora crassa*. *Eukaryot. Cell* **2011,** *10*, (6), 820-831.

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

(17) Ebikeme, C.; Hubert, J.; Biran, M.; Gouspillou, G.; Morand, P.; Plazolles, N.; Guegan, F.; Diolez, P.; Franconi, J. M.; Portais, J. C.; Bringaud, F. Ablation of succinate production from glucose metabolism in the procyclic trypanosomes induces metabolic switches to the glycerol 3-phosphate/dihydroxyacetone phosphate shuttle and to proline metabolism. *J. Biol. Chem.* **2010,** *285*, (42), 32312-32324.

(18) Flores-Valverde, A. M.; Horwood, J.; Hill, E. M. Disruption of the steroid metabolome in fish caused by exposure to the environmental estrogen 17 alpha-ethinylestradiol. *Environ. Sci. Technol.* **2010,** *44*, (9), 3552-3558.

(19) Yap, I. K. S.; Brown, I. J.; Chan, Q.; Wijeyesekera, A.; Garcia-Perez, I.; Bictash, M.; Loo, R. L.; Chadeau-Hyam, M.; Ebbeis, T.; De Iorio, M.; Maibaum, E.; Zhao,

L. C.; Kesteloot, H.; Daviglus, M. L.; Stamler, J.; Nicholson, J. K.; Elliott, P.; Holmes, E. Metabolome-wide association study identifies multiple biomarkers that discriminate north and south Chinese populations at differing risks of cardiovascular disease INTERMAP study. *J. Proteome Res.* **2010,** *9*, (12), 6647-6654.

(20) Spraul, M.; Schutz, B.; Rinke, P.; Koswig, S.; Humpfer, E.; Schafer, H.; Mortter, M.; Fang, F.; Marx, U. C.; Minoja, A. NMR-based multi parametric quality control of fruit juices: SGF Profiling. *Nutrients* **2009,** *1*, (2), 148-155.

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

(22) Moore, G. R.; Ratcliffe, R. G.; Williams, R. J. P. NMR and the Biochemist. *Essays Biochem.* **1983,** *19*, 142-195.

(23) Nicholson, J. K.; Buckingham, M. J.; Sadler, P. J. High-resolution ^{1}H-NMR studies of vertebrate blood and plasma. *Biochem. J.* **1983,** *211*, (3), 605-615.

(24) Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y. L.; Ebbels, T. M. D.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat. Protoc.* **2010,** *5*, (6), 1019-1032.

(25) Nicholson, J. K.; Wilson, I. D.; Lindon, J. C. Pharmacometabonomics as an effector for personalized medicine. *Pharmacogenomics* **2011,** *12*, (1), 103-111.

(26) 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.

(27) Gavaghan, C. L.; Li, J. V.; Hadfield, S. T.; Hole, S.; Nicholson, J. K.; Wilson, I. D.; Howe, P. W. A.; Stanley, P. D.; Holmes, E. Application of NMR-based metabolomics to the investigation of salt stress in maize (*Zea mays*). *Phytochem. Anal.* **2011,** *22*, (3), 214-224.

(28) Lee, D. Y.; Fiehn, O. High quality metabolomic data for *Chlamydomonas reinhardtii*. *Plant Methods* **2008,** *4*, 13.

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

(30) Malz, F., Quantitative NMR in the solution state NMR. In *NMR spectroscopy in pharmaceutical analysis*, Holzgrabe, U.; Wawer, I.; Diehl, B., Eds. Elsevier: Oxford, 2008; pp 43-62.

(31) 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.

(32) Jacobsen, N. E. *NMR spectroscopy explained: Simplified theory, application, and examples for organic chemistry and structural biology*. John Wiley and Sons, Inc: Hoboken, 2007.

(33) Szantay, C., Jr.; Beni, Z.; Balogh, G.; Gati, T. The changing role of NMR spectroscopy in off-line impurity identification: A conceptual view. *TrAC - Trends Anal. Chem.* **2006,** *25*, (8), 806-820.

(34) Webb, A. Increasing the sensitivity of magnetic resonance spectroscopy and imaging. *Anal. Chem.* **2012,** *84*, (1), 9-16.

(35) Styles, P.; Soffe, N. F.; Scott, C. A.; Cragg, D. A.; Row, F.; White, D. J.; White, P. C. J. A high-resolution NMR probe in which the coil and preamplifier are cooled with liquid helium. *J. Magn. Reson.* **1984,** *60*, (3), 397-404.

(36) Fratila, R. M.; Velders, A. H., Small-volume nuclear magnetic resonance spectroscopy. In *Annu Rev Anal Chem*, Cooks, R. G.; Yeung, E. S., Eds. 2011; Vol. 4, pp 227-249.

(37) Jones, C. J.; Larive, C. K. Could smaller really be better? Current and future trends in high-resolution microcoil NMR spectroscopy. *Anal. Bioanal. Chem.* **2012,** *402*, (1), 61-68.

(38) Ebel, A.; Dreher, W.; Leibfritz, D. Effects of zero-filling and apodization on spectral integrals in discrete Fourier-transform spectroscopy of noisy data. *J. Magn. Reson.* **2006,** *182*, (2), 330-338.

(39) Rabenstein, D. L.; Millis, K. K.; Strauss, E. J. Proton NMR spectroscopy of human blood plasma and red blood cells. *Anal. Chem.* **1988,** *60*, (24), A1380-A1391.

(40) 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.

(41) Pauli, G. F.; Goedecke, T.; Jaki, B. U.; Lankin, D. C. Quantitative H-1 NMR. Development and potential of an analytical method: An update. *J. Nat. Prod.* **2012,** *75*, (4), 834-851.

(42) 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. Pharm. Biomed. Anal.* **2010,** *52*, (5), 645-651.

(43) Hoult, D. I. Solvent peak saturation with single-phase and quadrature Fourier transformation. *J. Magn. Reson.* **1976,** *21*, (2), 337-347.

(44) 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., Ser A* **1995,** *117*, (2), 295-303.

(45) Wu, P. S. C.; Otting, G. SWET for secure water suppression on probes with high quality factor. *J. Biomol. NMR* **2005,** *32*, (3), 243-250.

(46) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscropy of aqueous solutions. *J. Biomol. NMR* **1992,** *2*, (6), 661-665.

(47) Hwang, T. L.; Shaka, A. J. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Reson., Ser A* **1995,** *112*, (2), 275-279.

(48) Claridge, T. D. *High-resolution NMR techniques in organic chemistry*. 2 ed.; Elsevier: Oxford, 2009.

(49) Xi, Y.; Rocke, D. M. Baseline correction for NMR spectroscopic metabolomics data analysis. *BMC Bioinformatics* **2008,** *9*.

(50) 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.

(51) Bales, J. R.; Higham, D. P.; Howe, I.; Nicholson, J. K.; Sadler, P. J. Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. *Clin. Chem.* **1984,** *30*, (3), 426-432.

(52) Ye, T.; Zheng, C.; Zhang, S.; Gowda, G. A. N.; Vitek, O.; Raftery, D. "Add to subtract": A simple method to remove complex background signals from the ^{1}H nuclear magnetic resonance spectra of mixtures. *Anal. Chem.* **2012,** *84*, (2), 994-1002.

(53) Carr, H. Y.; Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* **1954,** *94*, (3), 630-638.

(54) Meiboom, S.; Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* **1958,** *29*, (8), 688-691.

(55) Lucas, L. H.; Larive, C. K.; Wilkinson, P. S.; Huhn, S. Progress toward automated metabolic profiling of human serum: Comparison of CPMG and gradient-filtered NMR analytical methods. *J. Pharm. Biomed. Anal.* **2005,** *39*, (1-2), 156-163.

(56) Shanaiah, N.; Desilva, M. A.; Gowda, G. A. N.; Raftery, M. A.; Hainline, B. E.; Raftery, D. Class selection of amino acid metabolites in body fluids using chemical derivatization and their enhanced C-13 NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, (28), 11540-11544.

(57) Ye, T.; Mo, H. P.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S. C.; Raftery, D. Chemoselective N-15 tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. *Anal. Chem.* **2009,** *81*, (12), 4882-4888.

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

(59) 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.

(60) Rabenstein, D. L.; Keire, D. A., Quantitative chemical analysis by NMR. In *Modern NMR Techniques and Their Application in Chemistry*, Popov, A. I.; Hallenga, K., Eds. Marcel Dekker, Inc: New York, 1991.

(61) Diehl, P.; Sykora, S.; Vogt, J. Automatic analysis of NMR spectra: An alternative approach. *J. Magn. Reson.* **1975,** *19*, (1), 67-82.

(62) Mercier, P.; Lewis, M. J.; Chang, D.; Baker, D.; Wishart, D. S. Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *J. Biomol. NMR* **2011,** *49*, (3-4), 307-323.

(63) Davis, R. A.; Charlton, A. J.; Godward, J.; Jones, S. A.; Harrison, M.; Wilson, J. C. Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemom Intell. Lab. Syst.* **2007,** *85*, (1), 144-154.

(64) De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Rietzschel, E. R.; De Buyzere, M. L.; Langlois, M. R.; Bekaert, S.; Martins, J. C.; Van Criekinge, W. Evaluation of standard and advanced preprocessing methods for the univariate analysis of blood serum H-1-NMR spectra. *Anal. Bioanal. Chem.* **2010,** *398*, (4), 1781-1790.

(65) 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.

(66) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabonomics. *Anal. Chem.* **2006,** *78*, (13), 4281-4290.

(67) Weckwerth, W. Metabolomics in systems biology. *Annu. Rev. Plant Biol.* **2003,** *54*, 669-689.

(68) Mason, M. E.; Johnson, B.; Hamming, M. C. Volatile components of roasted peanuts. Major monocarbonyls and some noncarbonyl components. *J. Agric. Food Chem.* **1967,** *15*, (1), 66-&.

(69) Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* **2011,** *7*, (3), 307-328.

(70) Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* **2011,** *6*, (10), 1483-1499.

(71) Meng, J.; Zhang, X.; Wu, H.; Bu, J.; Shi, C.; Deng, C.; Mao, Y. Morphine-induced conditioned place preference in mice: Metabolomic profiling of brain tissue to find "molecular switch" of drug abuse by gas chromatography/mass spectrometry. *Anal. Chim. Acta* **2012,** *710*, 125-130.

(72) Zhang, A. H.; Sun, H.; Wang, X. J. Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Anal. Bioanal. Chem.* **2012,** *404*, (4), 1239-1245.

(73) Strehmel, N., Hummel, J., Erban, A., Strassburg, K., Kopka, J. Retention index thresholds for compound matching in GC-MS metabolite profiling. *J. Chromatogr. B* **2008,** *871*, 182-190.

(74) Kováts, E. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helv. Chim. Acta* **1958,** *41*, (7), 1915-1932.

(75) 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.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009,** *37*, D603-D610.

(76) 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*, (8), 1635-1638.

(77) Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass. Spectrom.* **1999,** *10*, (8), 770-781.

(78) Kim, H. J.; Park, K. J.; Lim, J. H. Metabolomic analysis of phenolic compounds in buckwheat (Fagopyrum esculentum M.) sprouts treated with methyl jasmonate. *J. Agric. Food Chem.* **2011,** *59*, (10), 5707-5713.

(79) Delatte, T. L.; Selman, M. H. J.; Schluepmann, H.; Somsen, G. W.; Smeekens, S. C. M.; de Jong, G. J. Determination of trehalose-6-phosphate in Arabidopsis seedlings by successive extractions followed by anion exchange chromatography-mass spectrometry. *Anal. Biochem.* **2009,** *389*, (1), 12-17.

(80) Khreit, O. I. G.; Grant, M. H.; Zhang, T.; Henderson, C.; Watson, D. G.; Sutcliffe, O. B. Elucidation of the Phase I and Phase II metabolic pathways of (+/-)-4 '-methylmethcathinone (4-MMC) and (+/-)-4 '-(trifluoromethyl)methcathinone (4-TFMMC) in rat liver hepatocytes using LC-MS and LC-MS2. *J. Pharm. Biomed. Anal.* **2013,** *72*, 177-185.

(81) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **2007,** *363*, (2), 185-195.

(82) Dunn, W. B.; Ellis, D. I. Metabolomics: Current analytical platforms and methodologies. *TrAC - Trends Anal. Chem.* **2005,** *24*, (4), 285-294.

(83) Alvarez-Sanchez, B.; Priego-Capote, F.; de Castro, M. D. L. Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC - Trends Anal. Chem.* **2010,** *29*, (2), 120-127.

(84) Buszewski, B.; Noga, S. Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique. *Anal. Bioanal. Chem.* **2012,** *402*, (1), 231-247.

(85) Spagou, K.; Wilson, I. D.; Masson, P.; Theodoridis, G.; Raikos, N.; Coen, M.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Nicholson, J. K.; Want, E. J. HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies. *Anal. Chem.* **2011,** *83*, (1), 382-390.

(86) Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids, and other polar compounds. *J. Chromatogr.* **1990,** *499*, 177-196.

(87) Cubbon, S.; Bradbury, T.; Wilson, J.; Thomas-Oates, J. Hydrophilic interaction chromatography for mass spectrometric metabonomic studies of urine. *Anal. Chem.* **2007,** *79*, (23), 8911-8918.

(88) Vailaya, A.; Horvath, C. Retention in reversed-phase chromatography: partition or adsorption? *J. Chromatogr. A* **1998,** *829*, (1-2), 1-27.

(89) 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*, (1-2), 90-98.

(90) Kraak, J. C.; Jonker, K. M.; Huber, J. F. K. Solvent-generated ion-exchange systems with anionic surfactants for rapid separations of amino acids. *J. Chromatogr.* **1977,** *142*, (NOV), 671-688.

(91) Horvath, C.; Melander, W.; Molnar, I.; Molnar, P. Enhancement of retention by ion-pair formation in liquid chromatography with nonpolar sationary phases. *Anal. Chem.* **1977,** *49*, (14), 2295-2305.

(92) Cecchi, T. Ion pairing chromatography. *Crit. Rev. Anal. Chem.* **2008,** *38*, (3), 161-213.

(93) Kojima, H.; Inagaki, M.; Tomita, T.; Watanabe, T.; Uchida, S. Improved separation and characterization of lipopolysaccharide related compounds by reverse phase ion pairing-HPLC/electrospray ionization-quadrupole-mass spectrometry (RPIP-HPLC/ESI-Q-MS). *J. Chromatogr. B* **2010,** *878*, (3-4), 442-448.

(94) Zhao, Y.; Liu, G.; Liu, Y.; Yuan, L.; Hawthorne, D.; Shen, J. X.; Guha, M.; Aubry, A. Improved ruggedness of an ion-pairing liquid chromatography/tandem mass spectrometry assay for the quantitative analysis of the triphosphate metabolite of a nucleoside reverse transcriptase inhibitor in peripheral blood mononuclear cells. *Rapid Commun. Mass Spectrom.* **2013,** *27*, (3), 481-488.

(95) Wahl, V.; Ponnu, J.; Schlereth, A.; Arrivault, S. p.; Langenecker, T.; Franke, A.; Feil, R.; Lunn, J. E.; Stitt, M.; Schmid, M. Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science* **2013,** *339*, (6120), 704-707.

(96) Thurman, E. M.; Ferrer, I., Comparison of quadrupole time-of-flight, triple quadrupole, and ion-trap mass spectrometry/mass spectrometry for the analysis of emerging contaminants. In *Liquid Chromatography/Mass Spectrometry, MS/MS and Time of Flight MS*, American Chemical Society: 2003; Vol. 850, pp 14-31.

(97) Gross, J. H. *Mass Spectrometry*. Springer-Verlag: Berlin, 2004; p 507.

(98) Skoog, D. A.; West, D. M.; Holler, F. J.; Crouch, S. R. *Analtyical Chemistry: An Introduction*. 7th ed.; Harcourt: Orlando, 2000.

(99) Heumann, K. G. Isotope dilution mass spectrometry of inorganic and organic substances. *Fresenius Zeitschrift Fur Analytische Chemie* **1986,** *325*, (8), 661-666.

(100) 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*, (3), 779-787.

(101) Xia, J.; Mandal, R.; Sinelnikov, I. V.; Broadhurst, D.; Wishart, D. S. MetaboAnalyst 2.0-a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **2012,** *40*, (W1), W127-W133.

(102) Du, F.; Ruan, G.; Liu, H. Analytical methods for tracing plant hormones. *Anal. Bioanal. Chem.* **2012,** *403*, (1), 55-74.

(103) Arsenault, J. C. *A Beginner's Guide to SPE*. Waters Corporation: Milford, 2012.

(104) Ali, K.; Maltese, F.; Fortes, A. M.; Pais, M. S.; Choi, Y. H.; Verpoorte, R. Monitoring biochemical changes during grape berry development in Portuguese cultivars by NMR spectroscopy. *Food Chem.* **2011,** *124*, (4), 1760-1769.

(105) Rolcik, J.; Recinska, J.; Bartak, P.; Strnad, M.; Prinsen, E. Purification of 3-indolylacetic acid by solid phase extraction. *J. Sep. Sci.* **2005,** *28*, (12), 1370-1374.

(106) Bruce, S. J.; Rochat, B.; Beguin, A.; Pesse, B.; Guessous, I.; Boulat, O.; Henry, H. Analysis and quantification of vitamin D metabolites in serum by ultra-performance liquid chromatography coupled to tandem mass spectrometry and high-resolution mass spectrometry - a method comparison and validation. *Rapid Commun. Mass Spectrom.* **2013,** *27*, (1), 200-206.

(107) Bayle, M. L.; Wopereis, S.; Bouwman, J.; Van Ommen, B.; Scalbert, A.; Pujos-Guillot, E. Semi-targeted metabolomic approaches to validate potential markers of health for micronutrients: analytical perspectives. *Metabolomics* **2012,** *8*, (6), 1114-1129.

(108) Exarchou, V.; Godejohann, M.; van Beek, T. A.; 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*, (22), 6288-6294.

(109) Agnolet, S.; Jaroszewski, J. W.; Verpoorte, R.; Staerk, D. H-1 NMR-based metabolomics combined with HPLC-PDA-MS-SPE-NMR for investigation of standardized Ginkgo biloba preparations. *Metabolomics* **2010,** *6*, (2), 292-302.

(110) Limtiaco, J. F. K.; Jones, C. J.; Larive, C. K. Characterization of heparin impurities with HPLC-NMR using weak anion exchange chromatography. *Anal. Chem.* **2009,** *81*, (24), 10116-10123.

(111) Rezzi, S.; Vera, F. A.; Martin, F. P. J.; Wang, S.; Lawler, D.; Kochhar, S. Automated SPE-RP-HPLC fractionation of biofluids combined to off-line NMR spectroscopy for biomarker identification in metabonomics. *J. Chromatogr. B* **2008,** *871*, (2), 271-278.

(112) Defernez, M.; Kemsley, E. K. The use and misuse of chemometrics for treating classification problems. *TrAC - Trends Anal. Chem.* **1997,** *16*, (4), 216-221.

(113) Trygg, J. O2-PLS for qualitative and quantitative analysis in multivariate calibration. *J. Chemom.* **2002,** *16*, (6), 283-293.

(114) Trygg, J.; Holmes, E.; Lundstedt, T. Chemometrics in metabonomics. *J. Proteome Res.* **2007,** *6*, (2), 469-479.

(115) Ali, K.; Maltese, F.; Toepfer, R.; Choi, Y. H.; Verpoorte, R. Metabolic characterization of Palatinate German white wines according to sensory attributes, varieties, and vintages using NMR spectroscopy and multivariate data analyses. *J. Biomol. NMR* **2011,** *49*, (3-4), 255-266.

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

(117) Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos Mag* **1901,** *2*, (7-12), 559-572.

(118) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002,** *16*, (3), 119-128.

(119) Trygg, J.; Wold, S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *J. Chemom.* **2003,** *17*, (1), 53-64.

(120) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. Assessment of PLSDA cross validation. *Metabolomics* **2008,** *4*, (1), 81-89.

(121) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Anal. Chem.* **2005,** *77*, (5), 1282-1289.

(122) Blaise, B. J.; Navratil, V.; Domange, C.; Shintu, L.; Dumas, M. E.; Elena-Herrmann, B.; Emsley, L.; Toulhoat, P. Two-dimensional statistical recoupling for the identification of perturbed metabolic networks from NMR spectroscopy. *J. Proteome Res.* **2010,** *9*, (9), 4513-4520.

(123) Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Evaluation of full-resolution J-resolved H-1 NMR projections of biofluids for metabonomics information retrieval and biomarker identification. *Anal. Chem.* **2010,** *82*, (5), 1811-1821.

(124) Sands, C. J.; Coen, M.; Maher, A. D.; Ebbels, T. M. D.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Statistical total correlation spectroscopy editing of H-1 NMR spectra of biofluids: Application to drug metabolite profile identification and enhanced information recovery. *Anal. Chem.* **2009,** *81*, (15), 6458-6466.

(125) Twaddle, N. C.; Churchwell, M. I.; Vanlandingham, M.; Doerge, D. R. Quantification of deuterated bisphenol A in serum, tissues, and excreta from adult Sprague-Dawley rats using liquid chromatography with tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010,** *24*, (20), 3011-3020.

(126) Do, N. M.; Olivier, M. A.; Salisbury, J. J.; Wager, C. B. Application of quantitative (19)F and (1)H NMR for reaction monitoring and in situ yield determinations for an early stage pharmaceutical candidate. *Anal. Chem.* **2011,** *83*, (22), 8766-8771.

(127) Kichik, N.; Tarrago, T.; Giralt, E. Simultaneous (19)F NMR screening of prolyl oligopeptidase and dipeptidyl peptidase IV inhibitors. *Chembiochem* **2010,** *11*, (8), 1115-1119.

(128) Reid, D. G.; Murphy, P. S. Hydrophobicity predicts in vivo F-19 magnetic resonance detectability of fluorinated psychiatric drugs: Simple test for likely success in clinical pharmacokinetic studies. *Drug Dev. Res.* **2008,** *69*, (5), 279-283.

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

(130) Lane, A. N.; Fan, T. W. M.; Xie, Z.; Moseley, H. N. B.; Higashi, R. M. Isotopomer analysis of lipid biosynthesis by high resolution mass spectrometry and NMR. *Anal. Chim. Acta* **2009,** *651*, (2), 201-208.

(131) Sonko, B. J.; Schmitt, T. C.; Guo, L.; Shi, Q.; Boros, L. G.; Leakey, J. E. A.; Beger, R. D. Assessment of usnic acid toxicity in rat primary hepatocytes using C-13 isotopomer distribution analysis of lactate, glutamate and glucose. *Food Chem. Toxicol.* **2011,** *49*, (11), 2968-2974.

(132) Yu, T. Y.; Singh, M.; Matsuoka, S.; Patti, G. J.; Potter, G. S.; Schaefer, J. Variability in C-3-plant cell-wall biosynthesis in a high-CO_2 atmosphere by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **2010,** *132*, (18), 6335-6341.

(133) Ardenkjaer-Larsen, J. H.; Fridlund, B.; Gram, A.; Hansson, G.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K. Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100*, (18), 10158-10163.

(134) Lerche, M. H.; Meier, S.; Jensen, P. R.; Hustvedt, S. O.; Karlsson, M.; Duus, J. O.; Ardenkjaer-Larsen, J. H. Quantitative dynamic nuclear polarization-NMR on blood plasma for assays of drug metabolism. *NMR Biomed.* **2011,** *24*, (1), 96-103.

(135) Meier, S.; Jensen, P. R.; Duus, J. O. Real-time detection of central carbon metabolism in living *Escherichia coli* and its response to perturbations. *FEBS Lett.* **2011,** *585*, (19), 3133-3138.

(136) Huala, E.; Dickerman, A. W.; Garcia-Hernandez, M.; Weems, D.; Reiser, L.; LaFond, F.; Hanley, D.; Kiphart, D.; Zhuang, M. Z.; Huang, W.; Mueller, L. A.; Bhattacharyya, D.; Bhaya, D.; Sobral, B. W.; Beavis, W.; Meinke, D. W.; Town, C. D.; Somerville, C.; Rhee, S. Y. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **2001,** *29*, (1), 102-105.

(137) Jackson, M. T. Conservation of rice genetic resources: the role of the International Rice Genebank at IRRI. *Plant Mol.Biol.* **1997,** *35*, (1-2), 61-67.

(138) Arbona, V.; Iglesias, D. J.; Talon, M.; Gomez-Cadenas, A. Plant phenotype demarcation using nontargeted LC-MS and GC-MS metabolite profiling. *J. Agric. Food Chem.* **2009,** *57*, (16), 7338-7347.

(139) Steinfath, M.; Strehmel, N.; Peters, R.; Schauer, N.; Groth, D.; Hummel, J.; Steup, M.; Selbig, J.; Kopka, J.; Geigenberger, P.; van Dongen, J. T. Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotechnol. J.* **2010,** *8*, (8), 900-911.

(140) Setter, T. L.; Waters, I. Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. *Plant Soil* **2003,** *253*, (1), 1-34.

(141) Xu, K.; Xu, X.; Fukao, T.; Canlas, P.; Maghirang-Rodriguez, R.; Heuer, S.; Ismail, A. M.; Bailey-Serres, J.; Ronald, P. C.; Mackill, D. J. SUB1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **2006,** *442*, (7103), 705-708.

(142) Bailey-Serres, J.; Fukao, T.; Ronald, P.; Ismail, A.; Heuer, S.; Mackill, D. Submergence tolerant rice: SUB1's journey from landrace tomodern cultivar. *Rice* **2010,** *3*, (2-3), 138-147.

(143) Bailey-Serres, J.; Voesenek, L. Flooding stress: Acclimations and genetic diversity. *Annu. Rev. Plant Biol.* **2008,** *59*, 313-339.

(144) Fukao, T.; Xu, K. N.; Ronald, P. C.; Bailey-Serres, J. A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* **2006,** *18*, (8), 2021-2034.

(145) Fukao, T.; Bailey-Serres, J. Ethylene - A key regulator of submergence responses in rice. *Plant Sci.* **2008,** *175*, (1-2), 43-51.

(146) Fukao, T.; Yeung, E.; Bailey-Serres, J. The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell* **2011,** *23*, (1), 412-427.

CHAPTER TWO

Differential Metabolic Regulation Governed by the Rice *SUB1A* Gene during Submergence Stress and Identification of Alanylglycine by ^{1}H-NMR Spectroscopy

Based on a paper published in Journal of Proteome Research

J. Proteome Res. 2012, 11, 320-330

Acknowledgements: For their contributions to this research, I would like to thank the following people: Dr. Szabolcs Béni (Semmelweis Unviersity) for his help in sample preparation and analysis and Dr. Takeshi Fukao for his and assistance with growing the plant material required for this research and for our critical discussions concerning the biological significance of the results.

Abstract:

Although the genetic mechanism of submergence survival for rice varieties containing the rice *SUBMERGENCE1A* (*SUB1A)* gene has been elucidated, the downstream metabolic effects have not yet been evaluated. In this chapter, the metabolomes of *Oryza sativa* ssp. *japonica* cv. M202 and cv. M202(*Sub1*) were profiled using ^{1}H NMR spectroscopy to compare the metabolic effect of submergence stress and recovery on rice in the presence or absence of *SUB1A.* Significant changes were observed in the NMR resonances of compounds in pathways important for carbohydrate metabolism. The presence of *SUB1A* in M202(*Sub1*) was correlated with suppression of carbohydrate metabolism in shoot tissue, consistent with the role of *SUB1A* in limiting starch catabolism to fuel elongation growth. The absence *of SUB1A* in M202 was correlated

with greater consumption of sucrose stores and accumulation of amino acids that are synthesized from glycolysis intermediates and pyruvate. Under submergence conditions, alanine, a product of pyruvate metabolism, showed the largest difference between the two varieties, but elevated levels of glutamine, glutamate, leucine, isoleucine, threonine, and valine were also higher in M202 compared with the M202(*Sub1*) variety. The identification and characterization of alanylglycine (AlaGly) in rice is also reported. After 3 d of submergence stress, AlaGly levels decreased significantly in both genotypes but did not recover within 1 d of de-submergence with other the metabolites evaluated. The influence of *SUB1A* on dynamic changes in the metabolome during complete submergence provides new insights into the functional roles of a single gene in invoking a quiescence strategy that helps stabilize crop production in submergence-prone fields.

2.1 Introduction

Exploration of the metabolic response of rice to submergence and reoxygenation stress was undertaken with the goal of providing new insights into metabolic regulation governed by the *SUB1*A gene. Through transcript and biochemical analyses, the *SUB1*A gene has been linked to lower carbohydrate consumption under submerged conditions resulting in a quiescence strategy that enables the plant to endure prolonged flash flooding events (Figure 1.6).[1-3] For rice plants that do not contain the *SUB1*A gene, submergence leads to increased carbohydrate consumption and shoot elongation, with the plant attempting to outgrow the flood waters to resume oxidative respiration and photosynthesis.[1-3]

Recently, Jung et al.[4] and Mustroph et al.[5] reported microarray analyses of mRNA transcripts in M202 and M202(*Sub1*) plants under submerged and control conditions. These near-isogenic lines differed significantly in the submergence responsive expression of over 800 genes. Both studies showed variation in genes associated with carbohydrate metabolism, alcoholic fermentation and the biosynthesis of various amino acids. Although *SUB1*A coordinates accumulation of various transcripts associated with acclimation responses to submergence, the influence of the gene on metabolic adjustments remains unclear.

The goal of this chapter is to use NMR-based metabolomics to evaluate the downstream effects of *SUB1*A on the metabolic profile of rice plants under control and submerged conditions, and during submergence recovery. These experiments will provide insights into metabolic regulation during low oxygen stress in rice plants with and without the *SUB1*A gene, thereby providing insights into the complex submergence response of submergence tolerant and intolerant rice. Metabolite levels were measured in both the M202 and M202(*Sub1*) genotypes in two time course experiments performed for 0, 1, 2, 3 d, or 0, 3, 7, 12 d of submergence in both genotypes. Metabolite levels were also measured after a 1 d recovery period to probe whether *SUB1*A affects the reestablishment of a normal metabolic profile upon de-submergence.

2.2 Materials and Methods

2.2.1 Chemicals.

Deuterium oxide, sodium-3-trimethylsilylpropionic acid – d_4 (TMSP), acetic acid – d_4, and ammonium deuteroxide – d_5 were purchased from Cambridge Isotope

Laboratories, Inc. (Andover, Massachusetts, USA). Pure water (18 MΩ) was obtained by filtration with a Millipore filtration system (Millipore, Billerica, MA, USA). All other solvents used for this study were HPLC grade (≥ 99% purity). Methanol (MeOH), glacial acetic acid, and D,L-alanine glycine (AlaGly) were obtained from Fisher Scientific (Pittsburgh, PA, USA). Ammonium hydroxide was obtained from EMD Chemicals (Merck, Darmstadt, GER). All other solvents were obtained from Sigma-Aldrich (St. Louis, MO, USA). Metabolite standards were purchased from the following companies: shikimate, glucose and sucrose from Sigma-Aldrich, valine (Val), leucine (Leu), isoleucine (Ile), glutamate (Glu), aspartate (Asp), asparagine (Asn), alanine (Ala), threonine (Thr), serine (Ser), and tyrosine (Tyr) from Fisher Scientific, and glutamine (Gln) from MP Biomedicals (Solon, OH, USA).

2.2.2 Growth Conditions and Plant Materials

Rice was grown and submergence-stressed as reported by Fukao, et al.[6] Briefly, the seeds were surface sterilized with 1% (v/v) sodium hypochlorite and 0.2% (v/v) Tween-20, and rinsed thoroughly using deionized (DI) water. The seeds were soaked in DI water overnight in the dark and placed on moist paper in a Pyrex dish covered with plastic wrap for 5 days. Following germination, seeds were transplanted into pots with soil (25 plants per pot) and grown for 12 days (until the 3-leaf stage) in greenhouse conditions at 30 °C. The survival rate of the transplanted seedlings was greater than 99%.

2.2.3 Submergence Treatment and Plant Harvest

Prior to submergence treatment, six 12 1 liter trash cans were filled with DI water and allowed to equilibrate overnight in the greenhouse to the ambient temperature (~30 °C). Once the seedlings reached the three fully-expanded leaf stage, two submergence experiments were performed. In the long-term submergence experiment plants were submerged for 0, 3, 7, and 12 d or were subjected to 12 d submergence and 1 d recovery. In the short-term submergence experiment, plants were submerged for 0, 1, 2, 3 d or were subjected to 3 d submergence and 1 d recovery. For the recovery, plants were de-submerged and were allowed to recover for 24 h on the greenhouse bench. For tissue harvest, plants were de-submerged and shoot tissue was immediately harvested, rinsed in DI water, flash frozen in liquid nitrogen and stored at -80 °C. Control tissues (the 0 d treatment from both experiments) were harvested at the beginning of submergence treatment. In the short-term experiment, an additional 4 d no-submergence control was harvested with the 3 d submergence and 1 d recovery plants. All submergence treatments began at 1 PM and all plants were harvested at 1 PM on the day of treatment completion. Prior to analysis, shoot tissue was ground to a fine powder by mortar and pestle under liquid nitrogen, lyophilized overnight until dry, and stored at -80 °C. There were a total of six biological replicates for each condition, with the plants from one pot (n=25) combined in a tissue pool to create one biological replicate.

2.2.4 Metabolite Extraction

2.2.4.1 Metabolite Extraction for Samples from the Long-term Submergence Treatment

Metabolites were extracted from 20 mg dry weight (D.W.) of lyophilized tissue in 80/20 MeOH/H_2O using a procedure similar to that described by Kaiser et al. (2009).[7] The aqueous component of the extraction solvent contained 100 mM acetic acid adjusted to pH 4.7 with ammonium hydroxide. A 1.0 mL aliquot of the MeOH/H_2O extraction buffer was added to the dried tissue and agitated for 1 m at 300 rpm using a platform shaker. Samples were centrifuged at 12 000 x g for 4 m and 800 µL of supernatant transferred to a clean 1.5 mL Eppendorf micro-centrifuge tube (the pellet was discarded). Samples were centrifuged under vacuum overnight until dry using a Thermo-Savant SC110 model speed vacuum equipped with a RVT400 refrigerated vapor trap and a GP110 gel pump. Dried samples were stored at -20 °C prior to reconstitution and analysis. Immediately prior to NMR analysis, samples were dissolved in 700 µL of D_2O reconstitution buffer containing 100 mM deuterated acetic acid adjusted to pD 7.6 with ammonium deuteroxide – d_5. The solution pD was calculated from the pH meter reading (pH*) using the equation pD = pH* + 0.4.[8] The reconstitution buffer also contained 175 µM TMSP as a ^{1}H NMR chemical shift reference. Prior to analysis, a liquid-liquid extraction was performed with 100 µL $CDCl_3$ to remove residual lipids.[7] The sample was briefly vortexed and then centrifuged for 2 m at 5 000 x g to break the emulsion after which a 620 µL aliquot of the D_2O phase was transferred to a 5 mm NMR tube for analysis.

2.2.4.2 Metabolite Extraction for Samples from the Short-term Submergence Treatment.

The extraction for the short-term treatment was performed as described above except that 45 mg of dry tissue was extracted with 1.5 mL of the 80/20 MeOH/H_2O extraction buffer. After centrifugation for 2 m at 5 000 x g to break the emulsion, 1.0 mL of the supernatant was transferred to a clean vial for drying by speed vacuum. The remaining steps prior to analysis were performed as described above. This extraction method differed from the protocol used for the long-term submergence experiment because more tissue was available for extraction, allowing increased metabolite concentrations and shorter NMR analysis times.

2.2.5 NMR Analysis

2.2.5.1 Data Acquisition for Metabolite Analysis

All samples were analyzed with a 14.1 T Bruker Avance NMR spectrometer tuned to detect ^1H resonances at 599.84 MHz. ^1H NMR spectra were collected with a 5 mm inverse broadband probe with xyz gradients using the Bruker-defined wet pulse program (wet) to reduce the intensity of the residual water resonance.[9] Digital quadrature detection (DQD) was used with a transmitter frequency set on the water resonance. The probe was tuned and matched manually and the magnetic field homogeneity was optimized using up to 28 shims. The sample temperature was maintained at 298 K. Samples were locked using D_2O and spectra were acquired without spinning. The TMSP line-widths after apodization with an exponential function equivalent to 0.5 Hz line broadening were 1.6 ± 0.2 Hz and 1.4 ± 0.1 Hz for the long-term and short-term submergence experiments, respectively. Free induction decays (FIDs) were collected

into 32 768 points and zero-filled to 65 536 points. For the long-term submergence experiments, a spectral width of 11.97 ppm was collected with a 10.7 μs 90° pulse. A relaxation delay of 1.5 s was used and 16 dummy scans were followed by coaddition of 1200 transients for a total experiment time of 1.30 h at a temperature of 298 K. For the short-term submergence experiment, the same parameters were applied except a 10.5 μs 90°excitation pulse was used to collect 16 dummy scans followed by 640 co-added scans for a total experiment time of 0.70 h.

2.2.5.2 Data Acquisition for Metabolite Identification

For metabolite identification, 200 mg of rice (D.W.) were extracted (as described for the short-term experiment) and the extracts combined to form a concentrated rice sample that was analyzed by NMR. Metabolite identification was facilitated by acquisition of two-dimensional NMR spectra which provided homonuclear and heteronuclear correlations.

The total correlation spectroscopy (TOCSY) spectrum was measured using the Bruker pulse sequence mlevphpr with presaturation of the residual water resonance during the 2.0 s relaxation delay for a total experiment time of 12.33 h.[10] A ^{1}H transmitter frequency of 4.696 ppm was used with spectral widths of 11.97 ppm in F1 and F2 excited by an 11.38 μs 90°excitation pulse. Phase-sensitive data was acquired using DQD according to the States-TPPI method.[11] Co-addition of 32 FIDs for each of 512 increments was preceded by 16 dummy scans with 8192 data points per FID (complex). A trim-pulse of 2.5 ms was used with a spin-lock duration of 80 ms. Spectra were processed using a sine-bell shaped window function phase-shifted by 90° and were

zero-filled to yield a 8192 x 2048 data matrix prior to Fourier transformation. The phase of each spectrum was adjusted manually and automatic baseline correction performed using a 5th order polynomial function.

The multiplicity-edited ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectrum was collected using DQD according to the echo-antiecho method[12-15] using the Bruker pulse program hsqcedetgpsip.2 with ^{13}C GARP decoupling during acquisition applied with a ±14 kHz decoupling bandwidth. The 11.97 ppm ^{1}H spectral width was excited with a 9.25 µs ^{1}H 90° pulse and the 140 ppm ^{13}C spectral width was excited by a ^{13}C 90° pulse of 19.5 µs. The ^{1}H offset frequency was 4.696 ppm, the ^{13}C offset frequency was 80 ppm, and the J_{CH} was set to 145 Hz. The relaxation delay was 1.5 s and 16 dummy scans preceded co-addition of 80 FIDs for each of 512 increments. The total experiment time for the metabolite extract was 19 h 5 m. Spectra were zero-filled to give a 4096 x 1024 data matrix, apodized using a sine-bell window function phase-shifted 90°, Fourier transformed and baseline, and phase corrected as described above.

The ^{1}H-^{13}C heteronuclear multiple bond correlation (HMBC) spectrum was collected using DQD according to the QF method using the Bruker pulse program hmbcgplpndqf.[16, 17] A 9.38 µs ^{1}H 90° pulse excited the 11.97 ppm ^{1}H spectral width and a 19.5 µs ^{13}C pulse excited the 260 ppm spectral width. The ^{1}H and ^{13}C offsets were 4.700 and 100.00 ppm, respectively. The J_{CH} was set to 159 Hz and the long range coupling set at 4.59 Hz. A relaxation delay of 1.5 s and 16 dummy scans preceded the co-addition of 96 FIDs for each of 512 increments for a total experiment time of

1 d 2 h 5 m. Spectra were zero-filled to give an 8192 x 1024 data matrix, and apodized, Fourier transformed, baseline and phase corrected as described for the HSQC experiment.

For identification of AlaGly, a multiplicity-edited ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC spectrum was collected for a 5 mM solution of the racemic AlaGly compound in the same deuterated buffer described in section 2.2.4.1. The experiment was performed similarly to that for the concentrated rice extract as described above except co-addition of 16 FIDs followed 16 dummy scans for each of 256 increments for a total experiment time of 1 h 54 m for the HSQC experiment and 2 h 10 m for the HMBC experiment.

2.2.5.3 NMR Data Processing and NMR Prediction

One-dimensional NMR spectra were processed using MestReNova version 6.1.1 (Mestrelab Research S.L., Santiago de Compostela, Spain). Due to the complex nature of the spectra, automatic phasing was followed by manual adjustment as needed and manual baseline correction using the multipoint baseline correction feature in MestReNova. Baseline points were selected in spectral regions not containing any peaks and the segments algorithm of the multipoint baseline correction feature was selected to connect the points with a flat line that the program defines as a baseline value of zero. Peak integration for ^{1}H-NMR was performed using the line fitting function available on MestReNova with a 0.05 Hz and 10.00 Hz lower and upper width constraints, respectively, a 13.0% position constraint, and a maximum of 200 iterations with a local minima filter of 5. 2D NMR spectra were processed using Bruker TopSpin ver. 1.3 (Bruker Biospin Corporation, Billerica, MA, USA). ^{1}H and ^{13}C spectra predictions for

AlaGly were calculated using the ACDlabs prediction software (Advanced Chemistry Development, Inc, Toronto, Canada).

2.2.6 Data Analysis

2.2.6.1 Principal Component Analysis

Principal component analysis (PCA) was carried out using Minitab 15 (Minitab Inc, State Park, PA, USA). Bucket integration (binning) for PCA was performed manually with integral bins spanning 0.02 ppm from 0.5 - 9.0 ppm.[28] The data was mean-centered using Excel. Each spectrum was normalized to the sum of the bins ranging from 0.5 - 9.0 ppm and excluding regions containing HOD (4.54 - 5.00 ppm), methanol (3.34 - 3.38 ppm), and a minor buffer contaminant identified by analysis of a buffer blank (1.22 - 1.28 and 0.88 - 0.92 ppm). Because sucrose was by far the most abundant metabolite in the spectrum, bins corresponding to the sucrose resonances were removed prior to multivariate analysis to better evaluate variance contributions of other metabolites.

2.2.6.2 Univariate Statistical Analysis

Univariate data analysis was carried out using the integrals of the resonances of individual metabolites from the peak-fitted data and normalized to the sum of the entire spectrum (as described in section 2.2.6.1). Fold changes were calculated by averaging the results for at least five biological replicates of each variety per time point and taking the ratio of the averages determined for the M202 and M202(*Sub1*) varieties. The significance of the fold changes were evaluated using SPSS Statistics 19 (IBM, Armonk, NY, USA). Trajectory plots were generated using Origin 7.5 (OriginLab, Northampton, MA, USA) and used in Figures 2.2 and 2.5 to represent the relative area of a specific

metabolite over the course of the treatment period. The error bars included in these plots represent the standard deviation of the mean. Averages and standard deviations for the trajectory plots were calculated in Excel (Microsoft, Redmond, WA, USA).

2.3 Results and Discussion

Two experiments were carried out to evaluate the metabolic differences between the shoot tissue of the M202(*Sub1*) and M202 rice varieties during and one day following submergence stress. Initially, a long-term submergence experiment was conducted to evaluate differences in metabolite reconfiguration over an extended time period and to establish how early in the stress significant differences in the metabolite levels of the two varieties could be detected. The long-term experiment consisted of a 0 d control, 1, 3, 7, and 12 d submergence or 12 d submergence with 1 d of recovery. After observing marked differences in relative metabolite levels as early as day 3 of the submergence treatment, we performed a shorter time-course consisting of a 1 d control (1dc), 1, 2, and 3 d of submergence or 3 d of submergence with 1 d recovery. The shorter time-course was designed to observe the more immediate submergence response of the two varieties, and because M202 did not survive being submerged for 12 d, this experiment also provided the opportunity to evaluate recovery of both the M202 and M202(*Sub1*) varieties following de-submergence. The short-term submergence experiment also included a terminal control harvested with the 1 d recovery samples. The purpose of this 4 d control (4dc) was to account for metabolite differences as a function of plant growth over the course of the treatment period.

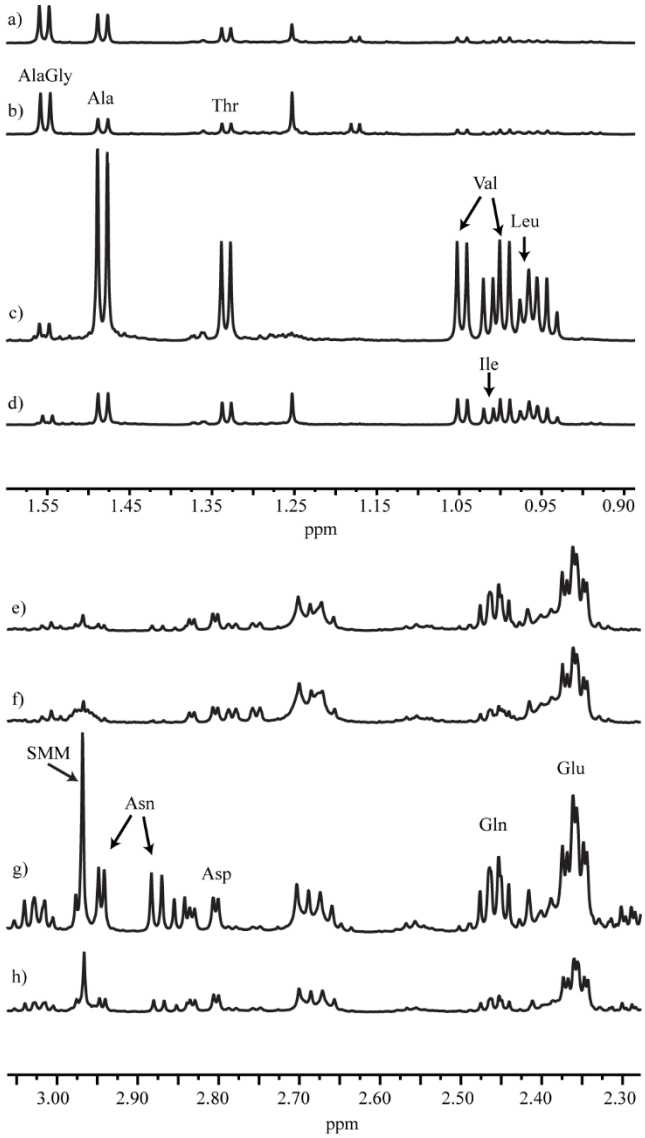


Figure 2.1. Selected regions from representative ^{1}H NMR spectra comparing the M202(*Sub1*) and M202 rice varieties at different submergence time points: (a) and (e) M202 1 d control (1dc), (b) and (f) M202(*Sub1*) 1dc, (c) and (g) M202 at 3 d submergence, and (e) and (h) M202(*Sub1*) at 3 d submergence. The well-resolved resonances of selected metabolites are labeled.

The evaluation of metabolite reconfiguration as a consequence of submergence stress in the presence or absence of *SUB1*A was accomplished using NMR spectroscopy. Figure 2.1 shows representative spectra from the short-term submergence experiment comparing the two nearly genetically identical varieties. Two regions of the spectrum were chosen to represent the effects of submergence on plants with and without the *SUB1*A gene. The first region contains the resonances of Ala, Ile, Thr, and Val (Figure 2.1a-d) while the second region contains Asn, Asp, Glu, Gln, and S-methylmethionine (SMM) (Figure 2.1 e-h). The panels compare the 1 d control (Figure 2.1 a, b, e and f) and 3 d submergence treatment samples (Figure 2.1 c, d, g and h) of the two genotypes. Metabolic differences are clearly distinguishable as a consequence of submergence and as determined by genotype in the representative spectra for metabolites such as Ala, Gln, SMM, Thr, and Val demonstrating both the impact of submergence and an effect of the presence of the *SUB1*A gene. Although the control spectra of the two genotypes are similar, the response of the M202 variety to submergence stress produces a significant elevation in the levels of several metabolites compared to M202(*Sub1*). These data demonstrate that *SUB1*A has an observable impact on metabolic regulation during submergence stress even without data manipulation (such as normalization).

2.3.1 Evaluation of the Data Normalization Method

Quantitative interpretation of the metabolomics data measured for the two submergence experiments was accomplished using PCA and by direct comparison of resonance integrals. The data was sum-normalized prior to statistical analysis to compensate for differences in extraction efficiency and other experimental parameters,

for example receiver gain or dilution.[18] To ensure that the sum-normalization method was robust, non-normalized ratios of the integrals measured for the anomeric resonances of sucrose and glucose (α anomer only) were calculated for each replicate. After averaging the six replicates for each experiment and calculating the standard deviation from the raw integrated data, the ratios were plotted as a function of treatment (Figure 2.2a) and compared to the sum-normalized metabolite abundance plots (Figure 2.2b and c). Analysis of the sum-normalized data for the individual metabolites showed a decrease of available sucrose in both varieties (Figure 2.2b), although the decrease was less severe in the submergence tolerant M202(*Sub1*) variety, consistent with an earlier report.[6] Sucrose decreased continuously in M202 shoot tissue, whereas the decline in M202(*Sub1*) was significantly attenuated after an initial spike in glucose levels (Figure 2.2c). The evaluation of the sucrose:glucose ratio (Figure 2.2a) revealed that the relative levels of these two carbohydrates was maintained during submergence in both genotypes but increased dramatically in response to de-submergence in the M202(*Sub1*) variety. In this comparative analysis, the sum-normalization of the metabolite abundance data set provided an unbiased evaluation of metabolic changes.

