Novel Methods for Identifying and Quantifying Metabolites in Complex Biological Extracts by Multidimensional Nuclear Magnetic Resonance Spectroscopy

by

Ian A. Lewis

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2010

To my fiancée Jennifer

TABLE OF CONTENTS

ACKNOWLEDGMENTS

GLOSSARY

1D: one-dimensional

2,3-BPG: 2,3-Bisphosphoglycerate

2D: two-dimensional

Bioanalytical Metabolomics: comprehensive quantitative analysis of metabolites in complex biological samples

BMRB: BioMagResBank, a repository of NMR data of metabolites collected under standardized conditions

D_2O: deuterium oxide, a solvent that is frequently used for NMR-based analysis of aqueous extracts

DETOCS: Difference Edited Total Correlation Spectroscopy, a homonuclear ^1H NMR pulse program that separates signals from ^{13}C- and ^{12}C-bound protons into distinct spectra

DSS: (4,4-dimethyl-4-silapentane-1-sulfonic acid), a compound used to reference chemical shifts in NMR spectra

EDTA: ethylenediaminetetraacetate, a chelator that binds to Fe(III) with high affinity

FMQ by NMR: fast metabolite quantification by nuclear magnetic resonance spectroscopy, a method for quantifying metabolites in NMR spectra and strategy for shortening the acquisition times of multidimensional experiments

GE: glycolytic enzymes

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, an organic buffer that is used as an internal pH reference in NMR-based bioanalytical metabolomics

HMDB: Human Metabolome Database, a repository of MS and NMR data and bioinformatics tools related to human metabolism

HSQC: Heteronuclear Single Quantum Coherence, a heteronuclear NMR pulse program that is frequently used for acquiring 2D ^{1}H-^{13}C NMR spectra of metabolites

MES: 2-(N-morpholino)ethanesulfonic acid, an organic buffer that is used as an internal concentration reference for NMR-based bioanalytical metabolomics

Metabolomics: The study of metabolites in complex biological extracts

MMC: Madison Metabolomics Consortium, a collection of University of Wisconsin laboratories that contributed to the metabolite standards libraries used by the BMRB and MMCD

MMCD: Madison Metabolomics Consortium Database, a bioinformatics tool for identifying metabolites in MS and NMR spectra

MS: mass spectrometry

NMR: nuclear magnetic resonance spectroscopy

PPP: pentose phosphate pathway

RBC: red blood cell

rNMR: an open source software package for NMR data analysis

ROI: region of interest, a subsection of an NMR spectrum

TOCSY: Total Correlation Spectroscopy, a homonuclear NMR pulse program that is frequently used for collecting 2D ^{1}H-^{1}H NMR data of metabolites

Novel Methods for Identifying and Quantifying Metabolites in Complex Biological Extracts by Multidimensional Nuclear Magnetic Resonance Spectroscopy

Ian A. Lewis

Under the supervision of Professor John L. Markley

University of Wisconsin-Madison

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful tools available for identifying and quantifying metabolites in simple mixtures. However, NMR's analytical utility is largely negated by signal overlap, which is inherent to NMR spectra of complex biological extracts. Consequently, NMR-based metabolomics studies are rarely able to identify the individual components of mixtures. Over the past ten years, roughly 1,000 articles have been published on NMR-based metabolomics. These studies generally rely on multivariate statistics for deciphering the overlapped spectra. I have developed an alternative strategy, bioanalytical metabolomics, which capitalizes on state-of-the-art multidimensional NMR to minimize resonance overlap. The bioanalytical metabolomics strategy allows up to 90% of the NMR observable metabolites to be identified and accurately quantified. Moreover, this approach allows hypothesis-driven research to be conducted on larger scale than was previously possible. In this thesis, I present the tools and techniques that I have developed for solving practical problems associated with the new technology and provide examples of hypothesis-driven research conducted with the bioanalytical metabolomics strategy.

CHAPTER 1

Introduction

Metabolomics is the study of metabolites in complex biological extracts. Metabolomics differs from traditional bioanalytical chemistry and natural products chemistry in scope; metabolomics encompass all observable metabolites whereas traditional methods target a predefined subset compounds. Metabolomics is generally used for biomarker discovery, classifying samples, and other hypothesis-independent studies. Although the word "metabolomics", and its synonym "metabonomics", are relatively new (Nicholson, Lindon et al. 1999; Fiehn, Kopka et al. 2000), the concept these words represent has been in active use since the early 1970's (Pauling, Robinson et al. 1971).

Nuclear magnetic resonance (NMR) spectroscopy is a popular analytical tool for metabolomics. NMR is desirable because it provides an unbiased window into the composition of mixtures; the intensities of NMR signals are linearly proportional to concentration, NMR can detect almost any soluble organic compound present at sufficient concentrations, NMR can determine the molecular structures of compounds, and it can differentiate between isomers. Until recently, however, technological limitations have prevented metabolomics studies from fully capitalizing on these inherent strengths of NMR.

NMR spectra of complex biological extracts contain hundreds to thousands of signals. These signals are distributed over a finite bandwidth that is dictated by the design of the NMR pulse program and the physical properties of the nuclei detected in the experiment. In one dimensional (1D) ^{1}H NMR spectra, the most popular experiment for NMR-based metabolomics, metabolite signals are confined to a narrow bandwidth (~10 ppm). As a result, the majority of signals overlap with resonances from other compounds. Overlapping NMR signals are difficult to assign and are quantitatively unreliable (Lewis, Schommer et al. 2007). Consequently, most metabolomics studies do not attempt to identify and quantify the individual components of extracts.

The established NMR-based approach for analyzing spectra relies on multivariate statistics for interpreting overlapped data. Principle components analysis, or any one of a variety of multivariate techniques (Lindon, Holmes et al. 2001), is used to generate spectral fingerprints associated with a particular biological condition. These multivariate signatures are useful for classifying unknown samples or can be analyzed directly to find spectral biomarkers for disease. Although the established statistics-based methods for analyzing data have proven themselves to be reproducible (Dumas, Maibaum et al. 2006), they provide little information on the molecular composition of extracts, and thus, do not benefit from NMR's analytical strengths.

A wide variety of multidimensional NMR pulse sequences have been developed by biomolecular NMR spectroscopists to mitigate problems associated with peak overlap in complex spectra. Multidimensional NMR is of obvious utility to metabolomics, but has been avoided because of serious practical challenges associated with this technique. Although two dimensional (2D) ^{1}H-^{1}H and ^{1}H-^{13}C NMR experiments significantly reduce the number of overlapped resonances, the traditional methods for interpreting 2D NMR data are labor intensive, time consuming, and require considerable NMR expertise (Lewis, Schommer et al. 2009). Furthermore, the intensities of multidimensional NMR signals are affected by a wide range of factors unrelated to the concentrations of molecules (Lewis, Schommer et al. 2007; Lewis, Karsten et al. 2010; Lewis, Shortreed et al. 2010). The difficulty in assigning and quantifying multidimensional NMR data has made this technique essentially inaccessible to the metabolomics community.

The goal of my Ph.D. research was to develop a practical multidimensional NMR strategy for identifying and quantifying metabolites in complex biological extracts. In collaboration with various members of the Madison Metabolomics Consortium (MMC), I (1) helped develop an extensive collection of NMR spectra of metabolites (Markley, Anderson et al. 2007), (2) participated in the creation of a bioinformatics tool for identifying metabolites (Cui, Lewis et al.

2008), (3) devised a general method for measuring molar concentrations of molecules by multidimensional NMR (Lewis, Schommer et al. 2007), (4) wrote a NMR data analysis software package (Lewis, Schommer et al. 2009), (5) developed an NMR pulse sequence and analytical strategy for quantifying metabolites on the basis of isotopic enrichment (Lewis, Karsten et al. 2010), (6) devised a simple method for standardizing differential T_1 relaxation of NMR signals (Lewis, Karsten et al. 2010), and (7) invented a device for extracting metabolites from biological tissues (Appendix II).

These tools have dramatically simplified comprehensive metabolic analyses and allow metabolomics to capitalize on state-of-the-art NMR technology. Moreover, these tools have been integrated sufficiently to make comprehensive metabolic analysis a feasible alternative to the statistics-based metabolomics (Lewis, Shortreed et al. 2010). This comprehensive approach, which I refer to as bioanalytical metabolomics, allows hypothesis-driven research to be conducted in the context of metabolomics.

This thesis is a compilation of publications that document the most successful tools and techniques I developed over the course of my PhD research. For each chapter, I have provided a brief description of the role I played and the significance of the research in the context of bioanalytical metabolomics. I have also included three studies that I conducted on erythrocyte metabolic regulation (Chapters 8-10). These studies serve as examples of hypothesis-driven research conducted using the bioanalytical metabolomics strategy.