2.3.2 Global Analysis of NMR Metabolite Profiles Comparing Submergence-stressed M202 and M202(*Sub1*) Rice

Global metabolic differences between the NMR spectra measured for extracts of the control, submergence stressed and de-submerged M202 and M202(*Sub1*) plants were first explored by PCA of the sum-normalized bins. These bins comprised 356 variables

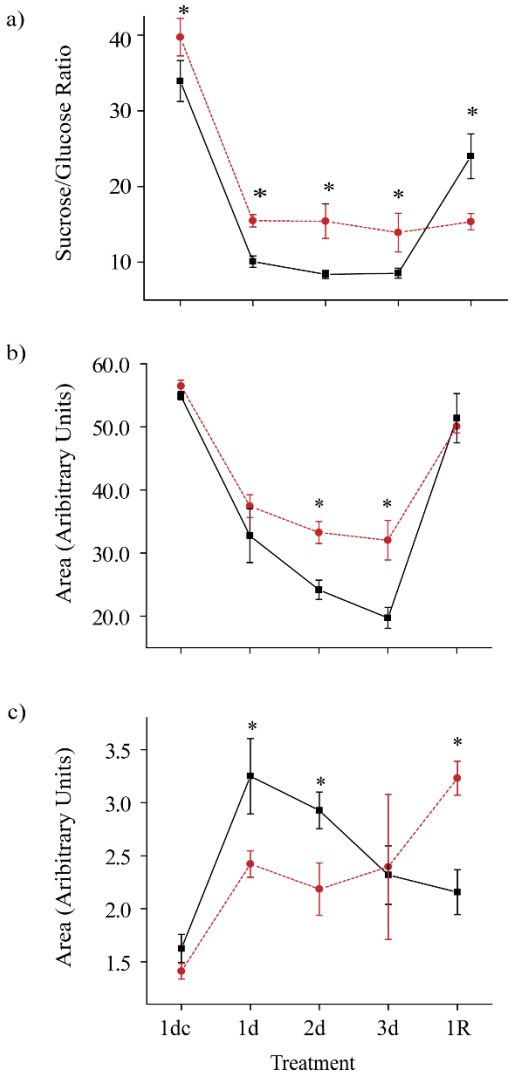


Figure 2.2. Relative abundance of sucrose and glucose over the time course of short-term submergence and recovery in the two rice genotypes, M202 (■) and M202(*Sub1*) (●). (a) The trajectory plot shows the sucrose/glucose ratios determined from the raw integrated NMR data. The trajectory plots for (b) sucrose and (c) glucose show the relative abundance of each metabolite after sum normalization. Each point is the average of at least five biological replicates with the error bars representing the standard deviation. Asterisks represent a significant difference between the varieties at the 95% confidence interval. Treatment time points: 1 d control (1dc), 1, 2, 3 d submergence (1d, 2d, 3d), and 3 d submergence + 1 d recovery (1R).

after removal of the solvent, contaminant and sucrose regions as described in section 2.2.6.1. The data analysis presented below explores both the effect of submergence as well as the *SUB1*A gene during submergence stress on the metabolic response of rice shoot tissue.

2.3.2.1 PCA Analysis of the Long-term Submergence Data

Figure 2.3a shows the scores plots for the long-term submergence experiment. PC1 and PC2 represented 65% and 17% of the variance, respectively, and were used for data interpretation. The long-term submergence experiment scores plot (Figure 2.3a) contains eight distinct groups that represent the control samples and the individual treatments. Of the stress treatment samples, only the 12 d submergence and 12 d submergence followed by 1 d recovery of the M202 variety grouped together. The overlap of the control time points from the two varieties indicates minimal variance at the metabolic level due to the presence or absence of the *SUB1A* gene under non-stress growth conditions. The loadings plot (Figure 2.3b) shows contributions to PC1 and PC2 due to regions of the NMR spectrum that differ in extracts of plants subjected to long-term submergence either between the two varieties of rice or between treatments within the same genotype. Loadings contributing significantly to PC1 are due to metabolites such as Asp (2.85, 2.87 ppm), Asn (2.95 ppm), and Asn/SMM (2.97 ppm). The most significant contributions to PC2 are due to the integral regions 3.41 and 3.71, 3.73 ppm, most likely reflecting changes in glucose levels.

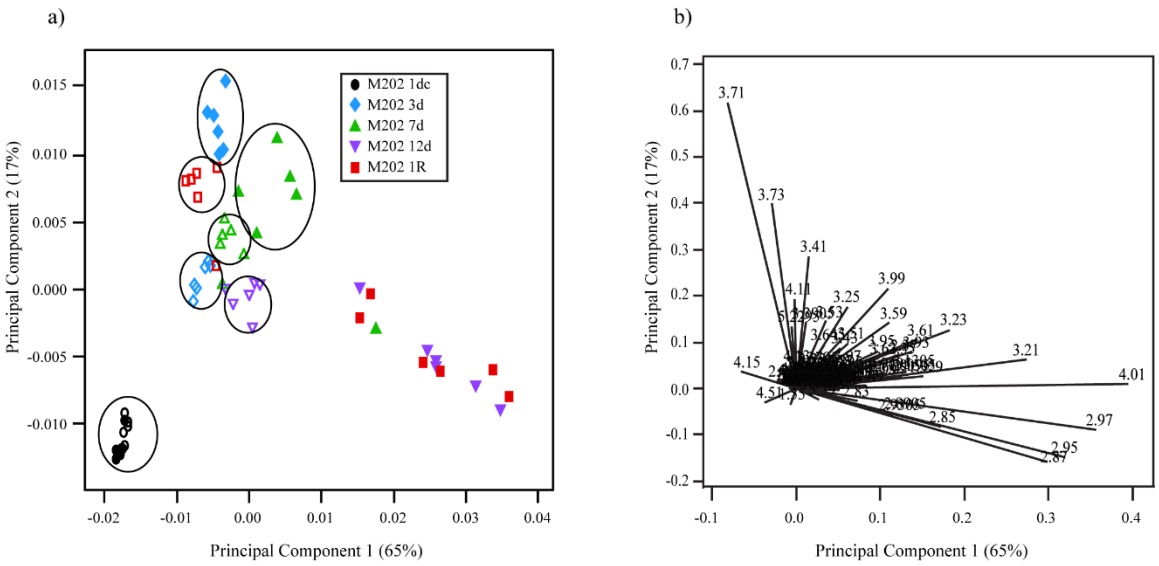


Figure 2.3. (a) The scores plot showing the first two principal components for the long-term experiment. The legend identifies the treatments for the M202 variety. A similar scheme but with open symbols represents the M202(*Sub1*) variety. (b) The loadings plot showing the variables contributing most to the variance along the first two principal components.

2.3.2.2 PCA Analysis of the Short-term Submergence Data

For the short-term submergence experiment, PC1 and PC2 represented 82% and 7% of the variance, respectively. The scores plot (Figure 2.4a) contains nine distinct groups that represent the control samples for both rice varieties and individual treatments for each variety. As for the long-term submergence experiment, all controls overlapped indicating minimal variance between the two varieties in the absence of submergence stress. Additionally, the initial (1dc) and end (4dc) controls overlapped, confirming that observed metabolic changes were a consequence of submergence stress and not the progression in plant development. The loadings plot (Figure 2.4b) identified several bins contributing to separation along PC1 and PC2. The bins representing the largest contributions to PC1are attributed to the resonances of metabolites such as Ile (0.94) Val (0.99, 1.05 ppm), Leu/Ile (0.97 ppm), Ala (1.47, 1.49 ppm), Glu (2.35, 2.37 ppm), and Asn/SMM (2.97 ppm). As was observed for the long-term submergence experiment, the most significant loadings for PC2 are the regions 3.41, 3.71, and 3.73 ppm containing the resonances of glucose. In both experiments, interaction of the treatment with the variety was responsible for the greatest amount of variance. Furthermore, the PCA results clearly showed a change in the overall metabolome in response to submergence and de-submergence that was distinct in the two varieties differing only in the presence or absence of the *SUB1*A gene.

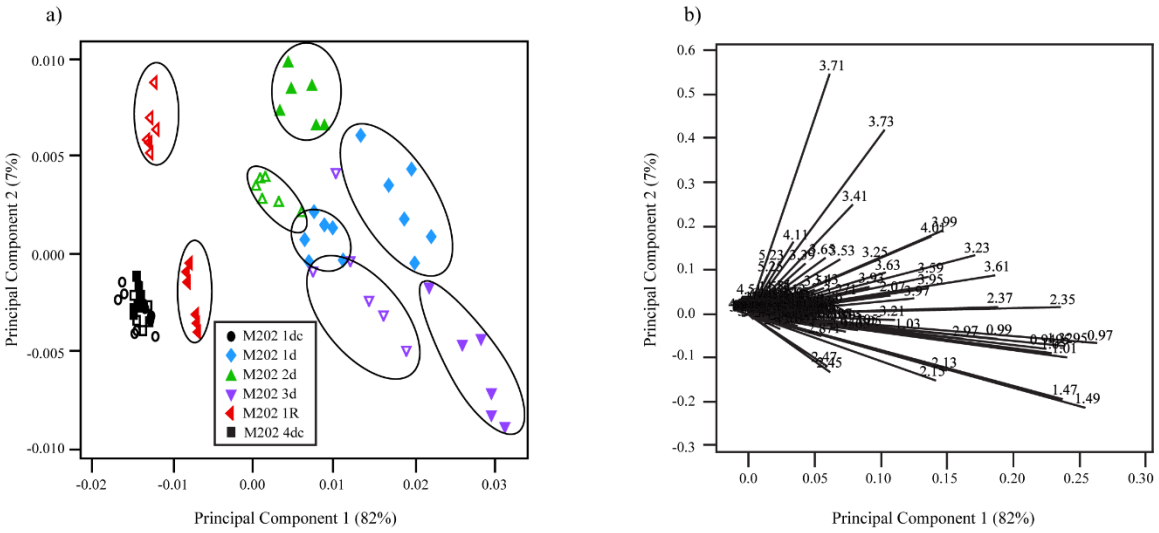


Figure 2.4. (a) The scores plot showing the first two principal components from the short-term experiment. The legend identifies the treatments for the M202 variety. A similar scheme but with open symbols represents the M202(*Sub1*) variety. (b) The loadings plot showing the variables contributing most to the variance along the first two principal components.

2.3.3 The Effects of Submergence Stress on Carbohydrate Consumption

Carbohydrate consumption and elongation growth has been shown to be a survival strategy with costs that outweigh benefits in a transient deep flooding event, causing more rapid exhaustion of energy reserves and ultimately death of varieties that do not contain the *SUB1*A gene.[6] Indeed, the data presented here provide an expanded perspective of the complex metabolic response to submergence stress in the submergence intolerant M202 and submergence tolerant M202(*Sub1*) varieties.[6] To evaluate differences in carbohydrate metabolism in the two genotypes under the treatment conditions in the 3 d submergence experiment, the normalized areas of the sucrose and glucose (α only) anomeric resonances were plotted (Figure 2.2 b and c). Differences between the metabolite levels in M202 and M202(*Sub1*) that are significant at the 95% confidence interval are indicated by an asterisk. The results in Figure 2.2 are also presented in tabular format as fold changes relative to the control for each variety (Table 2.1) and as fold changes comparatively between the two varieties (Table 2.2), with significance at the 95% confidence interval indicated by an asterisk. Our results showed a consistent distinction between the two genotypes in the changes in the abundance of sucrose during the submergence. Here we show that short-term submergence promotes a rapid decrease in sucrose levels by the first day of submergence, followed by either no further change in M202(*Sub1*) or continued reduction in M202 shoot tissue (Figure 2.2b, Tables 2.1 and 2.2). This finding is consistent with the previous proposal that *SUB1*A acts to restrict the breakdown of leaf starch and soluble carbohydrates during submergence to limit the availability of energy reserves necessary to fuel elongation

Table 2.1. Ratios of quantified metabolites comparing the treatments to the control for the same genotype. The treatments include one (1d), two (2d), and three (3d) day submergence treatments, and a three day submergence treatment followed by one day of recovery (1R). The ratios are obtained from the average of at least five biological replicates.

	M202(Sub1)				*M202*			
	1d	*2d*	*3d*	*R1*	*1d*	*2d*	*3d*	*1R*
AlaGly	1.6*	1.3	0.8	0.6*	1.2	1.0	0.6*	0.5*
Ala	3.2*	4.7*	15.2*	1.0	3.1*	7.2*	9.6*	3.4*
Asn	NQ	NQ	NQ	NQ	NQ	NQ	NQ	NQ
Asp	3.3*	2.9	3.4	0.6	2.3*	2.5	3.0	1.1*
Gln	1.7*	1.5*	1.6*	2.3*	2.0*	1.8*	1.4*	1.3*
Glu	2.2*	2.6*	3.1*	1.0*	1.9*	2.3*	2.5*	1.2*
Glucose	1.6	1.8*	4.5*	0.7*	0.8*	1.5*	2.2*	1.4*
Ile	NQ	NQ	NQ	NQ	NQ	NQ	NQ	NQ
Leu	NQ	NQ	NQ	NQ	NQ	NQ	NQ	NQ
Ser	2.2*	2.3*	3.2*	0.8*	1.9*	2.6*	2.7*	1.0
Shikimate	1.3	1.2	0.3	0.9	1.4	1.6	0.4	0.3
SMM	NQ	NQ	NQ	NQ	NQ	NQ	NQ	NQ
Sucrose	0.7*	0.6*	0.3*	0.9*	0.6*	0.4*	0.4*	0.9
Thr	5.0*	6.7*	14.6*	1.0	3.8*	6.2*	8.5*	1.6
Tyr	NQ	NQ	NQ	NQ	NQ	NQ	NQ	NQ
Val	10.5*	16.3*	31.6*	1.1*	8.8*	17.4*	23.6*	2.6*

Values marked with an asterisk indicate significant differences between the control and treated plants at the 95% confidence interval. Metabolites not detectable in the controls are represented by NQ (not quantifiable) in the reported fold change values.

Table 2.2. Ratios of the quantified metabolites measured for M202 *versus* M202(*Sub1*) at specific treatment time points in the short-term submergence experiment. The treatments include a control (1dc), one (1d), two (2d), and three (3d) day submergence treatments, and a three day submergence treatment followed by one day of recovery (1R). The ratios are obtained from the average of at least five biological replicates.

	M202/M202(Sub1)				
	1dc	*1d*	*2d*	*3d*	*1R*
AlaGly	1.3	1.0	1.0	0.8	1.0
Ala	1.6	1.6*	2.5*	2.3*	5.5*
Asn	ND	0.6	1.1	1.5	ND
Asp	1.1	0.8	0.9	1.0	1.8
Gln	2.0*	1.0	1.7*	2.0*	4.1*
Glu	1.3	1.0	1.1*	1.1*	1.5*
Glucose	1.2	1.3*	1.3*	1.0	0.7*
Ile	ND	1.1	1.5*	1.4*	4.5
Leu	ND	1.0	1.2*	1.2	2.3
Ser	1.2	1.1	1.4*	1.3	1.4
Shikimate	0.8	0.9	1.1	0.4	0.3
SMM	ND	0.7	0.9	1.2	1.9
Sucrose	1.0	0.9	0.7*	0.6*	1.0
Thr	1.7	1.3	1.6*	1.7*	2.8*
Tyr	ND	ND	1.5	1.4	2.1
Val	1.3	1.1	1.4*	1.4*	3.3*

Values marked with an asterisk indicate significant differences between the varieties at a 95% confidence interval. Metabolites not detectable in both varieties are represented by ND.

growth.[6] Evaluation of the recovery period after short-term submergence revealed a remarkable ability of both varieties to replenish sucrose to levels comparable to those of control plants, indicating that submergence stress did not irreversibly damage the capacity to accumulate sucrose upon recovery in either variety via phosphorylation. The levels of glucose were greater during submergence than during growth in air in both genotypes, with higher accumulation in M202 at day 1 and day 2 by a fold change of 1.3 compared to the M202(*Sub1*) genotype (Table 2.2, Figure 2.2c). Interestingly, glucose levels reached a maximum after 1 d of submergence and then declined over the next 2 d of submergence in M202 until de-submergence. This may reflect the more pronounced degradation of starch during submergence in M202. In fact, starch reserves in aerial tissue were more quickly consumed in M202 than M202(*Sub1*) during submergence, consistent with higher levels of transcripts encoding enzymes associated with starch and sucrose catabolism.[6] During the recovery period, relative glucose levels remained stable in the M202 variety and increased in M202(*Sub1*). The rise in glucose in M202(*Sub1*) shoots during recovery can be indicative of a higher level of photosynthesis after de-submergence in the submergence tolerant variety, consistent with the finding by Fukao and co-workers that M202(*Sub1*) shoots maintain significantly higher amounts of chlorophyll *a* and *b* during submergence.[6] These differences in the dynamics of sucrose and glucose accumulation in the two genotypes results in significant differences in the ratios of these two metabolites over the treatment time course (Figure 2.2a). Altogether, these data emphasize the function of *SUB1*A in the restriction of carbohydrate consumption in shoot tissue during submergence stress.

2.3.4 Amino Acid Metabolism.

The normalized areas of the well-resolved resonances of the other quantified metabolites were plotted for both genotypes over the time-course of the 3 d submergence experiment (Figure 2.5). Differences between the metabolite levels in M202 and M202(*Sub1*) that are significant at the 95% confidence interval are indicated by an asterisk. With the exception of Gln, the relative levels of the metabolites measured for the controls are statistically indistinguishable for the two genotypes. For most metabolites, a consistent trend is observed; normalized area increases over the course of the submergence treatment and then decreases rapidly during the recovery period (Figure 2.5).

Waterlogging and hypoxia, a component of submergence stress,[19] can promote the accumulation of Ala and other free amino acids in various organs of Arabidopsis, rice, popular, potato, soybean, wheat and *Medicago truncatula.*[20-26] When oxidative respiration is limited in plant cells, pyruvate, a final product of glycolysis, is converted to the fermentation end products ethanol and lactic acid as well as succinate, γ-aminobutyrate (GABA), Ala and some other amino acids or their precursors.[2] It has been proposed that the prioritized conversion of pyruvate to Ala may contribute to cell survival under oxygen deprivation by limiting cytosolic acidification by lactate and loss of carbon skeletons through diffusion of ethanol out of cells.[2, 27-29] Our ^{1}H-NMR spectra revealed submergence-promoted accumulation of 11 amino acids in aerial tissue of M202 and M202(*Sub1*) (Figure 2.5). Transient rises of these amino acids including Ala, Ile, Leu, Ser, Thr, Tyr and Val were observed in response to submergence in both varieties but

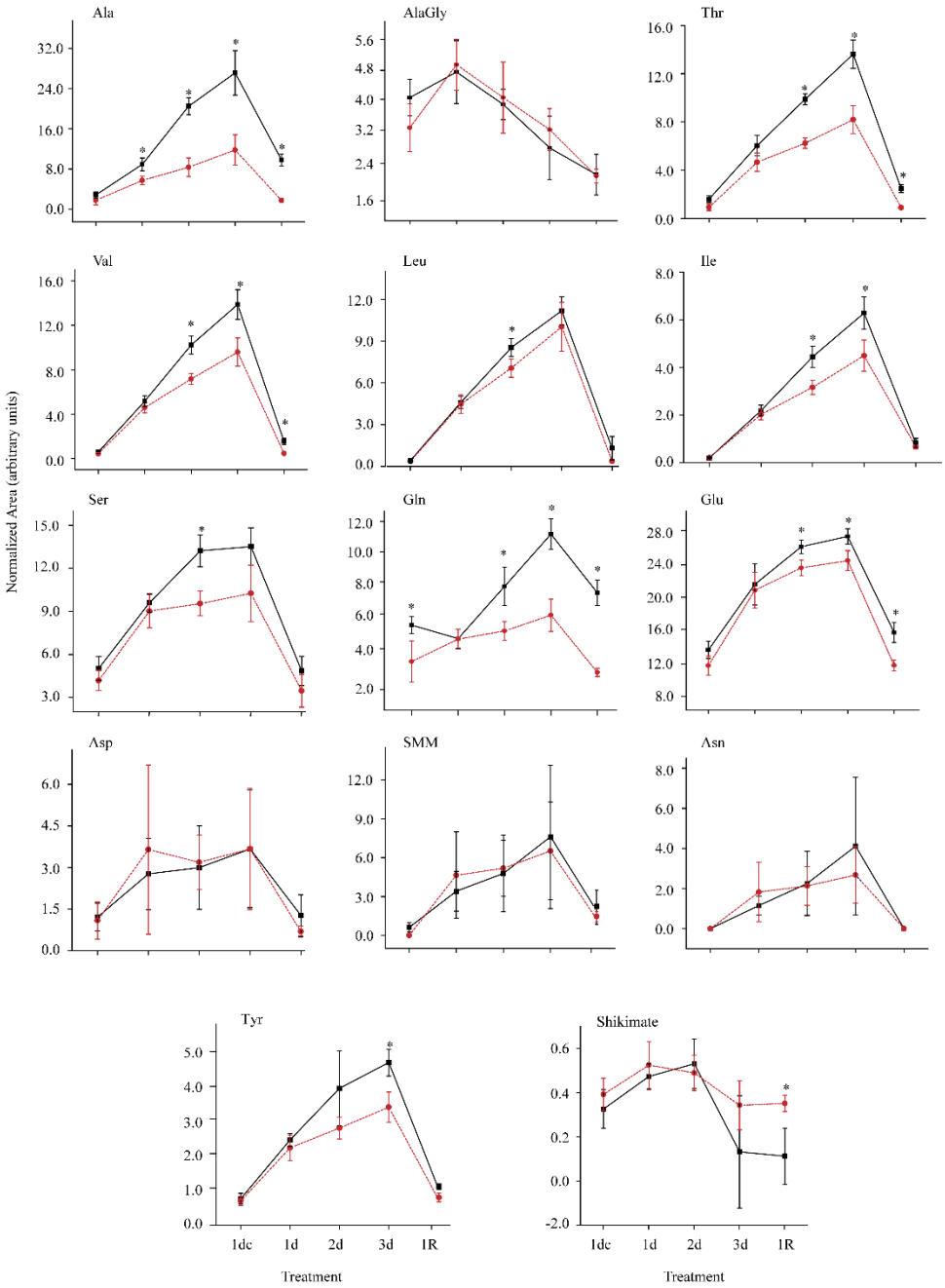


Figure 2.5. Abundance of detected metabolites over the time course of short-term submergence and recovery for M202 (■) and M202(*Sub1*) (●) by NMR. Each data point represents the average of metabolite levels determined from normalized peak areas of at least five biological replicates with the error bars indicating the standard deviation. Asterisks indicate a significant difference between the varieties at the 95% confidence interval. Treatment time points: 1 d control (1dc), 1, 2, 3 d submergence (1d, 2d, 3d), and 3 d submergence + 1 d recovery (1R).

were consistently less acute in M202(*Sub1*) shoot tissue, reaching 1.2-2.5 fold higher levels in M202 relative to M202(*Sub1*) (Table 2.2, Figure 2.2). Increases in a number of these amino acids have been reported in studies of metabolome adjustments in response to oxygen deprivation and other abiotic stresses.[30-33, 26, 25, 34] It is believed that these amino acids can be converted back to either glycolytic or TCA cycle intermediates for energy production upon recovery from oxygen deprivation.[2] Consistent with this hypothesis, levels of the 11 identified amino acids that accumulated during submergence dramatically decreased 1 d after recovery in both varieties (Figure 2.5). The rapid decline in the amino acids upon de-submergence provides support of the hypothesis that production of these metabolites, as an alternative to ethanol, allows carbon to be recycled upon reoxygenation [6] Although de-submergence led to a decline in the abundance of the amino acids, the significantly higher levels of these amino acids in M202 after 1d recovery indicates that there may be additional stress associated with de-submergence that is alleviated in the M202(*Sub1*) variety (Table 2.2).

The effects of the *SUB*1A gene on nitrogen metabolism can also be inferred from changes in Ala, Glu, Gln, Asn, and Asp. Although implicated as a carbon sink, Ala can also be used to store nitrogen as well. It has been reported that assimilated $^{15}NH_4$ is preferably accumulated in Ala in the green alga *Selenastrum minutum* under oxygen deficiency.[35] Ricoult and co-workers also showed the accumulation of nitrogen-labeled Ala instead of Asn occurred during anoxia in the model legume *Medicago truncatula.*[23] Ammonium is assimilated into Gln and sequentially converted into Glu. In addition, Glu and Asn can be synthesized from Gln and Asp by Asn synthetase. Based on results of this

study, the *SUB1*A gene may also be involved in the regulation of nitrogen assimilation during submergence. Submergence increased the levels of Glu, and Gln in aerial tissue of the two genotypes, but these amino acids were significantly more abundant in M202, whereas Asn and Asp accumulated similarly in M202 and M202(*Sub1*) (Figure 2.5). In the future, metabolic flux studies might be used to determine if the limitation of Gln and Glu accumulation in M202(*Sub1*) provides another example of an energy-saving mechanism mediated by *SUB1A*, due to the requirement for ATP consumption in the production of Glu through ammonium assimilation. A pathway diagram summarizing the detected components and their relationship to central carbon metabolism are provided in Figure 2.6, showing the differences of each metabolite as a function of stress and genotype. The relative abundance of each metabolite is represented as bar graphs and statistical significance at the 95% confidence interval indicated by an asterisk. The effect of submergence on the detected metabolites and their respective biochemical pathways indicates that central carbon metabolism is severely affected in both genotypes but less profoundly in the M202(*Sub1*) genotype. Figure 2.6 also shows the relationship of Ala accumulation with pyruvate and Glu metabolism, indicating a possible role for Ala as both a carbon and nitrogen store. Furthermore, the accumulation of Ala by Glu catabolism can have an influence on the TCA cycle due to the production of 2-oxoglutarate, a TCA cycle intermediate.

SMM, a cationic metabolite, was found to accumulate equally in both genotypes during submergence. However other sulfur containing metabolites, such as methionine, SMM's metabolic precursor, were not detected in this study. These results

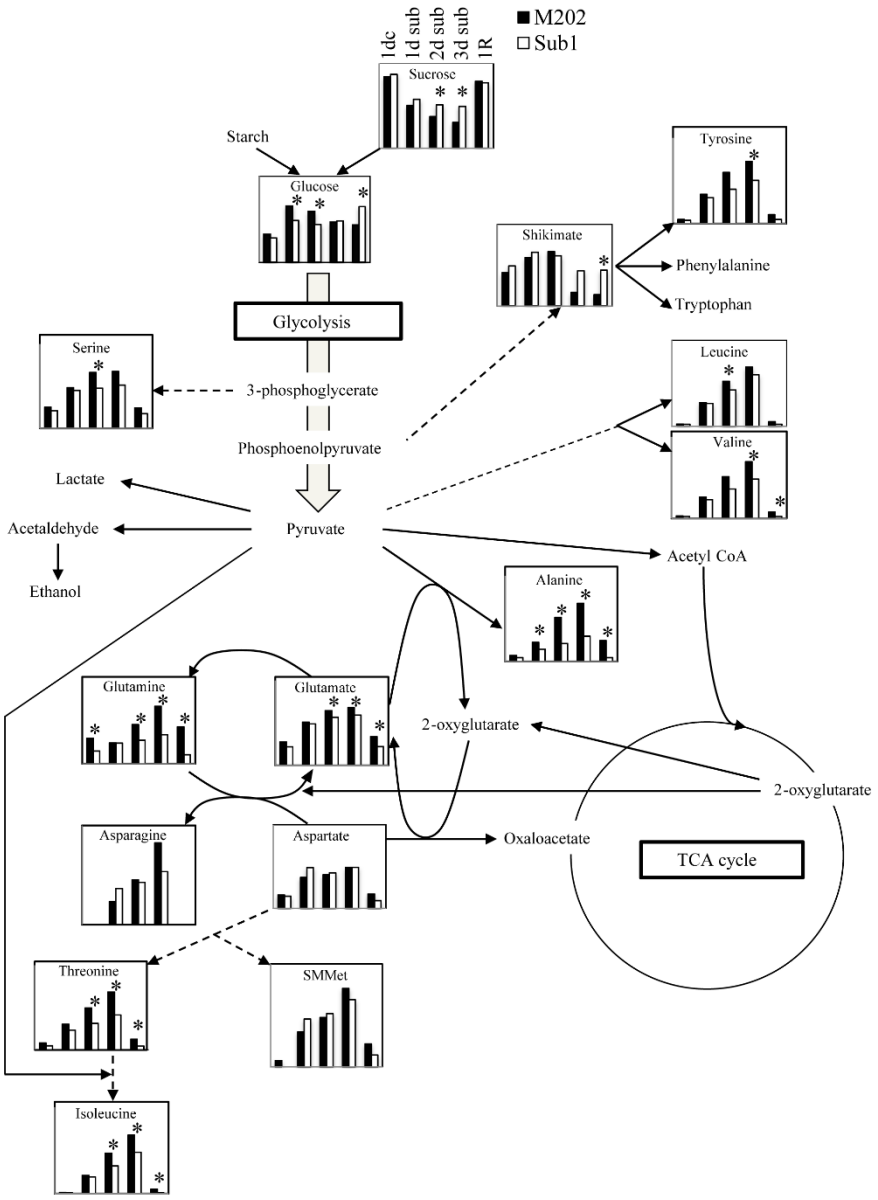


Figure 2.6. Pathway diagram with bar graphs representing relative metabolite abundance as a function of treatment. Each bar graph represents normalized average metabolite levels of at least 5 biological replicates at either 1dc, 1, 2, 3 d submergence (1d, 2d, 3d), and 3 d submergence + 1 d recovery (1R). Asterisks indicate a significant difference between the varieties at the 95% confidence interval.

are consistent with a report showing the accumulation of SMM over other sulfur containing metabolites, including glutathione, cysteine, and methionine in anoxic rice coleoptiles.[36] Menegus and co-workers also suggest that SMM functions as a storage form of methionine, allowing the regeneration of methionine and ultimately S-adenosylmethionine (AdoMet).[36] AdoMet, a versatile plant metabolite, is a methyl donor and plays a role in ethylene biosynthesis.[37] Other studies have elucidated a variety of possible functions for SMM in plants. Bourgis et al. (1999)[38] revealed that SMM was the primary phloem sulfur transporter in wheat, whereas Ko and co-workers demonstrated that SMM is a possible 1-aminocylcopropane-1-carboxylate (ACC) synthase substrate and inhibitor.[39] Since ACC is an ethylene precursor and *SUB1A* is an ethylene-induced gene, accumulation of SMM could act to limit the production of ethylene during the stress.[6] Although the precise functions of SMM are not known, the connection with other important metabolic pathways suggest it could play roles in the submergence response of rice.

2.3.5 Identification of Alanylglycine (AlaGly).

A well-resolved doublet at 1.55 ppm, as seen in Figure 2.1 a-d, was observed in all non-treated rice shoot tissue samples that could not be attributed to the expected organic or amino acids. Subsequent TOCSY (Figure 2.7), HMBC (Figure 2.8a), and HSQC (Figure 2.9a), experiments for a concentrated rice extract revealed carbon and proton chemical shifts for a molecule consistent with the dipeptide AlaGly. The TOCSY experiment (Figure 2.7) showed the methyl resonance correlating with a methine resonance.

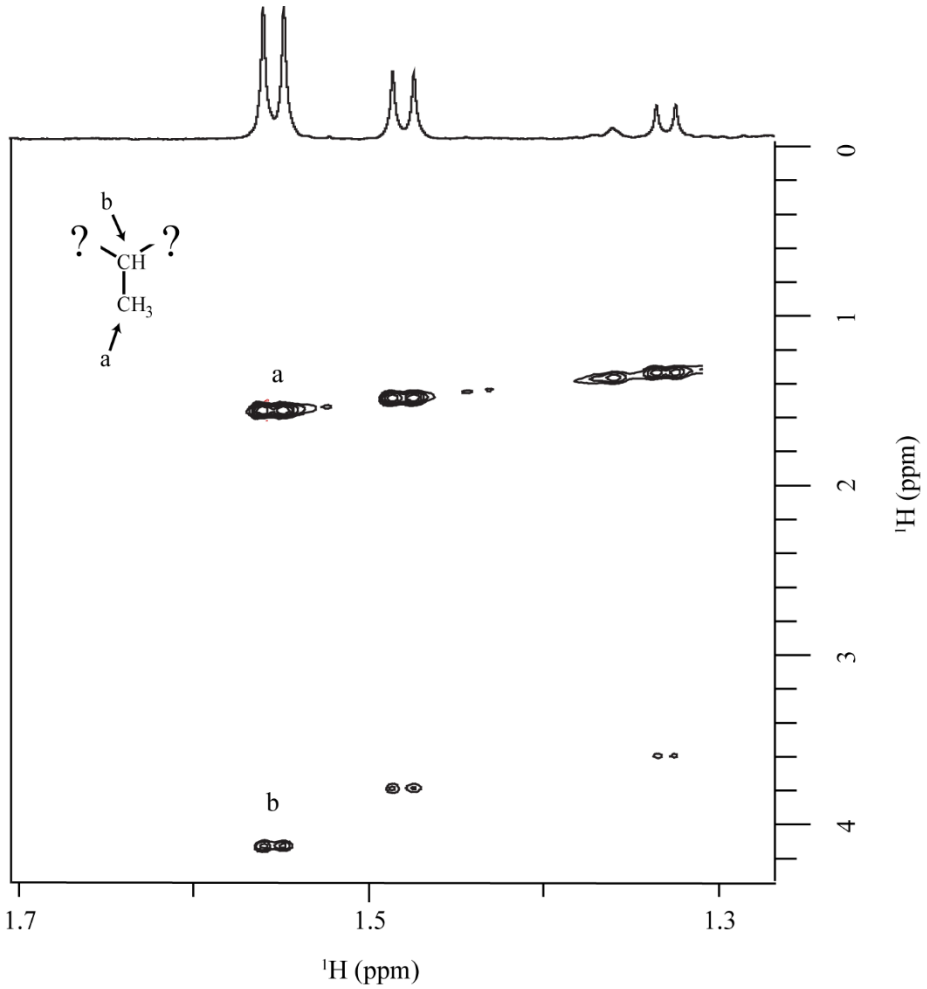


Figure 2.7. TOCSY spectrum of extracted rice shoot tissue containing AlaGly. The correlation of the methyl resonance (1.55 ppm) with the methine resonance (4.15 ppm) can be clearly seen and are labeled as "a" and "b", respectively, to correlate with a proposed structure of the compound (inset). The question marks represent unknown parts of the compound.

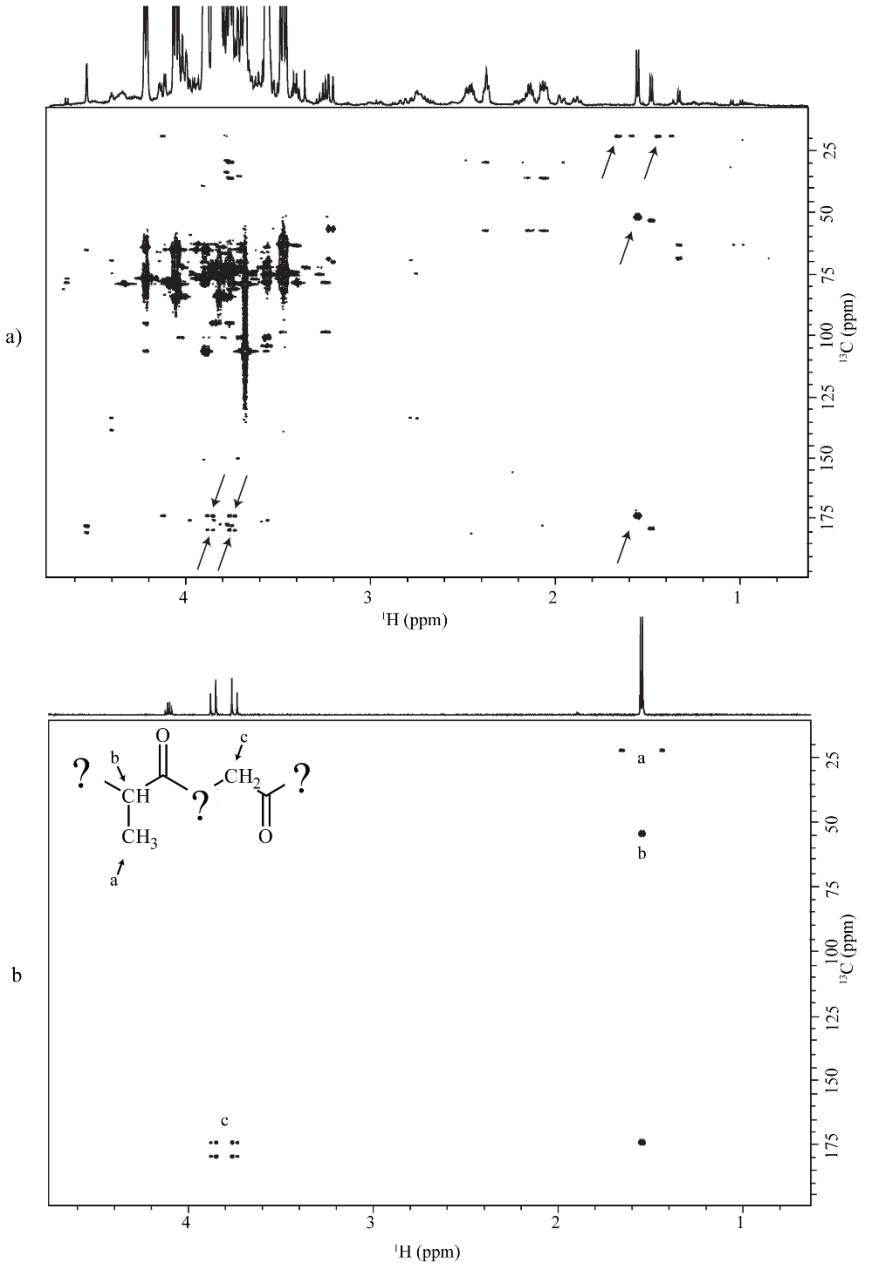


Figure 2.8. ^{1}H-^{13}C HMBC of (a) a rice sample extract containing alanylglycine and (b) racemic alanylglycine standard at 5 mM. The arrows in (a) indicate the alanylglycine resonances. The labels in (b) are included to correlate the identified resonances with their respective positions in AlaGly (inset). Question marks represent parts of the compound that could not be elucidated with the HSQC experiment.

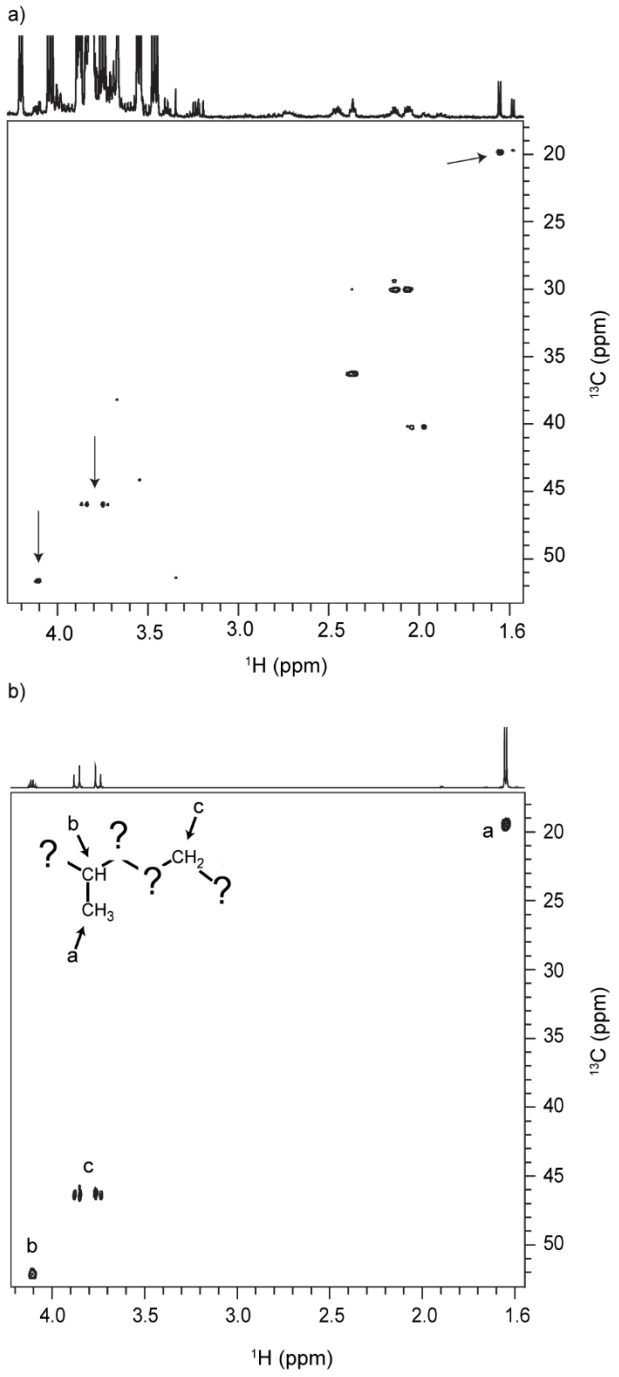


Figure 2.9. ^{1}H-^{13}C HSQC of (a) a rice sample extract containing alanylglycine and (b) racemic alanylglycine standard at 5 mM. The arrows in (a) indicate the alanylglycine resonances. The labels in (b) are included to correlate the identified resonances with their respective positions in AlaGly (inset). Question marks represent parts of the compound that could not be elucidated with the HSQC experiment.

The TOCSY experiment was followed by an HMBC experiment (Figure 2.8a), which showed the through-space correlation of the methyl and methine resonances with a carbonyl carbon, which also correlated with a resonance consistent with a methylene group. An HSQC experiment (Figure 2.9a) confirmed the presence of a methylene group. Predicted ^{1}H and ^{13}C NMR spectra for AlaGly closely matched the chemical shifts observed in the rice extract. The assignment of the doublet at 1.55 ppm to AlaGly was confirmed by spiking a tissue extract with a racemic AlaGly standard and by comparison of the HMBC (Figure 2.8b) and ^{1}H-^{13}C HSQC (Figure 2.9b) spectra of the unspiked extract and the standard.

Analysis of AlaGly normalized integrals during the short-term submergence experiment did not reveal a significant difference between the rice varieties as a consequence of submergence or de-submergence (Figure 2.5, Tables 2.1 and 2.2). AlaGly levels decreased during submergence and showed no sign of return to pre-stress levels after 24 h of recovery (Figure 2.5). The lack of AlaGly recovery upon desubmergence was notably distinct from the behavior of the amino acids and carbohydrates mentioned above.

D-Alanylglycine (D-AlaGly) was previously detected in rice leaf blades.[40] Additional experiments, using axenic treatment to ensure all organisms that could contribute D-amino acids or peptides were absent from the growth medium, revealed that D-AlaGly was indeed produced by the plant.[41] It was subsequently confirmed by Manabe and co-workers that D-AlaGly accumulation was limited to shoot tissue and required growth in light.[41]

To our knowledge, this is the first report of the detection and identification of AlaGly in a plant extract by NMR spectroscopy. Unfortunately, NMR is unable to distinguish between the D- and L-Ala enantiomers in AlaGly and further investigation is required to confirm the chirality of Ala in this dipeptide. Our experiments showed that under normal growth conditions, AlaGly was present in greater amounts than Ala in both varieties studied (Figure 2.1, Figure 2.5). During submergence stress, AlaGly levels decreased significantly and were not restored during the 1 d recovery period. No difference in relative AlaGly levels was observed between the two genotypes at any time point assayed. Because the submergence treatment was carried out under low-light conditions to simulate agricultural flooding, the reduction in AlaGly levels is consistent with the findings reported by Manabe.[42] During the 1 d recovery period, however, AlaGly levels were not observed to return to pre-submergence levels although light levels were higher and photosynthetic activity presumably increased. If AlaGly accumulation is directly related to photosynthesis, a difference between the two varieties after 24 h of recovery might be expected because of the better maintenance of chlorophyll content[6] and increase in glucose accumulation in the M202(*Sub1*) variety (Figure 2.5), which is likely to correspond to a greater capacity for photosynthesis upon de-submergence.

2.4 Conclusions

This study demonstrates the differential regulation of several metabolic pathways by the genetic determinant of submergence tolerance, *SUB1*A. Although previous studies have indicated differences in regulation of bulk carbohydrates and transcripts associated with metabolism in the M202(*Sub1*) and M202 varieties,[6, 4, 5] untargeted-metabolic

analysis has not yet been reported for these varieties. Our findings support the hypothesis that *SUB1*A invokes a quiescence strategy during submergence that limits the consumption of energy reserves to fuel shoot elongation growth. Consistently, accumulation of amino acids synthesized from glycolytic intermediates and pyruvate occurred to a lesser extent in M202(*Sub1*) under submerged conditions. Furthermore, metabolic regulation by *SUB1A* was evident upon de-submergence, indicating that the stress induced upon re-oxygenation is also affected by the presence or absence of the *SUB1A* gene as indicated by Fukao et al. (2011).[43] We also identified fluctuation in SMM accumulation during submergence stress and recovery. Although SMM accumulated equally in both varieties of rice, SMM can be connected to ethylene production, which is directly involved *in SUB1A* gene regulation in submergence tolerant rice and promotion of elongation growth in submergence intolerant rice.[44, 6] The utility of NMR for metabolic profiling was also demonstrated by the identification and characterization of AlaGly. NMR provides a non-destructive method for characterization of unknown and the identification of the unexpected metabolite AlaGly could have been much more difficult using alternative methods of analysis.

In chapter 3, the metabolome of the same vegetative rice tissue samples examined in this chapter are explored by GC-MS to achieve a deeper understanding of central carbon metabolism and energy production during and after submergence stress. Additionally, the differences and similarities between the results obtained using NMR and GC-TOF-MS are also examined demonstrating the benefits of applying multiple analytical techniques in metabolomics experiments.

2.5 References

(1) Bailey-Serres, J.; Fukao, T.; Gibbs, D. J.; Holdsworth, M. J.; Lee, S. C.; Licausi, F.; Perata, P.; Voesenek, L. A. C. J.; van Dongen, J. T. Making sense of low oxygen sensing. *Trends Plant Sci.* **2012,** *17*, (3), 129-138.

(2) Bailey-Serres, J.; Voesenek, L. Flooding stress: Acclimations and genetic diversity. *Annu. Rev. Plant Biol.* **2008,** *59*, 313-339.

(3) Voesenek, L.; Bailey-Serres, J. Plant Biology: Genetics of high-rise rice. *Nature* **2009,** *460*, (7258), 959-960.

(4) Jung, K. H.; Seo, Y. S.; Walia, H.; Cao, P. J.; Fukao, T.; Canlas, P. E.; Amonpant, F.; Bailey-Serres, J.; Ronald, P. C. The submergence tolerance regulator *SUB1*A mediates stress-responsive expression of AP2/ERF transcription factors. *Plant Physiol.* **2010,** *152*, (3), 1674-1692.

(5) 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.

(6) Fukao, T.; Xu, K. N.; Ronald, P. C.; Bailey-Serres, J. A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* **2006,** *18*, (8), 2021-2034.

(7) Kaiser, K. A.; Barding, G. A.; Larive, C. K. A comparison of metabolite extraction strategies for H-1-NMR-based metabolic profiling using mature leaf tissue from the model plant *Arabidopsis thaliana*. *Magn. Reson. Chem.* **2009,** *47*, S147-S156.

(8) Glasoe, P. K.; Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **1960,** *64*, (1), 188-190.

(9) Ogg, R. J.; Kingsley, P. B.; Taylor, J. S. WET, a T-1-insensitive and B-1-insensitive water-suppression method for *in-vivo* localized H-1-NMR spectroscopy. *J. Magn. Reson., Ser B* **1994,** *104*, (1), 1-10.

(10) Bax, A.; Davis, D. G. MLEV-17-based two-dimenional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **1985,** *65*, (2), 355-360.

(11) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. Rapid recording of 2D NMR-spectra without phase cycling - Application to the study of hydrogen-exchange in proteins. *J. Magn. Reson.* **1989,** *85*, (2), 393-399.

(12) Kay, L. E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992,** *114*, (26), 10663-10665.

(13) Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J. Magn. Reson.* **1991,** *93*, (1), 151-170.

(14) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sorensen, O. W.; Griesinger, C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed-field gradients. *J. Biomol. NMR* **1994,** *4*, (2), 301-306.

(15) Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **1993,** *31*, (3), 287-292.

(16) Bax, A.; Summers, M. F. ^{1}H and ^{13}C assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *J. Am. Chem. Soc.* **1986,** *108*, (8), 2093-2094.

(17) Summers, M. F.; Marzilli, L. G.; Bax, A. Complete ^{1}H and ^{13}C assignments of coenzyme-B12 through the use of new two-dimensional NMR experiments. *J. Am. Chem. Soc.* **1986,** *108*, (15), 4285-4294.

(18) 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.

(19) Pedersen, O.; Rich, S. M.; Colmer, T. D. Surviving floods: leaf gas films improve O(2) and CO(2) exchange, root aeration, and growth of completely submerged rice. *Plant J.* **2009,** *58*, (1), 147-156.

(20) 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.

(21) Miyashita, Y.; Dolferus, R.; Ismond, K. P.; Good, A. G. Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in Arabidopsis thaliana. *Plant J.* **2007,** *49*, (6), 1108-1121.

(22) Narsai, R.; Howell, K. A.; Carroll, A.; Ivanova, A.; Millar, A. H.; Whelan, J. Defining core metabolic and transcriptomic responses to oxygen availability in rice embryos and young seedlings. *Plant Physiol.* **2009,** *151*, (1), 306-322.