CHAPTER 2

Novel NMR and MS approaches to metabolomics

Adapted from:

Ian A. Lewis, Michael R. Shortreed, Adrian D. Hegeman and John L. Markley. Novel NMR and MS approaches to metabolomics. <u>Handbook of Metabolomics</u> (in press)

This chapter provides an overview of my up-to-date protocols, most firmly established analytical tools, and my current thinking on modern bioanalytical metabolomics. I originally wrote this section as a practical guide to the numerous technological advances of recent years. With the exception of the MS section, which was contributed by Dr. Shortreed and Dr. Hegeman, the material presented here was derived from my own research. Several sections in this chapter overlap with material from Chapters 3-5, I have included the overlapping material here because it is presented in a more general context and because some of my methods have evolved since their original publication.

2.1. Abstract

Identifying and quantifying metabolites in complex biological samples is one of the most challenging aspects of metabolomics. Recently, several important advances in databases, software, instrumentation, and laboratory techniques have greatly simplified the most laborious tasks of metabolite identification and have made quantification more reliable. These technological advances have made bioanalytically-oriented studies a feasible alternative to the statistics-based methods commonly used for metabolomics. We discuss the tools that have become most important in our own research and comment on emerging technologies that may play an important role in future studies. In addition, we provide practical guidelines for designing studies and give the step-by-step protocols used in our lab for sample preparation, metabolite identification, and accurate quantification of molecules.

2.2. Bioanalytical metabolomics

All metabolomics studies involve elements of natural products chemistry, analytical chemistry, and statistics. The degree to which each of these elements is weighted influences experimental design and the type of data that is ultimately derived from a study. Currently, most metabolomics investigations emphasize statistics. Spectroscopic data derived from these studies are deciphered by using sophisticated multivariate tools, and potential biomarkers are identified on the basis of their statistical significance (Lindon, Holmes et al. 2007). Although multivariate analyses are effective for classifying samples, they do not provide a transparent mechanism for identifying and quantifying individual metabolites.

Bioanalytical metabolomics is an emerging strategy that emphasizes comprehensive metabolite assignment, accurate measures of concentration, and transparent data analyses that minimize the use of statistics. Although these characteristics are similar to traditional metabolism research (Radda and Seeley 1979; Shulman, Brown et al. 1979), bioanalytical

metabolomics differs from traditional studies in scope. Unfractionated biological samples contain hundreds (NMR) to thousands (MS) of observable signals. Whereas traditional methods restrict analyses to a few pre-determined metabolites, bioanalytical metabolomics attempts to identify and quantify all of the observable signals. Until recently, this type of comprehensive metabolite profiling was too labor intensive to be practical in routine analyses. However, the tools and techniques discussed in this chapter have dramatically simplified the laborious aspects of data analysis and have made bioanalytical metabolomics a practical alternative to statistics-based studies.

2.3. Sample preparation

Consistent sample preparation is an important component of bioanalytical metabolomics. Although a wide variety of techniques are effective, no protocol is appropriate for all metabolites. Extraction conditions, such as solvent temperature and hydrophobicity, directly affect the molecules that can be identified and quantified in a study. Aqueous solvents extract hydrophilic molecules; non-polar solvents extract hydrophobic metabolites; harsh conditions (acids, bases and boiling) promote unwanted chemistry at labile functional groups; gentle conditions (cold methanol-water) are less effective for denaturing proteins. In short, every procedure alters one's perception of *in vivo* metabolism. The goal is to find a reproducible method that preserves the metabolites relevant to a particular study. In this section, we present some general guidelines for minimizing technical error and provide the sample preparation protocol we use for routine NMR-based metabolomics studies of aqueous metabolites.

2.3.1 Methods for minimizing technical error

Metabolite concentrations observed in tissue extracts and biological fluids can vary considerably across a dataset. These sample-to-sample differences originate from both natural variability and technical error. Whereas biological variance is essential for interpreting metabolic differences between samples, technical error is simply an obstacle to meaningful data analysis. Although there are many sources of error in metabolomics studies, the main contributors are inconsistent sample preparation and technical shortcomings in analytical equipment (e.g. resonance overlap, ion suppression, and imprecision in peak picking). Careful experimental design can control these sources of error.

Extraction solution conditions. Variation in the amount biological material used to prepare each sample is one of the primary sources of technical error in metabolomics. Inconsistency at this level is directly proportional to quantitative error in the final analysis. Sample-to-sample differences also affect solution conditions (osmolarity, pH, etc.) of the extraction buffer and analytical medium. These second-order effects alter the extraction efficiencies of sparingly soluble compounds and complicate data analysis.

A direct method for controlling variation in sample size is to aliquot tissues on the basis of mass. This approach is most effective in studies involving large sample sizes and becomes increasingly error-prone as sample size diminishes. Animal tissues can be weighed directly on an analytical balance whereas plant samples generally require prior lyophilization to standardize water content. An alternative strategy, which is appropriate for cell cultures, is to prepare samples with uniform optical density and aliquot samples on the basis of volume. We have found this method to be less consistent, but considerably more convenient, than centrifuging cell suspensions and aliquoting samples on the basis of mass.

The volume of solvent present in the extraction medium can also contribute to technical error, particularly in extractions involving volatile organic solvents or high temperatures. Solvent loss due to evaporation can be minimized by using sealed reaction vials. We have found 22 mL screw-top vessels to be convenient for parallel extractions; they allow 16 or more samples to be incubated, centrifuged, filtered and lyophilized simultaneously with conventional laboratory instrumentation.

Preparation of analytical solutions. Solution conditions in the analytical medium are another major contributor to technical error. Salinity, pH, and the concentrations of metal ions affect the sensitivity of NMR spectrometers and the efficacy bioinformatics-based resonance assignments. The strategy for standardizing these conditions is analogous to the methods used for standardizing extractions; samples need to be prepared with a consistent solute to solvent ratio.

The most reliable method for ensuring consistent analytical conditions is to analyze raw extracts without any additional sample preparation. Unfortunately, metabolites present in unconcentrated extracts are generally too dilute for NMR analyses. Furthermore, extractions typically employ [1]H-containing solvents, which are another potential source of error in NMR analyses (e.g. spectral overlap, radiation damping, and receiver overflow resulting from incomplete solvent suppression). Consequently, NMR-based studies often prepare analytical solutions from dried extracts dissolved at relatively high concentrations in perdeuterated solvents. Though necessary, this strategy introduces some additional complications into sample preparation and data analysis.

We employ two strategies for preparing dried extracts: 1) dissolving all samples in a fixed volume of solvent and 2) preparing solutions on a mass to volume basis. Of the two approaches, the fixed volume method is preferable, because it requires the least amount of

sample handling. In studies involving easily standardized extractions, the fixed volume approach yields technical error of less than 5% in our hands. The alternative mass-to-volume approach is appropriate for studies involving large samples (>40 mg extract) with major sample-to-sample differences in salinity. In these cases, we prepare analytical solutions from dried extracts dissolved in 8-17 µL of perdeuterated solvent per mg of sample.

Titrating pH in analytical solutions. Rigorous pH control is an essential component of bioanalytical metabolomics. All of the existing tools for bioinformatics-based metabolite identification and resonance assignment require samples to match the solution conditions used in the public databases (Cui, Lewis et al. 2008). Minor deviations in pH (0.01) alter NMR chemical shifts and pH-dependent exchange broadening alters the intensities of metabolites that are near their pK_a. These effects can have a significant impact on the amount of time required to assign spectra and the reliability of quantitative analyses.

In our experience, simply adding a buffer to the analytical medium (up to 25 mM HEPES) is insufficient for controlling pH to the degree required for automated resonance assignment. Each sample must be hand titrated using a pH meter equipped with a small electrode. We adjust sample pH using concentrated acid or base (~1 M DCl or ~1M NaOD) to minimize titration-related volume changes. For a trained technician, titrating 400 µL sample to 7.400 ± 0.004 requires about three minutes. Although titration is tedious, it saves considerable time overall because it allows labor-intensive data analysis to be replaced with automated resonance assignments.

Internal standards. The use of internal standards is one of the simplest means of ensuring data quality. Although highly reproducible data can be collected without an internal standard, small changes in NMR line shape, sample dilution, and salt concentration affect the intensities of NMR signals. Standards control for these variables and allow biologically relevant variation to

be distinguished from technical error. Furthermore, clever use of internal standards allows one to calculate metabolite concentrations relative to the amount of starting material.