(23) Ricoult, C.; Cliquet, J. B.; Limami, A. M. Stimulation of alanine amino transferase (AlaAT) gene expression and alanine accumulation in embryo axis of the model legume *Medicago truncatula* contribute to anoxia stress tolerance. *Physiol. Plant.* **2005,** *123*, (1), 30-39.

(24) 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.

(25) van Dongen, J. T.; Froehlich, 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.* **2009,** *103*, (2), 269-280.

(26) Shingaki-Wells, R. N.; Huang, S. B.; Taylor, N. L.; Carroll, A. J.; Zhou, W. X.; Millar, A. H. Differential molecular responses of rice and wheat coleoptiles to anoxia reveal novel metabolic adaptations in amino acid metabolism for tissue tolerance. *Plant Physiol.* **2011,** *156*, (4), 1706-1724.

(27) Good, A. G.; Muench, D. G. Purification and characterization of an anaerobically induced alanine aminotransferase from barley roots. *Plant Physiol.* **1992,** *99*, (4), 1520-1525.

(28) Greenway, H.; Gibbs, J. Mechanisms of anoxia tolerance in plants. II. Energy requirements for maintenance and energy distribution to essential processes. *Funct. Plant Biol.* **2003,** *30*, (10), 999-1036.

(29) Menegus, F.; Cattaruzza, L.; Chersi, A.; Fronza, G. Differences in the anaerobic lactate-succinate production and in the changes of cell sap pH for plants with high and low resistance to anoxia. *Plant Physiol.* **1989,** *90*, (1), 29-32.

(30) Charlton, A. J.; Donarski, J. A.; Harrison, M.; Jones, S. A.; Godward, J.; Oehlschlager, S.; Arques, J. L.; Ambrose, M.; Chinoy, C.; Mullineaux, P. M.; Domoney, C. Responses of the pea (Pisum sativum L.) leaf metabolome to drought stress assessed by nuclear magnetic resonance spectroscopy. *Metabolomics* **2008,** *4*, (4), 312-327.

(31) Gavaghan, C. L.; Li, J. V.; Hadfield, S. T.; Hole, S.; Nicholson, J. K.; Wilson, I. D.; Howe, P. W. A.; Stanley, P. D.; Holmes, E. Application of NMR-based metabolomics to the investigation of salt stress in maize (*Zea mays*). *Phytochem. Anal.* **2011,** *22*, (3), 214-224.

(32) Narsai, R.; Rocha, M.; Geigenberger, P.; Whelan, J.; van Dongen, J. T. Comparative analysis between plant species of transcriptional and metabolic responses to hypoxia. *New Phytol.* **2011,** *190*, (2), 472-487.

(33) Whittaker, A.; Martinelli, T.; Farrant, J. M.; Bochicchio, A.; Vazzana, C. Sucrose phosphate synthase activity and the co-ordination of carbon partitioning during sucrose and amino acid accumulation in desiccation-tolerant leaf material of the C-4 resurrection plant Sporobolus stapfianus during dehydration. *J. Exp. Bot.* **2007,** *58*, (13), 3775-3787.

(34) 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.

(35) Vanlerberghe, G. C.; Joy, K. W.; Turpin, D. H. Anaerobic metabolism in the N-limited green-alga *Selenastrum minutum*. *Plant Physiol.* **1991,** *95*, (2), 655-658.

(36) Menegus, F.; Lilliu, I.; Brambilla, I.; Bonfa, M.; Scaglioni, L. Unusual accumulation of S-methylmethionine in aerobic-etiolated and in anoxic rice seedlings: An H-1-NMR study. *J. Plant Physiol.* **2004,** *161*, (6), 725-732.

(37) Ravanel, S.; Gakiere, B.; Job, D.; Douce, R. The specific features of methionine biosynthesis and metabolism in plants. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95*, (13), 7805-7812.

(38) Bourgis, F.; Roje, S.; Nuccio, M. L.; Fisher, D. B.; Tarczynski, M. C.; Li, C. J.; Herschbach, C.; Rennenberg, H.; Pimenta, M. J.; Shen, T. L.; Gage, D. A.; Hanson, A. D. S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell* **1999,** *11*, (8), 1485-1497.

(39) Ko, S.; Eliot, A. C.; Kirsch, J. F. S-Methylmethionine is both a substrate and an inactivator of 1-aminocyclopropane-1-carboxylate synthase. *Arch. Biochem. Biophys.* **2004,** *421*, (1), 85-90.

(40) Yamauchi, M.; Ohashi, T.; Ohira, K. Occurence of D-alanylglycine in rice leaf blades. *Plant Cell Physiol.* **1979,** *20*, (3), 671-673.

(41) Manabe, H.; Yamauchi, M.; Ohira, K. Studies on D-amino acids in rice plants - Behaviors of D-alanylglycine in rice seedlings. *Plant Cell Physiol.* **1981,** *22*, (2), 333-336.

(42) Manabe, H.; Ohira, K. Effect of light irradiation on the D-alanylglycine content in rice leaf blades. *Plant Cell Physiol.* **1983,** *24*, (6), 1137-1142.

(43) Fukao, T.; Yeung, E.; Bailey-Serres, J. The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell* **2011,** *23*, (1), 412-427.

(44) Fukao, T.; Bailey-Serres, J. Ethylene - A key regulator of submergence responses in rice. *Plant Sci.* **2008,** *175*, (1-2), 43-51.

CHAPTER THREE

Comparison of GC-MS and NMR for Metabolite Profiling of Rice Subjected to Submergence Stress

Based on a paper published in Journal of Proteome Research

J. Proteome Res. 2013, 12, 898-909

Acknowledgements: For their contributions to this research, I would like to thank the following people: Dr. Szabolcs Béni (Semmelweis Unviersity) for his help in sample preparation and analysis and Dr. Takeshi Fukao for his guidance and assistance with growing the plant material required for this research and for his critical input concerning the biological significance of the results.

Abstract:

In this chapter, a comparative analysis of GC-MS and ^{1}H NMR results was conducted for M202 and M202(*Sub1*) rice plants stressed by up to 3 days of submergence and allowed 1 d of post-submergence recovery. Most metabolomics studies are conducted using a single analytical platform. Each platform, however, has inherent advantages and disadvantages that can influence the analytical coverage of the metabolome. In this work, a more thorough analysis of the plant stress response was possible through the use of both ^{1}H NMR and GC-MS results. Several metabolites, such as S-methyl methionine and the dipeptide alanylglycine, were only detected and quantified by ^{1}H NMR. The high dynamic range of NMR, compared with that of the GC-TOF-MS instrument used in this study, provided broad coverage of the metabolome in a single experiment. The

sensitivity of GC-MS facilitated the quantitation of sugars, organic acids, and amino acids, some of which were not detected by NMR, and provided additional insights into the regulation of the TCA cycle. The combined metabolic information provided by ^{1}H NMR and GC-MS was essential for understanding the complex biochemical and molecular response of rice plants to submergence and recovery.

3.1 Introduction

In Chapter 2, we reported the results of a ^{1}H NMR metabolomics study of extracts of M202 and M202(*Sub1*) seedlings designed to probe metabolic reconfiguration as result of up to 12 d of complete submergence and during a 1 d post-submergence recovery period. The NMR spectra of the intolerant M202 plants reflected the rapid consumption of sucrose following submergence (Figure 2.2).[1] Compared with the controls, initial glucose levels for the intolerant variety rose dramatically and then gradually decreased over the 3 d submergence period (Figure 2.2, Table 2.1). A corresponding increase in other metabolites associated with pyruvate and glucose metabolism including alanine (Ala), threonine (Thr), valine (Val), leucine (Leu), and other amino acids was also observed (Figure 2.5). These NMR results corroborated the increased carbohydrate consumption associated with the escape response of submerged M202 plants. In contrast, in experiments with the submergence tolerant M202(*Sub1*) variety, the extent of carbohydrate consumption and resulting amino acid accumulation was less dramatic, reflective of the quiescence strategy associated with rice varieties containing the *SUB1*A gene.

S-methyl methionine (SMM) was identified in the ^{1}H NMR spectra of extracts of both rice varieties (Figure 2.5).[2, 1] SMM levels decreased significantly over the course of the experiment, although no significant difference was observed for the two genotypes investigated. SMM has been implicated as a possible regulator of ethylene biosynthesis as well as in sulfur transport and methionine storage.[3, 2, 4, 5] The ^{1}H NMR resonances of the dipeptide alanylglycine (AlaGly) were also detected in the spectra of extracts of both tissue types (Figure 2.7-2.8). Although differences in AlaGly content in the two varieties were not evident during submergence, the dipeptide was one of the more abundant metabolites detected. A GC-MS study of the metabolism of rice (*Oryza sativa* cv. Amaroo) embryos under aerobic and anaerobic germination conditions (dark, 30 °C) failed to report detection of either SMM or AlaGly.[6] The same report describes stress-induced changes of a variety of organic acids and tricarboxylic acid (TCA) cycle intermediates that were not detectable in our ^{1}H NMR experiments.[6] To rationalize the differences in metabolites reported in these two studies, this work compares the results of the independent yet complementary ^{1}H NMR and GC-MS techniques in profiling the response of the commercially grown submergence intolerant M202 and to that of the cross-bred submergence tolerant M202(*Sub1*) rice variety under normal growth conditions, submergence stress, and recovery following submergence. In addition, through a comparative analysis of both genotypes, a better understanding of the relative attributes of ^{1}H NMR and GC-MS for metabolomics studies can be discerned.

3.2 Materials and Methods

3.2.1 Materials and Reagents

N-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) in 1% trimethylchlorosilane for metabolite derivatization was purchased from Thermo Fisher Scientific (Waltham, MA). Fatty acid methyl esters (FAMEs) for use as retention index markers and methoxyamine hydrochloride (MeOX) were purchased from Supelco (Sigma-Aldrich Corp, St. Louis, MO). Metabolite standards were purchased from Sigma-Aldrich, Fisher Scientific (Pittsburgh, PA), and MP Biomedicals (Solon, OH). Water (18 MΩ/cm) was obtained using a Millipore filtration system (Millipore, Billerica, MA). Pyridine (+99% purity) was purchased from Acros Organics (Thermo Scientific, West Palm Beach, FL). Chloroform, obtained from Mallinckrodt Laboratory Chemicals (Phillipsburg, NJ), and methanol (Fisher Scientific) were of at least ACS grade.

3.2.2 Growth Conditions and Plant Materials

Oryza sativa ssp. *japonica* cv. M202 and cv. M202(*Sub1*) were grown and stressed for up to 3 d as described by Fukao, et al. and Barding, et al (Chapter 2.2.3).[7, 1] Surface sterilization of the seeds was accomplished with 0.2% Tween-20 and 1% sodium hypochlorite, after which the seeds were rinsed thoroughly with deionized (DI) water. Seeds were soaked in the dark overnight in DI water and germinated on moist paper in a Pyrex dish covered with plastic wrap for 5 d. Germinated seeds were transplanted into 10 cm x 10 cm pots with soil (25 plants per pot) and grown until the 3-leaf stage (12 days) at 30 °C in a greenhouse at ambient light and temperature. Transplanted seeds were

fertilized with Peters Excel 21-5-20 liquid fertilizer. The seedlings had a survival rate greater than 99%.

3.2.3 Submergence Stress and Plant Harvest

Submergence stress was carried out in six 121 L trash cans containing DI water to minimize water salinity and equilibrated overnight to the ambient greenhouse temperature. Once the seedlings matured to the 3-leaf stage, they were submerged for 0 (initial control), 1, 2, or 3 d or subjected to 3 d submergence and 1 day recovery. A 3d submergence stress was chosen as this duration allows complete recovery of both genotypes. For the recovery sample, plants were desubmerged and placed on the greenhouse bench for 24 hr prior to harvesting. Submerged plants were removed from the trash cans and aerial tissue immediately harvested, rinsed in DI water, flash frozen in liquid nitrogen and stored at -80 °C. Control tissue was treated similarly, including a DI water rinse prior to flash freezing. Harvesting of all plants occurred at 1 p.m. Following harvest, samples were ground by mortar and pestle under liquid nitrogen to a fine powder, lyophilized overnight until dry, and stored at -80 °C. There were four biological replicates for each treatment, with each biological replicate consisting of a homogeneous tissue pool from the plants in one pot (n=25). Aliquots of each tissue pool were also analyzed using ^{1}H NMR as described in Chapter 2.[1]

3.2.4 Tissue Extraction

Metabolite extraction was carried out using 10 mg dry weight of lyophilized tissue with modification of the previously described method (Chapter 2.2.4.2).[1] Briefly, the dried tissue was extracted with 1.0 mL 80/20 MeOH/H_2O extraction solvent in a 1.5

mL Eppendorf microcentrifuge tube. The samples were agitated for 1 min at 300 rpm using a platform shaker and centrifuged at 12000 x g for 4 min. A 50 µL aliquot of supernatant was transferred to a 350 µL glass GC flat bottom insert (Phenomenex, Torrence, CA) in a 1.5 mL Eppendorf tube and centrifuged under vacuum overnight using a Thermo-Savant SC110 model speed vacuum equipped with an RVT400 refrigerated vapor trap attached to a GP110 gel pump. Dried extracts were stored at -20 °C until derivatization for GC-MS analysis.

3.2.5 Sample Preparation

Dried metabolite extracts were treated according to Lee and Fiehn.[8] To each glass insert, 10 µL of 40 mg/mL MeOX in pyridine was added and the sample was shaken at 30 °C for 90 min. This was followed by addition of 4 µL from a standard mixture of FAMEs of a linear carbon chain length C8, C9, C10, C12, C14, C16, C18, C20, C22, C24, C26, C28, and C30. The standard mixture of FAMEs was prepared in chloroform with a C8-C16 concentration of 0.8 mg/mL and a C18-C30 concentration of 0.4 mg/mL. Samples were allowed to react with 90 µL of MSTFA at 37 °C for 30 min, after which the glass inserts were transferred directly to a wide-mouth crimp top vial (Phenomenex) and sealed with an 11 mm crimp cap. Samples were analyzed within 24 h of derivatization.

3.2.6 Sample Injection and Gas Chromatography

An Agilent 7683B Automatic Liquid Sampler equipped with a 10 µL Agilent syringe (Agilent Technologies, Santa Clara, CA) was used to inject 1 µL of derivatized sample, after which the syringe was cleaned 3 times in methylene chloride in preparation

for the next injection. The sample was injected into a splitless, single tapered MS-certified injector liner with glass wool (Agilent Technologies) heated at 230 °C and changed every 25 samples to ensure efficient analyte transfer to the GC column. The injector was operated in pulsed splitless mode with helium (99.998% purity, CalTool, Riverside, CA) pressure ramped to 35.0 PSI for 0.5 min, after which the flow was reduced and maintained at 1 mL/min for the duration of the separation. The separation was carried out using an Agilent 7890A gas chromatograph equipped with an Rtx-5sil MS column, 0.25 mm i.d. and 30 m length with an additional 10 m integrated guard column. The sample was introduced at an initial oven temperature of 60 °C held for 1 min, ramped at 10 °C/min to a final temperature of 320 °C, and held for 5 min. The total run time was 32 min. All instrument operations were controlled by Waters MassLynx software version 4.1 (Waters Corporation, Milford, MA).

3.2.7 Mass Spectrometry Data Acquisition and Analysis

The sample was introduced from the GC to the MS at a transfer line temperature of 320 °C. Electron impact ionization was used at 70 eV with a source temperature of 220 °C. The filament was turned on after 6.5 min and mass spectra were recorded from *m/z* 50 to 600 at a rate of 10 spectra s^{-1}. The detector was operated at 2700 V. An additional solvent delay was incorporated starting at 24.7 min and ending at 25.0 min to prevent ionization of sucrose, which is present at high abundance and would saturate the detector. Data were collected without the use of the dynamic range enhancement feature of the Waters GCT Premier to avoid compromising deconvolution and integration of the highly complex chromatograms. The data were collected and stored in Waters file format

(*.raw). Data were converted as collected from Waters file format to *.cdf format for processing with the Automated Mass Spectral Deconvolution and Identification System (AMDIS, NIST, Gaithersburg, MD). Deconvolution settings used a component width of 17 scans, high resolution, high sensitivity, and medium shape. Retention indices (RI's) were calculated for each sample by AMDIS using an internal standard library and calibration standard library. For compound identification, AMDIS queried the NIST 08 Mass Spectral Library. All compound identities were confirmed with RI and MS data obtained from the Golm Metabolome Database or checked against an in-house library generated using metabolite standards.[9]

MarkerLynx XS (Waters Corporation) was used for data preprocessing to collect integration values for identified metabolites. Peaks were detected without smoothing from an initial retention time of 7.00 min and final retention time of 32.00 min, with a low-mass cutoff of 73.5 Da, a high-mass cutoff of 600 Da, and a mass accuracy of 0.10 Da. The peak width at 5% was overestimated as 2 sec to ensure peak detection. A peak-to-peak baseline noise value of 1.0, a marker intensity threshold of 25 counts, and a mass and retention time window of 0.1 Da/min were also used. A noise elimination level of 3.0 was chosen and data was deisotoped. The results were exported to Excel (Microsoft, Redmond, WA) where the retention times and the extracted masses were matched with identified metabolites. One mass-retention time pair with the corresponding area for each metabolite was taken for data normalization and statistical analysis. The mass-retention time pair with the highest relative abundance was chosen to represent each metabolite (excluding the silylation-related fragments 73 and 147 *m/z*).

3.2.8 Statistical Analyses

The data were normalized by dividing the area of the individual components by the summed area for all the identified metabolites as described by Lee et al.[8] P-values were calculated in Excel using a 2-tailed Student's t-test comparing inter- and intra-genotype differences as a result of submergence stress or recovery. The GC-MS results for one member of the M202(*Sub1*) 1 day recovery (1R) sample set were excluded from the statistical analysis due to poor sensitivity and presence of peaks inconsistent with the rest of the data set. Trajectory plots were prepared with Origin 7.5 (OriginLab, Northampton, MA) using the average and standard deviations determined for each treatment. The integrated, sum-normalized area of mass-retention time pairs from the identified metabolites was used for principal components analysis (PCA) carried out with Minitab 15 (Minitab, Inc., State Park, PA) after mean-centering in Excel.

3.3 Results and Discussion

This study examines the differences in the GC-MS metabolic profiles measured for extracts of M202 and M202(*Sub1*) seedlings subjected to control, submergence stress and recovery conditions. These GC-MS results are then compared to the metabolic response observed for these genotypes with ^{1}H NMR, as described in Chapter 2.[1] In both sets of experiments aliquots of the same tissue samples were extracted and analyzed.

Figure 3.1 shows a region of the representative total ion chromatograms (TICs) measured for M202 rice extracts comparing the results for control, 3 d submergence, and 1 d post-submergence recovery samples. Several organic and amino acids are detected in this region of the TIC including aspartate (Asp), γ-aminobutyric acid (GABA), glycerate,

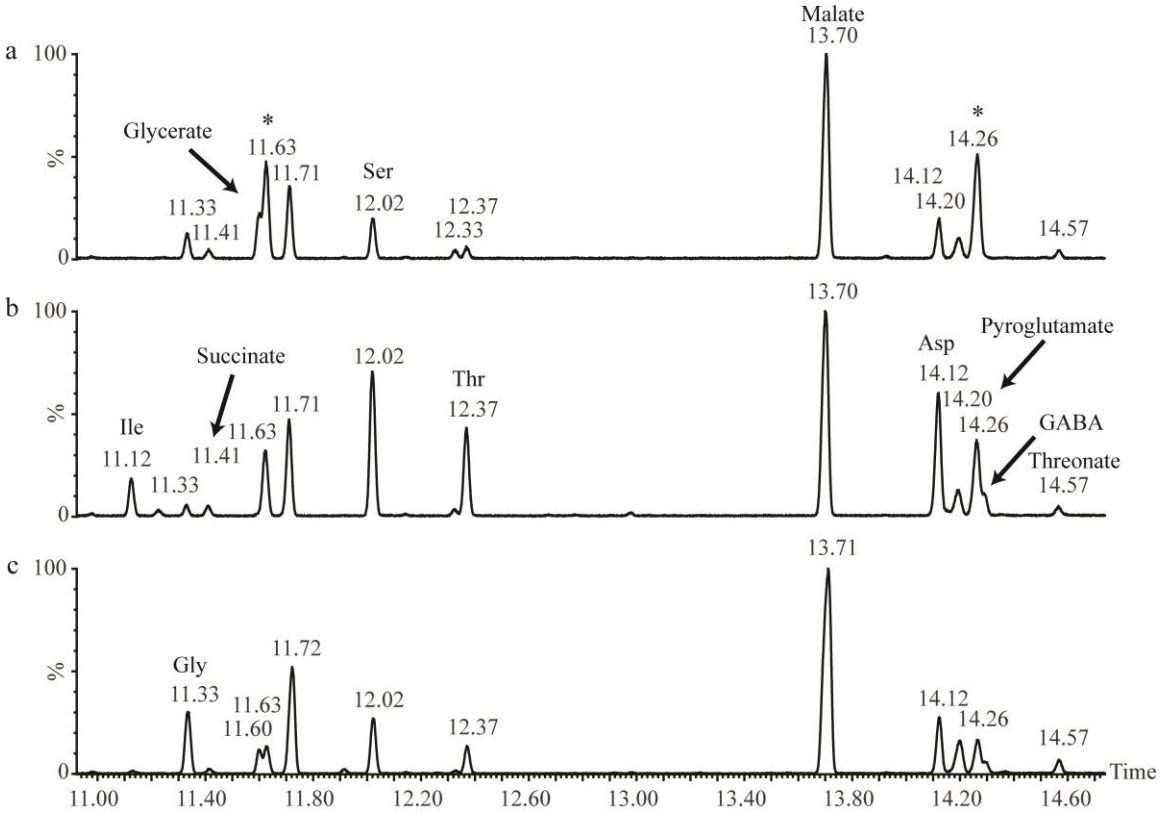


Figure 3.1 Representative GC-MS total ion chromatograms for the M202 variety a) control (day 0), b) 3 d submergence, and c) 3 d submergence + 1 d post-submergence recovery. Identified metabolites are labeled.

glycine (Gly), isoleucine (Ile), malate, pyroglutamate, serine (Ser), succinate, threonine (Thr), and threonate. Differences in relative metabolite abundance as a consequence of treatment are evident by visual inspection of Figure 3.1, with some peaks increasing in intensity, others decreasing, and some appearing to remain unchanged.

3.3.1 Metabolomics Analyses of the Stress Response

Global analysis of the metabolic changes due to submergence and recovery stresses was performed by PCA (Figure 3.2). PC1 and PC2 represent 63.1% and 17.9%, respectively, of the explained variance with distinct groupings observed for the two varieties during stress conditions. In the absence of stress the control treatments do not separate based on genotype in the PCA scores plot (Figure 3.2a). Within 1 d of submergence stress, both genotypes cluster to the right of the controls along PC1, whereas the tolerant and intolerant varieties differ along the second component. The distinction between M202 and M202(*Sub1*) in Figure 3.2a remains consistent over the course of the treatment, with the sample trajectory moving further along PC1 with duration of submergence and with the two varieties maintaining their separation in PC2. The data for the tissue samples after 1 d of recovery shift back towards the controls. The 1 d recovery samples of the submergence-tolerant M202(*Sub1*) variety cluster with the controls while the intolerant M202 samples are clustered below, well-separated from the controls in PC2.

Overall, the GC-MS PCA results in Figure 3.2 are consistent with the previously reported PCA analysis of the ^{1}H NMR data (Figure 2.4),[1] however there are fundamental differences between the two datasets that can be attributed to several factors. PCA of the

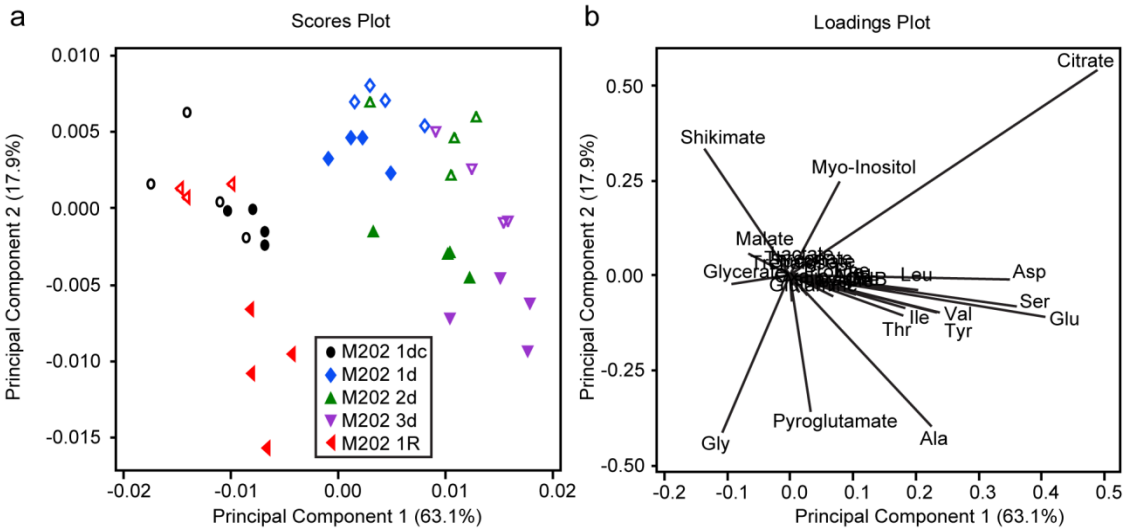


Figure 3.2 Scores plot (a) showing principal components 1 and 2. The legend identifies the individual treatments for the M202 variety. The same scheme but with open symbols was used to represent the M202(*Sub1*) samples. The loadings plot (b) shows the variables that contributed most to the variance along the first and second principal components.

NMR data used over 300 variables generated by integrating "bins" of the NMR spectrum in equidistant sections allowing use of the entire spectrum, with the exception of solvent peaks. As a result, more NMR variables were available to define the differences between the data sets. PCA of the GC-MS results was conducted only with the identified variables because unidentified peaks can result from derivatization side reactions. The variables representing the largest separation along PC1, PC2, or in both dimensions (e.g. citrate and shikimate) are clearly identified as important in the loadings plot (Figure 3.2b). Metabolites that contribute to separation primarily along PC1 are Asp, glutamate (Glu), Ile, Ser, Thr, and tyrosine (Tyr). Contributions to PC2 can be attributed to Ala, Gly, myo-inositol, and pyroglutamate. Despite the distinct grouping of the two varieties in Figure 3.2a, PCA is unable to define statistical differences or quantitative relationships between the varieties during stress. This level of information requires a direct comparison of the results obtained for each metabolite in samples measured at the different stress time points.

3.3.2 Metabolite Profiles Determined by GC-MS

The differences between metabolite levels in the two genotypes were examined to better understand the metabolic shifts resulting from submergence and recovery. Metabolite profiles for the control, submergence and recovery treatments were measured using GC-MS for 17 amino acids, 7 organic acids, myo-inositol, and trehalose and are summarized in the trajectory plots shown in Figure 3.3. These differences are also apparent in Table 3.1, which examines the fold changes of individual metabolites over time within and between genotypes.

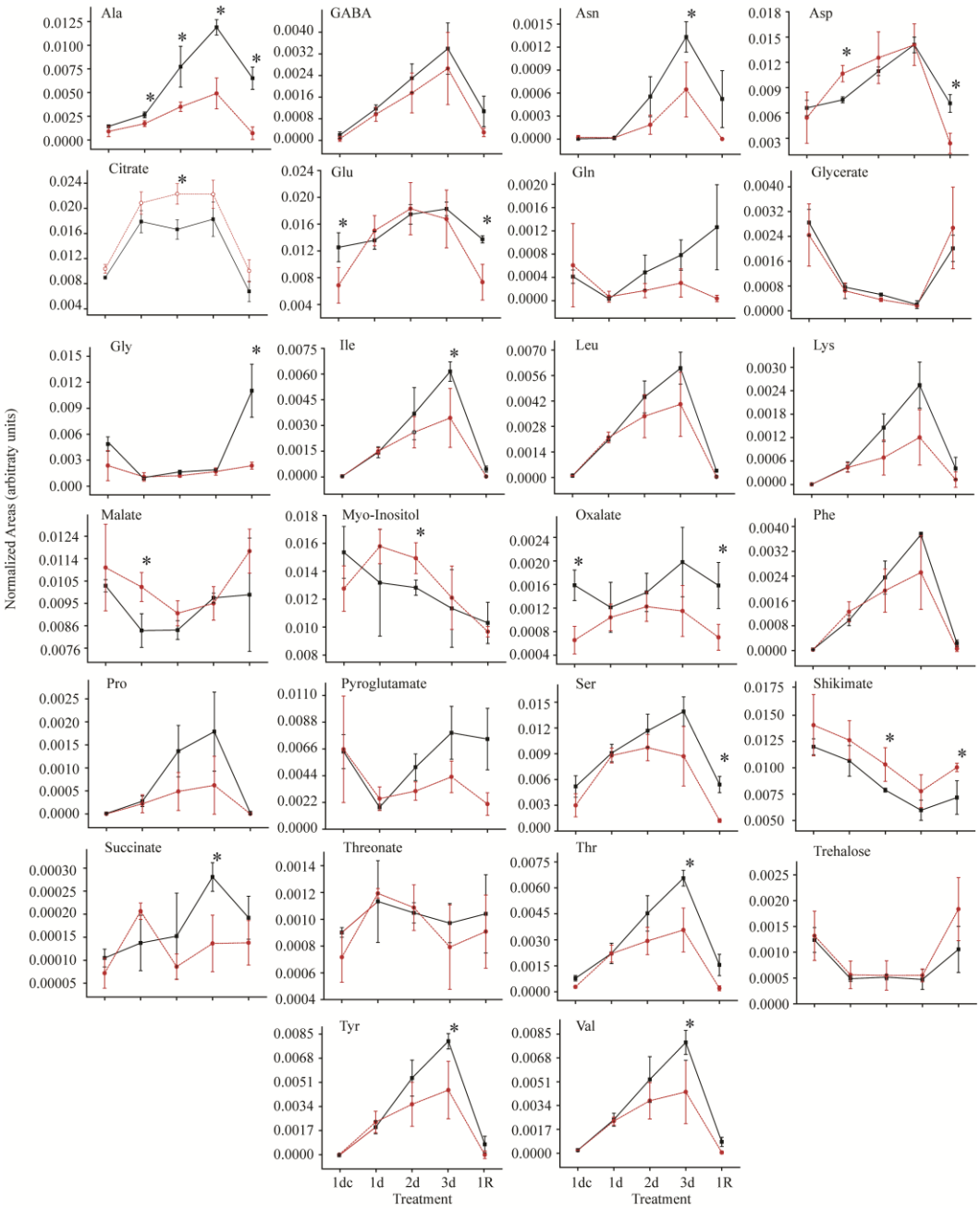


Figure 3.3. Trajectory plots from the GC-MS results representing the average normalized relative peak areas for M202 (■) and M202(*Sub1*) (●). Time points are connected using solid (M202) or dotted (M202(*Sub1*)) lines. Treatments are labeled as 1 d control (1dc), 1d, 2d, 3d submergence and 3d submergence + 1d post-submergence recovery (1R). Each data point represents the average of at least three biological replicates with error bars representing the standard deviation and asterisks indicating differences between then genotypes with a P-value < 0.05.

Table 3.1. A comparison of the treatment and control levels of metabolites as represented by fold changes for M202 and M202(*Sub1*) relative to the controls, and between the two varieties for each treatment condition. Asterisks represent significant differences at the 95% confidence limit. ND represents metabolite ratios that are not determined due to values below the limit of quantitation.

	M202				M202(*Sub1*)				M202/M202(*Sub1*)				
	1d:1dc	**2d:1dc**	**3d:1dc**	**1R:1dc**	**1d:1dc**	**2d:1dc**	**3d:1dc**	**1R:1dc**	**1dc**	**1d**	**2d**	**3d**	**1R**
Sugars													
Trehalose	0.4*	ND	ND	0.9	0.4	ND	ND	0.8	0.9	ND	ND	ND	0.6
Amino Acids													
Ala	2.1*	7.0*	8.7*	4.7*	1.9	3.3*	5.2	0.8	1.4	1.5*	3.0*	2.3*	8.3*
Asn	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.8	3.5	2.8*	ND
Asp	1.1	1.7*	2.1*	1.1	2.0	2.3*	2.6*	0.4	1.2	0.7*	0.9	1.0	3.0*
Gln	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Glu	1.0	1.4*	1.5*	1.1	2.2*	2.7*	2.6*	1.1	1.8*	0.9	1.0	1.0	1.9*
Gly	0.2*	0.4*	0.4*	2.5*	0.5	0.5	0.7	1.0	1.8	0.8	1.4	1.1	4.6*
Ile	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	1.7	1.8*	ND
Leu	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	1.3	1.5	ND
Lys	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	2.3	1.9	3.3
Phe	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.7	1.3	1.4	ND
Pro	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Ser	1.7*	2.4*	2.7*	1.0	2.9*	3.2*	3.1*	0.4	1.7	1.0	1.3	1.5	4.4*
Thr	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	1.7	1.8*	ND
Tyr	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.8	1.6	1.6*	ND
Val	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.0	1.6	1.8*	ND
Pyroglutamate	0.3*	0.8	1.2	1.2	0.4	0.5	0.7	0.3	1.0	0.7	1.6	1.8*	3.6*
Organic Acids													
Citrate	2.0*	1.9*	2.0*	0.8	2.0*	2.2*	2.2*	1.0	0.9	0.9	0.8*	0.8	0.7
Glycerate	0.3*	0.2*	0.1*	0.7*	0.3	0.1	0.1	1.1	1.2	1.2	1.5	1.3	0.8
GABA	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.1	1.3	1.1	ND
Malate	0.8*	0.8*	0.9	0.9	0.9	0.8	0.9	1.1	0.9	0.8*	1.0	1.0	0.8
Oxalate	0.6*	1.0	1.2	1.0	1.6	1.9*	1.4	1.1	2.4*	1.0	1.3	2.2	2.2*
Shikimate	0.9	0.7*	0.5*	0.6*	0.9	0.7	0.6	0.7	0.9	0.9	0.8*	0.7	0.7*
Succinate	1.2	1.9*	2.7*	1.8*	2.9*	1.2	1.6	1.9	1.5	0.6	2.3	2.4*	1.4
Threonate	1.4	1.1*	1.1	1.2	1.7*	1.5	1.3	1.3	1.3	1.1	0.9	1.0	1.1
Other Metabolites													
Myo-Inositol	0.8	0.9	0.7*	0.7*	1.2	1.2	1.0	0.8	1.2	0.8	0.9*	0.9	1.1

Similar to what was previously reported by ^{1}H NMR (Figure 2.5, Table 2.1),[1] the GC-MS analysis presented in Figure 3.3 shows that 15 of the 17 detected amino acids also accumulated during submergence and showed a return toward pre-treatment values after 24 hr of recovery. Significant differences (P value ≤ 0.05) between the genotypes were evident in the levels of 11 of the 17 detected amino acids either at some time point of the submergence treatment or after 24 hr recovery (Table 3.1). Amino acid metabolism reflects a stress response involving pathways related to elevated carbon catabolism through glycolysis and nitrogen utilization. Amino acids accumulated in both varieties during submergence stress however this trend was more prominent for the intolerant variety, supporting the previous observation that M202 has higher levels of starch consumption and sugar catabolism to generate ATP to fuel underwater elongation growth.

In contrast to the behavior of the majority of the amino acids, Gly levels in the intolerant variety increased 2.5 fold above non-stressed levels during recovery after re-oxygenation (Figure 3.3, Table 3.1). Gly is directly related to the metabolism of several amino acids, including Ser and Thr and the increase in Gly levels could result from the recycling of these amino acids.[10] However, the tolerant variety also accumulates Ser and Thr during submergence, although at lower amounts, and does not show a corresponding increase in Gly after 24 hr recovery. The stasis of Gly in samples of M202(*Sub1*) for all treatment conditions examined supports the assertion that increases in Gly levels in M202 may be a mechanistic response to re-oxygenation/dehydration stress. Because re-oxygenation and dehydration stresses are affected by the presence of *SUB1*A and

*SUB1*A-containing varieties are more tolerant to dehydration, the observation in Figure 3.3 that Gly does not increase during recovery in the tolerant cultivar could be associated with the better re-oxygenation tolerance of this genotype. This is consistent with reports of Gly accumulation in other plants as a result of increased photorespiration during dehydration.[11] There may be other reasons for this distinction in Gly accumulation during submergence recovery. A recent report suggests that Gly can act as a ligand for ligand-mediated gating of calcium in plants.[12] Calcium signaling is important for a variety of plant stress responses, including synergistic control of stomata opening with abscisic acid[13] and increased production of alcohol dehydrogenase, a key enzyme in anaerobic metabolism, during oxygen deprivation.[14, 15] Alternatively, Gly accumulation observed in the M202 samples is not a result of the catabolism of pathway-related metabolites but instead a response to re-oxygenation or dehydration stresses experienced by plants after prolonged submergence.[16]

Pyroglutamate is an uncommon amino acid related to glutathione metabolism and also commonly found at the *N*-termini of proteins.[17, 18] In both rice varieties, pyroglutamate levels decrease during the first day of submergence (Figure 3.3). Over the subsequent 2 days, levels of pyroglutamate in M202 return to pre-stress levels while in M202(*Sub1*) they remain lower than for the controls for the duration of the study, including the recovery period. This divergence in pyroglutamate levels for the two varieties could be indicative of differences in glutathione metabolism or protein turnover.

Energy production is a key component of plant survival, especially under submergence conditions where photosynthesis can be dramatically reduced as a result of

limited diffusion of carbon dioxide.[19] As a result of submergence stress, significant differences between the two genotypes were observed in the levels of the detected TCA cycle intermediates and components of related pathways (Figure 3.3). For the intolerant variety citrate, GABA, and succinate increased in abundance throughout the course of submergence and then decreased after re-oxygenation. Interestingly, citrate and succinate exhibited differing trends over the 3 d submergence treatment. After an initial accumulation within the first day of submergence, citrate plateaued and subsequently decreased upon re-oxygenation. Succinate gradually accumulated, with an apex at day 3 of submergence before decreasing during re-oxygenation. In contrast, malate decreased during the first day of submergence while the re-oxygenation time point is inconclusive due to high variance. The sudden accumulation and plateau of citrate without detecting a corresponding increase in isocitrate or α-ketoglutarate suggests that either the TCA cycle is regulated at citrate or that the flux through the subsequent intermediates is fast. The gradual increase of succinate, however, suggests a non-circular pathway for the TCA cycle through 2-oxoglutarate or oxaloacetate. Both are produced in the metabolism of pyruvate and Glu to Ala. GABA, which also gradually accumulates during stress is generated from 2-oxyoglutarate and has been reported to be produced during anaerobic metabolism.[20] GABA is also part of the photorespiration pathway, producing succinic semialdehyde, Gly, and Ala corroborating the accumulation of Gly in the intolerant variety.[21] The lack of change in Gly for the tolerant variety is still puzzling. The production of Gly requires glyoxylate which was not monitored in this study. Monitoring glyoxylate production in the two rice varieties as well as flux studies could provide a

better understanding of this difference in Gly production in the two rice varieties. Interestingly, succinate accumulation trends follow a similar trajectory as Ala and the other detected TCA cycle intermediates (Figure 3.3). In addition, the reduction of malate levels during submergence suggests a slowing of the TCA cycle, supporting the regulation of the TCA cycle at citrate and the accumulation of succinate through alternative pathways. Malate could also be converted to pyruvate possibly providing additional means of carbon storage during oxygen stress, however flux analysis is necessary to fully understand the role of the detected TCA cycle intermediates.[22] The results obtained for the TCA intermediates argue for future flux experiments. Since these experiments are complex and require expensive stable isotope labeled compounds, advance knowledge of the pathways of interest is critical to their success.

Although the trends for citrate and GABA in the tolerant variety are similar to those observed for M202, the disparity between the genotypes is evident in the time course plots for these TCA-related metabolites (Figure 3.3). Succinate increased drastically in M202(*Sub1*) samples after the first day of submergence, decreased back to initial levels after 2 d and then remained unchanged. The accumulation of succinate in M202 was 2.4 times that of M202(*Sub1*) by 3 d (Table 3.1), however, no statistically significant differences between the varieties were observed at the other points examined. Citrate and malate differed from other metabolites in that levels in M202(*Sub1*) were higher than those measured for the intolerant variety. The first day of submergence resulted in 0.8 times less malate for the intolerant variety compared to the tolerant variety. Similarly, citrate was 0.8 times less abundant in M202 compared to levels in the

M202(*Sub1*) samples (Table 3.1). Although these were the only time points where statistically significant differences (P value ≤ 0.05) between the two varieties were observed, this general trend is consistent throughout the course of the experiment. The higher levels of the TCA cycle intermediates detected in the tolerant variety could reflect the greater photosynthetic capacity of M202(*Sub1*) due to the presence of more chlorophyll during submergence as reported by Fukao et al.[16, 7]

Despite the sensitivity of GC-MS measurements we were unable quantify the sugars sucrose, glucose and fructose due to the method selected and the nature of the time-of-flight (TOF) instrument available for this work. Sucrose is the primary metabolite in rice extracts, dominating the other metabolites by a significant margin. NMR is able to accommodate a ~ 10^6 dynamic range compared to 10^4 for the Waters GCT Premier when using dynamic range enhancement feature, which was not employed in this study.[23, 24] To observe the less abundant compounds and take full advantage of the sensitivity of our GCT Premier TOF instrument, we did not attempt to quantify these soluble sugars (glucose, sucrose, and fructose) because of detector overload at the concentrations necessary to adequately quantify other metabolites. Also, because sucrose and glucose were easily quantified using ^{1}H NMR, there was not a strong motivation to attempt to measure these analytes by GC-MS Figure 2.2). However, glyceric acid, a sugar alcohol, and trehalose were detected by GC-MS and observed to decrease during submergence in both genotypes. During submergence, trehalose levels drop below the limit of quantification, preventing a metabolic comparison of the genotypes. The

metabolite profiles for glyceric acid in M202 and M202(*Sub1*) are indistinguishable suggesting that the levels of this compound are not impacted by the presence of *SUB1*A.

Narsai et al. explored the anaerobic germination of *Oryza sativa* cv. Amaroo (also a *japonica* sub-species) using a variety of "-omics" techniques, including metabolite profiling by GC-MS.[6] They reported changes in 166 metabolites during the course of their experiment, which is significantly more than the 30 metabolites reported in this study in both the NMR and GC-MS datasets (Table 3.2). There are several possibilities that might explain the different numbers of metabolites observed in these two studies. Plant growth stage at the time of harvest greatly influences the numbers of detected metabolites.[25] The anoxia study performed by Narsai et al. examined embryos grown in the dark in either anaerobic or aerobic conditions and separated from the endosperm within a few days of germination.[6] The study reported herein used seedlings sampled within 17 days after germination (12 days after planting) and grown under aerobic conditions until submergence stressed. Several reports have indicated that metabolic content varies during plant and fruit development, especially during the germination and development of young plants.[26, 25, 27, 28] The experiments of Narsai et al. were also performed using a different rice variety than used in this study.[6] Using GC-MS and ^{1}H NMR we have shown significant metabolic differences as a result of submergence stress for two varieties differing in the presence or absence of a single gene, *SUB1*A.[1] Differences between cultivars can be much more significant, resulting in a wider range of metabolites produced in response to anaerobic stress.

Table 3.2. Metabolites quantified by GC-MS and NMR. The "X" denotes detection of the metabolite by NMR, GC-MS, or both instruments.

Metabolites	NMR	GC-MS
Sugars		
Glucose	X	
Sucrose	X	
Trehalose		X
Amino Acids		
Ala	X	X
Asn	X	
Asp	X	X
Gln	X	X
Glu	X	X
Gly		X
Ile	X	X
Leu	X	X
Lys		X
Phe		X
Pro		X
Pyroglutamate		X
Ser	X	X
SMM	X	
Thr	X	X
Tyr	X	X
Val	X	X
Organic Acids		
Citrate		X
Glycerate		X
GABA		X
Malate		X
Oxalate		X
Shikimate	X	X
Succinate		X
Threonate		X
Other Metabolites		
AlaGly	X	
Myo-Inositol		X

Finally, the greater number of metabolites reported by Narsai et al. could be attributed in part to the quadrupole mass analyzer used in their study which has a wider linear quantitation range than our TOF instrument.[29] However, despite the greater number of metabolites detected in their study, neither SMM nor AlaGly, which we detected using ^{1}H NMR, were reported by Narsai and co-workers.

3.3.3 Comparison of the GC-MS and NMR Results

Because of the fundamental differences in the nature of GC-MS and NMR measurements, the results obtained using each platform contributed to our understanding of the complex metabolic response of rice to submergence stress. Table 3.2 lists the metabolites quantified in this study and the instrument used for their measurement. In total, four metabolites: sucrose, glucose, AlaGly, and SMM, were exclusively quantified by NMR.[1] Fourteen metabolites including several organic acids, TCA cycle intermediates, and amino acids were exclusively quantified by GC-MS, and 11 metabolites were determined by both instruments. Increased signal averaging could have been used to improve the sensitivity of NMR quantitation of the less abundant metabolites, however, this approach rapidly becomes impractical because improvement in S/N is proportional to the square root of the number of co-added transients.[30, 23] Furthermore, several compounds that were detected by GC-MS were not observed in the NMR spectrum due to problems of resonance overlap, for example the Gly resonance which is obscured by the more intense sucrose resonances.

The analytical and biological reproducibility of the GC-MS and ^{1}H NMR results can be evaluated by comparing the metabolite trajectories obtained for overlapping

metabolites quantified for M202 using both instruments. Overlaid in Figure 3.4 are trajectory plots from both NMR and GC-MS experiments for different extracts of the same samples of M202 tissue. Figure 3.5 shows a similar set of plots for M202(*Sub1*). Peak areas determined in both experiments are normalized, with the GC-MS data normalized to all identified metabolites (as described in section 3.2.8) and the NMR data sum normalized as described in Barding et al. (Chapter 2.2.6).[1] A strong correlation between the two data sets is observed in Figures 3.4 and 3.5, with the profiles for most metabolites displaying similar trends throughout the course of treatment. The profiles for Gln and shikimate show the greatest degree of difference between the datasets. From the GC-MS dataset, shikimate decreases during the course of submergence, whereas a decrease was not apparent in the NMR data until the 3 d time point. Although the trends are similar for both datasets, the relative abundance of shikimate in the NMR data set is low and the resonances for shikimate appear in a crowded region of the spectrum complicating the deconvolution process. In contrast, shikimate is easily detected and quantified by GC-MS making the measurements more reliable for qualitative and quantitative analysis. The opposite was true for Gln, where it was readily quantified by NMR but detected in relatively low abundance by GC-MS analysis.

Some differences between the datasets shown in Figures 3.4 and 3.5 can be attributed to the higher standard deviation of the GC-MS data set for those time points and to differences in data normalization. Although data from both techniques were sum normalized, GC-MS data was normalized to the sum total of all identified components whereas for the NMR results the integrated areas of both identified and unidentified

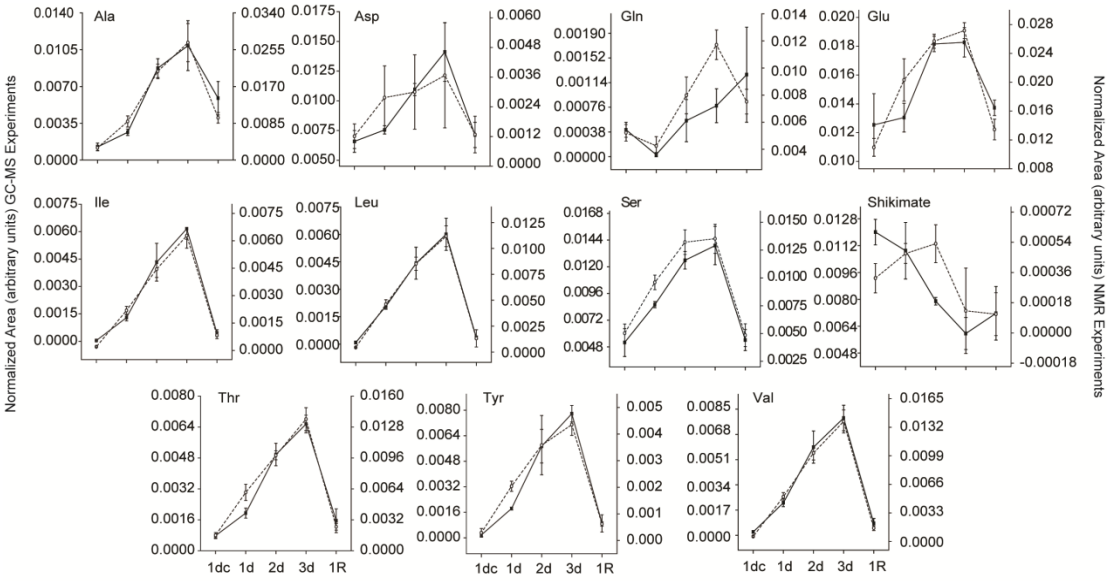


Figure 3.4. Trajectory plots comparing the normalized metabolic profiles measured using ^{1}H NMR (○) and GC-MS (■). Time points indicate controls (1dc), 1d, 2d, 3d submergence or 3d submergence followed by 1d post-submergence recovery (1R). Each data point represents the averaged normalized area of at least three biological replicates for the GC-MS data or five biological replicates for the NMR data with error bars indicating the standard deviations.