Our strategy is to extract 400 mg of dry weight tissue in 16 mL of water containing 167 µM of an internal standard. This approach allows us to relate metabolite concentrations observed in the NMR tube to the dry weight of the tissue. We typically use HEPES or MES as an internal standard because these compounds have multiple peaks that are well isolated from biological resonances in 2D ^{1}H-^{13}C NMR spectra. HEPES is convenient because it acts as an internal pH indicator (Fig. 2.1), but MES is probably a better concentration reference because of its lower pK_a. Neither of these compounds would be suitable for 1D ^{1}H or 2D homonuclear ^{1}H experiments because their resonances overlap with those of many biological compounds.
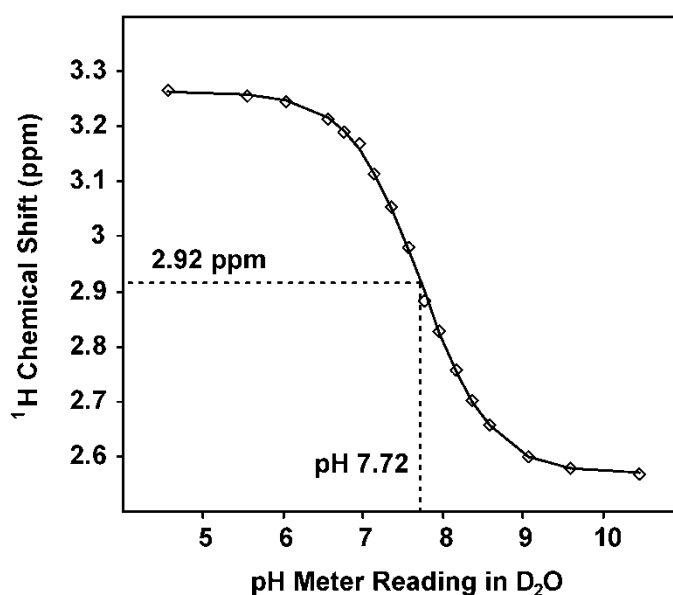
Figure 2.1. The ^{1}H chemical shifts of several HEPES peaks are pH sensitive and can be used as in internal pH reference. The titration curve was generated from a HEPES buffered saline solution (290 milliosmol) in D$_2$O. The dotted lines indicate the observed pKa of HEPES as an uncorrected meter reading in D$_2$O.

Although we have only applied the internal standard normalization strategy to aqueous extractions, the principle should be applicable to other types of extractions. In some cases, biological internal standards may already be present in the data. Plants grown in MES buffer, for example, accumulate MES proportionally to their dry mass (Fan, Lane et al. 2001; Lewis, Schommer et al. 2007). Non-metabolized biological internal standards are extremely useful provided that their normal *in vivo* concentrations can be calculated.

Entropy in sample order. Metabolite signals observed in either NMR or MS spectra can be influenced by the order of sample analysis. In the case of NMR data, systematic changes in shimming over the course of many samples results in different (usually worse) line shape. While these variations can, in theory, be corrected by normalization to an internal standard, data should always be collected in a random order to minimize systematic error. Furthermore, all NMR and MS data should be collected with technical replicates. Although technical replication is routine in the MS world, the NMR community has tended to avoid technical replication on the grounds that NMR yields high technical reproducibility. While NMR analysis can be made extremely reproducible through careful shimming, we have found that data collected with automated sample changers can have considerable variation in line shape. Technical replication and randomized sample order minimize the chance of systematic errors.

Errors in error bars. Careful scrutiny of the various studies that have quantified the speed of light since 1676 indicates that the value of c appears to have changed significantly over the last few centuries (Brown, Pais et al. 1995). Although these results could be interpreted as an exciting physical phenomenon, the more realistic explanation is that the error bars given on the light speed estimates were too low. Misleading error estimates are not unique to measurements of universal constants. In a landmark 1984 paper, Stuart Hurlbert showed that about half of the inferential statistics published in ecology between 1960 and 1984 were based on questionable data replication (Hurlbert 1984). This problem also applies to metabolomics. Confusing technical

replicates with biological replicates grossly misrepresents real biological variation and will inevitably lead to erroneous interpretations of statistical tests. Biologically meaningful investigations require careful experimental design with respect to replication. Although Hurlbert's paper was written for ecology, his recommendations for experimental design are directly applicable to the design of metabolomics experiments. Mehta and coworkers have also published a review on this topic that addresses the problems encountered in high-throughput studies (Mehta, Tanik et al. 2004).

2.3.2 Protocol for extracting high-abundance aqueous metabolites

The majority of research topics encountered in our laboratory involve studies of highly abundant water-soluble metabolites. The protocol presented here is a general method for preparing samples that is suited to NMR-based assays of amino acids, sugars, polyamines, polyols, and other thermally stable molecules found at high abundance in tissue extracts. We prefer this protocol because it is easy to parallelize, produces low technical error with relatively high yields, and results in a similar complement of NMR-observable metabolites from a wide variety of samples (Lewis, Schommer et al. 2007). The disadvantages of the protocol include its specificity to hydrophilic molecules and poor recovery of thermally unstable compounds (e.g. ATP, and other phosphorylated intermediates). However, many of the highly abundant metabolites found in tissue extracts are thermally stable and suitable to this preparation method. Of the amino acids we have observed, only glutamine (which undergoes thermal conversion to pyroglutamate) is significantly degraded.

1) Cryogenically homogenize tissue in a ball mill or mortar and pestle.
2) Lyophilize homogenized samples for 24 hours.
3) Aliquot 400 mg dry weight tissue samples into 22 mL screw-top reaction vessels.

4) Add 16 mL of boiling ddH$_2$O containing 167 µM HEPES or MES. HEPES or MES is used as an internal concentration reference and will have a final concentration of 1 to 5 mM in the final NMR analysis solution (depending on the volume of D$_2$O used).

5) Suspend sealed sample vials in a boiling water bath for 7.5 minutes.

6) Cool samples on ice for 10 minutes, then centrifuge reaction vials in a swing bucket centrifuge for 30 minutes at 1,000 x g.

7) Harvest supernatant and pressure filter the mixture through glass wool to remove any remaining particulate matter.

8) Microfilter the metabolite extract with a 3 kDa molecular weight cutoff spin concentrator to remove soluble proteins. Microfilters must be thoroughly washed prior to this step to remove glycerol from the membrane surface. We wash 25 mL spin concentrators by running 100 mL of water through the filters prior to use; even this amount of washing leaves detectable levels of glycerol in the final sample.

9) Lyophilize the metabolite filtrates to a dry powder.

10) Dissolve dried metabolite powder in a fixed volume of NMR solvent (D$_2$0 containing 500 µM DSS and 500 µM NaN$_3$). We use 800 µl for most tissues, but this volume is dependent the salinity of the tissue. The object with this step is to concentrate samples as much as possible without leaving a precipitate or generating excessively high salt concentrations (which will be evident from long 90º NMR pulse lengths). All samples related to a study must be prepared using the same volume of solvent.

11) Titrate samples with concentrated acid or base (~1 M DCl or ~1M NaOD) to a pH of 7.400 +/- 0.004.

2.4. Resources for NMR- and MS-based metabolomics

Traditional methods for identifying and quantifying molecules rely on visual inspection of data and hand assignment of signals. Although this approach is effective when applied to simple

mixtures, it is too labor intensive to be practical for comprehensive analysis of complex biological extracts. Modern bioanalytical metabolomics relies on bioinformatics, databases of metabolite standards, and specialty software, to make comprehensive analyses a more tractable challenge. Recent developments in these tools have dramatically improved the efficiency and reliability of analyses and have made bioanalytical metabolomics an increasingly popular research strategy. In this section, we focus on the tools we have found most effective in our own research; a more complete listing of useful resources can be found in Table 2.1.

Table 2.1. List of freely accessible metabolomics resources

Standards Initiative:
Metabolomics Standards Initiative (http://msi-workgroups.sourceforge.net/)
Small molecule databases:
ChemIDplus (http://chem.sis.nlm.nih.gov/chemidplus)
Human Metabolome Database (www.hmdb.ca)
Madison Metabolomics Consortium Database (http://mmcd.nmrfam.wisc.edu)
Metlin (http://metlin.scripps.edu)
PubChem (http://pubchem.ncbi.nlm.nih.gov)
Metabolic pathway databases:
BioCyc (www.biocyc.org)
ExPASy (www.expasy.ch/cgi-bin/search-biochem-index)
KEGG (www.genome.jp/kegg)
Reactome (www.reactome.org)
TAIR (www.arabidopsis.org)
UM-BBD (http://umbbd.msi.umn.edu)
Laboratory Information Management Systems (LIMS):
Sesame LIMS (www.sesame.wisc.edu)
SetupX (http://fiehnlab.ucdavis.edu/projects/binbase_setupx)
NMR and MS databases:
BioMagResBank (BMRB) (www.bmrb.wisc.edu)
Human Metabolome Database (www.hmdb.ca)
Madison Metabolomics Consortium Database (http://mmcd.nmrfam.wisc.edu)
Mass Spectrometry Database Committee (www.ualberta.ca/~gjones/mslib.htm)
NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)
NMR metabolomics database of Linkoping (http://www.liu.se/hu/mdl/main)
NMRShiftDB (http://www.nmrshiftdb.org)
NMR Data Analysis software:
MetaboMiner (http://wishart.biology.ualberta.ca/metabominer/)
rNMR (http://rnmr.nmrfam.wisc.edu/)

2.4.1. Databases and bioinformatics

For NMR-based studies, one of the most important advances in recent years has been the introduction of three libraries of experimental data collected on pure metabolite standards. Although several commercial and public NMR libraries have been in existence for years, the previous databases were either not curated, contained data collected under a variety of conditions (Steinbeck, Krause et al. 2003), or were not focused on biologically relevant molecules (Sadtler 1967). Furthermore, all of the earlier resources were restricted to 1D NMR data. In the last few years, the Madison Metabolomics Consortium (MMC), the Human Metabolome Project (HMP), and Bruker have expended considerable resources on collecting NMR data of standard compounds. These libraries are distinct from their predecessors in that they were collected under defined conditions, include a wide variety of NMR experiments and, in the case of the MMC and HMP data, are freely available. Between the MMC and HMP, spectra of about 1,000 metabolites are freely available over the web. These data have made it possible to replace many of the time consuming steps of metabolite identification and resonance assignment with bioinformatics and have paved the way for quantitative NMR-based metabolomics.

For MS-based researchers, the state of experimentally derived data libraries is more complicated. GC-MS researchers enjoy a high degree of standardization with respect to commercial instruments, instrumental settings, retention time reporting, and mass fragmentation. As a result, several large, high-quality, libraries (e.g. NIST Standard Reference Database) are available as well as specialized metabolomics libraries (e.g. BinBase from Oliver Fiehn's laboratory). LC-MS has yet to achieve the same level of standardization enjoyed by GC-MS. As a result, experimental libraries of LC-MS-observed retention times, exact masses, and fragmentation patterns are of more limited use. None the less, the HMP has made a

considerable effort in standardizing data collection and reporting of LC-MS data and has collected spectra of about 2,000 metabolites.

BioMagResBank (BMRB). The BMRB has served the NMR community for many years as the world repository of NMR data related to proteins and nucleic acids (Seavey, Farr et al. 1991; Doreleijers, Mading et al. 2003). Recently, the BMRB expanded its archives to include spectra of small molecules collected under standardized conditions (Markley, Anderson et al. 2007; Ulrich, Akutsu et al. 2008). The defining characteristics of the BMRB for metabolomics are as follows: the data are freely available, data entries are curated, all of the solvent and NMR spectral parameters are clearly defined, and the raw spectral data (in addition to peak-picked and processed spectra) can be downloaded. Having the raw spectral data at hand is important because it allows spectra of standards to be overlaid over extracts for hand verification of metabolite assignments.

Currently, the BMRB contains more than 5000 NMR spectra of 800 metabolites collected by the Madison Metabolomics Consortium (MMC).[1] These data were tailored to tissue-based metabolomics analyses in the design of their solvent conditions (D_2O, pH 7.400 ± 0.004, 50 mM $NaPO_4$ for water soluble metabolites; $CDCl_3$ for organic soluble metabolites) and have shown to be useful for identifying metabolites in complex ^1H-^{13}C NMR spectra of a variety of extracts (Lewis, Schommer et al. 2007). Although all of the data currently available from BMRB have been contributed by the MMC, the BMRB and Human Metabolome Project (HMP) are working toward incorporating the extensive HMP archives into the BMRB.

The BMRB offers several bioinformatics tools to enhance its function as a data repository. In contrast to the tools offered by the HMDB and MMCD, which are best suited to broader queries of the metabolite literature, the BMRB tools are primarily designed to make all of the archived

---

[1] Data collection efforts are ongoing; these numbers are current as of March 2010.

data easily accessible. An NMR peak query provides a rapid means of locating standards. Similarly, MS tools allow researchers to translate exact masses into molecular formulae or locate records on the basis of monoisotopic masses. In summary, the BMRB provides an extensive collection of freely available, high-quality NMR data coupled with an efficient query system. The BMRB database is available over the web at www.bmrb.wisc.edu.

Human Metabolome Database (HMDB). The human metabolome database was officially launched in January, 2007, and currently holds the honorable distinction of being the world's largest repository of NMR and MS data collected under standardized conditions relevant to metabolomics (Wishart, Tzur et al. 2007). As the name implies, the Human Metabolome Project's main focus is human metabolism. To this end, the Canadian group has amassed a prodigious database related to the biological significance, metabolic pathways, and physical properties of metabolites found in humans. The centerpiece of the HMDB is their collection of experimentally acquired MS and NMR data. Although there is approximately 30% overlap between the data collected by the MMC and HMP, the datasets were collected with different purposes in mind. The HMDB's main focus is biological fluids whereas the MMCD is aimed at analysis of tissue extracts. These different foci provide flexibility to the metabolomics community with respect to experimental design.

The HMDB offers a number of browsing and bioinformatics tools for accessing their data and searching the literature. One of the most useful HMDB features is their 'metabocard', a single web page containing approximately 90 data fields summarizing all of the HMDB information related a particular compound. This feature provides an efficient mechanism for learning about your recently identified metabolites. Similar to the MMCD, the HMDB provides direct links to other databases containing information relevant to a metabolite of interest. The links lead to a multitude of other resources specific to the metabolite in question. Although the HMDB also offers several bioinformatics tools for querying the database with experimental MS

or NMR data, we have found the main advantage of the database to be in browsing the metabolomics literature. The HMDB is unquestionably the most powerful resource available to the metabolomics community for this purpose. The HMDB is available over the web at http://www.hmdb.ca.

Madison Metabolomics Consortium Database (MMCD). The MMCD was initially developed as an in-house tool for identifying metabolites. The MMCD was released to the public in late 2006 as a bioinformatics resource for both MS and NMR based metabolomics (Cui, Lewis et al. 2008). In its first year of operation the MMCD received 91,000 visitors from around the globe. Public interest in the MMCD stems from its collection of 20,000 small molecules of biological interest gathered from electronic databases and the scientific literature. These data include the following: chemical formula, names and synonyms, structure, physical properties, NMR and MS data on pure compounds (when available), NMR chemical shifts determined by empirical and theoretical approaches, calculated isotopomer masses, information on the presence of the metabolite in different biological species, and extensive links to other databases.

In contrast to the HMDB, which excels at browsing metabolite data, the main advantage of the MMCD is in its bioinformatics capabilities. For identifying metabolites by name, the text search engine has a large collection of synonyms and automatically allows for fuzzy text matching. Users can also enter database ID numbers from a variety of other public resources (e.g. KEGG and CAS). For structure-based searches, the MMCD allows queries by molecular formula, string representation (e.g. SMILES and INCHI) or common structure files (e.g. .mol and .pdb). Alternatively, the structure can be drawn directly into a molecular graphics window. Users can combine as many as six structural criteria in a logical fashion to further refine the searches.

Perhaps the most powerful attributes of the MMCD are its metabolite assignment tools. Users can upload experimental NMR or mass spectra for bulk queries of the database. The NMR-based searches compensate for differences in magnetic field strength and filter search results on the basis of the overall patterns in the submitted peaks. NMR-based queries average about 95% sensitivity and 4% false discovery when analyzing 2D ^{1}H-^{13}C HSQC spectra of pure compound mixtures (Cui, Lewis et al. 2008). Although the MMCD's performance is diminished by chemical shift variation in real biological extracts, it is still one of the most effective automatic NMR-based metabolite identification tools in the public domain.

For mass-based searches, the MMCD is primarily designed for identifying metabolites by exact mass, although the MMCD can also handle LC-MS, and MS/MS data. Users can specify the ionization mode, mass accuracy, carbon and nitrogen isotopic composition, and allow for common adducts. Experimental LC-MS and MS/MS peak lists can be uploaded directly either as flat text files or in JCAMP-DX format for batch queries. Although sensitivity and false discovery rates are harder to estimate for MS based queries, MMCD users can expect between one and three matches for each mono-isotopic mass entry. As with most mass based queries, the efficacy of the search engine is primarily dictated by the mass accuracy and the mass range being queried. MMCD's main advantage for MS queries is that it allows users to restrict mass queries to known metabolites by using the biological filter provided under the miscellaneous search engine.

We recently upgraded the MMCD server and software to match heavy user traffic and will continue to expand the resource as demand increases. Currently, users can expect metabolite assignments on ^{1}H-^{13}C HSQC spectra in less than two seconds per spectrum. The MMCD is available over the web at http://mmcd.nmrfam.wisc.edu.

2.4.2. Software for NMR data analysis

Bioanalytical NMR-based metabolomics studies often require more than a thousand resonances assignments. Currently, this is a laborious task that is subject to human error and is difficult to document. Chemical shifts of many metabolites are subject to unpredictable variations resulting from uncontrolled differences in solution chemistry between samples. Consequently, resonances assigned in one spectrum cannot be transferred directly to other spectra. Although there are several effective software tools for identifying metabolites in complex NMR data (e.g. MMCD, metaboMiner, Chenomx, Bruker Amix, and BioRad KnowItAll), none of the existing tools were designed for assigning resonances across multiple spectra. As a result, every spectrum must be assigned individually. Using existing software tools, this task can require weeks of visual data inspection.