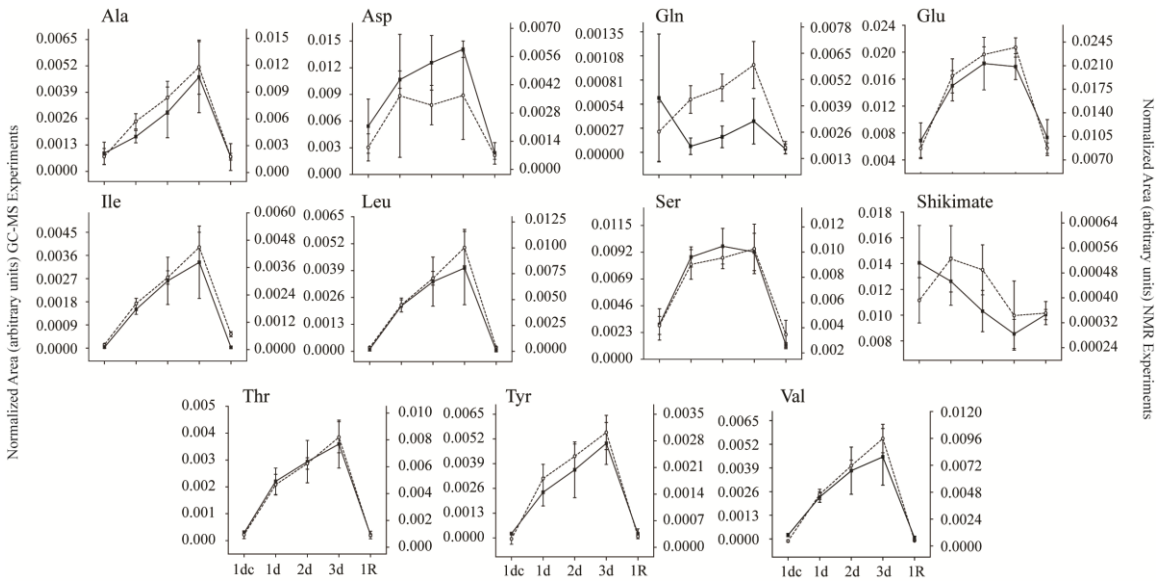


Figure 3.5. Trajectory plots comparing the normalized metabolic profiles measured using ^{1}H NMR (○) and GC-MS (■) for the M202(*Sub1*) tissue. Time points indicate controls (1dc), 1d, 2d, 3d submergence or 3d submergence followed by 1d post-submergence recovery (1R). Each data point represents the averaged normalized area of at least three biological replicates for the GC-MS data or five biological replicates for the NMR data with error bars indicating the standard deviation.

components were used. Despite the minor differences observed in Figure 3.4 and 3.5, these results show how datasets collected by two very different analytical methods can also provide confidence in the validity of data pre-processing and analysis.

For some analytes, such as Gln (Figure 3.6), differences in the NMR and GC-MS results can be attributed to differences in S/N. For both datasets, the Glu peak is present at higher intensity than Gln however, the Gln peak is of much lower relative intensity in the GC-MS data set and Gln was not detected in the 1 and 2 d samples. In contrast, the Gln resonances are readily detected in the NMR spectra for all time points evaluated. To better understand the differences between the datasets, S/N was measured for the Glu and Gln peaks for each data set (Table 3.3). At time points where Gln was detected by GC-MS, the S/N was ~5% that of Glu. When measured by NMR, the S/N of Gln was between 20 and 70% that of Glu and Gln was always present at quantifiable levels. Because Gln and Glu are important in plant nitrogen metabolism,[31] understanding the relationship between the two amino acids is important, giving ^{1}H NMR a relative advantage in their quantification.

3.3.4 Detection of AlaGly and SMM

Previously, we reported the identification of the dipeptide AlaGly in rice extracts by ^{1}H NMR analysis (Figure 2.5, 2.7-2.8).[1] Interestingly, AlaGly was not reported in recent GC-MS-based rice metabolite profiling studies, nor is it included in publicly available metabolomics libraries.[6, 32, 9] A cursory analysis of the GC-MS results of our rice extracts also failed to identify AlaGly, which contrasted with the pronounced intensity of AlaGly resonances in the ^{1}H NMR spectra measured for the same tissue

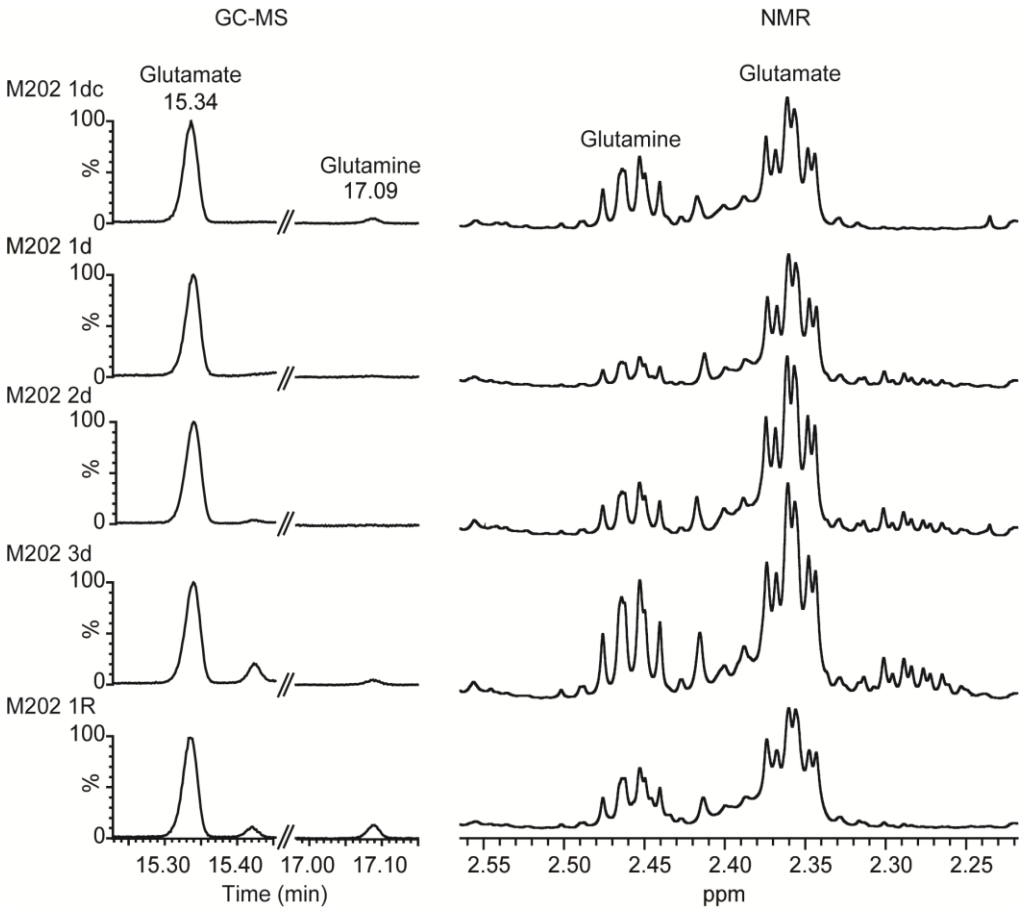


Figure 3.6. Stacked plots of the TICs from the GC-MS (left) and NMR (right) results showing the regions containing Glu and Gln. Only data from the extracts of the M202 variety are shown for simplicity but the data are representative of both genotypes. The time points indicate a control (1dc), 1d, 2d, 3d of submergence or 3d submergence followed by 1d post-submergence recovery (1R). For the GC-MS data, the TIC's are scaled to Glu.

Table 3.3: Signal to noise ratios determined from GC-MS and NMR data. For GC-MS data, S/N was calculated as the root mean square (RMS) using the MassLynx 4.1 software. For NMR data, the S/N was calculated using the Bruker Topspin 3.1 (Bruker Biospin, Billerica, MA). For noise calculation, a region of the ^{1}H spectral baseline was selected between 9.0 and 9.2 ppm that was free of signals and equal in width to the signal region. For each measurement, the average and standard deviation of at least 5 replicates for every time point are reported. The NC represents S/N not calculated due to a ratio below 10.

	GC-MS			NMR		
	Gln	**Glu**	**Gln/Glu**	**Gln**	**Glu**	**Gln/Glu**
M202 1dc	24 ± 5	526 ± 129	0.05 ± 0.02	156 ± 99	287 ± 181	0.54 ± 0.02
M202 1d	NC	779 ± 137	NC	47 ± 11	207 ± 52	0.23 ± 0.02
M202 2d	NC	707 ± 69	NC	102 ± 36	287 ± 64	0.35 ± 0.08
M202 3d	NC	1206 ± 312	NC	128 ± 46	229 ± 76	0.55 ± 0.05
M202 1R	30 ± 11	739 ± 172	0.04 ± 0.01	162 ± 68	219 ± 81	0.74 ± 0.14

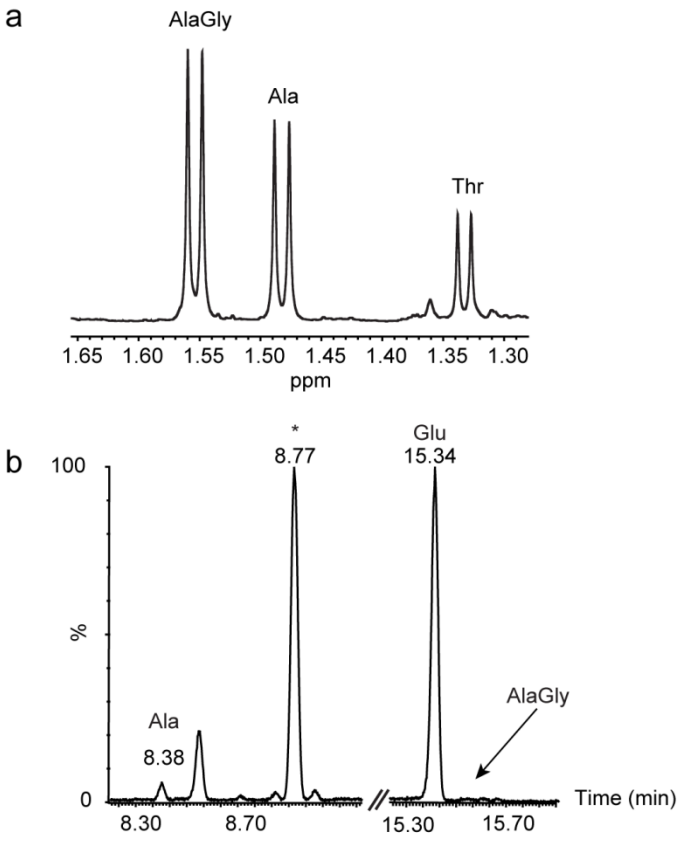


Figure 3.7. ^{1}H NMR spectrum (a) and GC-MS chromatogram (b) measured for extracts of the same M202 control rice tissue showing the differences in the response of the two analytical platforms for Ala and AlaGly. AlaGly is present in greater abundance than Ala in this sample as judged by the relative intensity of the NMR resonances but is not detectable in the GC-MS TIC. The retention time indicated on the chromatogram in (b) shows the expected elution time based injection of an AlaGly standard. An asterisk indicates the FAMEs retention index marker in (b).

samples. Figure 3.7 shows a representative ^{1}H NMR spectrum and GC-MS total ion chromatogram (TIC). Because of the inherently quantitative nature of NMR, comparison of the relative integrals of the methyl resonances for AlaGly (1.66) and Ala (1.00) in Figure 3.7a provides a direct quantitative relationship of the relative concentrations of the two compounds in the extract. Although GC-MS is generally considered to have greater sensitivity than NMR, AlaGly is not detected by GC-MS in Figure 3.7b. Because AlaGly could not be identified by deconvolution and library matching, a standard was derivatized and injected to establish the expected GC elution time for comparison with the rice tissue extracts. Several factors could be responsible for much lower detection sensitivity for AlaGly including low volatility, hydrogen bonding to the injector liner or poor derivatization efficiency.

SMM is another metabolite that we readily detected by ^{1}H NMR but did not observe by GC-MS analysis of rice extracts (Figure 2.5). SMM can play an important role in ethylene production as well as sulfur transport and has been previously shown to accumulate in anoxic rice seedlings.[5, 2, 3] Although NMR analysis of the submergence treated tissue did not reveal significant changes in the levels of SMM in the two genotypes, its absence from the GC-MS data was enigmatic. The positive formal charge of SMM could reduce its volatility and increase its ability to bond with various parts of the inlet and column. Another factor in its low sensitivity could be excessive fragmentation by the electron ionization (EI) source. Comparison of the mass spectra of AlaGly, Ala, and SMM (Figure 3.6) reveals that AlaGly and Ala both produce the same major fragment ions (m/*z* 73, 75, and 116) with differences between the molecules

observable in the very low abundance ions. In contrast, SMM is heavily fragmented, producing many ions at relatively high abundance with the m/z 176 ion as the most intense peak.

3.4 Conclusions

The goal of untargeted metabolomics experiments is to sample as much of the metabolite chemical space as possible. As this study has shown, this is best accomplished using multiple analytical platforms. The GC-MS results both corroborated and complemented our previous NMR study. Metabolites exclusively detected by GC-MS analysis include the amino acids Phe, Pro, pyroglutamate, the TCA intermediates citrate, malate, and succinate as well as the organic acids GABA, glycerate, oxalate, shikimate, and threonate. These metabolites provided additional support for carbohydrate regulation by *SUB1*A during submergence as evident through pyruvate metabolism to Ala, and changes in the levels of GABA, TCA cycle intermediates and other amino acids. Possible points of regulation of the TCA cycle and evidence for alternative, non-cyclic pathways were revealed by GC-MS through the rapid elevation of citric acid and gradual accumulation of GABA and succinate. A possible biomarker for dehydration or re-oxygenation stress was also evident in the intolerant variety through accumulation of Gly, which may affect cellular signaling pathways during the recovery period. The increase in relative abundance of pyroglutamate may also be an indicator of oxidative stress during and after submergence due to its role in glutathione metabolism.

Throughputs of both platforms were comparable in terms of the time required for measurements and data processing, while the advantages and disadvantages of each

platform were complementary. NMR is inherently quantitative, universally detects organic compounds and has a high dynamic range. However, NMR is generally considered to be less sensitive than GC-MS and in the analysis of whole plant extracts it can be difficult to resolve signals for some compounds due to resonance overlap. Although GC-MS was able to detect 14 compounds not observed by NMR, three important metabolites Gln, AlaGly and SMM were more readily detected and quantified by ^{1}H NMR. The dynamic range limitations of the TOF mass analyzer available for this work reduced somewhat the utility of GC-MS however the dynamic range of the measurements could be effectively expanded by using split injections or performing measurements of extracts at differing degrees of dilution. Taken together, these results demonstrate that the complementary use of ^{1}H NMR and GC-MS can facilitate a more thorough exploration the metabolome of the biological system in question.

In chapter 4, a targeted analysis of secondary metabolism is conducted on extracts from rice tissue. Weak anion-exchange solid phase extraction (WAX-SPE) is used to select for anionic compounds in polar extracts with the purpose of specifically targeting phosphorylated mono- and disaccharides. RPIP-UPLC-MS analysis is used to analyze the WAX-SPE eluent, allowing the identification and quantitation of several metabolites involved in secondary metabolism and energy utilization. Additionally, NMR and GC-MS analysis are combined with RPIP-UPLC-MS analysis of rice tissue to provide a more detailed understanding of the metabolic reconfigurations occurring during the recovery period following desubmergence.

3.5 References

(1) Barding, G. A.; Fukao, T.; Beni, S.; Bailey-Serres, J.; Larive, C. K. Differential metabolic regulation governed by the rice *SUB1*A gene during submergence stress and identification of alanylglycine by ^{1}H NMR spectroscopy. *J. Proteome Res.* **2012,** *11*, 320-330.

(2) Ko, S.; Eliot, A. C.; Kirsch, J. F. S-Methylmethionine is both a substrate and an inactivator of 1-aminocyclopropane-1-carboxylate synthase. *Arch. Biochem. Biophys.* **2004,** *421*, (1), 85-90.

(3) Menegus, F.; Lilliu, I.; Brambilla, I.; Bonfa, M.; Scaglioni, L. Unusual accumulation of S-methylmethionine in aerobic-etiolated and in anoxic rice seedlings: An H-1-NMR study. *J. Plant Physiol.* **2004,** *161*, (6), 725-732.

(4) Ravanel, S.; Gakiere, B.; Job, D.; Douce, R. The specific features of methionine biosynthesis and metabolism in plants. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95*, (13), 7805-7812.

(5) Bourgis, F.; Roje, S.; Nuccio, M. L.; Fisher, D. B.; Tarczynski, M. C.; Li, C. J.; Herschbach, C.; Rennenberg, H.; Pimenta, M. J.; Shen, T. L.; Gage, D. A.; Hanson, A. D. S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell* **1999,** *11*, (8), 1485-1497.

(6) Narsai, R.; Howell, K. A.; Carroll, A.; Ivanova, A.; Millar, A. H.; Whelan, J. Defining core metabolic and transcriptomic responses to oxygen availability in rice embryos and young seedlings. *Plant Physiol.* **2009,** *151*, (1), 306-322.

(7) Fukao, T.; Xu, K. N.; Ronald, P. C.; Bailey-Serres, J. A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* **2006,** *18*, (8), 2021-2034.

(8) Lee, D. Y.; Fiehn, O. High quality metabolomic data for *Chlamydomonas reinhardtii*. *Plant Methods* **2008,** *4*, 13.

(9) 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*, (8), 1635-1638.

(10) Goto, S.; Bono, H.; Ogata, H.; Fujibuchi, W.; Nishioka, T.; Sato, K.; Kanehisa, M. Organizing and computing metabolic pathway data in terms of binary relations. *Pacific Symposium on Biocomputing '97* **1997**, 175-186186.

(11) Martinelli, T.; Whittaker, A.; Masclaux-Daubresse, C.; Farrant, J. M.; Brilli, F.; Loreto, F.; Vazzana, C. Evidence for the presence of photorespiration in desiccation-sensitive leaves of the C-4 'resurrection' plant Sporobolus stapfianus during dehydration stress. *J. Exp. Bot.* **2007,** *58*, (14), 3929-3939.

(12) Dubos, C.; Huggins, D.; Grant, G. H.; Knight, M. R.; Campbell, M. M. A role for glycine in the gating of plant NMDA-like receptors. *Plant J.* **2003,** *35*, (6), 800-810.

(13) De Silva, D. L. R.; Hetherington, A. M.; Mansfield, T. A. Synergism between calcium ions and abscisic acid in preventing stomatal opening. *New Phytol.* **1985,** *100*, (4), 473-482.

(14) Subbaiah, C. C.; Sachs, M. M. Molecular and cellular adaptations of maize to flooding stress. *Ann. Bot.* **2003,** *91*, (2), 119-127.

(15) Bailey-Serres, J.; Chang, R. Sensing and signalling in response to oxygen deprivation in plants and other organisms. *Ann. Bot.* **2005,** *96*, (4), 507-518.

(16) Fukao, T.; Yeung, E.; Bailey-Serres, J. The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell* **2011,** *23*, (1), 412-427.

(17) Kumar, A.; Bachhawat, A. K. Pyroglutamic acid: throwing light on a lightly studied metabolite. *Curr. Sci.* **2012,** *102*, (2), 288-297.

(18) Schilling, S.; Stenzel, I.; von Bohlen, A.; Wermann, M.; Schulz, K.; Demuth, H. U.; Wasternack, C. Isolation and characterization of the glutaminyl cyclases from *Solanum tuberosum* and *Arabidopsis thaliana*: implications for physiological functions. *Biol. Chem.* **2007,** *388*, (2), 145-153.

(19) Bailey-Serres, J.; Voesenek, L. Flooding stress: Acclimations and genetic diversity. *Annu. Rev. Plant Biol.* **2008,** *59*, 313-339.

(20) Bailey-Serres, J.; Fukao, T.; Gibbs, D. J.; Holdsworth, M. J.; Lee, S. C.; Licausi, F.; Perata, P.; Voesenek, L. A. C. J.; van Dongen, J. T. Making sense of low oxygen sensing. *Trends Plant Sci.* **2012,** *17*, (3), 129-138.

(21) Shelp, B. J.; Mullen, R. T.; Waller, J. C. Compartmentation of GABA metabolism raises intriguing questions. *Trends Plant Sci.* **2012,** *17*, (2), 57-59.

(22) 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*, (8), 462-470.

(23) Jacobsen, N. E. *NMR spectroscopy explained: Simplified theory, application, and examples for organic chemistry and structural biology*. John Wiley and Sons, Inc: Hoboken, 2007.

(24) Gross, J. H. *Mass Spectrometry*. Springer-Verlag: Berlin, 2004; p 507.

(25) Lombardo, V. A.; Osorio, S.; Borsani, J.; Lauxmann, M. A.; Bustamante, C. A.; Budde, C. O.; Andreo, C. S.; Lara, M. V.; Fernie, A. R.; Drincovich, M. F. Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. *Plant Physiol.* **2011,** *157*, (4), 1696-1710.

(26) Howell, K. A.; Narsai, R.; Carroll, A.; Ivanova, A.; Lohse, M.; Usadel, B.; Millar, A. H.; Whelan, J. Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. *Plant Physiol.* **2009,** *149*, (2), 961-980.

(27) Muhlemann, J. K.; Maeda, H.; Chang, C. Y.; San Miguel, P.; Baxter, I.; Cooper, B.; Perera, M. A.; Nikolau, B. J.; Vitek, O.; Morgan, J. A.; Dudareva, N. Developmental changes in the metabolic network of snapdragon flowers. *Plos One* **2012,** *7*, (7), e40381.

(28) Wang, S. F.; Wang, X. F.; He, Q. W.; Liu, X. X.; Xu, W. L.; Li, L. B.; Gao, J. W.; Wang, F. D. Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Rep.* **2012,** *31*, (8), 1437-1447.

(29) Thurman, E. M.; Ferrer, I., Comparison of quadrupole time-of-flight, triple quadrupole, and ion-trap mass spectrometry/mass spectrometry for the analysis of emerging contaminants. In *Liquid Chromatography/Mass Spectrometry, MS/MS and Time of Flight MS*, American Chemical Society: 2003; Vol. 850, pp 14-31.

(30) Barding, G.; Salditos, R.; Larive, C. Quantitative NMR for bioanalysis and metabolomics. *Anal. Bioanal. Chem.* **2012,** *404*, (4), 1165-1179.

(31) Kusano, M.; Tabuchi, M.; Fukushima, A.; Funayama, K.; Diaz, C.; Kobayashi, M.; Hayashi, N.; Tsuchiya, Y. N.; Takahashi, H.; Kamata, A.; Yamaya, T.; Saito, K. Metabolomics data reveal a crucial role of cytosolic glutamine synthetase 1;1 in coordinating metabolic balance in rice. *Plant J.* **2011,** *66*, (3), 456-466.

(32) 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.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009,** *37*, D603-D610.

CHAPTER FOUR

The Effect of Re-oxygenation on the Primary and Secondary Metabolomes of *SUB1A* Containing Rice and the Quantitation of Trehalose-6-Phosphate by RPIP-UPLC-MS

Acknowledgements: I would like to acknowledge the following people for their contributions to this project: Dr. Szabolcs Béni (Semmelweis University) for his help in sample preparation, Ngoc Huynh for her assistance in sample preparation and data collection, Dr. Takeshi Fukao for the discussion regarding experimental approach, Elaine Yeung for her help planting and maintaining the rice, and Sumukh Sathnur for his assistance with planting, harvesting, and NMR data collection/analysis.

Abstract:

In this chapter, NMR, GC-MS and UPLC-MS were used to interrogate the metabolic differences introduced the presence or absence of *SUB1A* in the primary and secondary metabolomes of the shoots of rice (*Oryza sativa* ssp. *japonica*) in response to submergence stress, re-oxygenation, and the diurnal cycle. For these experiments, the M202 and M202(*Sub1*) rice varieties were submerged for 3 d at midday and allowed to recover until dusk, midnight, dawn, and midday 24 h post re-oxygenation. Simultaneously, control plants were harvested at each time point to elucidate the contributing effects of the diurnal cycle on the results. Although few differences between genotypes were detected under control conditions, distinctions in shoot metabolite levels between the genotypes during re-oxygenation were observable at night for several

metabolites, including methionine, glycine, histidine, serine, and shikimate. These results suggest that either re-oxygenation, low-light to darkness, or a combination of the two had a significant effect on metabolism. Additionally, changes in the levels of trehalose-6-phosphate (T6P), a key signaling molecule responsible for carbon sensing, was monitored over the course of the treatment by reverse-phase ion-pair ultrahigh performance chromatography – mass spectrometry (RPIP-UPLC-MS) and found to be differentially regulated during the night. The metabolic information obtained through this re-oxygenation time course study provides greater insights into the molecular response of rice following de-submergence.

4.1 Introduction

Diel effects on plant metabolism are well documented to influence a wide range of metabolic pathways, including glycolysis, nitrogen assimilation, tricarboxylic acid (TCA) cycle, reactive oxygen species (ROS) removal, and photorespiration.[1, 2] During submergence stress, the rice *SUB1A* gene has been shown to regulate carbon consumption, nitrogen metabolism, photorespiration, and the TCA cycle by decreasing the gibberellic acid response to catabolize leaf starch reserves.[3-5] Additionally, the *SUB1A* gene has been hypothesized to improve plant tolerance to ROS, which are commonly encountered during reoxygenation upon de-submergence and other abiotic stresses.[6] Interestingly, presence of the *SUB1A* gene is also known to increase survivability of rice subjected to prolonged darkness by limiting the breakdown of chlorophyll as well as through the conservation of carbon stores.[7] Prolonged darkness is

commonly encountered by crops during submergence stress due to turbid water, suggesting that *SUB1A* is not only induced by the entrapment of ethylene under complete submergence, but other stresses encountered by the plant, such as extended darkness. This would reduce shoot photosynthesis driving the consumption of leaf starch reserved. Whether *SUB1A* is regulated by the diurnal cycle, however, is not well understood, and the influence of SUB1A on metabolism during normal day and night conditions might prove valuable.

In Chapters 2 and 3, the effects of re-oxygenation after 24 h of recovery were interrogated in the submergence tolerant and intolerant rice varieties.[3, 4] Although re-oxygenation allows the plant to regain anaerobic metabolism, it also increases ROS production and causes dehydration, a secondary stress resulting from the submergence event.[6] Unique metabolic differences between the M202 and M202(*Sub1*)varieties were detected and could be related to photosynthesis, photorespiration, and carbon fixation at the recovery time point. We hypothesized that these differences during recovery may reflect a secondary stress induced by factors associated with de-submergence. To further investigate the metabolic differences between the M202 and M202(*Sub1*) variety during the first 24 h of re-oxygenation, plants were submerged for 3 d, desubmerged at midday, and allowed to recover until dusk, midnight, dawn, or midday 24 h following de-submergence. To simultaneously evaluate the effect of *SUB1A* on the diurnal cycle of rice plants as well as the influence of the diurnal cycle on de-submergence, control plants were harvested at the corresponding time intervals. Figure 4.1 shows the experimental design for submergence and recovery. Tissue harvest began at the 3 d submergence time

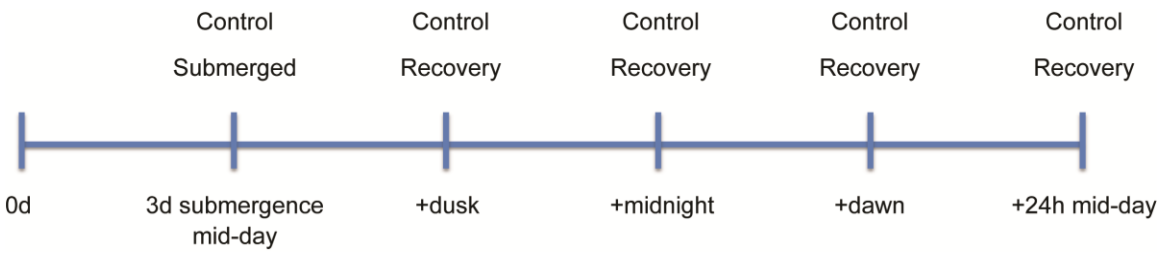


Figure 4.1. The experimental approach used to interrogate the effects of submergence and the diurnal cycle. Control and stressed plants were harvested at 3 d of submergence and at each proceeding time point for each of the M202 and M202(*Sub1*) varieties.

point and continued for 24 h, resulting in 5 time points samples for the stress-treated and diurnal control plants.

Trehalose-6-phosphate (T6P) is a secondary metabolite linked to carbon sensing and flowering.[8-11] The loss of function of one gene responsible for synthesis of T6P, *TPS1*, results in an embryonic lethal mutation in the model plant *Arabidopsis thaliana*, indicating a significant physiological role for the phosphorylated disaccharide.[12] Compared with non-*SUB1A* containing rice, Fukao and coworkers found that continuous expression of the *SUB1A* gene in transgenic accessions resulted in a significant flowering delay, however it is unknown whether this delay is the result of decreased levels of T6P due to sequestration of carbon flux or to perturbation of a different pathway.[7] Because of the importance of T6P in carbon sensing and flowering, understanding its relationship with the acclimation response in submergence tolerant and intolerant rice would further our understanding of the complex role of *SUB1A*.

Due to its low abundance, reliable quantitative analysis of T6P in plant tissue extracts requires a targeted approach to minimize matrix effects, and increase selectivity and sensitivity. Methods reported for T6P analysis include anion-exchange chromatography coupled with mass spectrometry (AEC-MS) and hydrophilic interaction chromatography coupled with mass spectrometry (HILIC-MS).[13, 14] The high salt content required for AEC-MS, however, makes the method undesirable as special instrumentation is needed to remove the non-volatile salts prior to MS.[14] HILIC-MS was determined to be preferred over AEC-MS because it is a more universal method, easily transferrable across laboratories, and improved sensitivity by a factor 20 compared with AEC-MS.[13]

HILIC, however, requires extensive column equilibration with the target sample matrix to produce a reproducible separation, which is disadvantageous when analyzing samples that require constant cleaning and re-equilibration.[15] Matrix effects can also lead to problems in HILIC separations when the nature of the matrix varies with sample type, a problem that can be significant in metabolomics experiments.

An attractive alternative to AEC-MS and HILIC-MS is reverse-phase ion-pair (RPIP)-UPLC-MS, which has been used extensively for the analysis of anionic oligosaccharides.[16-21] RPIP-UPLC utilizes an ion-pairing reagent (IPR) that has an ionic functional group that can ion-pair with the target compound and a hydrophobic tail that interacts with the column stationary phase (Figure 1.3). In this study, a RPIP-UPLC-MS method was developed for the determination of T6P and other anionic secondary metabolites to elucidate their physiological roles in the plant response to submergence stress, re-oxygenation, and the diurnal cycle.

4.2 Materials and Methods

4.2.1 Reagents

Acetic acid-d_4, ammonium deuteroxide -d_5, and deuterium oxide were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). NMR chemical shift reference 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and derivatization reagents methoxyamine hydrochloride (MeOX) and pyridine (+99% purity) were purchased from Sigma-Aldrich (St. Louis, MO). *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) in 1% trimethylchlorosilane was obtained from Thermo Fisher Scientific (Waltham, MA). Metabolite standards were obtained from MP Biomedicals (Solon, OH), Fisher

Scientific (Pittsburgh, PA), and Sigma-Aldrich. Water (Burdick and Jackson, VWR, Radnor, PA) chloroform (Mallinckrodt Laboratory Chemicals, Phillipsburg, NJ), methanol, acetic acid, and ammonium hydroxide (Fisher Scientific) were of at least ACS grade. Weak anion exchange solid phase cartridges (60 mg, 3cc) for LC-MS analysis were purchased from Bonna-Agela Technologies (Wilmington, DE). Dibutylamine was obtained from Fluka (Sigma-Aldrich).

4.2.2 Rice Growth Conditions

Rice was grown and submergence stressed as previously described.[3-5] The seeds were sterilized with 1% (v/v) sodium hypochlorite and 0.2% Tween-20 followed by rinsing with deionized (DI) water. Seeds were then immersed in overnight in DI water in the dark and germinated on moist paper towels in a Pyrex dish covered with plastic wrap for 5 d. After germination, seeds were transplanted into pots with soil using a planting density of 20 plants per pot and grown until the 3-leaf stage in a greenhouse at ambient temperature. Seedling survival after transplantation was >99%.

4.2.3 Treatment and Harvest

Submergence stress was carried out using six 121 L trash cans filled with DI water and allowed to equilibrate overnight to the ambient temperature of the greenhouse (~30 °C) prior to submergence. At the three-leaf stage (approximately 12 d), at midday seedlings were submerged for 3 d. For re-oxygenation, plants were transferred from submergence conditions to the greenhouse bench. Plants were harvested either upon de-submergence or at the end of the designated recovery period of dusk, midnight, dawn, and midday 24 h after de-submergence. Submergence and recovery time points were

determined based on the time of the sunrise and sunset. For tissue harvest, plants were de-submerged, aerial tissue was removed, rinsed in DI water, and flash frozen in liquid nitrogen and stored at -80 °C. Control plants were grown simultaneously and harvested with the corresponding treatment at either de-submergence or the individual recovery time points. Harvested plant tissue was ground to a powder by mortar and pestle under liquid nitrogen and lyophilized overnight until dry. Samples were then stored at -80 °C until extracted.

4.2.4 Metabolite Extraction and Solid Phase Extraction

4.2.4.1 Metabolite Extraction for NMR and GC-MS

Metabolite extraction for NMR and GC-MS analysis was carried out as previously described in sections 2.2 and 3.2.[3, 4] Briefly, 1.5 mL of 80:20 MeOH:H_2O was added to 50 mg of dried rice tissue in 2.0 mL Eppendorf tubes, shaken for 5 min on a platform shaker and centrifuged at 12 000 g for 4 min. For GC-MS analysis, 30 μL aliquots were delivered into glass inserts (350 μL, Phenomenex, Torrence, CA) supported by 1.5 mL Eppendorf tubes and for NMR analysis, 1 mL aliquots were transferred to clean 1.5 mL Eppendorf tubes. Samples were then centrifuged under vacuum until dry and stored at -20 °C until analyzed.

4.2.4.2 Metabolite Extraction for UPLC-MS Analysis

The sample extraction and solid phase extraction (SPE) protocols were adapted from a method developed by Delatte and coworkers.[14] Briefly, 1 mL of a 3:8 chloroform:acetonitrile mixture was added to 20 mg of dried plant tissue and shaken on a platform shaker for 5 min to break up cell walls. Two liquid/liquid extractions were

carried out by adding separately 0.8 mL of water (for a total of 1.6 mL). The samples were shaken for 5 min on a platform shaker, centrifuged at 12 000 x *g* for 4 min, and the aqueous layer transferred to a 2.0 mL Eppendorf tube. Because acetonitrile is miscible with water, partitioning of acetonitrile from the organic phase into the aqueous phase was expected and resulted in an additional 0.2 mL of the aqueous phase being transferred to the 2.0 mL Eppendorf tube for a final volume of 1.8 mL Prior to UPLC-MS analysis, samples were subjected to weak-anion exchange SPE. SPE cartridges were conditioned with 1.0 mL of MeOH followed by 1.0 mL water containing 5 % NH_4OH. The dried extracts were reconstituted in 2.0 mL of water, loaded onto the cartridge and washed with 1.0 mL of water followed by 1.0 mL of methanol. Cartridges were allowed to dry and then eluted into a 2.0 mL Eppendorf tube with 2.0 mL of 80/20 H_2O/MeOH containing 5 % formic acid and dried by vacuum centrifugation for UPLC-MS analysis.

4.2.5 NMR Analysis of Metabolite Extracts

Samples were dissolved in 700 µL of a deuterated 100 mM ammonium acetate buffer adjusted to a pD of 7.6 containing 175 µM DSS as a chemical shift reference (0 ppm). The pD of the solution was determined from the pH meter reading (pH^*) with the equation $pD = pH^* + 0.4$.[22] To remove lipids, a liquid-liquid extraction was performed with 100 µL $CDCl_3$, centrifuged at 5000 x *g* to break the emulsion, and 620 µL of the aqueous phase transferred to a clean NMR tube for analysis.

Samples were analyzed with a 14.1 T Bruker Avance NMR spectrometer tuned to detect ^1H resonances at 599.69 MHz. Spectra were collected acquired using a 10.0 µs 90° pulse with a 5 mm inverse broadband probe using digital quadrature detection (DQD)

with the transmitter set on the HOD frequency and sample temperature maintained at 298 K. The Bruker-defined wet pulse program (wet) was used to reduce the intensity of the HOD resonance. Magnetic field homogeneity was optimized using up to 28 shims and the probe was manually tuned and matched. Spectra were required without spinning and the spectrometer was locked using D_2O. A relaxation delay of 1.5 s and a spectral width of 11.97 ppm were used giving a total experiment time of 41 min for which 16 dummy scans and 1048 transients were co-added. The DSS line-width was less than 0.8 Hz for each sample before apodization with an exponential function equivalent to 0.5 Hz line broadening. Free induction decays (FIDs) were collected into 32 768 points and zero filled to 65 536 points.

4.2.6 GC-MS Analysis of Metabolite Extracts

All procedures for GC-MS analysis and data processing were performed as described in the materials and methods section of Chapter 3 (Chapter 3.2).[3]

4.2.7 RPIP-UPLC-MS Analysis of Metabolite Extracts

The sample extracts prepared for GC-MS analyses were also used for the RPIP-UPLC-MS measurements. Samples were reconstituted in 200 µL of water and 20 µL injected onto the column. UPLC separations were performed on a 2.1 x 100 mm Acquity UPLC HSS T3 column with a 1.8 µm particle size (Waters Corporation, Milford, MA). A guard column of the same stationary phase was used in-line with the analytical column. The separation was performed on an Acquity Ultra Performance LC (Waters Corporation) with a binary solvent system for gradient elution; solvent A was aqueous 10 mM DBA buffer adjusted to a pH of 7.4 with acetic acid and solvent B was 10 mM DBA

buffer in MeOH adjusted to a pH meter reading of 8.0 with acetic acid. The solvent gradient used an initial isocratic step of 98 % solvent A for 5 min, followed by a 3.5 % increase of solvent B over the next 8 min, after which the faction of solvent B was increased by 10.8 % for 6 min until reaching 95 %. The fraction of solvent B was maintained at 95 % for 2 min, following a return to 98 % A over 1 min. The column was allowed to equilibrate for 8 min between injections. A Waters ESI quadruple time-of-flight mass spectrometer operated in negative mode was used to measure total ion chromatograms (TICs) using the following parameters: desolvation temperature, 250 °C; source temperature, 120 °C; cone voltage, 35 V; capillary voltage, 2800 kV; extractor voltage, 1V; radio frequency lens, 0.5 V; m/z range, 80-800; scan time of 0.9 sec with an inter-scan delay 0.1 s.

4.2.7.1 Data Processing

NMR and GC-MS data pre-processing was carried out as described in sections 2.2.5.3 and 3.2.7. LC-MS data pre-processing was carried out with MassLynx (Waters Corporation) to collect integration and peak height values for identified metabolites. For adenosine monophosphate (AMP), 6-phosphogluconate (6PG), and 3-phosphoglycerate (3PG) integrated peak areas were determined by integrating the extracted ion chromatogram (XIC) for each compound. Prior to integration, Savitzky Golay smoothing with a window size of ±3 scans 2 smooths was applied to each XIC and automatic integration performed by MassLynx. Because of partial peak overlap of the different isomers of phosphorylated sugars, peak heights for glucose-6-phosphate (G6P), sucrose-6-phosphate (S6P), and trehalose-6-phosphate (T6P) were extracted from mass spectra.

Prior to data extraction, the mass spectrum of each compound was smoothed as described for integrated data, the top 90% of the peak was centered with a minimum peak width at half height of 2 channels and the peak heights in counts determined. All data was transferred to an Excel spreadsheet for statistical analysis.

4.2.8 Statistical Analysis

Statistical analysis was carried out using either the peak areas (for NMR, GC-MS, and selected LC-MS analytes) or peak heights (for G6P, S6P and T6P) as indicated in section 4.2.7.1 using the freely available program R for statistical analysis. Data normalization was only used for data collected by NMR. The averages of the results from at least 5 biological replicates were used to determine fold changes and significance (p-value of 0.05) was determined by an analysis of variance (ANOVA) using the R package oneway.test.[23] Trajectory plots were generated in Origin 7.5 (OriginLab Corporation, Northampton, MA).

4.3 Results and Discussion

An analytical approach incorporating GC-MS, NMR, and RPIP-UPLC-MS was used to elucidate the effects of re-oxygenation after submergence stress on the M202 and M202(*Sub1*) rice varieties. Additionally, metabolite profiling of both genotypes was carried out to elucidate the interaction of the diurnal cycle and the *SUB1A* gene. Plants from both genotypes were de-submerged at midday and allowed to recover until dusk, midnight, dawn, or midday 24 h post submergence, and interrogated for differences in the levels of primary and secondary metabolites. Additionally, circadian differences between

the two genotypes were investigated in the absence of stress to determine differences between M202 and M202(*Sub1*).

4.3.1 NMR and GC-MS Analysis of Primary Metabolism

4.3.1.1 The Influence of the Diurnal Cycle on Rice Metabolism

The differences in metabolite levels between M202 and M202(*Sub1*) were examined to better understand the influence of the diurnal cycle at midday, dusk, midnight, dawn, or midday +24 h on metabolites in plants that possess or lack the *SUB1A* gene. Metabolically, energy is produced through photosynthesis in the light and through aerobic respiration in the dark. The metabolite profiles for control plants of both genotypes were measured by GC-MS and NMR and represented as trajectory plots in Figures 4.2 and 4.3, respectively, by plotting the area (GC-MS) or normalized area (NMR) over the course of the diurnal cycle. Additionally, the effect of the diurnal cycle on primary metabolism is summarized in Tables 4.1 and 4.2 as fold changes between and within genotypes, with significant differences (p-value < 0.05) represented by an asterisk. A total of 19 amino acids, 11 organic acids, 3 phytosterols, 2 sugars and 1 dipeptide were detected using both GC-MS and NMR. Of the 19 amino acids, 10 were quantified using both instruments and were highly correlated, with small differences between the trajectory plots attributable to differences in sensitivity in the two measurements. Sucrose, glucose, alanylglycine (AlaGly), S-methyl methionine (SMM), and γ-amino butyrate (GABA) were exclusively quantified by NMR while β-alanine (β-ala)*, pyroglutamate, glycine (Gly), histidine (His)*, lysine (Lys), methionine (Met)*, phenylalanine (Phe), tryptophan (Trp), tyrosine (Tyr), campesterol, cirate, fumarate, γ-

hydroxybenzoate, malate, malonate*, myo-inositol, shikimate, β-sitosterol, stigmasterol, succinate, and threonate were exclusively quantified by GC-MS. The compounds indicated with an asterisk were below the limit of quantitation (LOQ) for the control samples but were detected at higher levels in the treated plants as discussed in section 4.3.1.2.

Metabolite levels are known to be affected by diel conditions in plants, but regulation at this temporal level by the *SUB1A* gene has not been investigated.[24-26] In experiments using the *Oryza sativa* ssp. *japonica* cv. Zhonghua 10 variety, carbohydrate assays showed that levels of glucose, fructose, and sucrose peaked in leaf tissue at the end of the light cycle and decline during the dark cycle.[2] Our results indicate a slightly different trend in glucose levels for the M202 and M202(*Sub1*) varieties as compared with Zhonghua 10. As in the study by Wang and coworkers (citation), sucrose accumulated throughout the course of the day, peaking at the end of the light cycle, and declined during the dark cycle (Figure 4.3, Table 4.2). This is consistent with the hypothesis that starch produced during photosynthesis is catabolized to sucrose at night to maintain energy homeostasis through the production of glucose and fructose.[27] In contrast, glucose levels increased during the dark cycle in M202 and M202(*Sub1*). This result was the opposite of the observations by Wang and coworkers, who found that glucose decreased.[2] The reasons for the differences between the two experiments could be attributed to data normalization or differences between the varieties, however further experiments are required to resolve this.

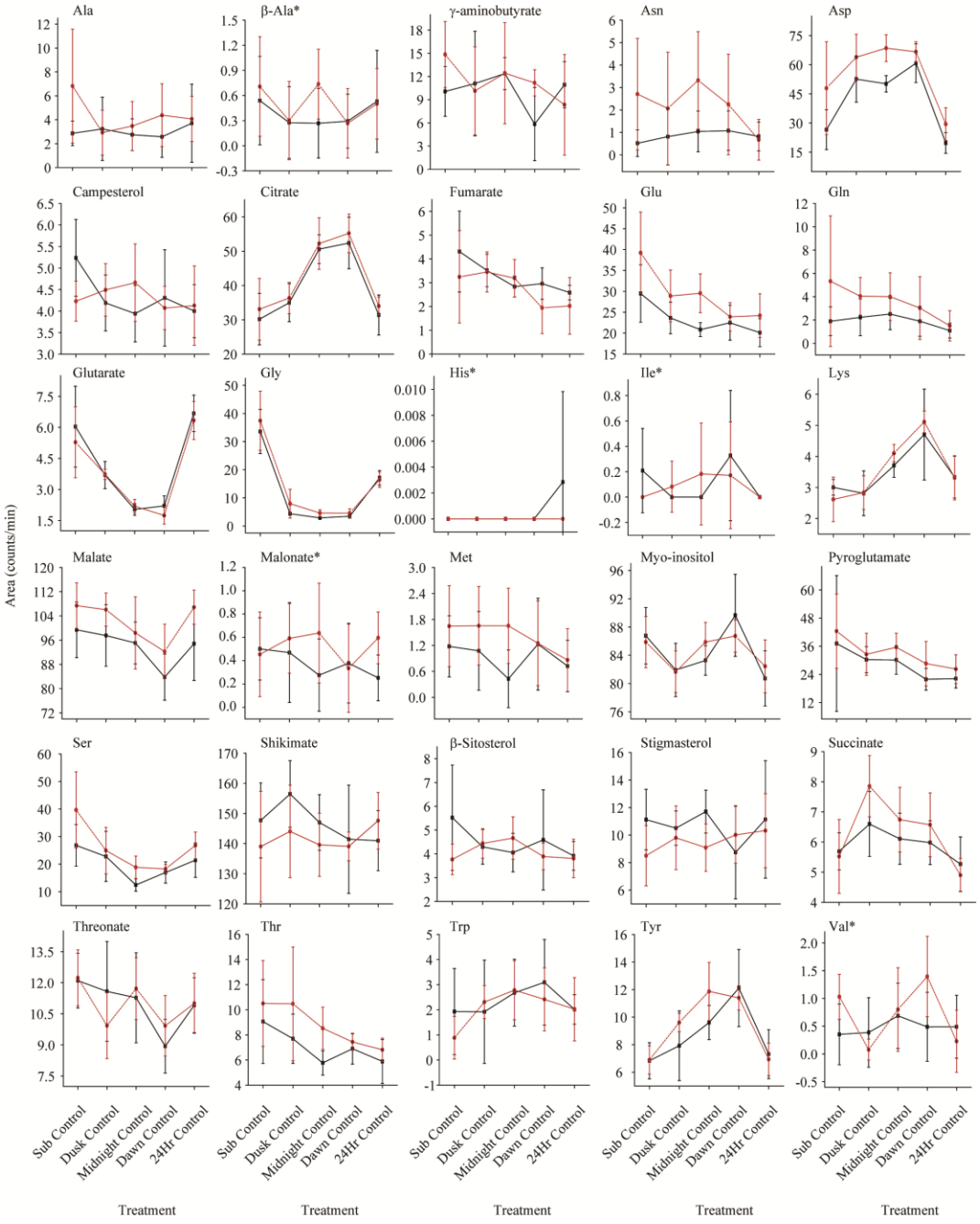


Figure 4.2. Trajectory plots from the GC-MS data representing the average area (counts/min) for M202 (■) and M202(*Sub1*) (●). The time points are connected using black (M202) or red (M202(*Sub1*)) lines. The treatments are labeled as Sub Control, Dusk Control, Midnight Control, Dawn Control, and 24Hr Control to represent control tissue at each time point. Each data point is the replicate of at least 5 biological replicates and error bars represent standard deviation. Metabolites below the limit of quantitation are indicated with an asterisk.

Table 4.1. A comparison of the levels of metabolites in extracts of M202 and M202(*Sub1*) controls determined by GC-MS. The comparison between varieties is represented by fold changes for each diurnal interval in the absence of submergence. All comparisons were made to the true midday time point. Asterisks indicate significant differences at the 95% confidence limit. ND represents metabolite ratios not determined because signals were below the limit of quantification. The "C" at the end of the treatment label is used to indicate non-submerged tissue harvested from control plants.