To make comprehensive NMR data analysis more feasible, we developed an open source software (rNMR) (Lewis, Schommer et al. 2009) written for the R statistical software environment. rNMR operates on a fundamentally different principle from existing NMR tools; rather than assigning peaks, rNMR extracts user-defined regions of interest (ROIs) from spectra. Unlike peak lists, which are static summaries containing limited information, ROIs contain all of the NMR data present with a defined set of chemical shift ranges and can be visually inspected. rNMR displays ROIs extracted from hundreds of samples side-by-side, and allows users to dynamically manipulate the size and placement of ROIs while simultaneously visualizing all of the NMR data related to an assignment. This strategy allows thousands of resonances to be visually inspected in a few minutes. Moreover, rNMR allows users to correct assignment errors at any stage of an analysis by simply adjusting the bounds of the affected ROI. Because all resonance assignments are made within the context of a defined chemical shift range, rNMR enforces consistent resonance assignments across hundreds of samples while maintaining flexibility to variations in chemical shift.

In addition to simplifying resonance assignment procedures, rNMR also makes quantitative analyses more transparent. Quantitative algorithms are based directly on the ROI data displayed to users, and the underlying NMR data behind any data point can be examined by simply clicking on the appropriate ROI. Because rNMR generates quantitative data on the fly from raw NMR spectra and a table containing the boundaries of each ROI, any rNMR analysis can be replicated by other researchers.

We initially developed rNMR as an in-house tool to solve practical problems encountered in our own research. Since its initial development, rNMR has expanded to include a broad range of peak picking, data visualization, and metabolite assignment tools that simplify data analysis. In addition, rNMR's architecture and licensing (general public license version 3) give users the freedom to customize and redistribute the program. The rNMR program, extensive help documentation, instructional videos, compiled standards data from the BMRB, and example data sets are available free of charge from http://rnmr.nmrfam.wisc.edu.

2.5. Mass spectrometry methods for identifying and quantifying metabolites

Metabolomics researches fall into two categories: those who use mass spectrometry (MS), and those who wish that NMR had the sensitivity of MS. Although the respective advantages of NMR and MS are well known, it is worth mentioning that despite the clear superiority of MS with respect to detection limit, MS has two fundamental challenges: non-uniform ionization efficiencies and translating identified masses into specific metabolites.

Direct analysis of metabolic extracts by MS has been reported (Vaidyanathan, Rowland et al. 2001; Aharoni, Ric de Vos et al. 2002), but MS analysis typically requires some form of fractionation to reduce ionization artifacts. Traditionally, this is handled by online coupling of liquid (LC) or gas chromatograph (GC) with the mass spectrometer. GC-MS is used extensively for profiling non-polar compounds and derivatives of some polar molecules (Fiehn, Kopka et al.

2000; Koek, Muilwijk et al. 2006). Many of the technical challenges associated with GC-MS based research have been resolved, but LC-MS is becoming increasingly popular for metabolomics analyses because of its compatibility with a wider range of biological compounds (Tolstikov and Fiehn 2002; Wang, Zhou et al. 2003; Dalluge, Smith et al. 2004; von Roepenack-Lahaye, Degenkolb et al. 2004; Lafaye, Labarre et al. 2005; Wu, Mashego et al. 2005; Want, O'Maille et al. 2006). The topics addressed in this section cover several new LC-MS compatible methods that have made identifying and quantifying metabolites a more tractable problem.

Quantification with selective isotope labeling. The analytical precision of ESI-MS is primarily limited by two related variables: ionization efficiency and matrix effects (Stokvis, Rosing et al. 2005). Ionization efficiency, or the percent of a molecular species that are ionizable, depends on a number of instrumental factors, molecular characteristics and solution conditions. Instrumental factors, particularly pressure and temperature at the ion source, are difficult to control and can produce significant run-to-run and day-to-day variations. Matrix effects, which occur when ions other than the target compound compete for charge, are also problematic because minor changes in the matrix can have a pronounced effect on ionization efficiency. Although chromatography can reduces these problems, observed peak intensities in metabolomics studies are inevitably influenced by factors other than metabolite concentrations (Roy, Anderle et al. 2004).

One method for improving the quantitative performance of MS is to relate observed signals to isotopically labeled (^{2}H, ^{13}C, ^{15}N, or ^{18}O-substituted) internal standards for each of the target molecules. Some care should be taken with ^{2}H labeled compounds to ensure that labels are limited to non-labile atoms and that the perdeuterated positions do not interact with the chromatographic columns (perdeuteration can lead to chromatographic shifts) (Pan, Kora et al. 2006; Yang, Mirzaei et al. 2006). The main advantage of internal isotopic standards is that they co-elute with their unlabeled counterparts, and thus can be used to normalize variation in

ionization efficiency and matrix effects. In practice, this entails dividing peak intensities (or volumes) of unlabeled compounds by the intensities of the corresponding labeled standards. Accurate concentrations of the unlabeled compounds can then be calculated by multiplying the normalized signals by the known concentrations of the standards. The primary disadvantage of this approach is that an isotopically labeled standard is needed for every compound of interest, which becomes difficult in comprehensive studies because of the price (~$100/mg) and potential unavailability of labeled standards.

There are several alternatives for larger scale projects in which absolute quantification can be replaced by relative abundance. One method involves *in vivo* isotopic labeling of a control sample using an economical substrate (e.g. *E. coli* grown on U-^{13}C glucose or acetate). A fixed amount of the labeled mixture produced *in vivo* is added to each of the test samples, and the relative abundances of metabolites are computed by comparing the signal from labeled molecules to their corresponding unlabeled counterparts (Mashego, Wu et al. 2004; Birkemeyer, Luedemann et al. 2005; Lafaye, Labarre et al. 2005; Wu, Mashego et al. 2005). This strategy works well for most small free living organisms (yeast, bacteria, and tissue cultures) and is applicable to some whole plants. However, the approach is limited in mammals because of the difficulty in achieving uniform isotopic labeling and the prohibitive expense.

An alternative strategy for calculating relative abundances of metabolites is to use selective chemistry to isotopically label molecules containing a particular functional group (Fig. 2.2) (Berry and Murphy 2005; Regnier and Julka 2006; Shortreed, Lamos et al. 2006; Lamos, Shortreed et al. 2007; Yang, Adamec et al. 2007). The general strategy for the selective chemistry approach is similar to the *in vivo* approach in that an isotopically labeled control mixture is used as a concentration reference sample for a series of test samples. Selective chemistry-based quantification requires each test mixture to be derivatized in the same manner as the isotope labeled control. To distinguish between molecules originating from the test and control mixture, test samples are derivatized with unlabeled reagents (natural abundance levels of isotopes), whereas control mixtures are derivatized with isotopically labeled reagents. The 'heavy' control mixtures are then mixed with 'light' test samples creating a composite mixture. When analyzed, the heavy and light derivatives co-elute from the LC-column and appear in the mass spectrum as pairs of peaks with a mass-shift equal to the difference in mass of the two isotopic labels (Fig. 2.3). The ratio of peak areas or intensities for each pair can then be used to compute the relative metabolite abundances in each of the test samples.
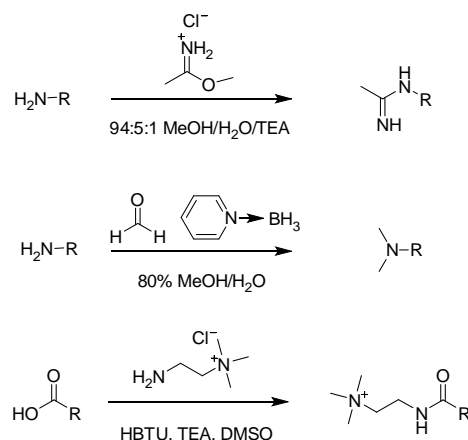
Figure 2.2. Isotopic labeling chemistry under different reaction conditions. Isotopic shifts resulting from differential labeling of amines with (top) methylacetimidate, where a 2 Da shift is produced from the two ^{13}C atoms; (middle) formalin, where primary amines acquire two ^{13}C methyl groups to produce a 2 Da shift, and (bottom) cholamine, where ^{2}H on all three methyl groups produce a 9 Da shift. (Abbreviations: MeOH, methanol; TEA, triethanolamine; HBTU, 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; and DMSO, dimethyl sulfoxide.)
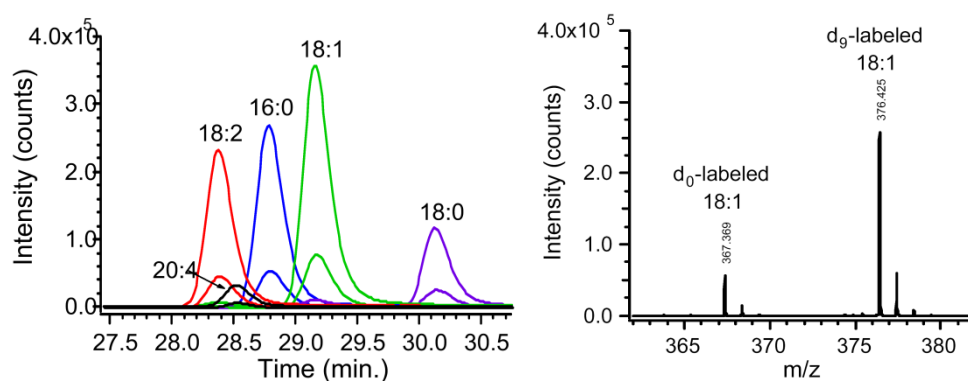
Figure 2.3. Representative extracted ion chromatograms and a mass spectrum for cholamine-labeled fatty acids. Light and heavy labeled fatty acids co-elute from reverse phase LC (left) and are easily distinguished by MS by their characteristic 9 Da shift (right). Amines labeled as shown in Fig. 2.2 yield analogous results.