	M202				M202(*Sub1*)				M202/M202(*Sub1*)				
	DuskC:SubC	MidnightC:SubC	DawnC:SubC	24hrC:SubC	DuskC:SubC	MidnightC:SubC	DawnC:SubC	24hrC:SubC	SubC	DuskC	MidnightC	DawnC	24hrC
Amino Acids													
Ala	1.1	1.0	1.0	1.4	0.5	0.6	1.0	0.8	0.5	1.2	0.9	0.5	0.9
β-Ala	0.5	0.6	0.6	1.2	0.5	1.3	0.4	0.9	0.9	0.9	0.4	1.2	1.2
Asn	ND	ND	ND	ND	0.7	1.4	0.9	0.3	ND	0.4	0.3*	0.5	1.3
Asp	2.0	2.0	2.5	0.9	1.6	1.7	1.7	0.8	0.6	0.8	0.8*	1.0	0.7
Gln	1.2	1.2	1.1	0.7	0.8	0.7	0.6	0.4	0.4	0.7	0.7	0.8	0.7
Glu	0.8	0.8*	0.8	0.8	0.8	0.8	0.7	0.8	0.9	0.8	0.8*	1.0	0.9
Gly	0.1*	0.1*	0.1*	0.6*	0.2*	0.1*	0.1*	0.5*	1.0	0.6	0.7	0.8	1.1
His	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Ile	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Lys	0.9	1.3	1.6*	1.2	1.2	1.7	2.3*	1.5	1.3	1.0	1.0	1.0	1.1
Met	0.9	0.4	1.3	1.0	1.2	1.2	1.2	0.5	0.7	0.5	0.2*	0.8	1.3
Phe	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Ser	0.9	0.5*	0.7*	0.9	0.7	0.5*	0.5*	0.8	0.8	0.9	0.7	1.0	0.9
Thr	0.9	0.7	0.8	0.7	1.1	0.9	0.8	0.8	1.0	0.8	0.7	1.0	0.9
Trp	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Tyr	1.2	1.5	1.9*	1.2	1.6	1.9	1.9*	1.2	1.1	0.8	0.9	1.1	1.1
Val	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Organic Acids													
Citrate	1.2	1.7*	1.8*	1.2	1.2	1.7*	1.9*	1.2	1.0	1.0	1.1	1.0	1.0
Fumarate	0.9	0.7	0.8	0.7	1.0	1.0	0.7	0.8	1.3	1.1	1.0	1.4	1.2
GHB	1.4	1.3	0.6	1.2	0.9	1.1	0.9	0.7	0.8	1.2	0.9	0.6	1.4
Glutarate	0.6	0.3*	0.4*	1.2	0.8	0.4*	0.4*	1.4	1.3	1.0	1.0	1.3	1.1
Malate	1.0	1.0	0.9	1.1	1.1	1.0	1.0	1.2	1.1	1.0	1.1	1.0	1.0
Malonate	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Shikimate	1.1	1.0	1.0	1.1	1.2	1.1	1.1	1.3	1.2	1.1	1.1	1.1	1.0
Succinate	1.2	1.1	1.1	1.0	1.6	1.3	1.4	1.1	1.2	0.9	1.0	1.0	1.2
Threonate	1.0	1.0	0.8	1.0	0.9	1.1	0.9	1.1	1.1	1.2	1.0	1.0	1.1
Phytosterols													
Campesterol	0.8	0.8	0.9	0.9	1.2	1.2	1.1	1.2	1.4	1.0	0.9	1.1	1.1
β-Sitosterol	1.0	1.1	0.8	1.1	1.3	1.2	1.3	1.5	1.5	1.1	1.4*	0.9	1.1
Stigmasterol	0.8	0.8	0.9	0.8	1.3	1.4	1.2	1.2	1.7	1.0	0.9	1.2	1.1
Other Metabolites													
Myo-Inositol	1.0	1.0	1.1	1.1	1.1	1.1	1.2	1.2	1.2	1.0	1.1	1.1	1.1

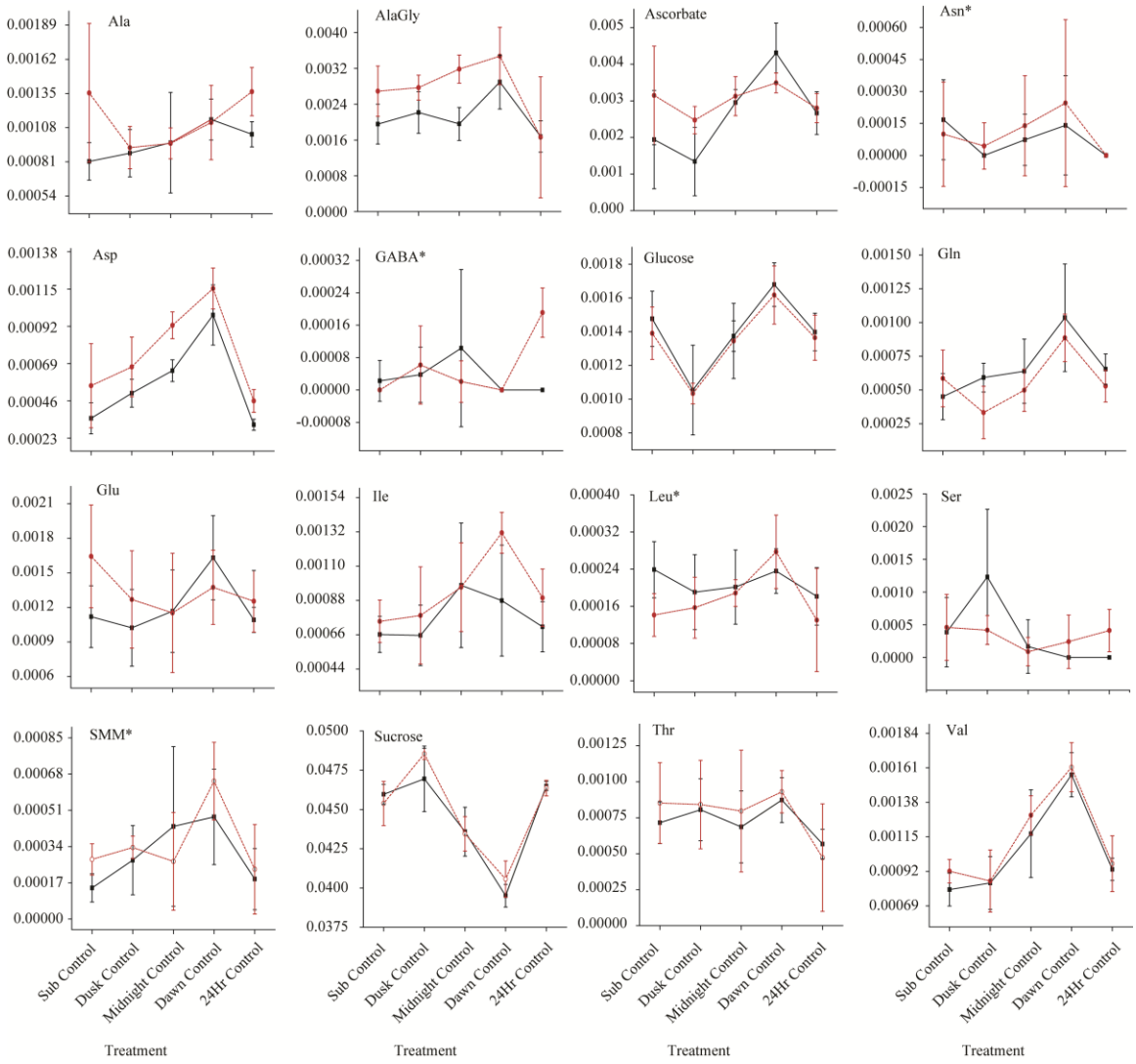


Figure 4.3. Trajectory plots of NMR data as relative normalized area for M202 (■) and M202(*Sub1*) (●). The time points are connected using black (M202) or red (M202(*Sub1*) lines. The treatments are labeled as Sub Control, Dusk Control, Midnight Control, Dawn Control, and 24Hr Control to represent control tissue sampled at each time point. Each data point is the average of at least 5 biological replicates and error bars represent the standard deviation. Metabolites below the limit of quantitation are indicated with an asterisk.

Table 4.2. A comparison of the levels of metabolites in extracts of M202 and M202(*Sub1*) controls determined by ^{1}H NMR. The comparison between varieties is represented by fold changes for each diurnal interval in the absence of submergence. All comparisons were made to the true midday time point. Asterisks indicate significant differences at the 95% confidence limit. ND represents metabolite ratios not determined due to insufficient signal for quantitation. The "C" at the end of the treatment label is used to indicate non-submerged tissue.

	M202				M202(*Sub1*)				M202/M202(*Sub1*)				
	DuskC:SubC	MidnightC:SubC	DawnC:SubC	24hrC:SubC	DuskC:SubC	MidnightC:SubC	DawnC:SubC	24hrC:SubC	SubC	DuskC	MidnightC	DawnC	24hrC
Sugars													
Glucose	0.7	0.9	1.1	0.9	0.7	1.0	1.2	1.0	1.1	1.0	1.0	1.0	1.0
Sucrose	1.0	0.9	0.9*	1.0	1.1	1.0	0.9*	1.0	1.0	1.0	1.0	1.0	1.0
Amino Acids													
Ala	1.1	1.2	1.4	1.3	0.7	0.7	0.8	1.0	0.6	1.0	1.0	1.0	0.8*
Asn	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Asp	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.8	0.7*	0.9	0.7
Glu	0.9	1.0	1.5	1.0	0.8*	0.7	0.8	0.8	0.7	0.8	1.0	1.2	0.9
Gln	1.3	1.4	2.3	1.5	0.6	0.9	1.5	0.9	0.8	1.8	1.3	1.2	1.2
Ile	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Leu	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Ser	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
SMM	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Thr	1.1	1.0	1.2	0.8	1.0	0.9	1.1	0.6	0.8	1.0	0.9	0.9	1.2
Val	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.0	0.9	1.0	1.0
Organic Acids													
Ascorbate	0.7	1.5	2.2	1.4	0.8	1.0	1.1	0.9	0.6	0.5	0.9	1.2	1.0
GABA	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Other Metabolites													
AlaGly	1.1	1.0	1.5	0.9	1.0	1.2	1.3	0.6	0.7	0.8	0.6*	0.8	1.0

With the exception of aspartate (Asp), glutamate (Glu), glutarate, glycine (Gly), lysine (Lys), and tyrosine (Tyr), metabolites were unchanged in response to the diurnal cycle. Interestingly, alanine (Ala) remained unchanged over the course of the dark and light conditions despite the presumed increase in glucose catabolism to pyruvate via glycolysis (Figure 4.3, Table 4.2), whereas Lys accumulated during the dark cycle and declined only during the light cycle. Both Ala and Lys are products of pyruvate metabolism, and in contrast to submergence stress where Ala is the preferred carbon storage metabolite (Chapters 2 and 3), Lys accumulates during the dark cycle and Ala is unchanged (Figure 4.3, Table 4.2). Similarly, Tyr and Asp accumulate in the absence of light with Tyr gradually peaking at dawn. Asp, a precursor to the TCA cycle intermediate 2-oxoglutarate, accumulated quickly between midday and dusk, remaining unchanged during the dark cycle, and then dropping by midday during the light cycle. The afternoon accumulation of Asp might be due to reduced TCA cycle activity through photosynthesis, effectively slowing glycolysis, which could affect energy production via the TCA cycle. However, it is still unclear if the TCA cycle is operates during light or dark conditions.[28] Flux studies are needed to understand the changes of Asp with respect to the TCA cycle. Citrate, a TCA cycle intermediate, also accumulated in the absence of light. Citrate, a feedback inhibitor of the TCA cycle, is the only detected TCA cycle intermediate to change significantly in response to changes in the light cycle (Figure 4.2 and Table 4.1). The accumulation of both citrate and Asp during darkness could indicate that the TCA cycle slows in the absence of photosynthesis. In contrast, levels of Gly and Ser decrease during darkness when compared to levels at midday (although only significantly for the

intolerant variety). Gly accumulation has been correlated in other plants with photorespiration but Gly can also function as a ligand in calcium signaling, which can have multiple functions in the plant stress response.[29-33] At dusk, the beginning of the dark cycle, Gly levels decrease rapidly by a 0.1 fold change compared with levels at midday and remained low until re-accumulating by midday (Figure 4.2, Table 4.1). Interestingly, Gly levels do not return to the same level as the previous midday measurement, an observation that may reflect uncontrollable greenhouse factors, such as light levels and temperature. The amino acids Ser and Thr are directly related to the metabolism of Gly but despite the pronounced changes in Gly levels, Ser and Thr levels changed more gradually, supporting the hypothesis that the changes in Gly levels are related to photorespiration.[34] Glu followed a similar trend, decreasing during the dark cycle, indicating that nitrogen accumulation might be affected by the diurnal cycle in rice.[2] Glu can be converted to 2-oxoglutarate, a TCA cycle intermediate, enabling the build-up of TCA cycle intermediates in anticipation of the imminent light cycle.[35] However, 2-oxoglutarate was not monitored during this study and further experiments are necessary to elucidate the role of Glu in nitrogen or carbon storage.

Diurnal differences between the genotypes were evident from the NMR data (Figure 4.3, Table 4.2) for Ala, AlaGly, and Asp and from the GC-MS data (Figure 4.2 and Table 4.1) for Asp and Glu. For Asp, the NMR data shows a gradual increase over the course of night, peaking at dawn and dropping at midday, with a statistically significant 0.7 fold change difference between M202 and M202(*Sub1*) occurring at midnight. Similarly, a statistically significant 0.7 fold change difference between M202

and M202(*Sub1*) was evident from the GC-MS data. In contrast to the trajectory plot generated from NMR data (Figure 4.3), GC-MS measurements of Asp showed a rapid increase at dusk, remained relatively unchanged at midnight and dawn, and subsequently decreased at midday after the resumption of photosynthesis. Despite the modest differences between the NMR and GC-MS trajectory plots, the accumulation and divergence of Asp coincides with the decrease in levels of Glu as measured by GC-MS. Because of high background and peak overlap in the NMR data in the spectral region that Asp proton resonances are detected, the GC-MS data are likely to be more reliable. Through a reversible transamination reaction, Glu is deaminated to form Asp.[36] The elevated levels of Glu and Asp in the tolerant variety compared with the intolerant variety can be attributed to nitrogen assimilation and TCA cycle intermediate production, respectively, and elevated levels of both metabolites in M202(*Sub1*) during low light levels suggest a distinction in ability to provide the precursors necessary to promote energy production.

The NMR and GC-MS trajectory plots for Ala provide no indication of genotype-dependent changes produced by the diurnal cycle. A statistically significant difference was detected at the 24 h control time point exclusively in the NMR data, with the M202 variety having 0.8-fold difference for Ala when compared with the M202(*Sub1*) variety (Table 4.2). This observation is inconsistent with the results presented in Chapters 2 and 3, where levels of Ala were determined to be statistically indistinguishable in the absence of stress (Figures 2.5 and 3.3). The reason for the genotypic differences in Ala levels in

this experiment is unclear and further experiments are needed to confirm the divergence between the two varieties during control conditions.

Differences between the genotypes for the dipeptide AlaGly during the dark cycle were also detected by NMR. AlaGly levels for the M202 variety were 0.6-fold lower than levels measured for M202(*Sub1*). As previously reported (Ch. 2), levels of AlaGly decrease in rice during extended dark cycles and subsequently recovery when re-exposed to light, linking the metabolite with photosynthesis through the experiments by Manabe and coworkers, where exogenous D-Ala was only incorporated into *D*-AlaGly during the light cycle.[37] These authors noted an increase in AlaGly content for up to 3 h after beginning the dark cycle.[37] The difference in AlaGly levels between the genotypes at the midnight time point remains unclear.

4.3.1.2 Metabolite Comparison After Re-oxygenation Following Submergence Stress

The effect of 1 d of recovery following submergence stress on the primary metabolism of the M202 and M202(*Sub1*) varieties was previously investigated by NMR and GC-MS.[3, 4] In Chapters 2 and 3, a divergence in the levels of several metabolites was observed for the tolerant and intolerant varieties during the recovery period. With the exception of glucose and sucrose, metabolite levels did not decrease as rapidly for the M202 variety compared with the M202(*Sub1*) variety (Table 2.1, Figure 2.5, Table 3.3, Figure 3.3). In both genotypes, sucrose levels completely converged by the 1 d recovery time point while glucose levels increased exclusively for the M202(*Sub1*) variety (Figure 2.3, Table 2.1). These data suggest that although sucrose levels are indistinguishable in the two genotypes within 1 d of reoxygenation, the effects on glucose metabolism and re-

assimilation of other accumulated metabolites may be differ. To better understand the effects of re-oxygenation after submergence stress, plants were desubmerged after 3 d of stress and harvested at the time of submergence, dusk, midnight, dawn, or +24 hr after desubmergence. Metabolite trajectories were generated from GC-MS measurements (Figure 4.4) and NMR measurements (Figure 4.5). The data from both GC-MS and NMR were also represented as fold changes within and between genotypes in Tables 4.3 and 4.4, respectively, with statistical significance ($P < 0.05$) represented by an asterisk.

Consistent with our previously reported results (Figures 2.5 and 3.3), 15 of the 18 detected amino acids accumulated in response to submergence, reflecting a stress response involving nitrogen assimilation and carbon catabolism (Figures 4.3 and 4.4, Tables 4.3 and 4.4). Although amino acid accumulation occurred in both genotypes, the trend was less evident for the M202(*Sub1*) variety, supporting earlier confirmation that the tolerant variety consumes less carbohydrate resources than the intolerant M202 variety.[5] During the recovery period, metabolite levels generally recovered moderately when sampled at dusk, remained unchanged when measured at the midnight and dawn time points, and continued after dawn with the resumption of photosynthesis. Despite the differences exhibited by the genotypes, some amino acids, including the branched-chain amino acids Val, Ile, and Leu as well as Gln, did not demonstrate a genotypic response comparable with what was previously reported after 3 d of submergence stress (Figures 2.5, 3.3).[3, 4] These results suggest that the submergence stress imposed in the previous and current experiments, although identical in length, may not have been of equivalent severity. This is not unexpected since treatments were performed in the greenhouse.

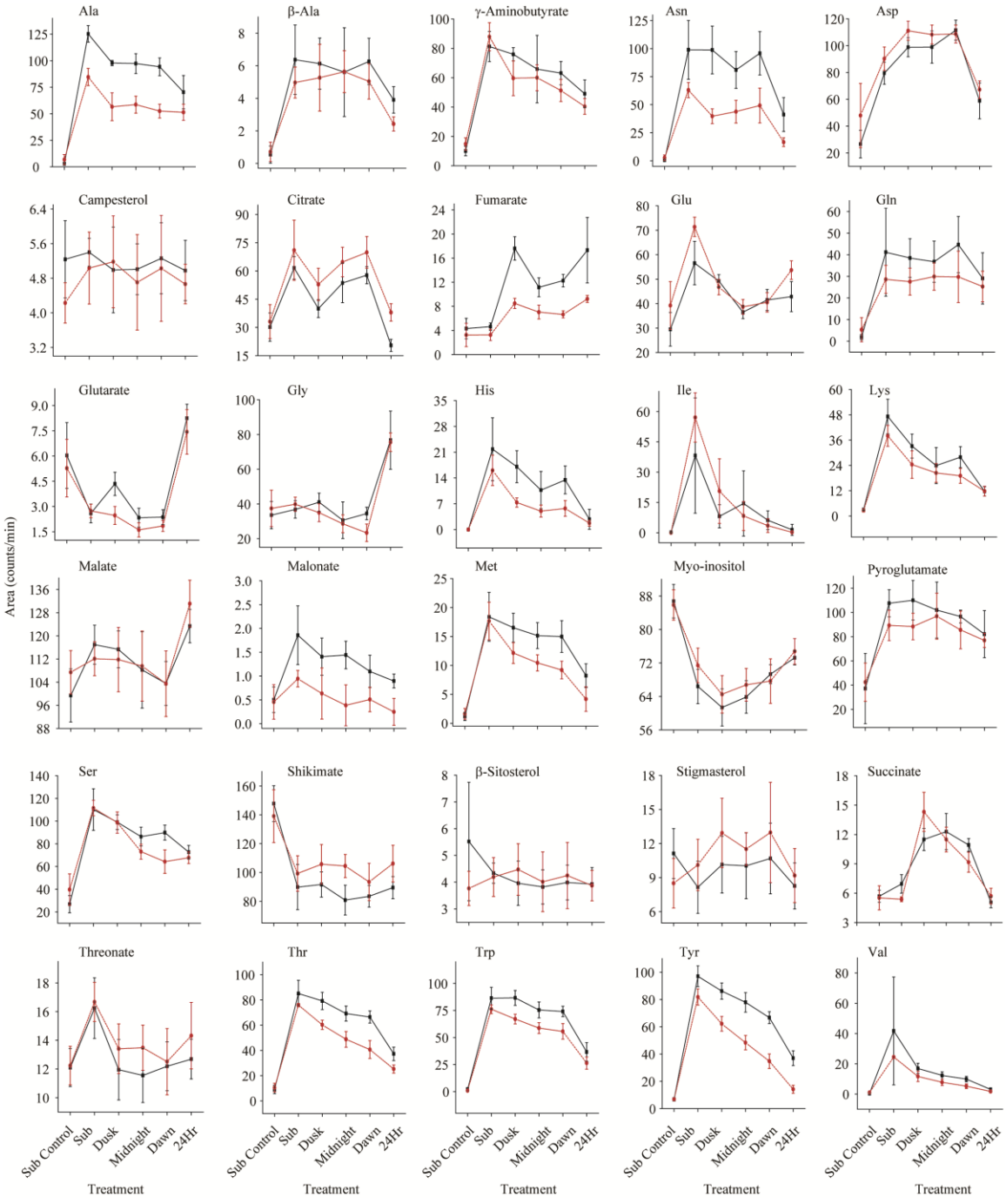


Figure 4.4. Trajectory plots from the GC-MS data representing the average area (counts/min) for M202 (■) and M202(*Sub1*) (●). The time points are connected using black solid (M202) or red dotted (M202(*Sub1*) lines. The treatments are labeled as the control (SubC), submerged (Sub), and recovery time points Dusk, Midnight, Dawn, and 24hr. Each data point is the average of at least 5 biological replicates and error bars represent standard deviation.

Table 4.3. A comparison of the levels of metabolites in M202 and M202(*Sub1*) determined by GC-MS analysis. The comparison between varieties is represented by fold changes samples collected immediately following the submergence treatment and the diurnal interval post desubmergence. The inter-genotype comparisons are made with the treated tissue and the initial control tissue. Asterisks indicate significant differences at the 95% confidence limit. ND represents metabolite ratios not determined due to insufficient signal for quantitation. The "C" at the end of the treatment label is used to indicate non-submerged tissue.

	M202					M202(*Sub1*)					M202/M202(*Sub1*)					
	Sub:SubC	Dusk:SubC	Midnight:SubC	Dawn:SubC	24hr:SubC	Sub:SubC	Dusk:SubC	Midnight:SubC	Dawn:SubC	24hr:SubC	SubC	Sub	Dusk	Midnight	Dawn	24hr
Amino Acids																
Ala	43.8*	34.2*	34.0*	32.9*	24.6*	12.4*	8.3*	8.6*	7.7*	7.5*	0.4	1.5*	1.7*	1.7*	1.8*	1.4
β-Ala	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.3	1.2	1.0	1.2	1.6*
Asn	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.6*	2.5*	1.8*	1.9*	2.5*
Asp	3.0*	3.7*	3.7*	4.2*	2.2	1.9*	2.3*	2.3*	2.3*	1.4	0.6	0.9	0.9	0.9	1.0	0.9
Glu	1.9*	1.7*	1.2	1.4	1.5	1.8*	1.2	1.0	1.0	1.4	0.8	0.8*	1.1	0.9	1.0	0.8*
Gln	21.7*	20.3*	19.4*	23.5*	15.3*	5.4*	5.2*	5.6*	5.6*	4.7*	0.4	1.4	1.4	1.2	1.5	1.1
Gly	1.1	1.2	0.9	1.0	2.3	1.1	0.9	0.8	0.6	2.0	0.9	0.9	1.2	1.1	1.5*	1.0
His	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.4	2.3*	2.1	2.3*	1.6
Ile	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.7	0.4	1.7	1.8*	6.9
Lys	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.2	1.4	1.2	1.5	1.0
Met	15.6*	14.0*	12.9*	12.7*	7.0*	10.7*	7.4*	6.3*	5.6*	2.5*	0.7	1.0	1.4*	1.5*	1.6*	2.0
Pyroglutamate	2.9*	3.0*	2.7*	2.6*	2.2*	2.1*	2.1*	2.3*	2.0*	1.8	0.9	1.2	1.2	1.1	1.1	1.1
Ser	4.1*	3.7*	3.2*	3.4*	2.7*	2.8*	2.5*	1.8*	1.6*	1.7*	0.7	1.0	1.0	1.2	1.4*	1.1
Thr	9.4*	8.7*	7.6*	7.3*	4.1*	7.2*	5.7*	4.6*	3.9*	2.4*	0.9	1.1	1.3*	1.4*	1.6*	1.5*
Trp	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.1	1.3*	1.3*	1.3*	1.4
Tyr	14.2*	12.6*	11.4*	9.8*	5.4*	11.9*	9.0*	7.0*	5.1*	2.1*	1.0	1.2*	1.4*	1.6*	1.9*	2.6*
Val	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.7	1.5	1.6	1.9*	1.8
Organic Acids																
Citrate	2.0*	1.3	1.8*	1.9*	0.7	2.1*	1.6	2.0*	2.1*	1.1	0.9	0.9	0.8	0.8	0.8	0.5*
Fumarate	1.1	4.1*	2.6*	2.8*	4.0*	1.0	2.6*	2.2*	2.1*	2.8*	1.3	1.4	2.1*	1.6*	1.8*	1.9*
GABA	8.1*	7.5*	6.5*	6.3*	4.9*	5.9*	4.0*	4.0*	3.5*	2.7*	0.7	0.9	1.3	1.1	1.2	1.2
Glutarate	0.4*	0.7	0.4*	0.4*	1.4*	0.5*	0.5*	0.3*	0.3*	1.4*	1.1	0.9	1.8*	1.5	1.3	1.1
Malate	1.2	1.2	1.1	1.0	1.2*	1.0	1.0	1.0	1.0	1.2*	0.9	1.0	1.0	1.0	1.0	0.9
Malonate	3.7*	2.8*	2.9	2.2*	1.8*	2.1	1.4	0.8	1.1	0.5	1.1	2.0*	2.2	3.7*	2.2	3.6*
Shikimate	0.6*	0.6*	0.5*	0.6*	0.6*	0.7*	0.8	0.8*	0.7*	0.8*	1.1	0.9	0.9	0.8*	0.9	0.8
Succinate	1.2	2.0*	2.2*	1.9*	0.9	1.0	2.6*	2.1*	1.7*	1.0	1.0	1.3*	0.8	1.1	1.2*	0.9
Threonate	1.3*	1.0	1.0	1.0	1.0	1.4*	1.1	1.1	1.0	1.2	1.0	1.0	0.9	0.9	1.0	0.9
Phytosterols																
Campesterol	1.0	1.0	1.0	1.0	1.0	1.2	1.2	1.1	1.2	1.1	1.2	1.1	1.0	1.1	1.0	1.1
β-Sitosterol	0.8	0.7	0.7	0.7	0.7	1.1	1.2	1.1	1.1	1.0	1.5	1.0	0.9	1.0	0.9	1.0
Stigmasterol	0.7	0.9	0.9	1.0	0.7	1.2	1.5	1.4	1.5	1.1	1.3	0.8	0.8	0.9	0.8	0.9
Other Metabolites																
Myo-Inositol	0.8*	0.7*	0.7*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.9*	1.0	0.9	1.0	1.0	1.0	1.0

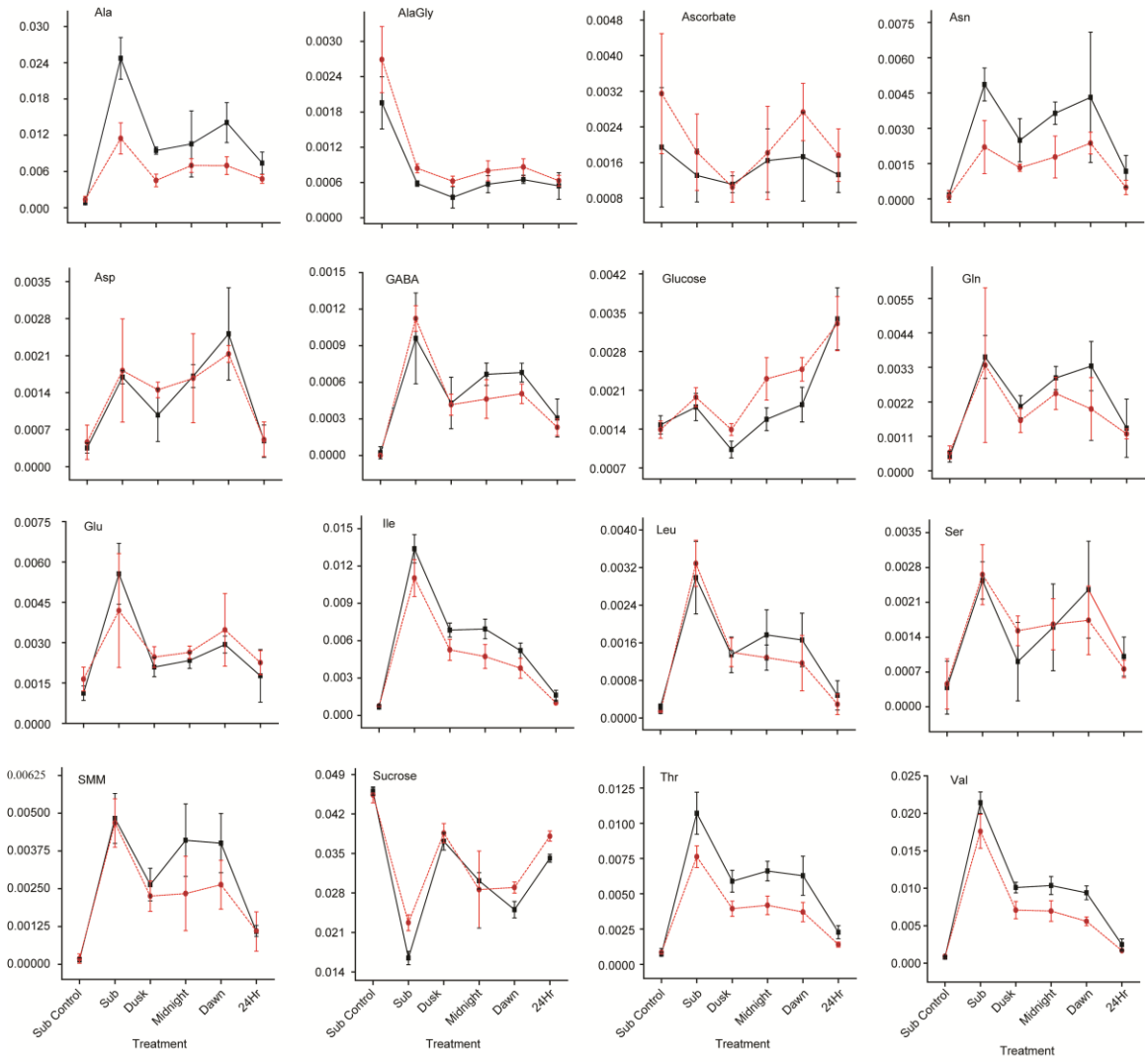


Figure 4.5. Trajectory plots from the NMR data representing the relative normalized average area for M202 (■) and M202(*Sub1*) (●) samples. The time points are connected using black (M202) or red (M202(*Sub1*) lines. The treatments are labeled as the control (SubC), submerged (Sub), and recovery time points Dusk, Midnight, Dawn, and 24hr. Each data point is the average of at least 5 biological replicates and error bars represent standard deviation.

Table 4.4. A comparison of the levels of metabolites between M202 and M202(*Sub1*) samples determined by NMR analysis. The comparison between varieties is represented by fold changes for submergence and diurnal interval post-desubmergence. The inter-genotype comparisons are made with the treated tissue and the initial control tissue. Asterisks indicate significant differences at the 95% confidence limit. ND represents metabolite ratios not determined due to insufficient signal for quantitation. The "C" at the end of the treatment label is used to indicate non-submerged tissue.

	M202					M202(*Sub1*)					M202/M202(*Sub1*)					
	Sub:SubC	Dusk:SubC	Midnight:SubC	Dawn:SubC	24hr:SubC	Sub:SubC	Dusk:SubC	Midnight:SubC	Dawn:SubC	24hr:SubC	SubC	Sub	Dusk	Midnight	Dawn	24hr
Sugars																
Glucose	1.2	0.7*	1.1	1.1	2.3*	1.4*	1.0	1.7*	1.8*	2.4*	1.1	0.9	0.7*	0.7*	0.7*	1.0
Sucrose	0.4*	0.8*	0.7*	0.5*	0.7*	0.5*	0.9*	0.6*	0.6*	0.8*	1.0	0.7*	1.0	1.1	0.9*	0.9*
Amino Acids																
Ala	30.4*	11.6*	13.0*	17.3*	9.1*	8.5*	3.3*	5.2*	5.1*	3.5*	0.6	2.2*	2.1*	1.5	2.0*	1.5
Asn	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	2.2*	1.9	2.0*	1.8	2.4
Asp	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	0.7	1.0	1.2	1.0
Glu	5.0*	1.9*	2.1*	2.6*	1.6	2.6	1.5*	1.6*	2.1	1.4	0.7	1.3	0.8	0.9	0.8	0.8
Gln	8.0*	4.5*	6.6*	7.4*	3.0	5.7	2.7*	4.2*	3.4*	2.0*	0.8	1.1	1.3	1.2	1.7	1.2
Ile	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.2	1.3*	1.5*	1.4	1.6
Leu	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	1.0	1.4	1.4	1.6
Ser	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.0	0.6	1.0	1.4	1.3
SMM	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.0	1.2	1.8	1.5	1.0
Thr	15.0*	8.2*	9.2*	8.8*	3.2*	9.0*	4.6*	4.9*	4.3*	1.6*	0.8	1.4*	1.5*	1.6*	1.7*	1.6*
Val	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.2	1.4*	1.5*	1.7*	1.5
Organic Acids																
Ascorbate	0.7	0.6	0.8	0.9	0.7	0.6	0.3	0.6	0.9	0.6	0.6	0.7*	1.1	0.9	0.6*	0.8
GABA	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.9	1.0	1.4	1.3	1.3
Other Metabolites																
AlaGly	0.3	0.2	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.2	0.7	0.7	0.6	0.7	0.8	0.9

Regardless of the physiological implications of experimental differences, both the submergence stress and re-oxygenation response in these experiments is consistent with the proposed quiescence model for carbon sequestration predominant in the tolerant variety.

Before, during, and after submergence Gly levels remained unchanged until 1 d post submergence when levels in both varieties increased by a ~1.5-fold change compared with measurements taken at the previous time points. Genotypic differences were only evident at the dawn, with levels in the M202 variety lower by a 1.5-fold change compared to the Gly levels measured in the M202(*Sub1*) variety. As discussed in Chapter 3 and section 4.3.1.1, alterations in Gly levels are most likely indicative of changes in photorespiration. As the recovery extended to the dawn time point, levels of Gly measured in the M202(*Sub1*) variety began to decrease. In context with the decrease in Gly levels during changes in the light cycle in the absence of stress (Figure 4.2, Table 4.1), the lower levels of Gly at the dawn point measured exclusively for the submergence tolerant variety might be indicative of a reacclimation to non-stressed conditions. The increase in Gly levels during the light cycle 1 d post submergence for both genotypes is inconsistent with what was observed previously (Chapter 3) further suggesting that the two stress conditions may not have been identical. More specifically, the better recovery of Gly, presumably due to photorespiration, indicates less severity of submergence stress in the diurnal experiment.

Monitoring changes in energy production as a result of submergence and desubmergence is key to understanding the metabolic reconfiguration attributed to the

presence or absence of the *SUB1A* gene. Four TCA cycle intermediates, succinate, citrate, fumarate, and malate were measured by GC-MS and changed as a result of either submergence, reoxygenation, or both. In both genotypes, citrate exclusively accumulated in both genotypes at 3 d of submergence, decreased moderately after desubmergence (at the dusk time point), and gradually re-accumulated during the dark cycle followed by a rapid decrease after dawn (Figure 4.4). The only significant difference between the two genotypes occurred after 24 h of recovery, with levels in the M202 variety measuring a 0.5-fold difference compared to M202(*Sub1*) (Table 4.3). The higher levels of citrate in the tolerant variety was opposite the trend for many of the other metabolites, suggesting that upon desubmergence the tolerant variety is better prepared to resume energy production through photosynthesis.[7] In contrast, succinate accumulates exclusively in the M202 variety during submergence, by a fold change of 1.3. By dusk after desubmergence, however, levels of succinate increase significantly in both genotypes. Although the difference is not significant, levels of succinate appear to be greater in the tolerant variety compared with the intolerant variety. From dusk to dawn, relative levels of succinate shift to higher levels measured in the intolerant variety (Table 4.3), supporting the hypothesis that the TCA cycle in the tolerant variety supports greater flux through the cycle, enabling more energy production after desubmergence during the night (i.e. when there is no photosynthesis). Similarly, fumarate does not change during submergence but accumulates during re-oxygenation in both varieties. Although fumarate does not appear to change in response to the diurnal cycle in the absence of stress (Figure 4.2, Table 4.1), the trends during re-oxygenation appear to be impacted by

the diurnal cycle. In both varieties, fumarate accumulates between desubmergence and dusk, increasing 4 and 2 fold for the intolerant and tolerant genotypes, respectively (Table 4.3). Fumarate levels subsequently decreased when measured at the midnight and dawn time points, and then increased when measured at 24 hr post desubmergence. The metabolism of fumarate is also related to Asp and Tyr.[38] As with the majority of the other measured amino acids, Tyr accumulates during submergence stress and decreases upon re-oxygenation. The decrease in Tyr coincides with the accumulation of fumarate (Figure 4.3), suggesting the degradation of Tyr to fumarate, however Tyr is also involved in protein metabolism and is the precursor of other important metabolites.[39] The amino acid Asp can also be catabolized to fumarate, but the levels of Asp are indistinct in the two genotypes. Differences in fumarate levels between the genotypes were opposite to what was observed for succinate and citrate, with fumarate content 2.1 times greater in the M202 variety compared with M202(*Sub1*), whereas succinate and citrate abundance was greater in the M202(*Sub1*) variety. The higher levels of fumarate in the intolerant variety can be correlated with the higher levels of other amino acids, including Tyr, which may permit greater production of fumarate. Fumarate may also accumulate to a greater extent in the M202 variety due to slower flux through the TCA cycle. As discussed in Ch. 1 and Ch. 3, metabolic flux experiments are required to better understand the changes to energy production due to the presence of the *SUB1A* gene during and after submergence stress, however these are beyond the scope of the current study.

Changes in Met, a metabolite responsible for sulfur transport, over the time course of the experiment were similar to the changes detected for other amino acids. Met accumulated during submergence, gradually declining during the dark cycle, followed by a rapid drop during the light cycle. Methionine levels in the two genotypes were significantly different at midnight, dawn, and 1 d post submergence, with levels in M202 ~1.5-fold higher relative to levels measured from the tolerant variety (Table 4.4). Met is a versatile metabolite, serving as a precursor to metabolites responsible for sulfur transport and storage, methylation, and a metabolic inhibitor of ethylene biosynthesis.[40-43] Approximately 80% of the flux of Met has been reported to be directed for S-adenosylmethionine (SAM) synthesis.[44] SMM, an uncommon amino acid generated from Met, was also observed to accumulate during submergence followed by a decrease when measured at dusk after de-submergence. During the dark cycle, it appears that differences between the two genotypes appear to diverge, with levels of SMM accumulating in the M202 variety and not the M202(*Sub1*) variety. These differences are not statistically significant, but the trend is similar to what is observed for Met. Because the phytohormone ethylene is responsible for the induction of the *SUB1A*, changes in Met and SMM could be important in hormone regulation.[5, 41]

Based on the GC-MS and NMR results, it is difficult to assign specific differences in the metabolic reconfiguration occurring during re-oxygenation to the presence or absence of the *SUB1A* gene. For some metabolites, a clear difference was observed between the genotypes during the course of re-oxygenation, particularly at the time points during the dark cycle. For example, shikimate, Met, and Ser diverge at either the

midnight or dawn time points following de-submergence. For other metabolites, the metabolite response appears to be indifferent between the genotypes or not affected by recovery. For example, malonate levels diverge as a result of submergence but genotypic differences remain consistent during the recovery. Similarly, genotypic differences are detected for Glu but during the recovery period, Glu levels converge. Without understanding how photorespiration or carbon and nitrogen flux differs between the genotypes and the pathways through which these processes impact metabolism, no firm conclusions can be made. Experiments by Fukao and coworkers showed that the *SUB1A* gene promotes survival when exposed to extended darkness, a common effect of flood conditions, with significant differences in starch, sucrose and fructose levels due to the presence or absence of the *SUB1A* gene after 1 d of darkness.[7] In the experiment presented herein, the plants were desubmerged at midday under ambient light in the greenhouse. This allowed the plants ~6 h of daylight to resume photosynthesis followed by a 12 h night. A longer dark cycle, for example, starting desubmergence at dusk instead of midday, might have a different physiological effect on the two genotypes. Additional experiments are necessary to further unravel the complex interaction of the *SUB1A* gene with metabolism.

4.3.2 RPIP-UPLC-MS Analysis of Phosphorylated Metabolites

The role of the *SUB1A* gene on secondary metabolism was investigated to better understand carbon allocation during and after submergence stress. Specifically, trehalose-6-phosphate (T6P) and other phosphorylated components of rice extract were profiled by RPIP-UPLC-MS. Because the phosphorylated monosaccharides

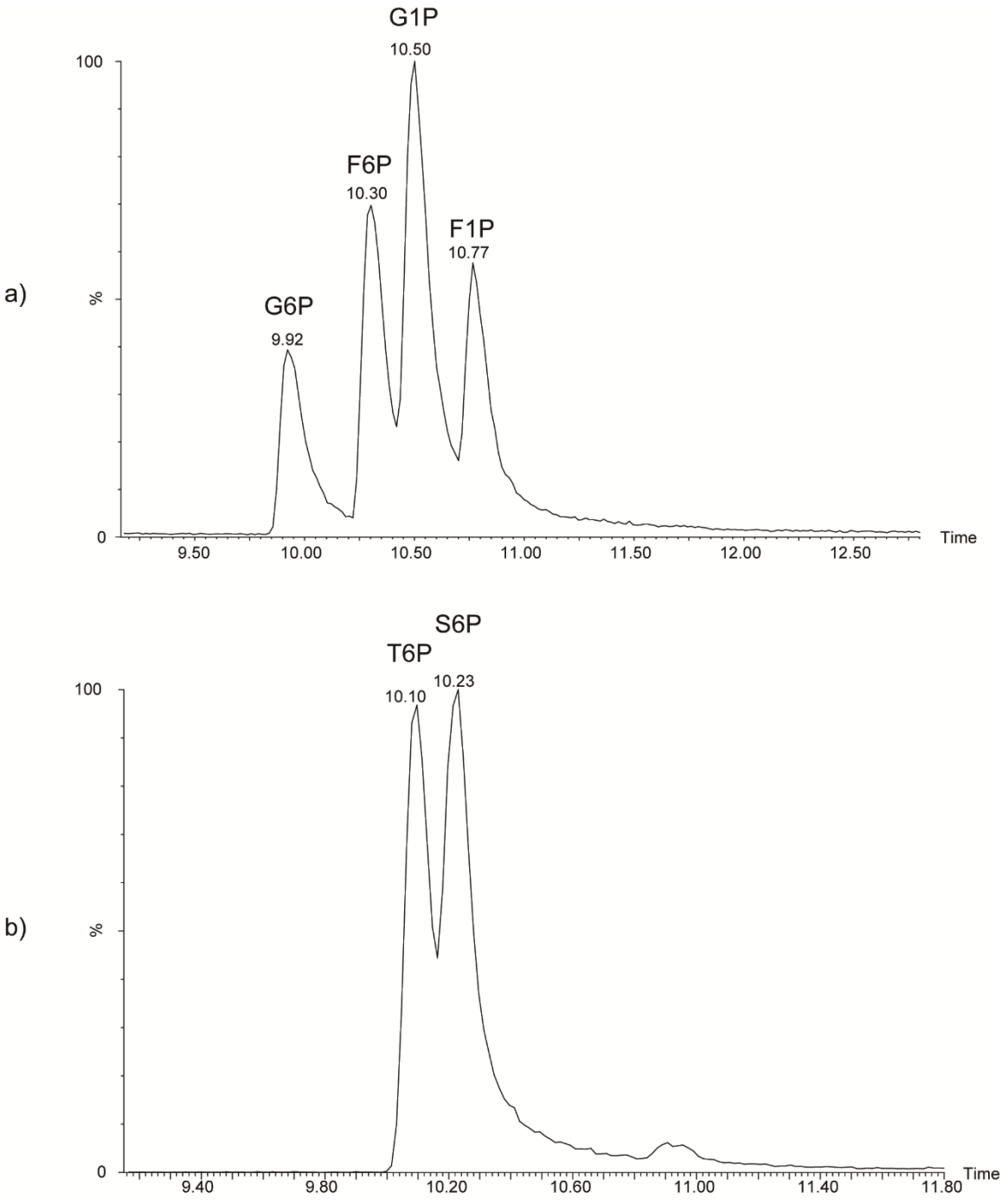


Figure 4.6. Extracted ion chromatograms obtained by RPIP-UPLC-MS from an injection of a 20 μM standard mixture of (a) phosphorylated monosaccharides G6P, F6P, G1P, and F1P and (b) phosphorylated disaccharides T6P and S6P.

glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), glucose-1-phosphate (G1P), and fructose-1-phosphate (F1P) have identical masses (*m/z* 259.14), RPIP-UPLC-MS was investigated for its ability to chromatographically resolve these isomers. Figure 4.6a shows the extracted ion chromatogram (XIC) of a standard mixture of these four phosphorylated monosaccharides. Although not baseline resolved, RPIP-UPLC separated the four phosphorylated monosaccharides, with G6P eluting first, followed by F6P, G1P, and finally F1P.

In this separation, an HSS T3 column (Waters Corp.) was used to separate the phosphorylated mono- and disaccharides because the column is more compatible with polar analytes. For analyte elution, 10 mM DBA was used in both the aqueous and organic mobile phases as an IPR because of the success others have had with low molecular weight compounds. Water was the aqueous mobile phase and methanol was used as the organic mobile phase and the pH adjusted to a pH meter reading of 7.4 and 8, respectively, after addition of DBA.[18] The mechanism driving the separation is not entirely clear. There are trends in the elution order but multiple mechanisms can be responsible for elution and a more detailed study is needed to explore the complex interactions encountered in RPIP.[17] The phosphorylated disaccharides T6P and Sucrose-6-phosphate (S6P) were reported by Delatte and coworkers to be detectable in Arabidopsis metabolite extracts.[14] Because these compounds have the same molecular weight (422.27 Da), a standard mixture was injected to determine their elution order. Figure 4.6b shows the separation of T6P and S6P. Although they do not baseline resolve,

T6P and S6P are sufficiently separated to allow independent identification and quantitation.

4.3.2.1 Metabolite Profiling of Phosphorylated Metabolites by RPIP-UPLC-MS.

Figure 4.7a shows a representative TIC taken from a metabolite extract of M202 shoot tissue at the midday control time point. Despite sample clean-up and targeted extraction of anionic compounds, the chromatogram is complex and a majority of the components could not be assigned based on mass spectra alone. Targeted metabolite profiling of phosphorylated compounds resulted in the identification and quantitation of six metabolites by RPIP-UPLC-MS, including T6P, S6P, G6P, adenosine-5'-monophosphate (AMP), 6-phosphogluconate (6PG), and 3-phosphoglycerate (3PG). Figure 4.7b and 4.7c show the XIC of *m/z* 421.27 and 259.14, respectively, with the identified peaks labeled. The phosphorylated disaccharides T6P and S6P are readily detected in rice tissue, in addition to a third peak eluting after S6P for which the identity is unknown. The same mass and similar elution time as T6P and S6P indicates that the third peak might be another phosphorylated disaccharide, such as maltose-6-phosphate or maltose-1-phosphate, both of which are products of starch and maltose catabolism, however standards are difficult to obtain or not available.[45] Lactose-1-phosphate (L1P) is another phosphorylated disaccharide with the same mass as T6P and S6P, however experiments with a standard determined that L1P is not the identity of the third peak (data not shown). Further experiments are needed to confirm the identity of the third peak.