The selective chemistry approach to quantification offers a number of advantages to the metabolomics community. Most importantly, it improves quantitative precision by normalizing variations in detection sensitivity resulting from ionization suppression (Constantopoulos, Jackson et al. 1999; Sterner, Johnston et al. 2000; Annesley 2003) and variability in retention time between runs (Pan, Kora et al. 2006). This is possible because the 'light' and 'heavy' metabolites co-elute within a single run and therefore have identical retention times and are electrosprayed from identical solution conditions. A second benefit of using a derivatization reagent is that it can help identify a metabolite by indicating the presence of a certain functional group. By employing a number of labeling strategies, one can target various classes of compounds. Furthermore, well-designed labeling reagents can improve chromatographic separation, enhance detection sensitivity and yield low coefficients of variation (Shortreed, Lamos et al. 2006; Guo, Ji et al. 2007; Lamos, Shortreed et al. 2007; Yang, Adamec et al. 2007). Although it is too early to judge the efficacy of this approach in metabolomics settings, selective chemistry promises to be a powerful, cost effective, tool in the MS metabolomics arsenal.

Isotope constrained formula assignments. High-resolution MS spectra of tissue extracts contain hundreds to thousands of peaks. Assigning identities to each of these features is far from trivial. Although common metabolites can be identified by GC-MS using the well developed commercial libraries, these libraries are of little use for novel compounds. Furthermore, existing LC-MS/MS literature (primarily from the HMDB) is of limited utility because of the large platform-dependent variability present in LC-MS systems. Currently, LC-MS analysis and novel compound identification require more extensive analytical techniques than those used for GC-MS.

One analytical strategy that can be used for identifying metabolites is to calculate elemental compositions of mass peaks obtained from high accuracy mass measurements. This approach is only feasible for compounds of low molecular weight and requires very high mass accuracy estimates, such as those obtained via FT-ICR MS. Unique molecular formula assignments of compounds less than 250 amu typically require a mass accuracy of 3 ppm. As mass increases, or mass accuracy decreases, the number of matching formulas balloons exponentially.

Computational and experimental constraints can be used to reduce the number of possible formulas for higher molecular weight species. Both natural abundance isotopic distribution (Kind and Fiehn 2006) and heuristically derived limits on elemental composition space (Kind and Fiehn 2007) are effective means of constraining molecular assignments. A more experimental approach is to create mixtures of uniformly substituted isotopomers and measure the mass shift associated with isotopic labeling (Rodgers, Blumer et al. 2000). For example, a spectrum can be collected for a mixture of unlabeled, fully ^{13}C labeled, and fully ^{15}N/^{13}C labeled versions of a molecule. Mass shifts observed in the spectrum of this mixture can then be used to calculate the number of carbon and nitrogen atoms present in a compound, and these, in turn, can be used to restrict the number of possible formula assignments (Hegeman, Schulte et al. 2007). Such

empirically determined constraints greatly simplify molecular formula calculations and allow unique formula assignments to be made at much higher mass values (Fig. 2.4).

Metabolomics applications of the isotopically constrained formula assignment method require uniform isotopic labeling. As mentioned previously, *in vivo* ^{15}N and/or ^{13}C isotopic labeling is relatively straightforward and cost effective in small free living organisms or tissue cultures (Beynon and Pratt 2005). However, some care should be taken to ensure that metabolites are uniformly labeled. Partial labeling results in excessive spectral complexity resulting from the various partially substituted isotopomers. We have recently added several computational resources to the BMRB as a tool for researchers who are interested in using isotope constrained assignments (Hegeman, Schulte et al. 2007).
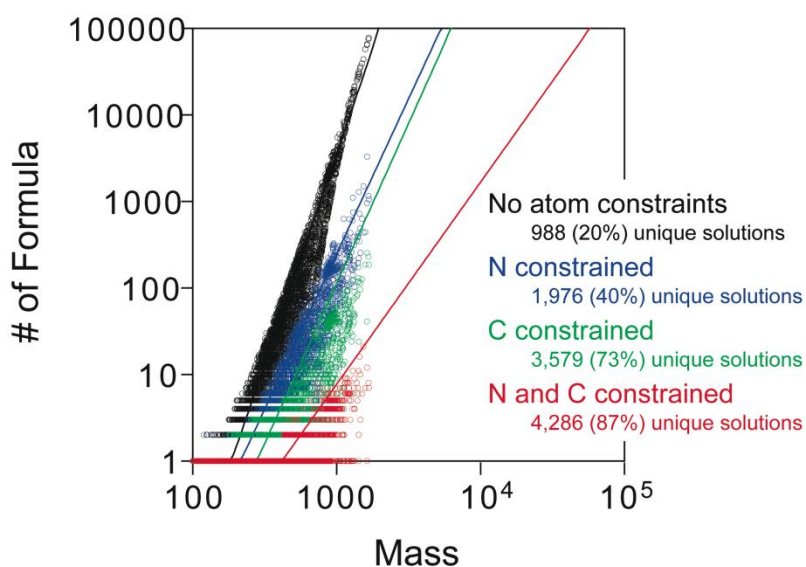
Figure 2.4. Number of calculated formulae with and without nitrogen and carbon constraints for 4,918 unique formulae derived from the BMRB database assuming mass accuracy of ±3 ppm.

2.6. NMR methods for identifying and quantifying metabolites

Although NMR is a relatively insensitive technique when compared to MS, NMR-based analyses enjoy several advantages over MS. Specifically, NMR peak intensities scale proportionally with concentration, NMR is sensitive to a wide variety of metabolite structures, and NMR analyses requires little to no sample preparation. Although NMR has become a popular tool for statistics-based metabolomics, NMR-based bioanalytical studies are relatively rare because of practical challenges in data analysis. However, the bioinformatics and software tools discussed in section 3 have largely eliminated these practical constraints.

In this section, we discuss several alternative strategies for collecting quantitatively reliable NMR data that can be used for bioanalytical metabolomics. A unifying theme of the techniques presented here is that they provide a mechanism for separating overlapped NMR signals. Overlapped signals, such as those found in 1D ^{1}H spectra of biological extracts, scale proportionally to the total overlapped spectral density and can neither be assigned nor quantified (Lewis, Schommer et al. 2007). Although 1D ^{1}H NMR is a reproducible chemometrics tool (Dumas, Maibaum et al. 2006), bioanalytical studies require well-defined spin systems for resonance assignments and isolated peaks for quantification (Lewis, Schommer et al. 2007). Currently, there are three general strategies for producing the sufficiently sparse spectra required for quantitative studies: mathematical deconvolution of spectra, multidimensional NMR, and selective pulse sequences.

2.6.1. Mathematical and statistical methods

As a rule, signals correlate very well with themselves. Not surprisingly, various peaks from the same compound are highly covariant across multiple spectra. The Nicholson laboratory recognized this several years ago and formalized the use of covariance matrices to group resonances of various compounds (Holmes, Cloarec et al. 2006). This approach has the

advantage of allowing investigators to use 1D ^{1}H NMR spectra, the fastest and most sensitive NMR experiment, and still provide a mechanism for dispersing resonances from individual compounds. However, the covariance approach does not provide a mechanism for quantification. As a result, quantitative estimates of peaks identified through covariance are still rooted in the dubious metric of overlapped 1D ^{1}H signals.

An alternative approach developed by Weljie and coworkers is to fit overlapped signals with modeled peaks and base quantitative estimates on the modeled data (Weljie, Newton et al. 2006). Resonance deconvolution has been used for decades in a wide variety of traditional NMR studies and is a well established method for separating overlapped peaks. The curve fitting method introduced by Welje et al. is probably the most reliable approach for quantifying metabolites from overlapped 1D ^{1}H NMR data and has become the preferred strategy for many bioanalytically-oriented researchers. However, resonance deconvolution is a finicky hand-manipulated process whose performance is affected by the skill of the person operating the software. A commercial implementation of Welje's method, Chenomx, removes some of this uncertainty. However, Chenomx is dependent upon matching metabolites with one of the standards present in their commercial library, and thus can be ineffective for some compounds. In addition, Chenomx is expensive and only supports 1D NMR analyses. Despite these disadvantages, resonance deconvolution is a viable method for deriving quantitative information from overlapped NMR spectra and may be an attractive alternative to researchers who are constrained to using 1D NMR.

2.6.2. Multidimensional NMR methods

One of the most effective methods for mitigating problems associated with resonance overlap is to use one of the myriad of multidimensional NMR experiments that have been developed by bimolecular NMR spectroscopists. In contrast to other methods for separating

overlapped resonances discussed here, multidimensional NMR has the significant advantage of contributing empirically-determined structural information about the observed resonances. This is of no small consequence to bioanalytical metabolomics studies, which must differentiate between many structurally similar metabolites.

Several groups have applied multidimensional NMR to metabolomics (Fan 1996; Fan, Lane et al. 2001; Viant 2003; Kikuchi, Shinozaki et al. 2004), but practical challenges have made this technique unpopular for routine studies. 2D pulse sequences require longer acquisition times, are less quantitatively robust, are prone to data artifacts, and require more NMR expertise than traditional 1D spectroscopy. In this section, we provide guidelines for mitigating these problems and discuss the trade-offs of various multidimensional NMR strategies.

Resolution versus sensitivity. Two categories of 2D NMR have proven effective in metabolomics: ^{1}H-^{1}H homonuclear and ^{1}H-^{13}C heteronuclear experiments. These categories have several inherent tradeoffs that must be considered in the experimental design phase of any bioanalytical metabolomics study. Homonuclear proton experiments are more sensitive because of the 100% natural abundance of ^{1}H and its favorable magnetogyric ratio. However, metabolite signals in ^{1}H-^{1}H experiments occupy a narrow bandwidth (roughly 10 ppm), and most ^{1}H-^{1}H pulse sequences produce multiple signals from each resonance (i.e. symmetrical cross-peaks, and diagonal peaks). The low bandwidth and signal redundancy result in resonance overlap problems. In contrast, ^{1}H-^{13}C sequences offer superior separation of metabolite signals because of carbon's larger bandwidth (roughly 180 ppm) and because most ^{1}H-^{13}C pulse sequences produce fewer signals per metabolite. Although modern ^{1}H-detected ^{13}C experiments are significantly more sensitive than the traditional direct detection methods used in 1D-^{13}C experiments, the low natural abundance of carbon (1.1%) limits metabolite investigation in unlabeled samples. The lower sensitivity of ^{1}H-^{13}C experiments can be partially mitigated by concentrating NMR samples (we prepare samples at up to 10 × higher concentrations than

found *in vivo*) or by isotopically enriching samples. However, sample limitation and biological constraints often make concentration or isotopic enrichment unfeasible. As a result, many NMR-based studies must either choose high sensitivity experiments at the cost of increased spectral overlap, or better signal separation at the cost of lower sensitivity. The correct choice is governed by the biological goals of a study, the amount of material that can be obtained, the concentrations of target metabolites, and the complexity of the mixture being investigated.

Additional practical considerations may influence the decision between 2D homonuclear ^{1}H and heteronuclear ^{1}H-^{13}C experiments. Studies involving 2D ^{1}H-^{1}H experiments require considerably longer NMR acquisition times to adequately resolve metabolite signals. In addition, chemical shift variations are more problematic in ^{1}H-^{1}H spectra because ^{1}H shifts are more affected by solution conditions than ^{13}C shifts. Consequently, bioinformatics-based assignments of ^{1}H-^{1}H data are less reliable, and metabolite identifications are more ambiguous. On the other hand, ^{1}H-^{13}C analyses at natural abundance ^{13}C levels requires 40 mg of metabolites per NMR sample (roughly 400 mg starting material). If sufficient starting material can be obtained for a single representative sample, then we recommend using ^{1}H-^{13}C analysis of a concentrated sample for metabolite identification purposes. These assignments can then be transferred to ^{1}H-^{1}H data for analyses of more dilute test samples. For investigators who are new to 2D NMR, we recommend learning the metabolite identification and quantification process using a biological model that allows all samples to be analyzed via ^{1}H-^{13}C HSQC (heteronuclear single quantum coherence) or HMQC (heteronuclear multiple quantum coherence). This recommendation is based on the relative ease of assigning and quantifying ^{1}H-^{13}C data.

Metabolite identification. The introduction of the BMRB, MMCD and HMDB databases have dramatically reduced the length of time required to assign NMR spectra. These resources allow researchers to submit peak lists from experimental data and return a list of possible metabolite identifications. Currently, the results of these queries must be verified by overlaying

spectra of standards (available from www.bmrb.wisc.edu) onto a representative extract. Step-by-step instructions for validating resonance assignments are given below. Although our methods may be adapted to some 1D NMR applications, the procedures described here primarily refers to the analysis of 2D ^{1}H-^{13}C or ^{1}H-^{1}H NMR spectra. As mentioned above, we have found that 2D ^{1}H-^{13}C HSQC data are easier to assign than ^{1}H-^{1}H experiments and recommend that new investigators learn the assignment process with ^{1}H-^{13}C data.

Protocol for metabolite identification

1) Collect a high-resolution sensitivity enhanced ^{1}H-^{13}C HSQC spectrum (e.g. Varian pulse sequence gHSQC) of a representative sample. This will likely require 512-2048 increments in the indirect dimension, four scans, and as long an acquisition time as the decoupling strategy allows. The goal of the initial spectrum is to produce one high-quality, unambiguous dataset with minimal peak overlap to be used for metabolite identification purposes. It is virtually impossible to resolve all of the signals, but most aliphatic signals are dispersed in 1028 indirect increments. It is important to match the sample's solvent conditions to those used by either the MMC or HMDB because bioinformatics-based assignments require predictable peak locations.

2) After a high quality spectrum has been collected, process the data with the minimum appropriate window function (excessive line broadening will mask J-coupling), reference the chemical shifts, peak-pick the data, and submit the peak list to the MMCD, HMDB, or BMRB. Accurate chemical shift referencing is critical for bioinformatics-based assignments.

3) Bring the experimental data into rNMR (Lewis, Schommer et al. 2009) for analysis (rNMR supports conversions of data in Bruker, Varian, or NMRpipe formats to the Sparky format used by rNMR) and download spectra of the possible matches from either the HMDB or BMRB. We have converted most of the BMRB ^{1}H-^{13}C HSQC and ^{1}H-^{1}H TOCSY standards

to Sparky format, and these data can be downloaded in bulk from http://rnmr.nmrfam.wisc.edu.

4) Overlay the spectra of each potential metabolite onto your high resolution spectrum of the representative extract. Reliable metabolite assignments must have all of the correct ^{1}H and ^{13}C chemical shifts, correct peak multiplicities, and intense peaks should show long-range [1]H-^{13}C coupling consistent with the standard (Fig. 2.5). Some tolerance can be given for chemical shift variation (+/- ~0.025 ppm), provided that all of the other criteria are fulfilled. A few resonances are more variable than others. Specifically, malate, citrate, and a number of aromatic resonances can have considerable variation despite careful pH titration. Ambiguous metabolite assignments can be checked by adding pure standards to the extract; correct assignments will show increases in peak intensity proportional to the amount of standard added. Important assignments (i.e. those that are shown to have significant changes between groups) should be validated by an independent analytical technique.
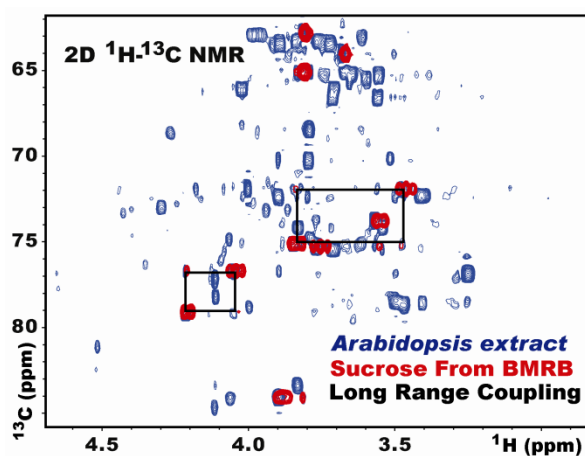
Figure 2.5. Two-dimensional ^{1}H-^{13}C HSQC NMR spectrum of sucrose from the BMRB (red) overlaid onto an aqueous whole-plant extract from *A. thaliana* (blue). Black boxes indicate long-range proton carbon couplings used to validate the assignment.