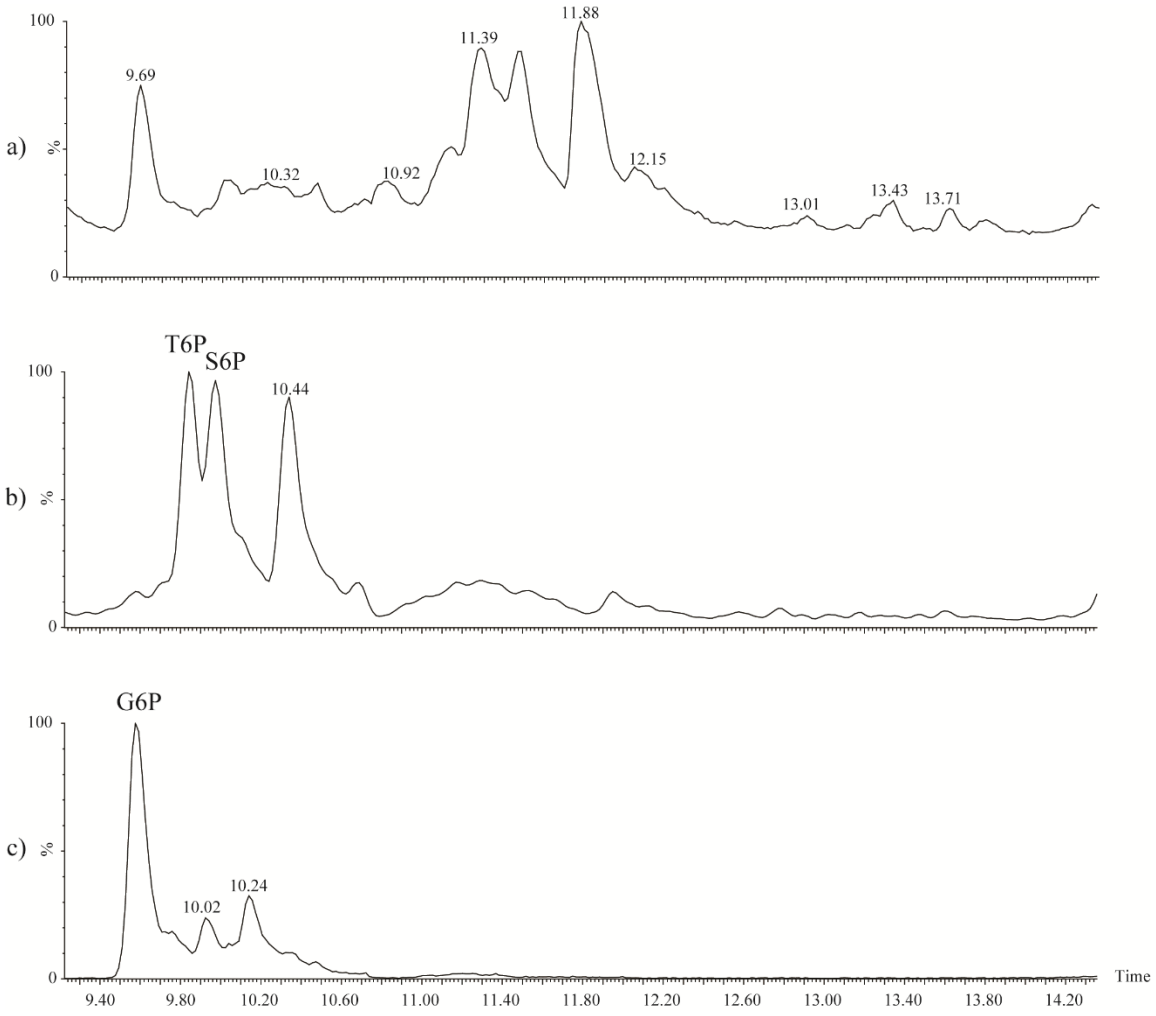


Figure 4.7. (a) Representative RPIP-UPLC-MS chromatogram of a 20 μL injection of a metabolite extract from M202 control shoot tissue. (b) Extracted ion chromatogram of *m/z* 421.27 showing T6P, S6P, and an unknown peak at retention time 10.44 min. (c) Extracted ion chromatogram of *m/z* 259.14 showing G6P and two minor peaks that could not accurately be identified.

The purpose of these experiments was to determine if T6P was differentially regulated in the shoot tissue of submergence tolerant and intolerant rice varieties. Figure 4.8 shows the trajectory plots taken from control tissue sampled at midday, dusk, midnight, dawn, or midday + 24hr. The relative T6P levels in the controls were statistically indistinguishable in the two genotypes, indicating that the presence of *SUB1A* did not affect T6P levels during the course of normal dark and light cycle (Figure 4.8, Table 4.5). For both genotypes, however, T6P levels increased between dusk and midnight, declined between midnight and dawn, recovering between dawn and noon, when compared with the initial midday time point. There was a fold increase of ~1.5 at dawn compared with the initial midday control (Table 4.5). One of the roles of T6P in plants is to act as a signaling molecule for carbon sensing. When sugar reserves are elevated, T6P levels accumulate and promote growth through the inhibition of SnRK1.[8, 46] Although sucrose levels are at their lowest point in the diurnal cycle, glucose levels increase during the dark cycle as sucrose is catabolized, reaching an apex at the dawn time point concomitantly with T6P (Figures 4.3 and 4.8, Table 4.2). Though G6P levels remain unchanged (Figure 4.8), elevated levels of glucose would enable flux through G6P and uridine diphosphate glucose (UDPG) producing T6P.[10] However, the increase in T6P immediately before dawn conflicts with a report by Wahl and coworkers showing T6P levels increasing and decreasing during the day/night cycle (respectively), suggesting that T6P changes in direct response to sucrose levels.[9] Further investigation is required to fully understand why the T6P trend we observe in rice is opposite to the results published by Wahl and coworkers in Arabidopsis. The response of AMP, 6PG,

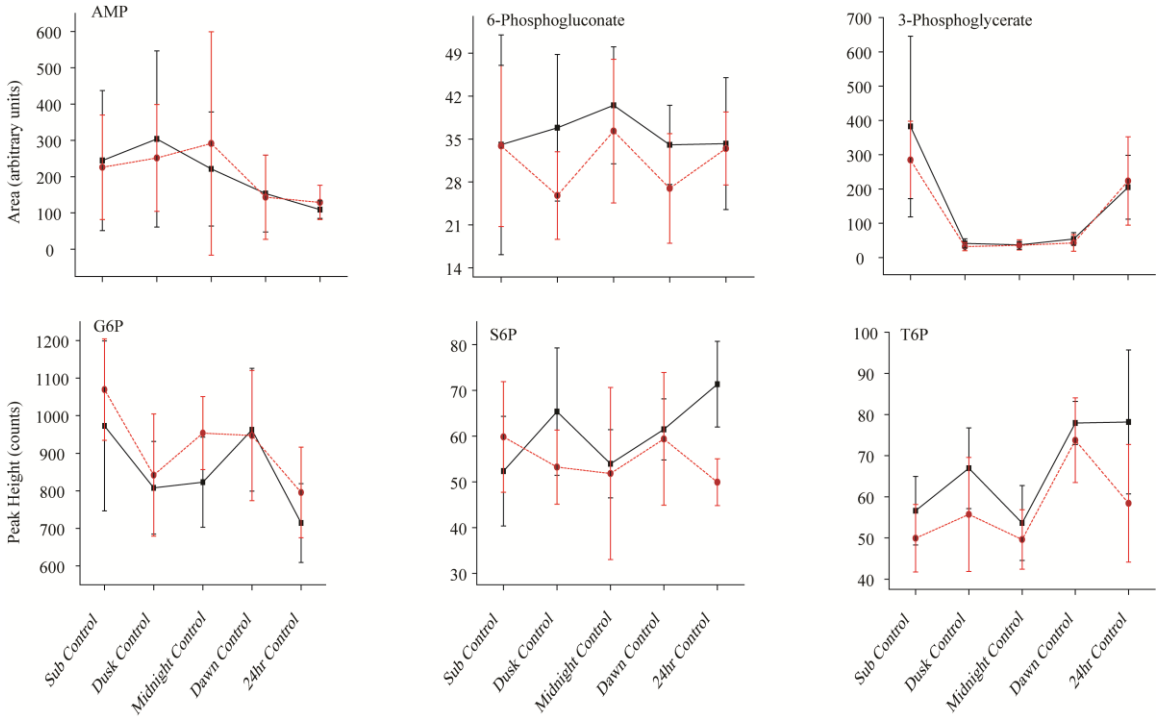


Figure 4.8. Trajectory plots from RPIP-UPLC-MS data representing either area (arbitrary units) or peak height (counts) for M202 (■) and M202(*Sub1*) (●). The time points are connected using black (M202) or red (M202(*Sub1*) lines. The treatments are labeled as Sub Control, Dusk Control, Midnight Control, Dawn Control, and 24Hr Control to represent control tissue at each time point. Each data point is the replicate of at least 5 biological replicates and error bars represent the standard deviation.

Table 4.5. A comparison of the levels of metabolites in M202 and M202(*Sub1*) controls determined by RPIP-UPLC-MS analysis. The comparison between varieties is represented by fold changes for each diurnal interval in the absence of submergence. Asterisks indicate significant differences at the 95% confidence limit. ND represents metabolite ratios not determined due to insufficient signal for quantitation. The "C" at the end of the treatment label is used to indicate non-submerged tissue.

	M202				M202(*Sub1*)				M202/M202(*Sub1*)				
	DuskC:SubC	MidnightC:SubC	DawnC:SubC	24hrC:SubC	DuskC:SubC	MidnightC:SubC	DawnC:SubC	24hrC:SubC	SubC	DuskC	MidnightC	DawnC	24hrC
T6P	1.2	0.9	1.4*	1.4	1.1	1.0	1.5*	1.2	1.1	1.2	1.1	1.1	1.3
S6P	1.2	1.0	1.2	1.4	0.9	0.9	1.0	0.8	0.9	1.2	1.0	1.0	1.4*
G6P	0.8	0.8	1.0	0.7	0.8	0.9	0.9	0.7	0.9	1.0	0.9	1.0	0.9
6-PG	1.2	1.0	1.0	0.8	1.1	0.8	1.0	1.0	1.4	1.1	1.3	1.0	1.1
3-PG	0.1*	0.1*	0.1*	0.5	0.1*	0.1*	0.2*	0.8	1.3	1.0	1.3	0.9	0.1
AMP	1.0	0.9	1.3	1.1	0.8	0.8	0.9	0.9	1.2	0.8	1.1	0.8	1.2

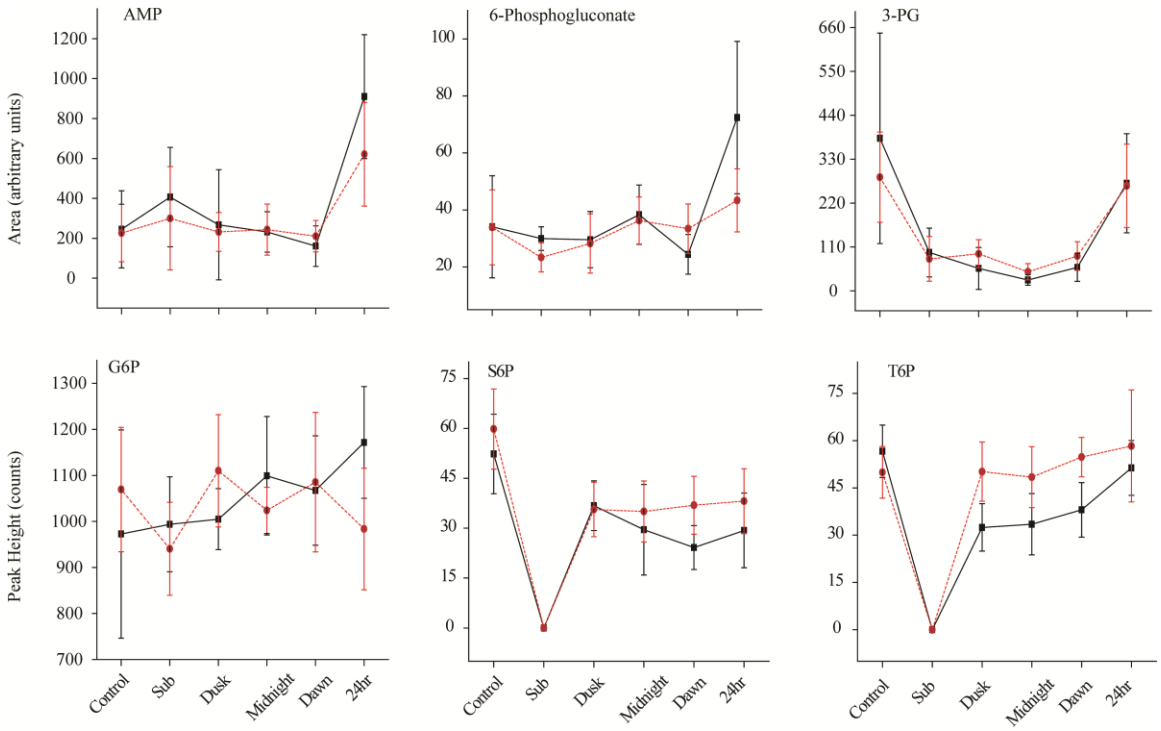


Figure 4.9. Trajectory plots from RPIP-UPLC-MS representing either area (arbitrary units) or peak height (counts) for M202 (■) and M202(*Sub1*) (●). The time points are connected using black (M202) or red (M202(*Sub1*)) lines. The treatments are labeled as the control (SubC), submerged (Sub), and recovery time points Dusk, Midnight, Dawn, and 24hr. Each data point is the replicate of at least 5 biological replicates and error bars represent the standard deviation.

Table 4.6 A comparison of the levels of metabolites in M202 and M202(*Sub1*) plants subjected to submergence as determined by RPIP-UPLC-MS analysis. The comparison between varieties is represented by fold changes for submergence and diurnal interval post desubmergence. Asterisks indicate significant differences at the 95% confidence limit. ND represents metabolite ratios not determined due to insufficient signal for quantitation. The "C" at the end of the treatment label is used to indicate non-submerged tissue.

	M202					M202(*Sub1*)					M202/M202(*Sub1*)					
	Sub:SubC	Dusk:SubC	Midnight:SubC	Dawn:SubC	24hr:SubC	Sub:SubC	Dusk:SubC	Midnight:SubC	Dawn:SubC	24hr:SubC	SubC	Sub	Dusk	Midnight	Dawn	24hr
T6P	ND	0.6*	0.6*	0.7*	0.9	ND	1	1	1.1	1.2	1.1	ND	0.6	0.7	0.7*	0.9
S6P	ND	0.7	0.6	0.5*	0.6*	ND	0.6	0.6	0.6	0.6	0.9	ND	1	0.8	0.7	0.8
G6P	1	1	1.1	1.1	1.2	0.9	1	1	1	0.9	0.9	1.1	0.9	1.1	1	1.2
6-PG	0.9	0.9	1.1	0.7	2.1	0.7	0.8	1.1	1.0	1.3	1.0	1.3	1.0	1.1	0.7	1.7
3-PG	0.3	0.1	0.1*	0.2*	0.7	0.3*	0.3*	0.2*	0.3*	0.9	1.3	1.2	0.6	0.6	0.7	1.0
AMP	1.7	1.1	0.9	0.7	3.7*	1.0	1.0	1.1	0.9	2.7	1.1	1.4	1.2	1.0	0.8	1.5

and S6P are largely unchanged by the diurnal cycle. Interestingly, S6P accumulates at the midday +24 h control time point in the intolerant variety, indicating that the M202 variety may experience higher levels of starch catabolism in response to re-exposure to light. Levels of S6P in the M202 variety increase by a fold change of 1.5 compared with those in the M202(*Sub1*) variety (Table 4.5). These differences are intriguing, as the same accumulation is not noted for the initial midday control despite the similar experiment conditions. Other experiments are needed to confirm and better resolve these findings. The metabolite 3-PG is a byproduct of glycolysis and carbon fixation.[47] Lower levels in both genotypes occurred at night during the absence of photosynthesis followed by a subsequent accumulation during the light cycle.

The effect of re-oxygenation on T6P levels was also investigated. Figure 4.9 shows the trajectory plots and Table 4.6 reports the fold changes between and within genotypes for T6P, S6P, G6P, 3PG, 6PG, and AMP before, during, and at several time points after submergence stress. At 3 d of submergence, levels of T6P decrease below the limit of detection for both genotypes, consistent with the upregulation of starch catabolism promoted by this stress.[5] By dusk after de-submergence, levels of T6P accumulate back to pre-stress conditions for the M202(*Sub1*) variety but to only 60% of their initial value for the M202 variety (Table 4.6). The difference between the two genotypes in the reaccumulation of T6P can be related to the relative levels of glucose. At the dusk, midnight, and dawn post de-submergence, glucose levels in the two genotypes diverged, with the intolerant variety decreasing by 0.7-fold compared to levels in the tolerant variety (Table 4.4). Similarly, the T6P levels are lower in the intolerant

variety by a fold change of 0.6 and 0.7 for the dusk and dawn time points, respectively. These data suggest that by producing more T6P and thus inhibiting the sucrose non-fermenting kinases (SnRK1) the intolerant variety is better able to return to normoxic conditions, including anabolism and normal plant growth.[8, 10, 46] An alternative explanation is that T6P levels return more quickly to near-normal levels in the tolerant variety due to better recovery of photosynthesis. This might be because photosynthesis upon desubmergence is better in the M202(*Sub1*) variety, however further experiments are needed. Another factor for consideration is that *SUB1A* has also been shown to influence flowering time when constitutively expressed in rice and Arabidopsis under non-stressed conditions.[48, 49] Similarly, T6P was shown by Wahl and coworkers to regulate flowering in Arabidopsis.[9] As a result, we can postulate that higher levels of T6P in *SUB1A* containing rice might also influence or regulate flowering time. It is unclear, however, if the differential regulation of T6P is a direct or indirect result of the *SUB1A* gene, with T6P levels increasing and decreasing in response to glucose levels or other stimuli.

The other metabolites quantified by RPIP-UPLC-MS were not differentially regulated, however there were trends correlated with submergence stress and recovery. Both AMP and 6PG accumulated after re-oxygenation only at after 1 d after de-submergence (Table 4.6). The metabolite 3PG responded similarly in both control and submergence conditions due to the cessation of photosynthesis during dark conditions. Interestingly, S6P had similar trends as T6P, decreasing below the limit of detection during submergence and increasing at the dusk time point, although at lower levels than

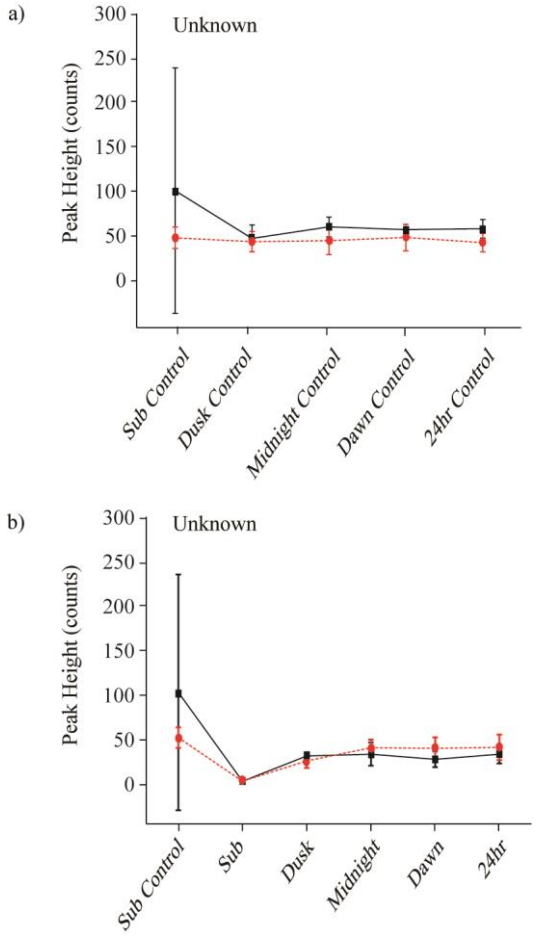


Figure 4.10. Trajectory plots from RPIP-UPLC-MS representing peak height (counts) for M202 (■) and M202(*Sub1*) (●). The time points are connected using black (M202) or red (M202(*Sub1*)) lines. (a) Measurements from the diurnal experiment with treatments labeled as Sub Control, Dusk Control, Midnight Control, Dawn Control or 24hr Control. (b) Measurements after submergence and recovery with treatments labeled as the control (Sub Control), submerged (Sub), and recovery time points Dusk, Midnight, Dawn, and 24hr. Each data point is the replicate of at least 5 biological replicates and error bars represent the standard deviation.

was measured at the control time point by a fold change of ~0.6 for both genotypes. A longer submergence may cause metabolite levels to diverge between the two genotypes.

The trajectory plots for the unknown peak from Figure 4.7b are shown in Figure 4.10. During the diurnal cycle, the levels of the unknown are largely indifferent, indicating that it is not strongly affected by photosynthesis (Figure 4.10a). In response to submergence, however, levels of the unknown dropped below the limit of detection and subsequently increased during the recovery period (Figure 4.10b). In both cases, levels of the unknown metabolite were not different between the genotypes. Although the metabolite was not identified, the decrease in levels of the metabolite during submergence correlates with the trends observed for both T6P and S6P (Figure 4.9). The trajectories for these three components are in stark contrast to most of the other detected metabolites which are unaffected by submergence (Figures 4.4, 4.5 and 4.9), further suggesting that the unknown metabolite may be a phosphorylated disaccharide. Figure 4.10 illustrates the power of metabolomics for biomarker discovery in that interesting trends in metabolites can be observed, even when their identity is not known a priori. This allows the investigator to focus efforts on the identification of those components that give a metabolic signature that is responsive to the experimental design. Although, the unknown compound was not identified in this work, future experiments to isolate the component in sufficient quantity for NMR characterization could provide new insights into the metabolic response of rice to submergence stress.

4.4 Conclusions

The work presented in this chapter further supports the benefits of combining NMR, GC-MS, and UPLC-MS to extend our understanding of the metabolic reconfiguration of *SUB1A* containing rice in response to submergence stress and the diurnal cycle. Although this study suggests that under normal light/dark cycles there is very little differentiation between the two genotypes, the day/night cycle does have an effect on the recovery of the two genotypes. Sucrose, for example, was different only during submergence and at the dawn and after 24 h of recovery. Differences in glucose metabolism were only apparent during levels of low light, with levels appreciably higher in the M202(*Sub1*) variety compared with levels in M202. The plant metabolite profiles reflect changes due to both desubmergence and diurnal cycle, however genotypic differences during the recovery period did not compare with those distinguished in Chapter 3, perhaps due to the differences in the experimental conditions. Additionally, targeting other classes of metabolites, such as flavonoids, is necessary to for a more comprehensive analysis secondary metabolism in response to oxidative stress upon reoxygenation. Further investigation will enhance our understanding of the physiological changes associated with submergence, re-oxygenation, and circadian rhythms that may be influenced by the *SUB1A* gene.

The quantitation of T6P and other phosphorylated secondary metabolites provided a deeper understanding of the role of sugar sensing in submergence survival. Higher T6P levels detected for the tolerant variety during desubmergence is further evidence of better carbon management in plants containing the *SUB1A* gene compared with the M202

variety. Additionally, the differences in T6P levels during re-oxygenation can be correlated with the flowering differences observed between *SUB1A* and non-*SUB1A* containing rice varieties.[50] Understanding the correlation of T6P with rice floral meristem development and ultimately flowering after submergence stress is not yet well understood. Additionally, a better separation of T6P and S6P is required for more confidence in the integration. Continued investigation of the molecular mechanisms involving T6P is necessary to better understand the differences experienced by *SUB1A* and non-*SUB1A* containing rice.

In chapter 5, the novel chemometric technique VIZR is introduced to interrogate the metabolic differences in urine ^{1}H NMR spectra due to the ingestion of non-dietary substances, such as ibuprofen, energy drinks, and alcohol. This program allows individual samples to be discriminated from a library of control urine samples and heat maps, projections, and a scoring approach will be used to identify spectral regions responsible for discrimination. Although urine samples were selected for the development and evaluation of VIZR, we anticipate that this approach can be extended to plant metabolite profiling studies such as those described in this chapter.

4.5 References

(1) Farre, E. M.; Weise, S. E. The interactions between the circadian clock and primary metabolism. *Curr. Opin. Plant Biol.* **2012,** *15*, (3), 293-300.

(2) Wang, Z. Z.; Wang, T. Dynamic proteomic analysis reveals diurnal homeostasis of key pathways in rice leaves. *Proteomics* **2011,** *11*, (2), 225-238.

(3) Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K. Comparison of GC-MS and NMR for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12*, (2), 898-909.

(4) Barding, G. A.; Fukao, T.; Beni, S.; Bailey-Serres, J.; Larive, C. K. Differential metabolic regulation governed by the rice *SUB1*A gene during submergence stress and identification of alanylglycine by ^{1}H NMR spectroscopy. *J. Proteome Res.* **2012,** *11*, 320-330.

(5) Fukao, T.; Xu, K. N.; Ronald, P. C.; Bailey-Serres, J. A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* **2006,** *18*, (8), 2021-2034.

(6) Fukao, T.; Yeung, E.; Bailey-Serres, J. The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell* **2011,** *23*, (1), 412-427.

(7) Fukao, T.; Yeung, E.; Bailey-Serres, J. The submergence tolerance gene SUB1A delays leaf senescence under prolonged darkness through hormonal regulation in rice. *Plant Physiol.* **2012,** *160*, (4), 1795-1807.

(8) Schluepmann, H.; Berke, L.; Sanchez-Perez, G. F. Metabolism control over growth: a case for trehalose-6-phosphate in plants. *J. Exp. Bot.* **2011,** *63*, (9), 3379-3390.

(9) Wahl, V.; Ponnu, J.; Schlereth, A.; Arrivault, S. p.; Langenecker, T.; Franke, A.; Feil, R.; Lunn, J. E.; Stitt, M.; Schmid, M. Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science* **2013,** *339*, (6120), 704-707.

(10) Delatte, T. L.; Sedijani, P.; Kondou, Y.; Matsui, M.; de Jong, G. J.; Somsen, G. W.; Wiese-Klinkenberg, A.; Primavesi, L. F.; Paul, M. J.; Schluepmann, H. Growth Arrest by Trehalose-6-Phosphate: An Astonishing Case of Primary Metabolite Control over Growth by Way of the SnRK1 Signaling Pathway. *Plant Physiol.* **2011,** *157*, (1), 160-174.

(11) Wingler, A.; Delatte, T. L.; O'Hara, L. E.; Primavesi, L. F.; Jhurreea, D.; Paul, M. J.; Schluepmann, H. Trehalose 6-Phosphate Is Required for the Onset of Leaf Senescence Associated with High Carbon Availability. *Plant Physiol. 158*, (3), 1241-1251.

(12) Eastmond, P. J.; van Dijken, A. J. H.; Spielman, M.; Kerr, A.; Tissier, A. F.; Dickinson, H. G.; Jones, J. D. G.; Smeekens, S. C.; Graham, I. A. Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. *Plant J.* **2002,** *29*, (2), 225-235.

(13) Torano, J. S.; Delatte, T. L.; Schluepmann, H.; Smeekens, S. C. M.; de Jong, G. J.; Somsen, G. W. Determination of trehalose-6-phosphate in Arabidopsis thaliana seedlings by hydrophilic-interaction liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2012,** *403*, (5), 1353-1360.

(14) Delatte, T. L.; Selman, M. H. J.; Schluepmann, H.; Somsen, G. W.; Smeekens, S. C. M.; de Jong, G. J. Determination of trehalose-6-phosphate in Arabidopsis seedlings by successive extractions followed by anion exchange chromatography-mass spectrometry. *Anal. Biochem.* **2009,** *389*, (1), 12-17.

(15) Spagou, K.; Wilson, I. D.; Masson, P.; Theodoridis, G.; Raikos, N.; Coen, M.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Nicholson, J. K.; Want, E. J. HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies. *Anal. Chem.* **2011,** *83*, (1), 382-390.

(16) Jones, C. J.; Beni, S.; Larive, C. K. Understanding the Effect of the Counterion on the Reverse-Phase Ion-Pair High-Performance Liquid Chromatography (RPIP-HPLC) Resolution of Heparin-Related Saccharide Anomers. *Anal. Chem.* **2011,** *83*, (17), 6762-6769.

(17) Jones, C. J.; Membreno, N.; Larive, C. K. Insights into the mechanism of separation of heparin and heparan sulfate disaccharides by reverse-phase ion-pair chromatography. *J. Chromatogr. A* **2010,** *1217*, (4), 479-488.

(18) Langeslay, D. J.; Urso, E.; Gardini, C.; Naggi, A.; Torri, G.; Larive, C. K. Reversed-phase ion-pair ultra-high-performance-liquid chromatography-mass spectrometry for fingerprinting low-molecular-weight heparins. *J. Chromatogr. A* **2013**.

(19) Korir, A. K.; Larive, C. K. Advances in the separation, sensitive detection, and characterization of heparin and heparan sulfate. *Anal. Bioanal. Chem.* **2009,** *393*, (1), 155-169.

(20) Korir, A. K.; Lirntiaco, J. F. K.; Gutierrez, S. M.; Larive, C. K. Ultraperformance ion-pair liquid chromatography coupled to electrospray time-of-flight mass spectrometry for compositional profiling and quantification of heparin and heparan sulfate. *Anal. Chem.* **2008,** *80*, (4), 1297-1306.

(21) Limtiaco, J. F. K.; Beni, S.; Jones, C. J.; Langeslay, D. J.; Larive, C. K. The efficient structure elucidation of minor components in heparin digests using microcoil NMR. *Carbohydr. Res.* **2011,** *346*, (14), 2244-2254.

(22) Glasoe, P. K.; Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **1960,** *64*, (1), 188-190.

(23) Team, R. C. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2012.

(24) Espinoza, C.; Degenkolbe, T.; Caldana, C.; Zuther, E.; Leisse, A.; Willmitzer, L.; Hincha, D. K.; Hannah, M. A. Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in arabidopsis. *Plos One* **2010,** *5*, (11).

(25) Gibon, Y.; Usadel, B.; Blaesing, O. E.; Kamlage, B.; Hoehne, M.; Trethewey, R.; Stitt, M. Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. *Genome Biol* **2006,** *7*, (8).

(26) Fukushima, A.; Kusano, M.; Nakamichi, N.; Kobayashi, M.; Hayashi, N.; Sakakibara, H.; Mizuno, T.; Saito, K. Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, (17), 7251-7256.

(27) Rolland, F.; Baena-Gonzalez, E.; Sheen, J., Sugar sensing and signaling in plants: Conserved and novel mechanisms. In *Annu. Rev. Plant Biol.*, Annual Reviews: Palo Alto, 2006; Vol. 57, pp 675-709.

(28) Araujo, W. L.; Nunes-Nesi, A.; Nikoloski, Z.; Sweetlove, L. J.; Fernie, A. R. Metabolic control and regulation of the tricarboxylic acid cycle in photosynthetic and heterotrophic plant tissues. *Plant Cell Environ* **2012,** *35*, (1), 1-21.

(29) Bailey-Serres, J.; Chang, R. Sensing and signalling in response to oxygen deprivation in plants and other organisms. *Ann. Bot.* **2005,** *96*, (4), 507-518.

(30) De Silva, D. L. R.; Hetherington, A. M.; Mansfield, T. A. Synergism between calcium ions and abscisic acid in preventing stomatal opening. *New Phytol.* **1985,** *100*, (4), 473-482.

(31) Dubos, C.; Huggins, D.; Grant, G. H.; Knight, M. R.; Campbell, M. M. A role for glycine in the gating of plant NMDA-like receptors. *Plant J.* **2003,** *35*, (6), 800-810.

(32) Martinelli, T.; Whittaker, A.; Bochicchio, A.; Vazzana, C.; Suzuki, A.; Masclaux-Daubresse, C. Amino acid pattern and glutamate metabolism during dehydration stress in the 'resurrection' plant Sporobolus stapfianus: a comparison between desiccation-sensitive and desiccation-tolerant leaves. *J. Exp. Bot.* **2007,** *58*, (11), 3037-3046.

(33) Subbaiah, C. C.; Sachs, M. M. Molecular and cellular adaptations of maize to flooding stress. *Ann. Bot.* **2003,** *91*, (2), 119-127.

(34) Goto, S.; Bono, H.; Ogata, H.; Fujibuchi, W.; Nishioka, T.; Sato, K.; Kanehisa, M. Organizing and computing metabolic pathway data in terms of binary relations. *Pacific Symposium on Biocomputing '97* **1997**, 175-186186.

(35) Forde, B. G.; Lea, P. J. Glutamate in plants: metabolism, regulation, and signalling. *J. Exp. Bot.* **2007,** *58*, (9), 2339-2358.

(36) Kirsch, J. F.; Eichele, G.; Ford, G. C.; Vincent, M. G.; Jansonius, J. N.; Gehring, H.; Christen, P. Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure. *J. Mol. Biol.* **1984,** *174*, (3), 497-525.

(37) Manabe, H.; Ohira, K. Effect of light irradiation on the D-alanylglycine content in rice leaf blades. *Plant Cell Physiol.* **1983,** *24*, (6), 1137-1142.

(38) Miflin, B. J.; Lea, P. J. Amino acid metabolism. *Ann. Rev. Plant Physio.* **1977,** *28*, 615 pp.-615 pp.

(39) Castelluccio, C.; Paganga, G.; Melikian, N.; Paul Bolwell, G.; Pridham, J.; Sampson, J.; Rice-Evans, C. Antioxidant potential of intermediates in phenylpropanoid metabolism in higher plants. *FEBS Lett.* **1995,** *368*, (1), 188-192.

(40) Bourgis, F.; Roje, S.; Nuccio, M. L.; Fisher, D. B.; Tarczynski, M. C.; Li, C. J.; Herschbach, C.; Rennenberg, H.; Pimenta, M. J.; Shen, T. L.; Gage, D. A.; Hanson, A. D. S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell* **1999,** *11*, (8), 1485-1497.

(41) Ko, S.; Eliot, A. C.; Kirsch, J. F. S-Methylmethionine is both a substrate and an inactivator of 1-aminocyclopropane-1-carboxylate synthase. *Arch. Biochem. Biophys.* **2004,** *421*, (1), 85-90.

(42) Menegus, F.; Lilliu, I.; Brambilla, I.; Bonfa, M.; Scaglioni, L. Unusual accumulation of S-methylmethionine in aerobic-etiolated and in anoxic rice seedlings: An H-1-NMR study. *J. Plant Physiol.* **2004,** *161*, (6), 725-732.

(43) Ravanel, S.; Gakiere, B.; Job, D.; Douce, R. The specific features of methionine biosynthesis and metabolism in plants. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95*, (13), 7805-7812.

(44) Joshi, V.; Joung, J.-G.; Fei, Z.; Jander, G. Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. *Amino Acids* **2010,** *39*, (4), 933-947.

(45) Ritte, G.; Heydenreich, M.; Mahlow, S.; Haebel, S.; Kotting, O.; Steup, M. Phosphorylation of C6- and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. *FEBS Lett.* **2006,** *580*, (20), 4872-4876.

(46) Nunes, C.; Primavesi, L. F.; Patel, M. K.; Martinez-Barajas, E.; Powers, S. J.; Sagar, R.; Fevereiro, P. S.; Davis, B. G.; Paul, M. J. Inhibition of SnRK1 by metabolites: Tissue-dependent effects and cooperative inhibition by glucose 1-phosphate in combination with trehalose 6-phosphate. *Plant Physiol. Biochem.* **2012,** *63*, 89-98.

(47) Bonacci, W.; Teng, P. K.; Afonso, B.; Niederholtmeyer, H.; Grob, P.; Silver, P. A.; Savage, D. F. Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences* **2012,** *109*, (2), 478-483.

(48) Fukao, T.; Bailey-Serres, J. Submergence tolerance conferred by Sub1A is mediated by SLR1 and SLRL1 restriction of gibberellin responses in rice. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, (43), 16814-16819.

(49) Peña-Castro, J. M.; van Zanten, M.; Lee, S. C.; Patel, M. R.; Voesenek, L. A. J. C.; Fukao, T.; Bailey-Serres, J. Expression of rice SUB1A and SUB1C transcription factors in Arabidopsis uncovers flowering inhibition as a submergence tolerance mechanism. *The Plant Journal* **2011,** *67*, (3), 434-446.

(50) Pena-Castro, J. M.; van Zanten, M.; Lee, S. C.; Patel, M. R.; Voesenek, L. A. J. C.; Fukao, T.; Bailey-Serres, J. Expression of rice SUB1A and SUB1C transcription factors in Arabidopsis uncovers flowering inhibition as a submergence tolerance mechanism. *Plant J.* **2011,** *67*, (3), 434-446.

CHAPTER FIVE

VIZR - An Automated Chemometric Technique for the Metabolic Discrimination of Biofluids

Acknowledgements: I would like to thank the following people for their contributions to this research: Daniel J. Orr, who started the development of this program and was vital to its progress and Sumukh M. Sathnur who's input and contributions to VIZR were instrumental to making it a useful and publication-quality program.

Abstract:

This chapter describes a new chemometric technique, VIZR (Visual Interpretation of *Z*-Score Ratios). VIZR is written in the open source code R and can be used to identify metabolic differences between individual biosamples and a control group. To demonstrate the capabilities of VIZR, 50 urine samples were collected from healthy volunteers; 41 samples were collected randomly following a normal dietary routine and 8 test samples were collected after dietary supplementation with ibuprofen, alcoholic beverages, or an energy drink. A ninth test sample was prepared by 50% dilution of a control sample. Sample analysis was conducted by ^{1}H NMR spectroscopy and the collected data was subjected to VIZR analysis, which successfully discriminated each of the 9 test samples from the 41 control samples. In addition, VIZR analysis revealed the NMR spectral regions responsible for the disparity between the individual test samples and the control group. The self-normalizing nature of the VIZR calculation provides a robust analysis independent of dilution effects, which is especially important in urine

analyses. Potential applications of VIZR include high-throughput data analysis for toxicological profiling, disease diagnosis, and biomarker identification in any type of biosample for which a control dataset can be established. Although demonstrated herein for the statistical analysis of ^{1}H NMR data, the VIZR program is platform independent and could be applied to digitized metabolic datasets acquired using other techniques including hyphenated mass spectrometry measurements.

5.1 Introduction

The metabolic analysis of complex biosamples, such as tissues and body fluids, through various analytical methods is a growing area of interest due to the desire to correlate alterations in metabolism with disease, toxicology and substance abuse.[1-4] Urine and blood plasma or serum are particularly attractive for metabolite analysis because they can be obtained simply, in high volume through minimally or noninvasive means.[5, 6] Urine has the additional advantages of being molecularly stable and provides a relatively simple matrix, with minimal levels of protein in most patients.[7, 8] As a result, urine is widely used for clinical analyses and has been used for metabolite profiling of maple syrup urine syndrome,[5] propionate metabolism disorders,[9] inflammatory bowel disease, toxicological profiling studies,[10-12] and biomarker discovery.[13-15]

The concentrations of metabolites in biofluids, especially urine, can fluctuate significantly based on the hydration of the subject, complicating spectral interpretation and correlation of metabolite levels between subjects. To accommodate the influence of dilution, a variety of normalization techniques have been implemented. Sum

normalization, a popular approach, is performed by dividing the spectrum into bins (but can also be accomplished using individual data points), integrating each bin, and summing the integrals of the bins (excluding bins containing solvents and impurities).[8, 16] The integrals of interest are then normalized to the sum, accounting for dilution differences between samples. A disadvantage of sum normalization is that biologically-induced changes in the levels of the most abundant metabolites can influence the normalization constant derived from summing the binned integrals and artificially impact the normalized values of less abundant components, even those unaffected by the treatment.[17] Another approach normalizes the levels of metabolites in urine samples to creatinine, however changes in creatinine metabolism can be indicative of kidney disease making creatinine normalization unreliable.[16, 18] To avoid bias introduced by these normalization techniques, Dieterle and coworkers introduced the probabilistic quotient normalization (PQN) method which uses the median of calculated quotients generated through comparison to a reference spectrum.[17] The most frequent quotient is used as the dilution factor. PQN has been widely implemented and shown to be a robust technique for dilution correction however it still requires the use of a reference spectrum to adjust the integrated values of the spectrum under interrogation.

Another challenge in metabolomics analyses is the statistical differentiation of samples from control and non-control groups. The typical metabolomics approach utilizes multivariate analysis (MVA) for evaluating differences between sample groups. Principal components analysis (PCA) is a popular unsupervised MVA method that identifies patterns within a dataset without *a priori* knowledge of the identity of the

individual samples.[19] Although useful for identifying global differences within a dataset, statistical significance cannot be deduced based on groupings in the PCA scores plot. Similarly, partial least squares analysis (PLS) is a supervised MVA technique that is used to determine the differences between predefined sample groups, however it also does not provide statistical differentiation.[20, 21] Untargeted, high-throughput toxicity profiling, disease diagnosis, or biomarker discovery requires the interrogation of a sample (or samples) from a single individual against a set of samples from a healthy population. Both PCA and PLS are unsuitable for determining differences within a dataset when there is a low sample size for the unhealthy individual, and analysis using these methods can lead to random correlations.[22] Artificial neural networks (ANN) have also been used for biomarker identification and disease diagnosis.[23-25] However, ANN uses a training set of samples representing both the healthy and unhealthy patients prior to sample interrogation, requiring replicates of both datasets to develop the neural networks. In the absence of well-defined differences between the patient populations, i.e., in experiments aimed at biomarker discovery, designing an accurate training set for the neural networks to classify the sample is difficult at best.

To better facilitate the high-throughput analysis of ^{1}H-NMR spectra of biofluids for toxicology studies, biomarker discovery and disease diagnosis, we have developed a novel chemometric method written in the open source code R for data normalization, statistical interpretation, and analysis of biofluid data sets. Termed VIZR (Visual Interpretation of *Z*-Score Ratios), this technique incorporates self-normalization and subsequent *z*-score analysis to interrogate differences between a single sample and a

reference population. Previous use of *z*-score analysis by Charlton and coworkers was described as an aid in detecting contaminated soft drinks using NMR.[26] Their approach used *z*-score analysis for feature selection prior to MVA, reducing the dimensionality of the data in question to more accurately determine the features that are different. They were able to accurately determine the presence of contaminants in soft drinks using the combination of feature selection using *z*-scores and MVA, demonstrating the potential of *z*-score analysis for component discrimination in complex mixtures.

Data analysis with VIZR is performed independently of MVA using a *z*-score and can confidently discriminate samples and the corresponding NMR spectral regions that set them apart from a control group. A similar approach, ratio analysis NMR spectroscopy (RANSY), introduced by Wei and coworkers identifies chemically-related resonances in a complex spectrum through ratios of peak heights or integrals and their coefficient of variation (CV).[27] RANSY calculates a pseudo NMR spectrum that contains only peaks that are highly correlated, yielding individual metabolite spectra. The RANSY technique is useful for the targeted identification of a resonance of interest in an NMR spectrum, but unlike VIZR does not distinguish differences between a single sample and a control group. The practical application of VIZR for analysis of ^{1}H NMR spectral data is demonstrated by investigating 50 urine samples (41 control samples, and 9 test samples) collected from 4 male and 3 female volunteers.

5.2 Materials and Methods

5.2.1 Materials and Reagents

Deuterium oxide was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). Sodium-3-trimethylsilyl-propanesulfonic acid-d_6 (DSS) was purchased from Isotec (St. Louis, MO). Monobasic and dibasic sodium phosphate and sterile specimen cups (90 mL) were purchased from Fisher Scientific (Pittsburgh, PA). Sodium azide was obtained from Sigma-Aldrich (St. Louis, MO).

5.2.2 Sample Collection and Preparation

To minimize pH-induced chemical shift changes, a pH 7.3 phosphate buffer similar to that described by Beckonert et al. was prepared using 200 mM Na_2HPO_4 and 36 mM NaH_2PO_4 in 25% D_2O, which provided the deuterium lock signal.[7] DSS and sodium azide were added to the buffer to achieve final concentrations of 1 mM.[7] Human urine sample collection was conducted in accordance with UC Riverside's Office of Research Integrity and the protocol was approved by the Human Research Review Board (HS 12-086). The health of the volunteers was based on their own admission. Volunteers were requested to report the use of pharmaceuticals (i.e., pain killers), energy drinks and alcohol throughout the course of the study and test samples were so indicated by the subjects. A 670 µL aliquot of urine was transferred immediately after collection to a 1.5 mL micro-centrifuge tube along with 330 µL of buffer and stored at -80°C until analysis.

5.2.3 ^{1}H NMR Analyses

^{1}H NMR spectra were measured using a Bruker Avance 14.1 T NMR spectrometer equipped with a 5 mm inverse broadband probe with xyz gradients tuned to

599.69 MHz. The magnetic field homogeneity was optimized using up to 28 shims and the probe was manually tuned and matched. Solvent suppression was accomplished using excitation sculpting using the Bruker-defined pulse program zgesp with the transmitter set on the water resonance.[28] The sample temperature was maintained at 298 K, locked using D_2O, and spectra acquired without spinning using digital quadrature detection (DQD). The line-width of DSS for all spectra was less than 1 Hz prior to apodization. Free induction decays were collected into 32 768 points. A spectral width of 11.67 ppm was used with an 11.0 µs 90° pulse. The acquisition of 16 dummy scans preceded the co-addition of 256 transients with a relaxation delay of 1.5 s for a 17.7 min experiment time.

5.2.4 NMR Data Processing

Spectra were processed using the ACDlabs Spectrus Processor (Advanced Chemistry Development, Inc. Toronto, CAN). Spectra were apodized with an exponential function equivalent to 0.5 Hz line broadening, zero-filled to 65 536 points, and following Fourier transformation referenced to 0 ppm using the resonance of DSS. The processed spectra were manually phased and subjected to automated baseline correction using a 2^{nd} order polynomial followed by minor manual adjustment using the manual baseline correction feature. The residual water resonance, urea resonance, and regions below 0.5 ppm and above 8.5 ppm were set as dark regions (regions to be excluded from analysis) to avoid the inclusion of non-quantifiable resonances such as urea and residual water, and buffer components. After bucketing at 0.02 ppm intervals

the processed spectra were exported as *.txt files containing peak intensity with the corresponding chemical shift.

5.2.5 Statistical Analyses using PCA and VIZR

Data processing was carried out using the freely available program R (version 2.15.1, Vienna, Austria) using a personal computer. PCA was conducted using the 'prcomp' function from the 'stats' package on the un-normalized binned data with and without the bins containing the creatinine resonances. PCA was also conducted after removal of the creatinine resonances on sum normalized and creatinine normalized binned data. VIZR is a package built in-house that utilizes a variety of functions found within the 'miscTools' package as well as functions in the R basic installation.[29, 30] The VIZR package is freely available by request to the authors. Statistical analyses (both VIZR and PCA) were conducted blindly to ensure an unbiased evaluation and data labels were added only after the statistical analysis was complete. For these experiments, samples from volunteers of both genders were included together in the control group. When the samples were treated independently by gender, VIZR analysis was unable to distinguish between male or female control samples (data not shown).

5.3 Results and Discussion

This chapter describes a chemometrics approach, VIZR, written in the open source code R, which uses the *z*-score for high-throughput discrimination of samples compared to a control group. The utility of this approach is demonstrated by analysis of a set of 50 urine samples. The VIZR package is designed to statistically differentiate a sample from a population based on discrete regions of the NMR spectrum, but it could

also be used for the analysis of other data formats, for example IR or fluorescence spectra, or chromatographic results. For statistical analysis, the data is divided into two groups: a reference library generated from a control group and a sample or test group of spectra, which are individually compared to the reference library. The VIZR package relies on preprocessed data, requiring each NMR spectrum to be identically zero filled, referenced, baseline/phase corrected, and integrated prior to VIZR analysis. For these experiments, data integration was performed through binning using equidistant bin lengths of 0.02 ppm. A variety of other integration techniques can be used (e.g. intelligent binning, peak fitting, etc.) providing that the data submitted for VIZR analysis is of equal dimensions and all datasets are properly aligned so that they have corresponding chemical shift values.

5.3.1 Data Normalization

Prior to statistical analysis, it is common for data to be normalized to account for differences in dilution. The VIZR method is unique in that it accounts for dilution by creating a ratio of each bin with every other bin, resulting in a two-dimensional square matrix of length n such that:

$$\begin{matrix} \frac{x_n}{x_1} & \cdots & \frac{x_1}{x_1} \\ \vdots & & \vdots \\ \frac{x_n}{x_n} & \cdots & \frac{x_n}{x_1} \end{matrix} \qquad \text{Eq. 1}$$

where (x_1 ... x_n) are the integrated areas of the corresponding bins (as represented by chemical shifts) in ascending order. These ratios establish a relationship between the area of one bin with each of the other bins, eliminating dilution effects.

5.3.2 Statistical Analysis and Generation of *Z*-score Matrices and Heat Maps

Each dataset is represented as a two-dimensional square matrix of length *n* in which every cell within the matrix is treated uniquely, meaning the subsequent calculations are performed on the values in each cell of the matrix. Once the data has been normalized, the population mean and standard deviation are calculated from the control group (or reference library) matrices. Finally, individual *z*-score matrices are obtained for each of the test group samples by subtracting the average matrix from each sample matrix and then dividing that value by the standard deviation matrix as shown in Eq. 2 [31]

$$z = \frac{x - \mu}{\sigma} \qquad \text{Eq. 2}$$

where *x* represents the matrix for the sample being interrogated, μ represents the population average matrix, σ is the matrix containing the population standard deviations for each cell in *x*, and the elements of the matrix *z* correspond to differences, in terms of the number of standard deviations, of each element of the sample matrix from the corresponding population mean. Interpretation of the matrices is simplified through the generation of heat maps based on the absolute value of the *z*-score allowing a visual interpretation of the data. The threshold and upper limit for coloration of the heat maps is user defined so that only significant differences between datasets, as arbitrarily defined

by the user, are highlighted in the VIZR plots. Once the threshold is defined, the intervals for coloration are calculated and defined by the VIZR package for the heat map display. For the experiments reported herein, a lower threshold of 13 was chosen such that the heat maps of the control group samples are mostly blank while statistically significant differences in the test samples are highlighted.

5.3.3 Generation of *Z*-score Projection Spectra and the *Z*-score Scatter Plot

To simplify the data presentation, each of the matrices can be converted into a pseudo one-dimensional NMR spectrum, termed a *z*-score projection. This is accomplished through summation of each row of the *z*-score matrix, creating a one-dimensional plot with the *y*-axis representing the summed *z*-score value for a given chemical shift along the *x*-axis. The *z*-score projection enables the visual evaluation of the spectral differences between a test sample and the control group average.

A summary of the entire dataset is generated by the *z*-score scatter plot. Data reduction is accomplished through summation of the absolute values of the *z*-score projection and the results are plotted as the total *z* value of each sample in a scatter plot. e average and standard deviation of the total *z* value of the control group is determined and 1σ, 2σ, and 4σ lines plotted horizontally to represent the number of standard deviations the test samples are from the control group mean.

A total of 50 urine samples were collected by 4 male and 3 female volunteers. Each sample was given a unique identifier and the volunteers recorded the designation of the samples as belonging to the control or test groups. The samples were analyzed blind,

and the control and test labels added only after the spectral measurements and statistical analyses were completed. The test group contains 8 urine samples designated by the subjects as resulting from ingestion of the following non-dietary substances in addition to their normal diet during the period prior to sample collection: ibuprofen (doses of 800 mg were taken by a male subject, 600 mg by one female subject, and two separate 400 mg doses by another female subject), alcoholic beverages (two pints consumed by a male subject and one pint by a female subject), or a low carbohydrate AMP^{TM} energy drink consumed by one male. The alcoholic beverages were ingested during a meal. The control group contains 41 urine samples collected at random intervals by the volunteers without the ingestion of these non-dietary substances, as reported by the individual. To evaluate the influence of sample dilution on the VIZR analysis, one sample from the control group was diluted by 50% and included as a ninth test group sample. PCA and VIZR were each evaluated to determine whether these methods could identify the effects of ibuprofen, alcoholic beverages or energy drinks on the metabolic profile of urine samples compared with the controls.

5.3.4 PCA of Urine Samples

The unsupervised MVA of the urine samples was conducted using PCA. PCA was selected because it is a useful tool in evaluating the global differences (or similarities) between samples within a dataset and is widely applied in metabolomics and metabolic profiling investigations.[32, 33] Figure 5.1 shows the PCA results for 50 urine samples with the bins containing the creatinine resonances removed. Without data normalization the scores plot (Figure 5.1a) shows poor grouping of the control and test

samples, indicating that PCA is unable to clearly distinguish differences between the groups. Additionally, the natural variation found within the male and female control samples obscures the differences in the test samples. This lack of discrimination by PCA is further emphasized by the loadings plot (Figure 5.1b), which describes the contribution of each bin to the separation of the dataset. The separation along PC1, representing 76% of the variance, can be attributed to bins containing the intense NMR resonances of citrate and components due to the energy drink and ibuprofen while the separation along PC2, representing 8% of the variance, is primarily due to bins corresponding to NMR resonances resulting from ingestion of ibuprofen and the energy drink (Figure 5.1b). With the exception of citrate, the metabolites identified by the loadings plots are primarily associated with the test group, even though PCA does not effectively segregate the test samples from the controls. Bins due to additional metabolites may contribute to the dispersion of the samples in the PCA scores plot (Figure 5.1a) but are difficult to identify due to crowding in the loadings plot (Figure 5.1b).

As illustrated in Figure 5.2, inclusion of the bins containing the creatinine peaks in the unnormalized data produced a different sample distribution in the PCA scores plot (Figure 5.2a), however, an improved grouping of the test samples is not obtained. The bins corresponding to the creatinine resonances dominate the loadings plot (Figure 5.2b) obscuring the effects of other variables. Because data normalization is known to have a profound impact on MVA results, we evaluated the effects of normalizing the data to the bins containing the creatinine resonances, a commonly employed strategy in urine metabolomics studies.[18, 16] Normalizing the dataset to creatinine (Figure 5.3) did not

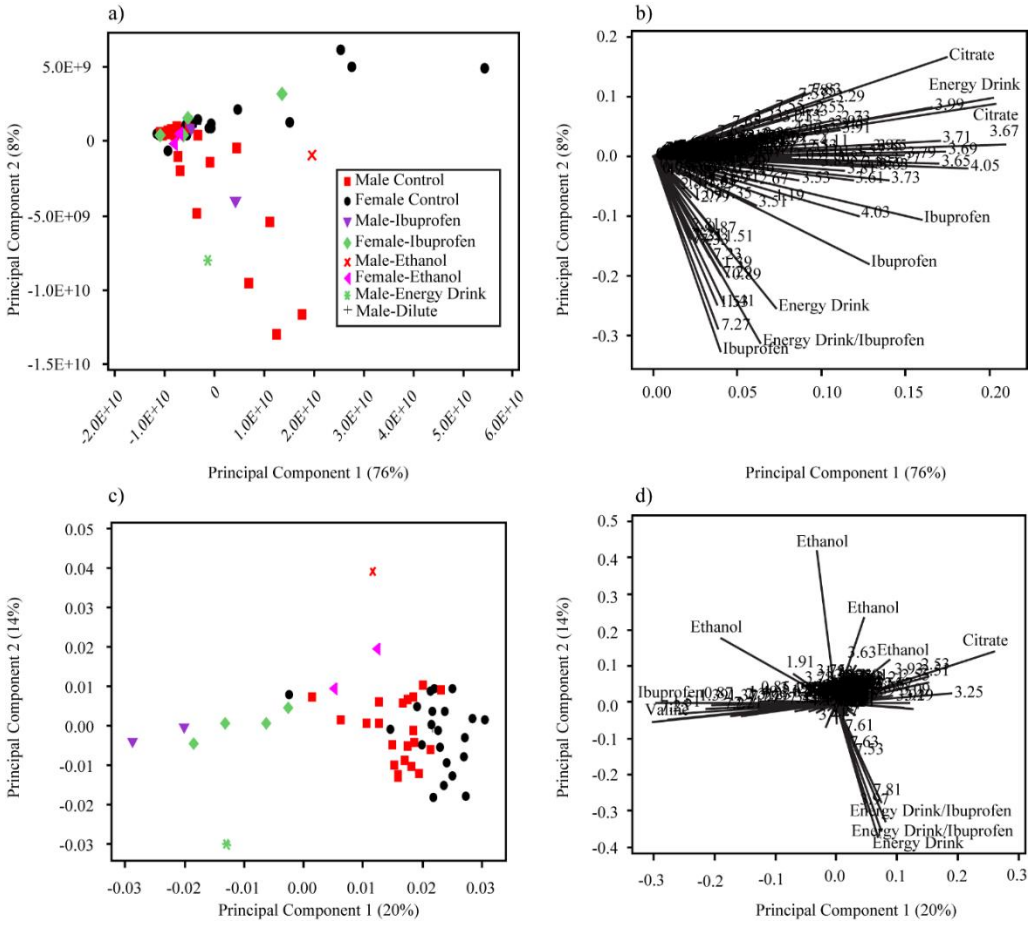


Figure 5.1. Scores and loadings plots from PCA of 50 different urine samples, 41 of which are control samples collected at random intervals after following a regular diet (as reported by the volunteers), 8 are test samples collected after dietary supplementation of ibuprofen, ethanol, or an energy drink, and a 9th test sample was generated by 50% dilution of a control sample. (a) Scores and (b) loadings plots from PCA performed on unnormalized data with the integral bins corresponding to creatinine removed. (c) Scores and (d) loadings plots from PCA performed on sum normalized data with the integral bins corresponding to creatinine removed. The legend in (c) is the same as that used in (a).

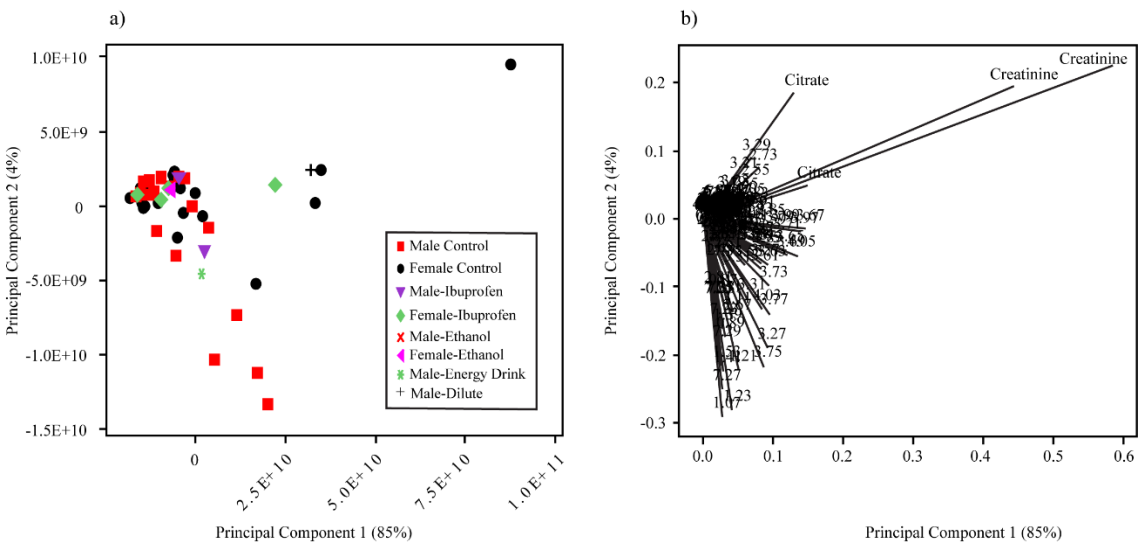


Figure 5.2. (a) Scores and (b) loadings plots from PCA of 50 different urine samples, 41 of which are control samples collected at random intervals after following a regular diet (as reported by the volunteers), 8 are test samples collected after dietary supplementation of ibuprofen, ethanol, or energy drinks, and a 9th test sample was generated by 50% dilution of a control sample. PCA was performed on the unnormalized data and the integral bins corresponding to creatinine are included.

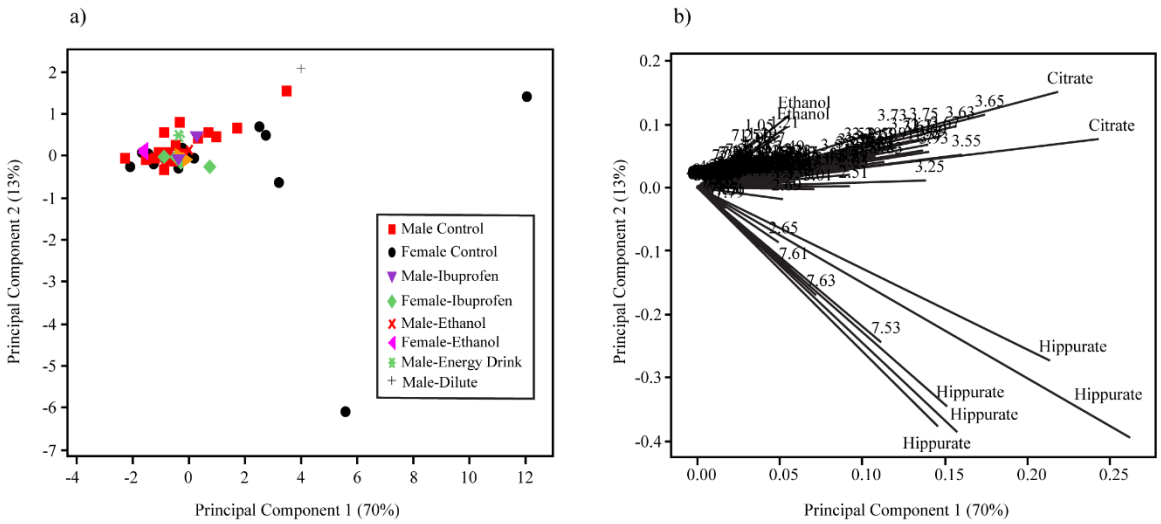


Figure 5.3. (a) Scores and (b) loadings plots from PCA of 50 different urine samples, 41 of which are control samples collected at random intervals after following a regular diet (as reported by the volunteers), 8 are test samples collected after dietary supplementation of ibuprofen, ethanol, or energy drinks, and a 9th test sample was generated by 50% dilution of a control sample. PCA was performed on creatinine normalized data and the integral buckets corresponding to creatinine are not included.

improve sample segregation. The scores plot (Figure 5.3a) shows overlap of most of the control and test samples with two outliers from the female control group while the loadings plot (Figure 5.3b) shows that citrate and hippurate were primarily responsiblefor sample dispersion. Although creatinine normalization is often used to compensate fordilution of urine samples, the results in Figure 5.3 suggest that it is an ineffective normalization approach for our data set.

Sum normalization of the data set was also examined and using this normalization method the test and control samples clustered separately in the PCA scores plot (Figure 5.1c). Additionally, the loadings plot (Figure 5.1d) highlights the spectral regions specific to the test samples responsible for the segregation of the scores plot. As shown in Figure 5.1, sum normalization prior to PCA can aid in sample discrimination, but this method also suffers from several drawbacks. Even for the sum normalized data, it was necessary to remove the bins containing the creatinine resonances prior to PCA to achieve segregation in the scores plot and to avoid having the loadings plot dominated by the bins containing the creatinine resonances (data not shown). Variation in creatinine concentration can be indicative of kidney health[18] and removal of these resonances reduces the effectiveness of PCA as a statistical platform for untargeted disease diagnosis. The need to remove the bins corresponding to the creatinine resonances to better discriminate the sample populations demonstrates influence that the most intense resonances can exert on the PCA results. A similar problem may be encountered in data sets that include diabetic patients whose urine samples often have high glucose levels that can mask peaks of less abundant metabolites.[34] In addition, in a large study designed for

biomarker discovery the identity of the test samples are not known and choosing a sample normalization method based on data segregation has obvious pitfalls. The comparatively small sample size of the test group may also have contributed to the poor discrimination in the PCA scores plot in Figure1a. As suggested by Broadhurst and Kell, PLS analysis is also expected to suffer from underrepresentation by the test group and lead to unreliable correlations due to the small sample group size.[22] Because of the limitations posed by MVA methods, VIZR was developed to efficiently discriminate samples that differ significantly from a representative population and identify the NMR spectral regions responsible for sample discrimination.

5.3.5 Analysis Using VIZR Heat Maps

VIZR analysis compares each sample to a control group and data interpretation is not dependent on *a priori* knowledge of the classification of the test sample. Figure 5.4 compares the VIZR heat maps generated for representative urine samples collected after ingestion of ethanol (Figure 5.4a) and ibuprofen (Figure 5.4b). These heat maps are visual representations of the *z*-score matrix of each sample calculated by the VIZR program as described in Eq. 2. Additional VIZR heat maps for a representative control sample and a urine sample collected after ingestion of an energy drink are shown in Figure 5.5. At a standard deviation cutoff of 13, the heat map for the control sample (Figure 5.5a) is blank and contains no colored regions that would indicate differences compared to the average matrix for the control sample library. In contrast, the urine samples collected following ingestion of ethanol (Figure 5.4a), ibuprofen (Figure 5.4b), or an energy drink (Figure 5.5b) show a clear divergence from the control samples as

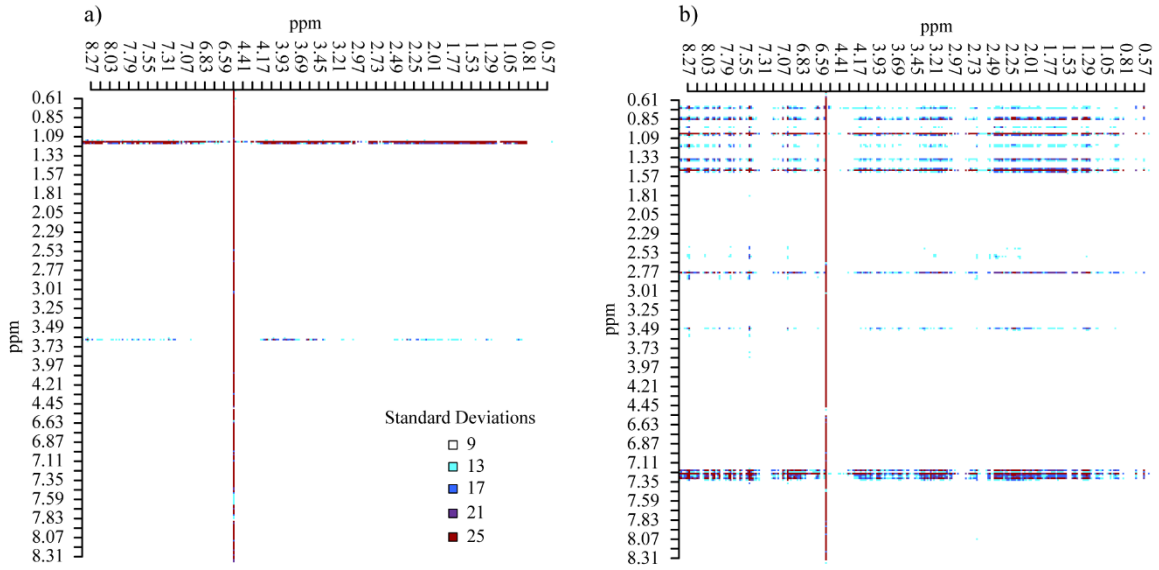


Figure 5.4. Heat maps reflecting *z*-score differences of urine samples after the ingestion of (a) an alcoholic beverage and (b) ibuprofen. The threshold for coloration (as indicated by the legend) was chosen according to the number of standard deviations the bin values are from the mean of the control samples: <13 white, 13 - 17 light blue, 17 - 21 dark blue, 21 - 25 purple, and >25 red. The red vertical line at 6.52 ppm in both figures is due to variation in the spectral baseline.

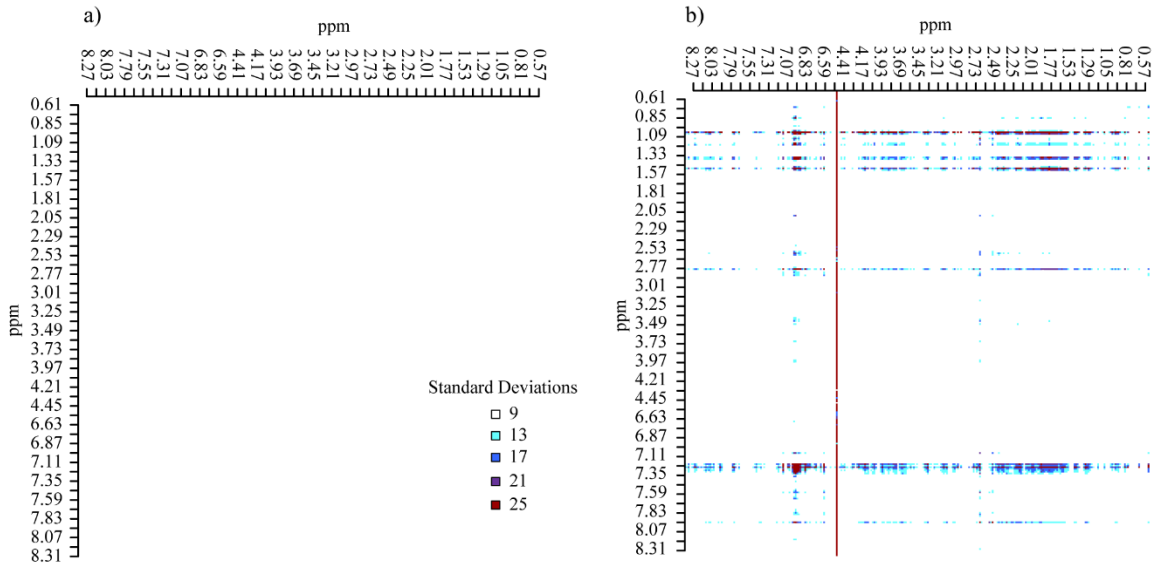


Figure 5.5. Heat maps reflecting z-score differences of (a) a representative control sample urine sample and (b) a urine sample taken after the ingestion of an energy drink. The threshold for coloration (as indicated by the legend) was chosen according to the number of standard deviations that bin values are from the mean of the control samples: <13 white, 13 - 17 light blue, 17 - 21 dark blue, 21 - 25 purple, and >25 red. The red vertical line in (b) is due to variation in the spectral baseline.

evidenced by blue and red coloration along the horizontal axis of the heat map. The vertical streaks observed, such as that at 6.52 ppm in Figure 5.4, result from baseline variation and are predominantly observed next to the discarded regions containing the residual water (4.5-5.0 ppm) and urea (5.0-6.5 ppm) resonances. The colored horizontal stripes at 1.2 and 3.6 ppm in the heat map in Figure 5.4a indicate that the intensity of these regions, which correspond to the chemical shifts of the CH_3 and CH_2 protons of ethanol, were statistically different from those of the average control population. The values used to generate the VIZR heat maps are calculated as a ratio of each bin to the bins of all other chemical shifts. For example, the heat map in Figure 5.4a contains a horizontal red band at 1.2 ppm for almost all chemical shift values across plot. This indicates that relative to almost all the other bins, the bin at 1.2 ppm due to the CH_3 protons of ethanol, is different from the control group at a standard deviation cutoff of 13. Similarly, the bin at 3.61 ppm, which corresponds to the chemical shift of the ethanol CH_2 resonance, is also different from most of the other bins of the control group, adding confidence that the differences highlighted are real. In Figure 5.4a the regions highlighted in the VIZR heat map are chemically related, although this will not necessarily be the case.

In contrast, the heat map in Figure 5.4b for the urine sample following ibuprofen ingestion is more complex showing numerous horizontal stripes in the methyl and aromatic chemical regions. Three of the horizontal bands at 0.9, 1.4 and 7.3 ppm can be attributed to excreted ibuprofen, however, the other regions highlighted in the VIZR heat map do not correlate with the spectrum of ibuprofen, indicating that these bins contain

resonances of ibuprofen metabolites or endogenous metabolites in biochemical pathways influenced by ibuprofen ingestion.[35-37] A similar observation can be made for the heat map of the urine sample after the ingestion of an energy drink (Figure 5.5b) which contains several bands in the aliphatic and aromatic chemical shift regions that are highlighted as statistically different from the control group average matrix. Many popular energy drinks contain compounds advertised to increase energy (such as guanine, taurine and caffeine) so it is not surprising that a higher abundance of aromatic and aliphatic resonances are observed in the urine NMR spectrum following ingestion of an energy drink.

5.3.6 VIZR *Z*-score Projections

An alternative way of viewing the output of the VIZR calculation is a *z*-score projection generated by summing each row of the sample *z*-score matrix. The *z*-score projection is effectively a calculated pseudo one-dimensional NMR spectrum with chemical shift plotted along the *x*-axis and the summed *z*-score value plotted along the *y*-axis. The peaks in the *z*-score projection highlight the regions of the test sample NMR spectrum that differ significantly from the average values of control sample spectra. Because absolute values are not used to calculate the z-score projections (as they are for the heat maps and the *z*-score scatter plot), a component absent in the sample but present in the controls would appear as a negative peak in the *z*-score projection. To illustrate the utility of this calculation, Figure 5.6 compares the averaged ^{1}H NMR spectrum of the control group (Figure 5.6a), the spectrum of a urine sample taken after ingestion of an alcoholic beverage (Figure 5.6b), the *z*-score projection (Figure 5.6c), and the spectrum

of ethanol dissolved in the same buffer as the urine samples (Figure 5.6d). Despite the complexity of the control and test spectra (Figure 5.6a and b, respectively) and the crowded nature of the spectral region between 3 and 4.5 ppm, VIZR analysis exclusively identified the CH_3 (1.18 ppm) and CH_2 (3.61 ppm) resonances of ethanol as being significantly different from those of the control group average (Figure 5.6a), emphasizing the capabilities of VIZR to identify differences even in crowded spectral regions.

The VIZR *z*-score projections similarly highlight the differences in urine samples collected after the ingestion of ibuprofen (Figure 5.7) and an energy drink (Figure 5.8). The *z*-score projection for the sample taken following ibuprofen (Figure 5.7c) reflects the excretion of the parent drug as well as the presence of NMR resonances for other metabolites in the aliphatic and aromatic regions of the spectrum (Figure 5.7b). As in the heat map presented in Figure 5.4b, these peaks could be due to the metabolic products of ibuprofen, such as the 2-[4-(2-carboxy-2-methylpropyl)-phenyl] propionic acid metabolite, oxpentifylline metabolite, or glucuronide conjugates as described by Wilson and Nichoslon[35-37] or endogenous metabolites whose levels were altered compared to the control average as a result of ibuprofen ingestion. Similarly, consumption of the energy drink resulted in significant changes in the aliphatic and aromatic regions of the NMR spectrum (Figure 5.8b) compared to average control spectrum (Figure 5.8a) and these differences are highlighted by the *z*-score projection (Figure 5.8c). When compared with the NMR spectrum measured for the energy drink (Figure 5.8d), the only resonances that are similar to the peaks found in the urine sample are at 1.13ppm, suggesting that many of the energy drink components are metabolized prior to excretion.

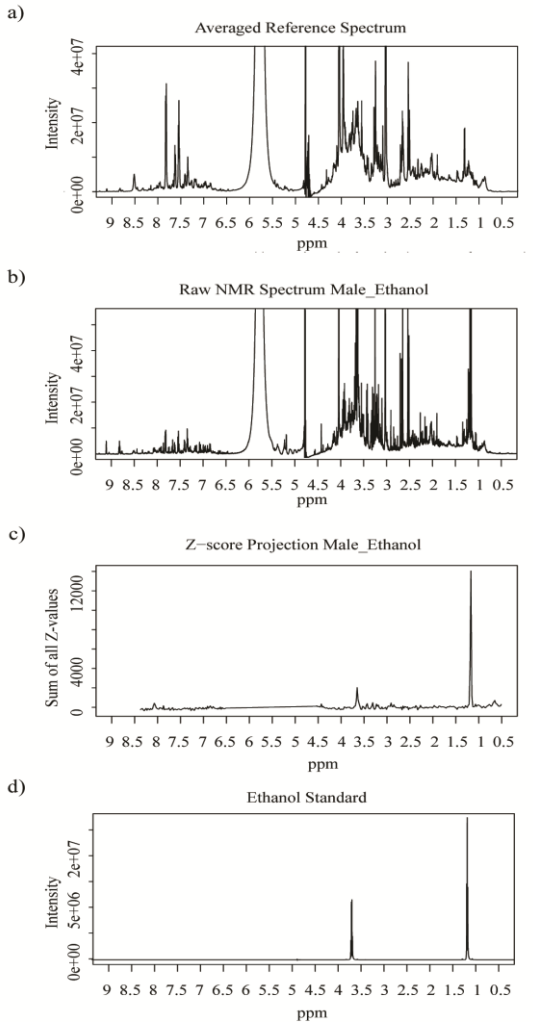


Figure 5.6. (a) The averaged NMR spectrum of the control group, (b) the spectrum of a urine sample after ingestion of an alcoholic beverage (ethanol), (c) the *z*-score projection of the urine ethanol sample, and (d) the ^{1}H NMR spectrum of ethanol dissolved in the same buffer as the urine samples. The *z*-score projection (c) of this sample highlights the CH_3 and CH_2 peaks of ethanol as significantly different compared to the average reference spectrum, even for the CH_2 peak which occurs in the crowded region of the spectrum. Figures (a), (b) and (c) were generated by VIZR and the NMR spectrum in (d) was added for comparison purposes.

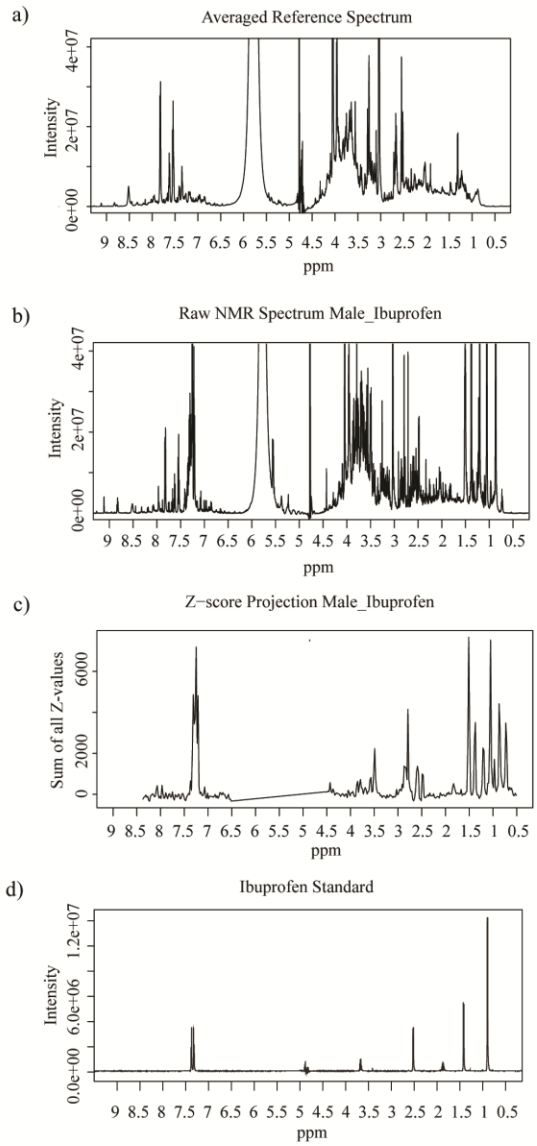


Figure 5.7. (a) The averaged NMR spectrum of the control group, (b) the spectrum of a urine sample after ingestion of ibuprofen, (c) the *z*-score projection of a urine sample after ingestion of ibuprofen, and (d) the ^{1}H NMR spectrum of ibuprofen dissolved in the same buffer as the urine samples. Figures a, b and c are generated by VIZR and the NMR spectrum in (d) was added for comparison purposes.

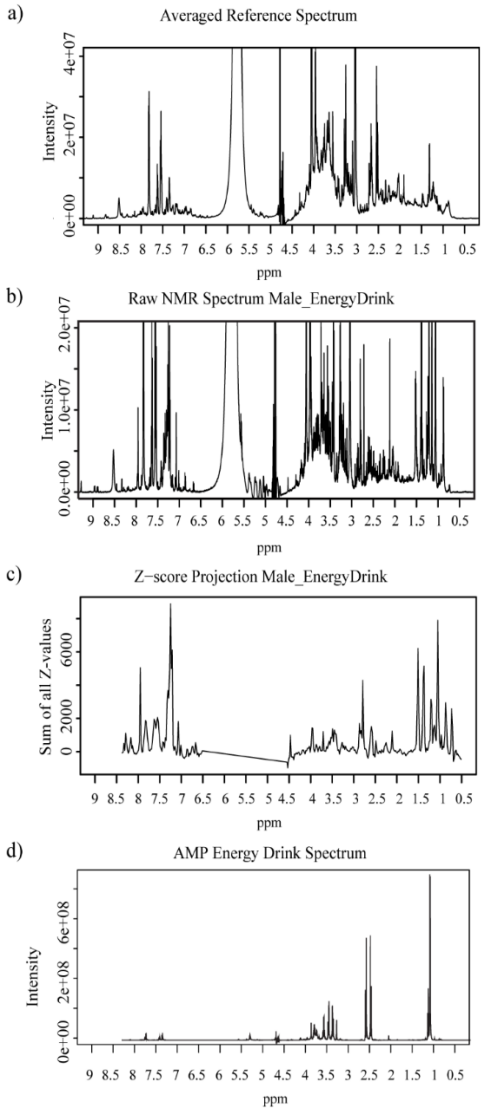


Figure 5.8. (a) The averaged NMR spectrum of the control samples, (b) the raw NMR spectrum of a urine sample taken after ingestion of an AMP^{TM} energy drink (c) the z-score projection of the urine sample after ingestion of an AMP^{TM} energy drink and (d) the spectrum of the AMP^{TM} energy drink diluted in the same buffer as the urine samples. Figures (a), (b), and (c) were generated by VIZR and the NMR spectrum in (d) was added for comparison purposes.

5.3.7 VIZR *Z*-score Scatter Plot

The VIZR heat map and *z*-score projection are useful for discovering the identity of resonances in the NMR spectrum (and by inference the chemical components of the sample) that differ significantly from those of the controls. In addition to this qualitative analysis, a quantitative statistical assessment is necessary to describe the extent of variation within the control group samples and the degree to which the individual test samples deviate from the control group average. Therefore, the VIZR program also calculates a *z*-score scatter plot that can be used for statistical significance testing for each member of the control and test groups. The *z*-score scatter plot is generated by plotting the sum of the absolute value of the *z*-scores from each *z*-score matrix as a function of the sample number, shown in Figure 5.9. Lines indicating the value of the summed *z*-score representing 1, 2, or 4 standard deviations (σ) from the control group mean are plotted in Figure 5.9 to facilitate significance testing. All of the control group samples lie within two standard deviations of the mean in Figure 5.9. Of the 41control urine samples tested, only 4 had a summed *z*-score value greater than one standard deviation of the population mean. Because the majority of the control group samples plotted within one standard deviation of the mean, a two standard deviation cutoff was applied in Figure 5 and all samples falling above the cutoff were investigated for specific metabolic aberrations. This analysis was facilitated by the use of VIZR heat maps and *z*-score projections to identify spectral regions responsible for differences between individual samples and the control group average. The advantage of using the sum of the *z*-score values to generate

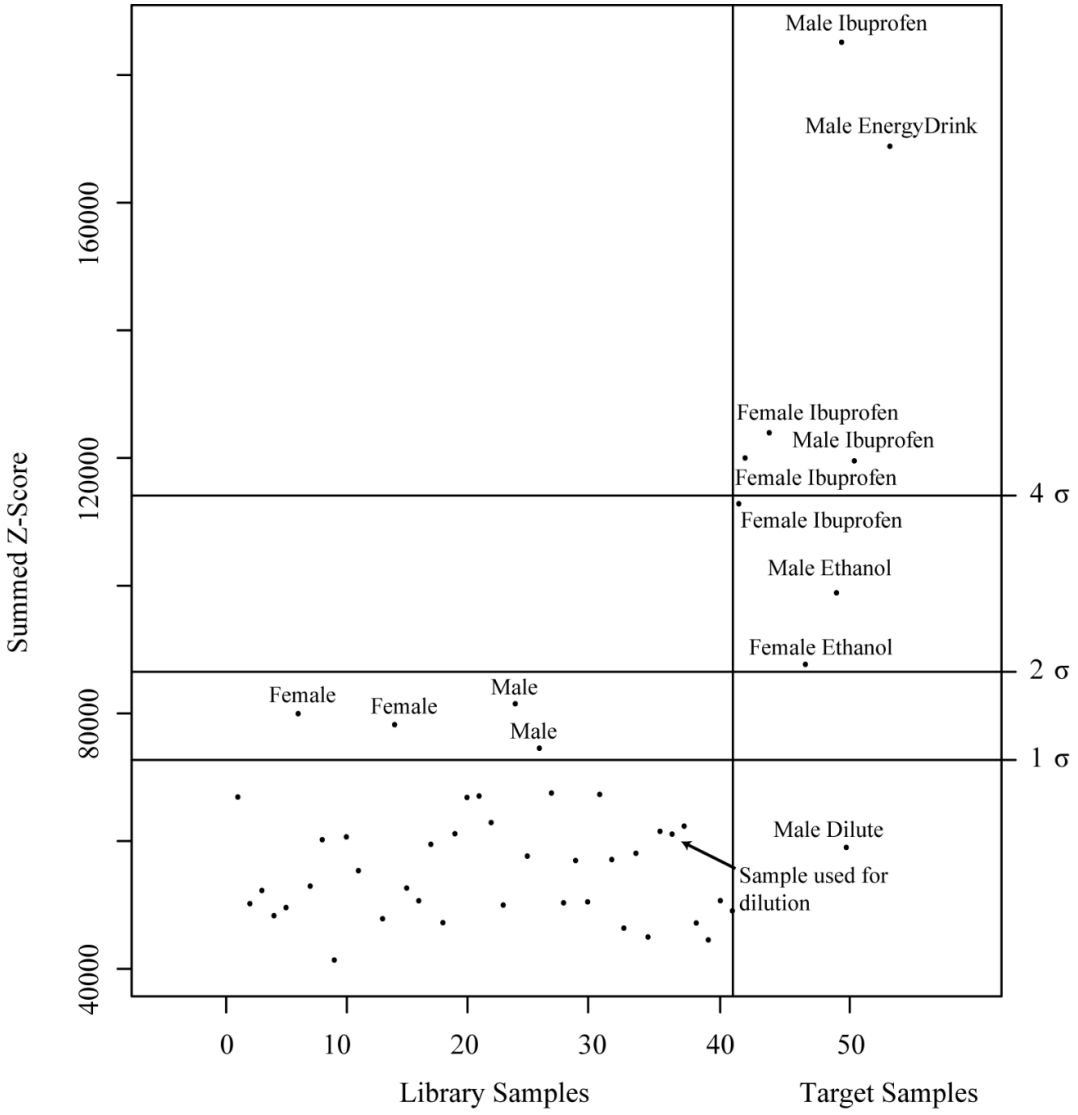


Figure 5.9. Scatter plot of the total *z*-value obtained from the matrices of the full data set. The *x*-axis represents the sample number, with the control samples on the left (up to 41) and the test samples on the right (42 to 50). The *y*-axis gives the total *z*-value for each sample. Horizontal lines representing 1, 2 and 4 standard deviations from the control sample mean are included. Selected labels have been added to facilitate sample identification.

the *z*-score scatter plot is that the sum does not depend on the magnitude of the individual *z*-score values for each chemical shift, but instead on the overall difference of each sample compared with the mean of the control group. For example, the *z*-score sum calculated for the urine samples collected after ingestion of ethanol lie more than two standard deviations away from the mean of the control population in Figure 5.9. The summed *z*-scores calculated from the spectra of test samples collected following ingestion of ibuprofen or an energy drink also plot more than 2 standard deviations awayfrom the population mean, with several samples different by more than 4 σ. It is important to note that all the test samples plot more than 2 standard deviations of the population mean and therefore are discriminated by VIZR as significantly different from the control group average.

The influence of sample concentration on VIZR analysis was investigated by diluting a control sample by 50% and including it as a test sample. The bin ratio normalization method used by the VIZR analysis accounts for dilution by calculating a ratio of each bin to all other bins. If a sample is diluted, the bin ratios should remain unchanged because the integrated values for each bin will decrease proportionally. When subjected to VIZR analysis, the summed *z*-score calculated for the diluted sample plotted in Figure 5.9 at a *y*-axis position nearly identical to the undiluted sample and well within one standard deviation of the control group mean. This result confirms that the bin ratio normalization method used by VIZR adequately accounts for dilution. Similarly, the heat

map and *z*-score projection of the diluted sample did not highlight any regions that were different from the control population (data not shown).

5.4 Conclusions

The novel chemometric program, VIZR, is a powerful new approach for the analysis of metabolic data sets based on the calculation of *z*-score matrices. The VIZR method combines bin ratio normalization with *z*-score analysis to compare data measured for a test sample to the average calculated for a representative population. In this study, VIZR was used to evaluate the metabolic profiles of test urine samples relative to a control population by analysis of ^{1}H NMR spectra. VIZR analysis successfully distinguished individual test samples collected after the ingestion of ibuprofen, alcoholic beverages or energy drinks from a group of 41 control urine samples. VIZR analysis identified specific spectral regions that differed in the test sample and control group spectra allowing expedited determination of component-level metabolic differences between the test and control groups without *a priori* knowledge of the test sample classification or the nature of the sample.

The VIZR program can readily be extended to problems related to disease diagnosis, biomarker detection, and toxicological profiling by identifying specific spectral regions containing metabolic abnormalities and to discriminate individual samples from a control group. For example, in toxicological studies, VIZR would be useful when searching for metabolic perturbations and catabolic intermediates of a dosed compound. Similarly, VIZR can aid biomarker identification and disease detection by discriminating individual samples with metabolic aberrations from a control group and

highlighting the spectral regions responsible for those differences. In a clinical setting, ranking the samples based on total *z*-score (Figure 5.9) enables fast and unbiased determination of samples that fall outside the normal metabolite profile, regardless of the number of metabolites that contribute to these differences. Because the VIZR technique is dependent on preprocessed data, it can be adapted to a variety of other data formats and analytical techniques including hyphenated MS measurements, electrophoretic separations, UV-Vis, IR, and fluorescence spectra.

5.5 References

(1) Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M. Urine metabolic signature of pancreatic ductal adenocarcinoma by H-1 nuclear magnetic resonance: Identification, mapping, and evolution. *J. Proteome Res.* **2012,** *11*, (2), 1274-1283.

(2) Zhang, A.; Sun, H.; Wang, X. Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Anal. Bioanal. Chem.* **2012,** *404*, (4), 1239-1245.

(3) Meng, J.; Zhang, X.; Wu, H.; Bu, J.; Shi, C.; Deng, C.; Mao, Y. Morphine-induced conditioned place preference in mice: Metabolomic profiling of brain tissue to find "molecular switch" of drug abuse by gas chromatography/mass spectrometry. *Anal. Chim. Acta* **2012,** *710*, 125-130.

(4) Aimetti, M.; Cacciatore, S.; Graziano, A.; Tenori, L. Metabonomic analysis of saliva reveals generalized chronic periodontitis signature. *Metabolomics* **2012,** *8*, (3), 465-474.

(5) Holmes, E.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Nicholson, J. K.; Lindon, J. C. 750 MHz H-1 NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *J. Pharm. Biomed. Anal.* **1997,** *15*, (11), 1647-1659.

(6) De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Rietzschel, E. R.; De Buyzere, M. L.; Langlois, M. R.; Bekaert, S.; Martins, J. C.; Van Criekinge, W. Evaluation of standard and advanced preprocessing methods for the univariate analysis of blood serum H-1-NMR spectra. *Anal. Bioanal. Chem.* **2010,** *398*, (4), 1781-1790.

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

(8) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* **2005,** *18*, (3), 143-162.

(9) Wikoff, W. R.; Gangoiti, J. A.; Barshop, B. A.; Siuzdak, G. Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clin. Chem.* **2007,** *53*, (12), 2169-2176.

(10) Ebbels, T. M. D.; Keun, H. C.; Beckonert, O. P.; Bollard, M. E.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Prediction and classification of drug toxicity using probabilistic modeling of temporal metabolic data: The Consortium on Metabonomic Toxicology screening approach. *J. Proteome Res.* **2007,** *6*, (11), 4407-4422.

(11) Nicholson, J.; Keun, H.; Ebbels, T. COMET and the challenge of drug safety screening. *J. Proteome Res.* **2007,** *6*, (11), 4098-4099.

(12) Schicho, R.; Shaykhutdinov, R.; Ngo, J.; Nazyrova, A.; Schneider, C.; Panaccione, R.; Kaplan, G. G.; Vogel, H. J.; Storr, M. Quantitative metabolomic profiling of serum, plasma, and urine by H-1 NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy Individuals. *J. Proteome Res.* **2012,** *11*, (6), 3344-3357.

(13) Gebregiworgis, T.; Powers, R. Application of NMR Metabolomics to Search for Human Disease Biomarkers. *Comb. Chem. High Throughput Screening* **2012,** *15*, (8), 595-610.

(14) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **2007,** *363*, (2), 185-195.

(15) Issaq, H. J.; Waybright, T. J.; Veenstra, T. D. Cancer biomarker discovery: Opportunities and pitfalls in analytical methods. *Electrophoresis* **2011,** *32*, (9), 967-975.

(16) 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.

(17) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabonomics. *Anal. Chem.* **2006,** *78*, (13), 4281-4290.

(18) Coresh, J.; Astor, B. C.; Greene, T.; Eknoyan, G.; Levey, A. S. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *Am. J. Kidney Dis.* **2003,** *41*, (1), 1-12.

(19) Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos Mag* **1901,** *2*, (7-12), 559-572.

(20) Trygg, J. O2-PLS for qualitative and quantitative analysis in multivariate calibration. *J. Chemom.* **2002,** *16*, (6), 283-293.

(21) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002,** *16*, (3), 119-128.

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

(23) Li, H.; Wang, L.; Yan, X. Z.; Liu, Q. J.; Yu, C. H.; Wei, H. D.; Li, Y. M.; Zhang, X. M.; He, F. C.; Jiang, Y. A Proton Nuclear Magnetic Resonance Metabonomics Approach for Biomarker Discovery in Nonalcoholic Fatty Liver Disease. *J. Proteome Res.* **2011,** *10*, (6), 2797-2806.

(24) McKenzie, J. S.; Donarski, J. A.; Wilson, J. C.; Charlton, A. J. Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011,** *59*, (4), 336-359.

(25) Ahmed, S.; Santosh, W.; Kumar, S.; Christlet, H. T. T. Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection. *J. Biomed. Sci.* **2009,** *16*, 12.

(26) Charlton, A. J.; Robb, P.; Donarski, J. A.; Godward, J. Non-targeted detection of chemical contamination in carbonated soft drinks using NMR spectroscopy, variable selection and chemometrics. *Anal. Chim. Acta* **2008,** *618*, (2), 196-203.

(27) Wei, S.; Zhang, J.; Liu, L.; Ye, T.; Gowda, G. A. N.; Tayyari, F.; Raftery, D. Ratio analysis nuclear magnetic resonance spectroscopy for selective metabolite identification in complex samples. *Anal. Chem.* **2011,** *83*, (20), 7616-7623.

(28) Hwang, T. L.; Shaka, A. J. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Reson., Ser A* **1995,** *112*, (2), 275-279.

(29) Henningsen, A.; Toomet, O. *miscTools: Miscellanneous tools and utilities*, R package version 0.6-12; 2011.

(30) Team, R. C. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2012.

(31) Cheadle, C.; Vawter, M. P.; Freed, W. J.; Becker, K. G. Analysis of microarray data using Z score transformation. *J. Mol. Diag.* **2003,** *5*, (2), 73-81.

(32) Hamersky, K. A.; Merrywell, C. E.; Fang, F.; Larive, C. K., Metabolic Profiling. In *NMR Spectroscopy in Pharmaceutical Analysis*, Holzgrabe, U.; Wawer, I.; Diehl, B., Eds. Elsevier: Oxford, 2008; pp 233-267.

(33) Barding, G.; Salditos, R.; Larive, C. Quantitative NMR for bioanalysis and metabolomics. *Anal. Bioanal. Chem.* **2012,** *404*, (4), 1165-1179.

(34) Ye, T.; Zheng, C.; Zhang, S.; Gowda, G. A. N.; Vitek, O.; Raftery, D. "Add to subtract": A simple method to remove complex background signals from the ^{1}H nuclear magnetic resonance spectra of mixtures. *Anal. Chem.* **2012,** *84*, (2), 994-1002.

(35) 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. Anal.* **1992,** *10*, (8), 601-605.

(36) Wilson, I. D.; Nicholson, J. K. Solid-phase extraction chromatography and nuclear magnetic resonance spectrometry for the identification and isolation of drug metabolites in urine. *Anal. Chem.* **1987,** *59*, (23), 2830-2832.

(37) Wilson, I. D.; Nicholson, J. K. Solid-phase extraction chromatography and NMR-spectroscopy (SPEC-NMR) for the rapid identification of drug metabolites in urine. *J. Pharm. Biomed. Anal.* **1988,** *6*, (2), 151-165.

CHAPTER SIX

Conclusions and Future Directions

6.1 Conclusions

The work presented herein describing the metabolic reconfiguration of *Oryza sativa* ssp. *japonica* cv. M202 and M202(*Sub1*) during abiotic stress will enable a more thorough understanding of the responses of plants to abiotic stressors. Initially correlated with prolonged submergence tolerance, the presence of the *SUBMERGENCE1A* (*SUB1A*) gene has also been shown to increase recovery from drought and tolerance of prolonged darkness. Ironically, plants also experience darkness and drought in conjunction with submergence stress and recovery, respectively, although each condition produces a specific molecular response.[1-4] The experiments described in this dissertation were performed to probe to how the presence or absence of the *SUB1A* gene in otherwise genetically identical cultivars influences metabolism of the shoot during and after submergence. In Chapter 2, two submergence time course experiments were conducted to explore the metabolic differences of the M202 and M202(*Sub1*) varieties as determined by NMR. A long term submergence time course, consisting of an initial control and 3, 7 and 12 d submergence or 12 d submergence + 1 d recovery, and a short term submergence time course, spanning an initial control and 1, 2, and 3 d submergence or 3 d submergence + 1 d recovery were performed. During the course of submergence, the tolerant M202(*Sub1*) variety displayed less carbon consumption as evident by higher levels of sucrose and lower levels of glucogenic amino acid accumulation as compared to the intolerant M202 variety. Additionally, we reported for the first time, identification of

the dipeptide alanylglycine (AlaGly) in rice extracts by NMR spectroscopy. Although significant differences were not observed between the two genotypes, levels of AlaGly decreased during submergence and did not show a recovery trend after 1 d post desubmergence, which is opposite to the trends observed for the majority of the detected amino acids. Previous reports indicate that levels of AlaGly may be affected by the diurnal cycle but additional experiments would be necessary to determine whether this factor could explain the results presented in Chapter 2.[5]

In Chapter 3, a deeper exploration of rice metabolism in the presence and absence of the *SUB1A* gene was carried out by GC-MS analysis using the tissue from the short term experiment discussed in Chapter 2. As with the NMR results, samples from the tolerant M202(*Sub1*) plants showed less amino acid accumulation than the intolerant variety, indicative of decreased catabolism of starch consistent with the quiescence strategy associated with expression of the *SUB1A* gene. Several metabolites not detectable by NMR were also quantified by GC-MS, including the TCA cycle intermediates citrate, malate, and succinate, which were differentially accumulated in the two genotypes. Relative levels of citrate and malate were greater in the tolerant variety than in the intolerant variety, which may reflect a higher flux through the TCA cycle under stress conditions. Chapter 3 also provides a comparison of the relative merits of NMR and GC-MS analyses for the rice tissue samples examined in this study. Although MS detection is generally more sensitive than NMR, hyphenation with GC requires metabolites to be volatile or made volatile through derivatization and not all metabolites are amenable to derivatization. AlaGly and S-methyl-methionine were detected by NMR

spectroscopy, but were not observed by GC-MS analysis. Similarly, glutamine, which had a high signal-to-noise ratio by NMR spectroscopy, was predominantly below the limit of quantitation when analyzed by GC-MS. These results highlight the benefits of analyses employing complementary techniques, such as GC-MS and NMR, in providing a more complete metabolic snapshot than can be obtained by a single analytical platform.

Chapter 4 expanded the exploration of the rice metabolic submergence response through NMR and GC-MS evaluation of the untargeted metabolic profiles of the submergence tolerant and intolerant rice varieties during a recovery time course. This chapter also considered dynamics in shoot metabolites over a twenty-four hour period. Plants were submerged for 3 d and upon desubmergence harvested immediately or at dusk, midnight, dawn, and +24 hr to evaluate the reoxygenation response. Distinctions observed between the genotypes as a result of submergence were consistent with the experiments in Chapters 2 and 3, however, the recovery profiles during the dark cycle suggested a significant genotypic influence. For several metabolites, the resulted indicated that either reoxygenation, darkness, or a combination of the two conditions invoked a separate stress response evident in the metabolic profiles. Control samples were also harvested at corresponding intervals to evaluate the influence of the diurnal cycle on recovery and to determine whether the genotypes experienced circadian differences. The diurnal patterns in the relative levels of the detected metabolites were mostly indifferent, suggesting that in the absence of stress, the two genotypes are similar. Finally, a targeted analysis of phosphorylated and other anionic metabolites was carried out using WAX SPE and RPIP-UPLC-MS. Specifically, the targeted experiments

allowed for the relative quantitation of trehalose-6-phosphate (T6P), a signaling metabolite involved in sensing carbon availability for ATP production. RPIP-UPLC-MS analysis demonstrated that T6P levels fell in both genotypes during submergence. During recovery, levels of T6P during the recovery time course were significantly higher in the tolerant variety. This result suggests that the tolerant variety is more adept at returning to normoxic growth conditions through inhibition of the SnRK1 pathway which drives carbon catabolism, thereby reinstating anabolic processes.[6-8] The greater levels of T6P in the tolerant variety and downstream relationship to SnRK1 pathway inhibition could support the higher rate of metabolite assimilation upon reoxygenation compared with the intolerant variety, however experiments evaluating transcript accumulation of related pathways as well as photosynthetic activity would be required to test this hypothesis.

Additionally, Chapter 2, 3, and 4 demonstrate the ability for rice plants, regardless of the presence or absence of the *SUB1A* gene, to utilize the metabolites that accumulate during submergence upon desubmergence. During recovery, levels for many of the organic acids and amino acids trend toward pre-stressed conditions (i.e. GABA, Citrate, Ala, Asn, Asp, Ile, Thr, and Val) The rapid consumption of Ala, by conversion back to pyruvate enables the plant to limit overall carbon loss during the stress, as compared to the production of ethanol, which can be lost due to efflux from the plant cell or by evaporation. The recovery of oxidative phosphorylation, particularly the activity of succinate dehydrogenase, is evident from the rapid accumulation of fumarate following desubmergence. However, the relative differences between the varieties for some of these amino acids increase during recovery, suggesting that the intolerant variety is less

capable of re-assimilating the accumulated amino acids than the tolerant variety (i.e. Ala, Asp, Gln, Glu, pyroglutamate, Ser, Thr, and Val). Thus, in addition to carbohydrate conservation during submergence stress, *SUB1A* may enable faster metabolite recycling during recovery. Together, these metabolic adjustments of *SUB1A*-containing plants provide a profound ability to survive an extended submergence stress.

Lastly, Chapter 5 presents a novel chemometric technique, Visualization of *Z*-score Ratios (VIZR), for the examination of biofluid data sets. Human urine samples were collected and evaluated by ^{1}H NMR spectroscopy for the effects produced by ingestion of ibuprofen, alcoholic beverages, and energy drinks. Because of the highly variable water content inherent in urine samples, a bin-ratio normalization method was developed to account for sample dilution. The VIZR technique was able to discriminate the metabolite profiles of urine samples obtained after ingestion of an alcoholic beverage, ibuprofen, or an energy drink. Furthermore, VIZR was able to identify the NMR spectral regions responsible for the disparity between normal samples and those taken after ingestion of the supplements. VIZR is not limited to NMR data but could readily be extended to the chemometric analysis of hyphenated MS or spectroscopic results. The VIZR technique has potential for high-throughput data analysis for disease diagnosis, toxicological profiling, and biomarker identification.

6.2 Future Directions

6.2.1 Absolute Quantitation and Metabolic Flux Analysis

As discussed in sections 1.2.2 and 1.3.3.2, absolute quantitation in metabolomics experiments can provide a more complete understanding of the biological system being

investigated.[9] For NMR spectroscopy, absolute quantitation requires well-resolved resonances of an internal standard. Since most measurements utilize an internal standard as a chemical shift reference, this does not pose a limitation. Indeed, a significant advantage of NMR over other techniques, including MS and spectroscopic methods, is that a single standard can be used for quantitation of every signal in the spectrum. Absolute quantitation by NMR spectroscopy requires the nuclei to be fully relaxed prior to each scan using a recycle period of at least 5 times the longest longitudinal (T_1) relaxation rate, limiting the throughput of NMR metabolomics experiments. Absolute quantitation with MS-based methods is more challenging. The most reliable method relies on addition of an isotopically labeled analog of each analyte directly into every sample analyzed. Isotopically labeled compounds, however, are expensive and are not available for all analytes of interest in a metabolomics experiments. Despite these challenges, the use of absolute quantitation in metabolomics studies allows for a direct comparison of metabolite levels between experiments without the need for normalization, and will likely become more prominent as the field matures.

Metabolomics experiments focus on the steady state change of metabolite levels in response to biotic or abiotic stimuli. However, most metabolites are connected through different biochemical pathways and the activation or inactivation of these pathways is important to understanding how the organism responds to the stress.[10] For example, alanine (Ala) is a product of pyruvate metabolism and an increase in glycolysis can result in an accumulation of Ala.[11] Similarly, Ala can also be a nitrogen sink. Vanlerberghe and coworkers demonstrated that during anaerobic conditions in green alga

almost all the assimilated ammonium was retained by Ala synthesis.[12] In our experiments, the accumulation of Ala is most likely driven by increased glycolysis during anaerobic metabolism.[11, 13] Changes in glutamine and glutamate, metabolites correlated with nitrogen metabolism, are also observed but it is unclear how nitrogen metabolism is affected and if Ala is acting as both a nitrogen and carbon sink. To understand how different biochemical pathways are affected, particularly nitrogen and carbon metabolism, metabolic flux analysis (MFA) will need to be carried out. Specifically, experiments using pre-exposure to ^{15}N labeled nitrate and ^{13}C labeled glucose can be carried out using either MS or NMR in submergence stressed plants. Subsequent metabolite analysis will reveal how and where the two genotypes store nitrogen during submergence stress. Additionally, by using ^{13}C labeled glucose, flux through glycolysis and the TCA cycle will provide a more detailed understanding how energy production is affected in the two genotypes during submergence stress.

6.2.2 Targeted Metabolite Profiling

The targeted profiling approach presented in Chapter 4 was effective for quantifying low levels of the T6P, however, it was necessary to use the ion count at the peak apex for quantitation rather than peak area because of poor peak shape and the inability to completely resolve T6P from sucrose-6-phosphate (S6P). To improve the quantitation of T6P, the chromatographic separation will have to be improved. One possibility is to explore the use of different ion pairing reagents to alter the retention characteristics of the analytes and reduce peak tailing.[14, 15] Alternative chromatographic methods could also be adapted for the targeted separation of T6P. Torano and coworkers

reported the use of HILIC for the analysis of T6P, however this method was optimized at pH 12.0 which is well above the recommended pH range of the column, potentially reducing the column lifetime.[16] Other columns and mobile phase additives could be examined for their ability to improve the HILIC separation for the desired analytes and a detailed exploration will likely be required to arrive at a suitable separation for T6P and other phosphorylated metabolites.[17]

Targeted profiling of other secondary metabolites will increase our understanding of the systemic response of rice and other crops to abiotic stress. For example, anthocyanins and other flavonoids are known to have antioxidant functions.[18] Fukao and coworkers suggested that reactive oxygen species (ROS), which are known to accumulate during reoxygenation after submergence and drought are mediated by the *SUB1A* gene.[3] Understanding how flavonoids and other phenolic compounds are regulated during and after submergence stress would provide insights into the greater capacity of *SUB1A* containing plants to tolerate ROS stress. Additionally, UDP-glucose, which is the biosynthetic precursor for T6P, has been implicated in the flavonoid biosynthesis pathways responsible for flower color.[19] Unraveling the complex biochemical relationships between secondary metabolites will enable a more complete understanding of the stress response and survival mechanisms of *SUB1A* rice.

6.2.3 Understanding the Effect of *SUB1A* on Metabolism in Rice during Drought and extended Darkness

Fukao and coworkers reported that the presence of the *SUB1A* gene confers greater tolerance to drought and extended darkness compared with non-*SUB1A*

containing rice varieties.[3] In Chapter 4, we reported negligible genotypic differences during the diurnal cycle under normal growth conditions, however, during submergence stress crops often endure a period of extended darkness due to water turbidity. To tease apart the effects of light stress and submergence stress, the metabolite profiles of M202 and M202(*Sub1*) could be evaluated during extended darkness. Additionally, the two genotypes should be evaluated for differences their metabolic response to drought stress. Fukao and coworkers observed that under extended drought, a majority of the plant tissue from both genotypes dies, but the recovery of the M202(*Sub1*) variety occurred from meristem, generating new leaves while the older leaves died. To further these studies, available genotypes that differ in the presence or absence of a constitutively expressed *SUB1A* gene might also be used.[1] Understanding how the metabolism of the two genotypes respond to drought will further our understanding of the systemic effect *SUB1A* has during different stresses.

6.2.4 Continued Development of Chemometric Methods

Increasing the availability and sophistication of data analysis software will increase the utility of analytical techniques for biomarker identification, disease diagnosis, and toxicological profiling. In Chapter 5 we reported the use of *z*-scores for discriminating individual test samples from a control group. In addition to discrimination, the VIZR program highlights the spectral regions that differ significantly from the control population and scores these differences for quantitative comparison. To better facilitate biomarker identification, VIZR analysis can be extended beyond flagging spectral regions of interest to identification of the specific metabolic components that are

responsible for these differences. The ratio of the resonance intensities of two different nuclei from the same compound is fixed for any concentration, providing that the spectra have a sufficient signal-to-noise ratio. Because the VIZR technique uses the bin-ratio normalization method, the *z*-score values for all resonances associated with a metabolite should be similar. A spectral library for all known metabolites can be generated, and the intensity-ratios for the resonances of the standards determined with corresponding chemical shift information. Extension of the VIZR program to automatically compare the intensity-ratios/ppm values of the library with the highlighted regions from a sample would provide putative metabolite identification.

The VIZR method can also be adapted for use with data sets produced by other analytical techniques, such as HPLC-MS. Existing metabolomics processing programs, such as MassLynx (Waters Corp, Milford, MA) and XCMS[20] automatically align and integrate mass spectral data across several experiments. The metabolite identities are conserved as mass-retention time pairs, so that the integrated area for a mass at a given retention time can be later correlated with a metabolite. VIZR can be easily adapted to utilize mass-retention pairs instead of chemical shift information for data visualization. The only requirement for VIZR analysis is that the data be two dimensional, containing both quantitative and spatial information. One reason that VIZR was written in the open-source code R was to allow for other investigators to adapt this approach and use it in ways that have not yet been envisioned. It will be interesting to see what applications this tool may find over the next decade.

6.3 References

(1) Fukao, T.; Bailey-Serres, J. Submergence tolerance conferred by Sub1A is mediated by SLR1 and SLRL1 restriction of gibberellin responses in rice. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, (43), 16814-16819.

(2) Fukao, T.; Xu, K. N.; Ronald, P. C.; Bailey-Serres, J. A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* **2006,** *18*, (8), 2021-2034.

(3) Fukao, T.; Yeung, E.; Bailey-Serres, J. The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell* **2011,** *23*, (1), 412-427.

(4) Xu, K.; Xu, X.; Fukao, T.; Canlas, P.; Maghirang-Rodriguez, R.; Heuer, S.; Ismail, A. M.; Bailey-Serres, J.; Ronald, P. C.; Mackill, D. J. SUB1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **2006,** *442*, (7103), 705-708.

(5) Manabe, H.; Ohira, K. Effect of light irradiation on the D-alanylglycine content in rice leaf blades. *Plant Cell Physiol.* **1983,** *24*, (6), 1137-1142.

(6) Delatte, T. L.; Sedijani, P.; Kondou, Y.; Matsui, M.; de Jong, G. J.; Somsen, G. W.; Wiese-Klinkenberg, A.; Primavesi, L. F.; Paul, M. J.; Schluepmann, H. Growth Arrest by Trehalose-6-Phosphate: An Astonishing Case of Primary Metabolite Control over Growth by Way of the SnRK1 Signaling Pathway. *Plant Physiol.* **2011,** *157*, (1), 160-174.

(7) Nunes, C.; Primavesi, L. F.; Patel, M. K.; Martinez-Barajas, E.; Powers, S. J.; Sagar, R.; Fevereiro, P. S.; Davis, B. G.; Paul, M. J. Inhibition of SnRK1 by metabolites: Tissue-dependent effects and cooperative inhibition by glucose 1-phosphate in combination with trehalose 6-phosphate. *Plant Physiol. Biochem.* **2012,** *63*, 89-98.

(8) Schluepmann, H.; Berke, L.; Sanchez-Perez, G. F. Metabolism control over growth: a case for trehalose-6-phosphate in plants. *J. Exp. Bot.* **2011,** *63*, (9), 3379-3390.

(9) Barding, G.; Salditos, R.; Larive, C. Quantitative NMR for bioanalysis and metabolomics. *Anal. Bioanal. Chem.* **2012,** *404*, (4), 1165-1179.

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

(11) Barding, G. A.; Fukao, T.; Beni, S.; Bailey-Serres, J.; Larive, C. K. Differential metabolic regulation governed by the rice *SUB1*A gene during submergence stress and identification of alanylglycine by ^{1}H NMR spectroscopy. *J. Proteome Res.* **2012,** *11*, 320-330.

(12) Vanlerberghe, G. C.; Joy, K. W.; Turpin, D. H. Anaerobic metabolism in the N-limited green-alga *Selenastrum minutum*. *Plant Physiol.* **1991,** *95*, (2), 655-658.

(13) Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K. Comparison of GC-MS and NMR for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12*, (2), 898-909.

(14) Jones, C. J.; Membreno, N.; Larive, C. K. Insights into the mechanism of separation of heparin and heparan sulfate disaccharides by reverse-phase ion-pair chromatography. *J. Chromatogr. A* **2010,** *1217*, (4), 479-488.

(15) Langeslay, D. J.; Urso, E.; Gardini, C.; Naggi, A.; Torri, G.; Larive, C. K. Reversed-phase ion-pair ultra-high-performance-liquid chromatography-mass spectrometry for fingerprinting low-molecular-weight heparins. *J. Chromatogr. A* **2013,** *In press*.

(16) Torano, J. S.; Delatte, T. L.; Schluepmann, H.; Smeekens, S. C. M.; de Jong, G. J.; Somsen, G. W. Determination of trehalose-6-phosphate in Arabidopsis thaliana seedlings by hydrophilic-interaction liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2012,** *403*, (5), 1353-1360.

(17) Buszewski, B.; Noga, S. Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique. *Anal. Bioanal. Chem.* **2012,** *402*, (1), 231-247.

(18) Reddy, A. M.; Reddy, V. S.; Scheffler, B. E.; Wienand, U.; Reddy, A. R. Novel transgenic rice overexpressing anthocyanidin synthase accumulates a mixture of flavonoids leading to an increased antioxidant potential. *Metab. Eng.* **2007,** *9*, (1), 95-111.

(19) Chen, W. H.; Hsu, C. Y.; Cheng, H. Y.; Chang, H.; Chen, H. H.; Ger, M. J. Downregulation of putative UDP-glucose: flavonoid 3-O-glucosyltransferase gene alters flower coloring in Phalaenopsis. *Plant Cell Rep.* **2011,** *30*, (6), 1007-1017.

(20) 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*, (3), 779-787.

Appendix 1

Script for Processing Metabolite Data

```
MetabMaster<-function(Dir,datafile,Genotype1,Genotype2){

        library(miscTools)
        library(gplots)
        setwd(Dir)
        datframe<-read.delim(datafile,header=T)
        header<-read.delim(datafile,header=F)
        header <- header[1,]
        ncol(datframe) -> n
        data5 <- NULL
        dataSummary <- NULL
        xu<-datframe$Treatment[!duplicated(datframe$Treatment)]
        length(xu) -> xn
        for (a in 1:xn){
        b<-xu[a]
        for (i in 4:n){
                d<-subset(datframe, datframe$Treatment==b)
                dAnova <- oneway.test(d[,i]~Genotype, data=d)
                data3<-dAnova$p.value
                data4<-as.matrix(data3)
                Q<-paste(Genotype1,Genotype2)
                row.names(data4)<-rownames(Q)
                colnames(data4)<-header[,i]
        data5 <- cbind(data5, data4)}
        data5 <- insertCol(data5, 1, b,cName="Treatment")
        dataSummary<-rbind(dataSummary,data5)
        data5 <- NULL}

        Genotype<-datframe$Genotype[!duplicated(datframe$Genotype)]
        xGenotype <- length(Genotype)
        Control<-subset(datframe, datframe$Treatment==xu[1])
        datSummary<-NULL
        dat1<-NULL
        dat2<-NULL
        s<-NULL
        for (z in 1:xGenotype){
                p<-NULL
                h<-Genotype[z]
```

```
                controlname<-xu[1]
                m<-subset(datframe, datframe$Genotype==h)
                p<-subset(Control, Control$Genotype==h)
                        for(a in 2:xn){
                        s<-NULL
                        b<-xu[a]
                        mm<-subset(m, m$Treatment==b)
                        s<-rbind(p,mm)
                                for(i in 4:n){
                                        mAnova<-oneway.test(s[,i]~Treatment, data=s)
                                        data2<-mAnova$p.value
                                        data2<-as.matrix(data2)
                                        dat1<-cbind(dat1, data2)}
                                        QR<-paste(controlname,b)
                                        row.names(dat1)<-QR
                                dat1<-insertCol(dat1, 1, h)
                                dat2<-rbind(dat2, dat1)
                                dat1<-NULL}
                                datSummary<-rbind(datSummary, dat2)
                                dat2<-NULL}

        dataSummary<-rbind(dataSummary, datSummary)
        write.table(dataSummary, "CompiledAnovaOutPvalues.txt", sep="\t",
col.names=NA, quote=F)

        ###Average and Standard Deviations of each datframe.data.data.data/Treatment
        dat2summary<-NULL
        dat3summary<-NULL
        dat4summary<-NULL
        for (z in 1:xGenotype){
                h<-Genotype[z]
                m<-subset(datframe, datframe$Genotype==h)
                for(a in 1:xn){
                        b<-xu[a]
                        mm<-subset(m, m$Treatment==b)
                                for (i in 4:n){
                                variable<-mm[,i]
                                Avg<-mean(variable, na.rm=T)
                                Avg<-as.matrix(Avg)
                                StDev<-sd(variable, na.rm=T)
                                colnames(Avg)<-header[,i]
                                allfiles <-cbind(Avg, StDev)
                                dat2summary<-cbind(dat2summary, allfiles)
```

```
                        allfiles<-NULL}
                        dat2summary<-insertCol(dat2summary, 1, mm[1,2],
cName="Treatment")
                        dat4summary<-rbind(dat4summary, dat2summary)
                        dat2summary<-NULL}
                dat4summary<-insertCol(dat4summary, 1, m[1,1],
cName="Genotype")
                dat3summary<-rbind(dat3summary, dat4summary)
                dat4summary<-NULL}
        write.table(dat3summary, "CompiledSummary.txt", sep="\t",
col.names=NA, quote=F)

            ###Generate table of ratios
    ncol.dat3summary<-ncol(dat3summary)
    ratios<-dat3summary[,seq(3,ncol(dat3summary),2)]
    factors<-dat3summary[,1:2]
    ratios<-cbind(factors,ratios)
    header2<-as.matrix(colnames(dat3summary))
    header3<-header2[seq(3,ncol(dat3summary),2),]
    ratios<-data.frame(ratios)
    nRatios<-ncol(ratios)
    ddRatio2<-NULL
    ddRatio3<-NULL
    for (a in 1:xn){
    b<-xu[a]
    for (i in 3:nRatios){
        d<-subset(ratios,ratios$Treatment==b)
        d1<-subset(d, d$Genotype==Genotype1)
        d2<-subset(d, d$Genotype==Genotype2)
        dd1<-as.matrix(d1[,i])
        dd1<-as.numeric(dd1)
        dd2<-as.matrix(d2[,i])
        dd2<-as.numeric(dd2)
        ddRatio<-dd1/dd2
        ddRatio<-as.matrix(ddRatio)
        colnames(ddRatio)<-header3[(i-2)]
    ddRatio2 <- cbind(ddRatio2, ddRatio)}
    paste(Genotype1,Genotype2)->Q
    rownames(ddRatio2)<-Q
    ddRatio2 <- insertCol(ddRatio2, 1, b,cName="Treatment")
    ddRatio3<-rbind(ddRatio3,ddRatio2)
    ddRatio2 <- NULL}
```

```
        ratioControl<-subset(ratios, ratios$Genotype==Genotype[1])
        wGenRatio1<-NULL
        wGenRatio2<-NULL
        allRatios<-NULL
        for (a in 1:xGenotype){
        gen<-Genotype[a]
        control1<-subset(ratioControl, ratioControl$Genotype==Genotype[a])
        gen1<-subset(ratios, ratios$Genotype==gen)
        gen1Control<-subset(gen1, gen1$Treatment==xu[1])
                for(b in 2:xn){
                treatment1<-subset(gen1, gen1$Treatment==xu[b])
                        for (i in 3:nRatios){
                        wGen1<-as.matrix(treatment1[,i])
                        wGen1<-as.numeric(wGen1)
                        wGen2<-as.matrix(gen1Control[,i])
                        wGen2<-as.numeric(wGen2)
                        wGenRatio<-wGen1/wGen2
                        wGenRatio<-as.matrix(wGenRatio)
                        colnames(wGenRatio)<-header3[(i-2)]
                        wGenRatio1 <- cbind(wGenRatio1, wGenRatio)}
                        rownames(wGenRatio1)<-Genotype[a]
                        wGenRatio1<- insertCol(wGenRatio1, 1, xu[b]:xu[1],
cName="Treatment")
                        wGenRatio2 <- rbind(wGenRatio2, wGenRatio1)
                        wGenRatio1<-NULL}
        allRatios<-rbind(allRatios, wGenRatio2)
        wGenRatio2<-NULL}
        allRatios2<-rbind(ddRatio3, allRatios)
        write.table(allRatios2,"Ratios.txt",sep="\t",quote=F, col.names=NA)



###Generate Plots (Trajectories)
dat<-read.table("CompiledSummary.txt", sep="\t", header=T)
forheader<-read.table("CompiledSummary.txt", sep="\t", header=F)
nn<-length(dat)
forheader<-forheader[,4:nn]
graph_title<-forheader[1,]
dat2<-subset(dat, dat$Genotype==Genotype1)
dat2<-dat2[,4:nn]
dat3<-subset(dat, dat$Genotype==Genotype2)
dat3<-dat3[,4:nn]
xaxis<-dat$Treatment[!duplicated(dat$Treatment)]
```

```
no<-length(dat3)
dat4<-NULL
for (i in seq(1,no,2)){
minval1<-dat2[,i]-dat2[,(i+1)]
minval1<-min(minval1)
minval2<-dat3[,i]-dat3[,(i+1)]
minval2<-min(minval2)
minval<-cbind(minval1,minval2)
minval<-min(minval)
maxval1<-dat2[,i]+dat2[,(i+1)]
maxval1<-max(maxval1)
maxval2<-dat3[,i]+dat3[,(i+1)]
maxval2<-max(maxval2)
maxval<-cbind(maxval1,maxval2)
maxval<-max(maxval)
g_range<-range(minval, maxval)
plot_dat<-plotCI(x=dat2[,i], pch=19, uiw=dat2[,(i+1)], lty=1, type="o", gap=0, xaxt="n",
barcol="black", xlab="Treatment",  ylim=g_range, ylab="Relative Normalized Area",
main=graph_title[1,i])
plot_dat<-plotCI(col=par("red"), barcol="red",x=dat3[,i], pch=21, uiw=dat3[,(i+1)],
lty=1, type="p", gap=0, xaxt="n",add=TRUE)
axis(1, at=1:length(xaxis), labels=xaxis)
lines(dat3[,i],lty=2, type="o", lwd=1)
legend(1,g_range[2],c(Genotype1,Genotype2),cex=0.8, pch=c(19,21), lty=1:2)
dat4<-paste("Trajectory_", graph_title[1,i], ".pdf",collapse=NULL)
a<-list(savePlot(dat4, type="eps", dev.cur()))
dev.copy(device=pdf, file=dat4)
dev.off()
dat4<-NULL}
}
```

Appendix 2

VIZR Script

```
vizr <-
function(Dir,fileR,fileS,writeref,writesam,lo,hi,ppm,negval,input){
setwd(Dir)
require(miscTools)
if (input=="mnova") {
read.table(fileR,sep="\t") -> dat
(dat[1,seq(3,ncol(dat),4)]+dat[1,seq(4,ncol(dat),4)])*0.5->e
rev(e)->f
dat[,seq(6,ncol(dat),4)]->z


}
if (input=="acd") {
read.table(fileR,sep="\t") -> z
z[1,c(-1,-2)]->nameR
z[-1,-1]->z

zz <- gsub("\\[|\\.\\.|\\]", "", as.character(z[,1]))
sapply(strsplit(zz, "  "), function(x)mean(as.numeric(x)))->z[,1]
z[order(z[,1],decreasing=TRUE),]->z
rm(zz)
t(z[,1])->e
rev(z[,1])->f
t(z)->z
z[-1,]->z
}

else {(stop ("invalid input type"))}


class(z)<-"numeric"
n<-ncol(z)
Col1<-matrix(f)
Col1<-insertRow(Col1,1,"")
Col1<-t(Col1)
Row1<-matrix(e)
data2<-NULL
data1<-NULL
```

```
write.table(z,"ReferenceFilesprocessed.txt",sep = "\t", quote = F)
DirR<-paste(Dir,"ReferenceMatrix",sep="/",collapse=NULL)
dir.create(DirR)
if(negval){ DirA<-paste(DirR,"Not_AbsVal",sep="/",collapse=NULL)
dir.create(DirA)
}
setwd(DirR)

ncol(z) -> n
nrow(z) -> m


if(writeref) {LstRatios<-list()
for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
data2<-NULL
}

data4<-paste("ReferenceMatrix_",d, ".txt",collapse=NULL)
write.table(data1, file=data4, sep="\t", row.names=F, col.names=F, quote=FALSE,
append=FALSE)
LstRatios[[d]]<-data1
data1<-NULL
}
}else {LstRatios<-list()
for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
data2<-NULL
}

LstRatios[[d]]<-data1
data1<-NULL
}
```

```
}
data_list<-LstRatios
data_files<-LstRatios

y<-Reduce("+",data_list)
g<-y/m
avg_matrix <- g

h<-as.matrix(g)
h<-insertRow(h,1,Row1)
h<-insertCol(h,1,Col1)

write.table(h, "Avg_Matrix.txt", sep="\t")

data_list2 <- NULL
for (i in 1:m){
data_list2[[i]] <- data_list[[i]]-g}
for (i in 1:m){
data_list2[[i]] <- data_list2[[i]]^2}

y <- Reduce("+", data_list2)
y <- y/(m-1)
y <- sqrt(y)
sd_matrix <- y

h<-as.matrix(y)
h<-insertRow(h,1,Row1)
h<-insertCol(h,1,Col1)

write.table(h, "SD_Matrix.txt", sep="\t")

for (b in 1:m){
for (i in 1:n){

for (a in 1:n){

data_files[[b]][(i),(a)]<-(data_files[[b]][(i),(a)])-(avg_matrix[(i),(a)])
}
}
}
data_files_write<-list()
for (b in 1:m){
for (i in 1:n){
```

```
for (a in 1:n){

data_files[[b]][(i),a]<-(data_files[[b]][(i),a])/(sd_matrix[(i),a])
}

}

as.matrix(nameR[b])->dat.name
if(negval){ setwd(DirA)
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreRefMatrix_notAbs_", dat.name, ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
setwd(DirR)
}
data_files[[b]]<-abs(data_files[[b]])
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreRefMatrix_Abs_", dat.name, ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
}

par(cex.axis=0.6, mar=c(5,4,4,5), xpd=TRUE)
for (b in 1:m){
as.matrix(nameR[b])->dat.name
data_files[[b]]->x
Bot<-as.character(f)
Side<-as.character(e)
length(f)->nS
Ticks<-seq(0,1,1/(nS-1))
j<-floor((2*0.06/ppm))
Ticks<-Ticks[seq(j,length(Ticks),j)]
Bot<-Bot[seq(j,length(Bot),j)]
Side<-Side[seq(j,length(Side),j)]

x[-1,-1]->x
```

```
hi->x[x>=hi]
HeatBrk<-seq(lo,hi,((hi-lo)/5))
MyCol<-gray((4:0)/4)
MyCol=c("#FFFFFF","#66FFFF","#3366FF","#663399","#990000")
data5<-paste("ZScoreRefColorHeatmap_",dat.name, ".pdf",collapse=NULL)
gsub(" ", "", data5)->data5
image(x, col=MyCol, breaks=HeatBrk, axes = FALSE)
axis(2,las=2, at=Ticks, labels = Side, tick = TRUE)
axis(1,las=2, at=Ticks, labels = Bot, tick = TRUE)
legend(x=1.0,y=0.8,legend=HeatBrk[-1],fill=MyCol,bty="n",cex=0.5)
dev.copy(device=pdf, file=data5, height=8, width=8)
dev.off()}

setwd(Dir)
DirS<-paste(Dir,"Samples",sep="/",collapse=NULL)
dir.create(DirS)
if(negval){ DirN<-paste(DirS,"Not_AbsVal",sep="/",collapse=NULL)
dir.create(DirN)
}

if (input=="mnova") {
read.table(fileS,sep="\t") -> dat
(dat[1,seq(3,ncol(dat),4)]+dat[1,seq(4,ncol(dat),4)])*0.5->e
rev(e)->f
dat[,seq(6,ncol(dat),4)]->z


}
if (input=="acd") {
read.table(fileS,sep="\t") -> z
z[1,c(-1,-2)]->nameS
z[-1,-1]->z
zz <- gsub("\\[|\\.\\.|\\]", "", as.character(z[,1]))
sapply(strsplit(zz, "  "), function(x)mean(as.numeric(x)))->z[,1]
z[order(z[,1],decreasing=TRUE),]->z
rm(zz)
t(z[,1])->e
rev(z[,1])->f
t(z)->z
z[-1,]->z
}

else (stop ("invalid input type"))
```

```
class(z)<-"numeric"
if (is.matrix(z)=="FALSE"){
as.matrix(z)->z
t(z)->z}
n<-ncol(z)
m<-nrow(z)
Col1<-matrix(f)
Col1<-insertRow(Col1,1,"")
Col1<-t(Col1)
Row1<-matrix(e)
data2<-NULL
data1<-NULL
setwd(DirS)
write.table(z,"SampleFilesprocessed.txt",sep = "\t", quote = F)

if(writesam){LstRatios<-list()

for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
data2<-NULL
}

data4<-paste("SampleMatrix_",d, ".txt",collapse=NULL)

write.table(data1, file=data4, sep="\t", row.names=F, col.names=F, quote=FALSE,
append=FALSE)
LstRatios[[d]]<-data1
data1<-NULL
}
}  else {LstRatios<-list()
for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
```

```
rbind(data1,data2)->data1
data2<-NULL
}

LstRatios[[d]]<-data1
data1<-NULL
}
}

data_files<-LstRatios

for (b in 1:m){
for (i in 1:n){

for (a in 1:n){

data_files[[b]][(i),(a)]<-(data_files[[b]][(i),(a)])-(avg_matrix[(i),(a)])
}
}
}
data_files_write<-list()
for (b in 1:m){
for (i in 1:n){

for (a in 1:n){

data_files[[b]][(i),a]<-(data_files[[b]][(i),a])/(sd_matrix[(i),a])
}

}
as.matrix(nameS[b])->dat.name
if(negval){ setwd(DirN)
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreMatrix_notAbs_", dat.name, ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
setwd(DirS)
}
data_files[[b]]<-abs(data_files[[b]])
data_files_write[[b]]<-data_files[[b]]
```

```
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreMatrix_Abs_", dat.name, ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
}

par(cex.axis=0.6, mar=c(5,4,4,5), xpd=TRUE)
for (b in 1:m){
as.matrix(nameS[b])->dat.name
data_files[[b]]->x
Bot<-as.character(f)
Side<-as.character(e)
length(f)->nS
Ticks<-seq(0,1,1/(nS-1))
j<-floor((2*0.06/ppm))
Ticks<-Ticks[seq(j,length(Ticks),j)]
Bot<-Bot[seq(j,length(Bot),j)]
Side<-Side[seq(j,length(Side),j)]

x[-1,-1]->x
hi->x[x>=hi]
HeatBrk<-seq(lo,hi,((hi-lo)/5))
MyCol=c("#FFFFFF","#66FFFF","#3366FF","#663399","#990000")
data5<-paste("ZScoreColorHeatmap_", dat.name, ".pdf",collapse=NULL)
gsub(" ", "", data5)->data5
image(x, col=MyCol, breaks=HeatBrk, axes = FALSE)
axis(2,las=2, at=Ticks, labels = Side, tick = TRUE)
axis(1,las=2, at=Ticks, labels = Bot, tick = TRUE)
legend(x=1.0,y=0.8,legend=HeatBrk[-1],fill=MyCol,bty="n",cex=0.5)
dev.copy(device=pdf, file=data5, height=8, width=8)
dev.off()}
dev.off()

}

sumZ <-
function(samDir,refDir,omit){
setwd(refDir)
list.files(pattern="ZScoreRefMatrix")->zscoreList_ref
length(zscoreList_ref)->nref
setwd(samDir)
```

```
list.files(pattern="ZScoreMatrix")->zscoreList_sam
length(zscoreList_sam)->nsam
a<-matrix(nrow=(nsam+nref),ncol=3)
setwd(refDir)
for(b in 1:nref){
read.table(zscoreList_ref[b])->x
file.name<-zscoreList_ref[b]
sub(".txt", "", file.name)->file.name
sub("ZScoreMatrix_notAbs_", "", file.name)->file.name
gsub(" ", "", file.name)->file.name
colSums(x,na.rm=TRUE)->y
sum(y)->z
a[b,1]<-"Reference"
a[b,2]<-file.name
a[b,3]<-z
}
setwd(samDir)

for (b in 1:nsam){
read.table(zscoreList_sam[b])->x
file.name<-zscoreList_sam[b]
sub(".txt", "", file.name)->file.name
sub("ZScoreMatrix_notAbs_", "", file.name)->file.name
gsub(" ", "", file.name)->file.name
as.matrix(x)->x
x[is.infinite(x)]<-NA
colSums(x,na.rm=TRUE)->y
sum(y)->z
a[b+nref,1]<-"Sample"
a[b+nref,2]<-file.name
a[b+nref,3]<-z
}
write.table(a,file="Matrix Total Z
Sum.txt",sep="\t",col.names=FALSE,row.names=FALSE)

(mean(as.numeric(a[1:nref,3]))+4*(sd(a[1:nref,3])))->cutoff
(mean(as.numeric(a[1:nref,3]))+2*(sd(a[1:nref,3])))->cutoff2
(mean(as.numeric(a[1:nref,3]))+1*(sd(a[1:nref,3])))->cutoff3
a[(1+nref):(nref+nsam),1:3]->d

if (is.logical(nrow(d)==NULL)){
write.table(d,file="Prioritized Total Z Sum.txt",sep="\t",col.names=FALSE)
}
```

```
else{
d[(as.numeric(d[,3])>cutoff),]->d
d[order(as.numeric(d[,3]),decreasing=TRUE),]->d
SDrow<-cbind("Reference","Mean+4SD",cutoff)
rbind(d,SDrow)->d
write.table(d,file="Prioritized Total Z Sum.txt",sep="\t",col.names=FALSE)
}


if(is.logical(omit)){
if(omit){
stop("omit must be FALSE or a string of values")
}
else{
xval<-c(1:nrow(a))
length(xval)->nn
(nn*0.1)+nn->nn
min(xval)->n
(n-(nn*0.1))->n
par(cex.axis=1, mar=c(5,4,4,5))
plot(x=xval,y=a[,3],pch=20,main="Total Z Sum Value vs Sample",ylab="Summed Z
Value",cex=0.7,xlab="", xlim=c(n,nn))
}
}

else{
a<-a[-omit,]
xval<-c(1:nrow(a))
length(xval)->nn
(nn*0.1)+nn->nn
min(xval)->n
(n-(nn*0.1))->n
par(cex.axis=1, mar=c(5,4,4,5))
plot(x=xval,y=a[,3],pch=20,main="Total Z Sum Value vs Sample",ylab="Summed Z
Value",cex=0.7,xlab="", xlim=c(n,nn))
}
abline(h=cutoff2)
abline(h=cutoff)
abline(h=cutoff3)
abline(v=nref)
axis(side=4,labels=c(expression(1*sigma),expression(2*sigma),expression(4*sigma)),at=
c(cutoff3,cutoff2,cutoff),cex.axis=0.5,las=2)
mtext("Reference Number",side=1,at=(0.5*nref),col="blue",line=3)
```

```
mtext("Sample Number",side=1,at=(nref+0.5*nsam),col="red",line=3)
text(x=xval, y=as.numeric(a[,3])*1.015,a[,2],cex=0.5)
dev.copy(device=pdf,file="Total Summed Z Plot.pdf",height=8,width=8)
dev.off()
dev.off()
}

sampleZ <-
function(Dir,fileS,writesam,refDir,negval,input) {
setwd(refDir)
require(miscTools)
M<-read.table("Avg_Matrix.txt",sep="\t")
avg_matrix<-M[-1,-1]
N<-read.table("SD_Matrix.txt",sep="\t")
sd_matrix<-N[-1,-1]
setwd(Dir)
read.table(fileS,sep="\t") -> dat
DirS<-paste(Dir,"Samples",sep="/",collapse=NULL)
dir.create(DirS)
if(negval){ DirN<-paste(DirS,"Not_AbsVal",sep="/",collapse=NULL)
dir.create(DirN)
}

if (input=="mnova") {
read.table(fileS,sep="\t") -> dat
(dat[1,seq(3,ncol(dat),4)]+dat[1,seq(4,ncol(dat),4)])*0.5->e
rev(e)->f
dat[,seq(6,ncol(dat),4)]->z


}
else {if (input=="acd") {
read.table(fileS,sep="\t") -> z
z[1,c(-1,-2)]->nameS
z[-1,-1]->z

zz <- gsub("\\[|\\.\\.|\\]", "", as.character(z[,1]))
sapply(strsplit(zz, "  "), function(x)mean(as.numeric(x)))->z[,1]
z[order(z[,1],decreasing=TRUE),]->z
rm(zz)
t(z[,1])->e
rev(z[,1])->f
t(z)->z
```

```
z[-1,]->z
}

else (stop ("invalid input type"))
}
setwd(DirS)
n<-ncol(z)
nrow(z)->m
Col1<-matrix(f)
Col1<-insertRow(Col1,1,"")
Col1<-t(Col1)
Row1<-matrix(e)
data2<-NULL
data1<-NULL
write.table(z,"SampleFilesprocessed.txt",sep = "\t", quote = F)

if(writesam){LstRatios<-list()

for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
data2<-NULL
}

data4<-paste("SampleMatrix_",d, ".txt",collapse=NULL)

write.table(data1, file=data4, sep="\t", row.names=F, col.names=F, quote=FALSE,
append=FALSE)
LstRatios[[d]]<-data1
data1<-NULL
}
}  else {LstRatios<-list()
for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
```

```
data2<-NULL
}

LstRatios[[d]]<-data1
data1<-NULL
}
}

data_files<-LstRatios

for (b in 1:m){
for (i in 1:n){

for (a in 1:n){

data_files[[b]][(i),(a)]<-(data_files[[b]][(i),(a)])-(avg_matrix[(i),(a)])
}
}
}
data_files_write<-list()
for (b in 1:m){
for (i in 1:n){

for (a in 1:n){

data_files[[b]][(i),a]<-(data_files[[b]][(i),a])/(sd_matrix[(i),a])
}

}
if(input=="acd"){as.matrix(nameS[b])->dat.name}
if(input=="mnova"){dat.name<-b}
if(negval){ setwd(DirN)
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreMatrix_notAbs_", dat.name, ".txt",collapse=NULL)
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
setwd(DirS)
}
data_files[[b]]<-abs(data_files[[b]])
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
```

```
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreMatrix_",b, ".txt",collapse=NULL)
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
}
}

refZ <-
function(Dir,fileR,writeref,negval,input){

setwd(Dir)
require(miscTools)
if (input=="mnova") {
read.table(fileR,sep="\t") -> dat
(dat[1,seq(3,ncol(dat),4)]+dat[1,seq(4,ncol(dat),4)])*0.5->e
rev(e)->f
dat[,seq(6,ncol(dat),4)]->z


}
else {if (input=="acd") {
read.table(fileR,sep="\t") -> z
z[1,c(-1,-2)]->nameR
z[-1,-1]->z

zz <- gsub("\\[|\\.\\.|\\]", "", as.character(z[,1]))
sapply(strsplit(zz, "  "), function(x)mean(as.numeric(x)))->z[,1]
z[order(z[,1],decreasing=TRUE),]->z
rm(zz)
t(z[,1])->e
rev(z[,1])->f
t(z)->z
z[-1,]->z
}

else (stop ("invalid input type"))
}
n<-ncol(z)
Col1<-matrix(f)
Col1<-insertRow(Col1,1,"")
Col1<-t(Col1)
Row1<-matrix(e)
data2<-NULL
```

```
data1<-NULL
write.table(z,"ReferenceFilesprocessed.txt",sep = "\t", quote = F)
DirR<-paste(Dir,"ReferenceMatrix",sep="/",collapse=NULL)
dir.create(DirR)
if(negval){ DirA<-paste(DirR,"Not_AbsVal",sep="/",collapse=NULL)
dir.create(DirA)
}
setwd(DirR)
ncol(z) -> n
nrow(z) -> m

if(writeref) {LstRatios<-list()
for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
data2<-NULL
}

data4<-paste("ReferenceMatrix_", d, ".txt",collapse=NULL)
write.table(data1, file=data4, sep="\t", row.names=F, col.names=F, quote=FALSE,
append=FALSE)
LstRatios[[d]]<-data1
data1<-NULL
}
}else {LstRatios<-list()
for (d in 1:m) {
for (a in n:1) {
for (b in 1:n) {
cbind(data2, ((z[d,a])/(z[d,b])))->data2

}
rbind(data1,data2)->data1
data2<-NULL
}

LstRatios[[d]]<-data1
data1<-NULL
}
}
```

```
data_list<-LstRatios
data_files<-LstRatios

y<-Reduce("+",data_list)
g<-y/m
avg_matrix <- g

h<-as.matrix(g)
h<-insertRow(h,1,Row1)
h<-insertCol(h,1,Col1)

write.table(h, "Avg_Matrix.txt", sep="\t")

data_list2 <- NULL
for (i in 1:m){
data_list2[[i]] <- data_list[[i]]-g}
for (i in 1:m){
data_list2[[i]] <- data_list2[[i]]^2}

y <- Reduce("+", data_list2)
y <- y/(m-1)
y <- sqrt(y)
sd_matrix <- y

h<-as.matrix(y)
h<-insertRow(h,1,Row1)
h<-insertCol(h,1,Col1)

write.table(h, "SD_Matrix.txt", sep="\t")

for (b in 1:m){
for (i in 1:n){

for (a in 1:n){

data_files[[b]][(i),(a)]<-(data_files[[b]][(i),(a)])-(avg_matrix[(i),(a)])
}
}
}
data_files_write<-list()
for (b in 1:m){
for (i in 1:n){
```

```
for (a in 1:n){

data_files[[b]][(i),a]<-(data_files[[b]][(i),a])/(sd_matrix[(i),a])
}

}
if(input=="acd"){
as.matrix(nameR[b])->dat.name
}
if(input=="mnova"){b->dat.name}
if(negval){ setwd(DirA)
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreRefMatrix_notAbs_",dat.name, ".txt",collapse=NULL)
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
setwd(DirR)
}
data_files[[b]]<-abs(data_files[[b]])
data_files_write[[b]]<-data_files[[b]]
data_files_write[[b]]<-insertRow(data_files_write[[b]],1,Row1)
data_files_write[[b]]<-insertCol(data_files_write[[b]],1,Col1)
data4<-paste("ZScoreRefMatrix_",dat.name, ".txt",collapse=NULL)
write.table(data_files_write[[b]], file=data4, sep="\t", row.names=F, col.names=F,
quote=FALSE, append=FALSE)
}

}

pspec <-
function(Dir,refDir,rfileR,rfileS,negval,mag,xax,type){
require(miscTools)
setwd(refDir)
as.matrix(read.table(rfileR,sep="\t",header=TRUE))->ref


ref[,-1]->ref
((rowSums(ref,na.rm=TRUE))/(ncol(ref)))->avgy
length(avgy)->n
colnames(ref)->refNames
refNames[-1]->refNames
ref[,1]->avgx
```

```
length(avgx)->n

setwd(Dir)
as.matrix(read.table(rfileS,sep="\t",header=TRUE))->rsam

rsam[,-1]->spy
colnames(spy)->samNames

ncol(spy)->m
Sxargs<-spy[,1]
length(Sxargs)->n
DirS<-paste(Dir,"Samples",sep="/",collapse=NULL)
if(negval){DirS<-paste(DirS,"Not_AbsVal",sep="/",collapse=NULL)}
setwd(DirS)


for (b in 2:m){
par(mfrow=c(3,1))
yargs<-spy[,b]
max(yargs)->maxval
maxval/mag->ycut
plot(x=Sxargs,y=yargs,type="l",main=paste("Raw NMR
Spectrum",samNames[b],collapse=NULL,sep="
"),xlab="ppm",ylab="Intensity",lwd=0.2,ylim=c(0,ycut),xlim=xax,xaxt="n")
axis(side=1, at=seq(0.5,9,0.5), labels=seq(0.5,9,0.5), tick=TRUE)

if(type=="ref"){
if(negval){data4<-paste("ZScoreRefMatrix_notAbs_",samNames[b],
".txt",collapse=NULL)
gsub(" ", "", data4)->data4
}
else{data4<-paste("ZScoreRefMatrix_",samNames[b], ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
}
}
if(type=="sam"){
if(negval){data4<-paste("ZScoreMatrix_notAbs_",samNames[b], ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
}
else{data4<-paste("ZScoreMatrix_",samNames[b], ".txt",collapse=NULL)
gsub(" ", "", data4)->data4
}
}
```

```
as.matrix(read.table(data4,sep="\t"))->x
xargs<-t(x[-1,1])
x[-1,-1]->x
x[is.infinite(x)]<-NA
yargs<-rowSums(x,na.rm=TRUE)
plot(x=xargs,y=yargs,type="l",main=paste("Z-score
Projection",samNames[b],collapse=NULL,sep=" "),xlab="ppm",ylab="Z
Sum",xaxt="n",xlim=xax)
axis(side=1, at=seq(0.5,9,0.5), labels=seq(0.5,9,0.5), tick=TRUE)
plot(x=avgx,y=avgy,type="l",main="Library Averaged Reference
Spectrum",xlab="ppm",ylab="Intensity",lwd=0.2,ylim=c(0,ycut),xaxt="n",xlim=xax)
axis(side=1, at=seq(0.5,9,0.5), labels=seq(0.5,9,0.5), tick=TRUE)
data6<-paste("ZSpectra_",samNames[b],".pdf",collapse=NULL)
dev.copy(device=pdf,file=data6,height=8,width=8)
dev.off()
}
}


heatZ <-
function(DirH,lo,hi,m,ppm,clr,type){
setwd(DirH)
if(type=="ref"){list.files(pattern="ZScoreRefMatrix")->name.list}
if(type=="sam"){list.files(pattern="ZScoreMatrix")->name.list}
par(cex.axis=0.6, mar=c(5,4,4,5), xpd=TRUE)
length(name.list)->m
for (b in 1:m){
data4<-name.list[b]
as.matrix(read.table(data4,sep="\t"))->x
Bot<-as.character(x[-1,1])
Side<-as.character(x[1,-1])
length(Side)->nS
Ticks<-seq(0,1,1/(nS-1))
j<-ceiling((2*0.06/ppm))
Ticks<-Ticks[seq(j,length(Ticks),j)]
Bot<-Bot[seq(j,length(Bot),j)]
Side<-Side[seq(j,length(Side),j)]

x[-1,-1]->x
hi->x[x>=hi]
HeatBrk<-seq(lo,hi,((hi-lo)/5))
if(clr){
MyCol=c("#FFFFFF","#66FFFF","#3366FF","#663399","#990000")
```

```
}
else{
MyCol<-gray((4:0)/4)

}

data5<-paste("Heatmap_",data4, ".pdf",collapse=NULL)


image(x, col=MyCol, breaks=HeatBrk, axes = FALSE)
axis(2,las=2, at=Ticks, labels = Side, tick = TRUE)
axis(1,las=2, at=Ticks, labels = Bot, tick = TRUE)
legend(x=1.0,y=0.8,legend=HeatBrk[-1],fill=MyCol,bty="n",cex=0.5)
dev.copy(device=pdf, file=data5, height=8, width=8)
dev.off()}
}
```