Calculating molar concentrations from 2D NMR spectra. Multidimensional pulse sequences have significant quantitative problems resulting from off-resonance effects, incomplete resonance transfer, complex relaxation pathways, decoupling artifacts and mixing times. These variables make peak intensities (and volumes) an unreliable metric for quantification (Fig. 2.6). Several laboratories are developing more quantitatively reliable versions of common bimolecular NMR pulse sequences (Heikkinen, Toikka et al. 2003; Koskela, Kilpelainen et al. 2005; Peterson and Loening 2007). Unfortunately, these efforts have yet to produce experiments that are sufficiently robust for bioanalytical metabolomics. However, we have developed a practical approach to metabolite quantification that allows researchers to use any NMR pulse sequence. The basic feature of our method is that we do not control quantification through pulse sequence design. Instead, we relate observed peak intensities to those of mixtures of pure standards of known concentration. Differences in relaxation rates and shimming are controlled by normalizing observed metabolite signals to an internal standard. The metabolite quantification strategy described here, and the fast data collection method described in the next section, are the components of our fast metabolite quantification (FMQ by NMR) approach, which is the basis of all of our laboratory's routine metabolomics work.
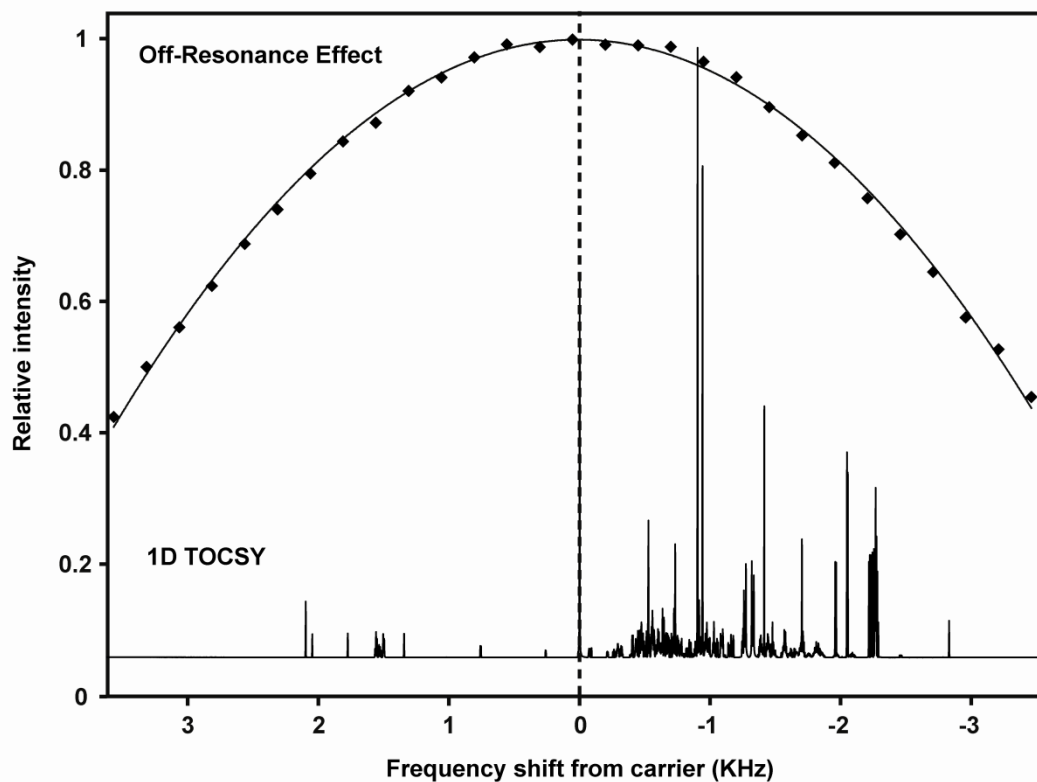
Figure 2.6. Off-resonance effects are one of the many factors that can influence peak intensities observed in NMR experiments. In this example, the intensities in ^{1}H-^{1}H TOCSY were measured as a function of ^{1}H frequency offset. The peak intensity is highest when the peak is closest to the transmitter frequency and lowers as the transmitter is tuned to higher or lower frequencies.

Metabolite Quantification Protocol

1) Identify all metabolites present in an extract (see **Protocol for metabolite identification**).

2) Prepare three mixtures containing all of the identified metabolites at 2, 5 and 10 mM. Although standardized metabolite intensities are linear well beyond the range of these standards (+/- 20 fold), the concentrations of standards should be adjusted to match the approximate range of expected concentrations. Mixtures should be prepared at high volume to minimize weighing errors. All standards and extracts should be titrated to pH 7.400 ± 0.004.

3) Include 5 mM of an internal standard in all of the standards mixtures and extracts to serve as an internal concentration reference. For ^{1}H-^{13}C NMR studies, we recommend HEPES or MES because both compounds have multiple isolated peaks that do not overlap with biological compounds. HEPES is convenient because it acts as an internal pH indicator, but MES is probably a more reliable concentration reference because of its lower pK_a.

4) Collect spectra of extracts and concentration reference samples under identical NMR acquisition conditions at the same time using the same instrument. Every sample, including the concentration reference samples, should be collected twice to produce two technical replicates for each sample. The sample order should be randomized. Each test sample should have at least three (preferably many more) independent biological replicates. From a statistical perspective, it is much better to have many independent biological replicates than to analyze many metabolites (large alpha corrections, such as Bonferroni correction, must be made in studies that use multiple comparisons).

5) Measure the peak intensities (area for 1D, peak height for 2D) of non-overlapped peaks from extracts and the concentration reference samples. Although peak area is a robust metric in 1D NMR, we have found 2D peak volumes to be considerably less reliable than peak heights when used in this procedure (regardless of the NMR analysis software).

6) Normalize signals observed in each spectrum to the average signal of the dispersed HEPES (MES) peaks. Raw intensities can be used without normalization, but any variations in salt concentration (or paramagnetic relaxation agents) between samples will be erroneously interpreted as differences in metabolite concentrations. If samples and standards are osmotically identical, then normalization is undesirable because it introduces some technical error. However, many NMR probes are highly salt sensitive, and even the standards may show strong non-linear effects of salt. If in doubt, it is better to normalize to the internal standard and accept a small increase in technical error.

7) Average normalized peak intensities across technical replicates.

8) Regress normalized peak intensities of the standards to produce a concentration versus peak intensity equation for each dispersed signal.

9) Calculate the observed concentrations for each normalized peak in the test samples using the equations derived from the standards samples.

10) Average concentration estimates across all dispersed peaks from each molecule.

The protocol described here produces concentration estimates with as little as 2.7% technical error from complex 2D NMR spectra (Lewis, Schommer et al. 2007). The main disadvantage of this approach is its dependence on standards, many of which are unavailable or are prohibitively expensive to use in the quantities required for this procedure. If the requisite standards can be obtained, however, then this strategy produces reliable quantitative information from most of the existing NMR pulse sequences. Moreover, the internal concentration references we use to control for differences in longitudinal (T_1) relaxation, and other variations between samples, can be introduced early in the sample preparation process to relate observed signal intensities to biologically relevant concentrations.

Fast multidimensional experiments. Although multidimensional NMR has seen qualitative applications to metabolomics for many years (Fan 1996; Fan, Lane et al. 2001; Viant 2003; Kikuchi, Shinozaki et al. 2004), very few studies have used the technique for quantitative purposes. One reason for this is that multidimensional NMR spectra take longer to collect than 1D ^{1}H NMR spectra, although the long acquisition times of 2D experiments have been somewhat exaggerated in the metabolomics community. Over the years, protein NMR spectroscopists have developed a wealth of techniques for reducing the requisite acquisition times of multidimensional experiments. These techniques range from mathematical approaches for non-linear sampling of data (Hyberts, Heffron et al. 2007), to pulse sequences that encode the indirect dimensions using gradients (Shrot and Frydman 2004) and optimization of pulse angles (Ross, Salzmann et al. 1997). All of these techniques are applicable to metabolomics, but the easiest way to save time is to be judicious in setting up the NMR acquisition parameters.

We have shown that a carefully adjusted 2D ^{1}H-^{13}C HSQC spectrum allows metabolites with concentrations over ~500 µM in the NMR tube (30−40 metabolites using our sample preparation methods) to be quantified in about ten minutes (Lewis, Schommer et al. 2007). The secret to our approach is in the prior identification of compounds. As discussed above, a single high resolution HSQC of a representative sample can be used to identify the abundant metabolites present in an extract. After the metabolites have been identified and the peaks have been assigned, one can capitalize on the assignments by reducing the number of indirect points and time to the minimum required to resolve two signals from each target molecule. These time savings can be achieved while maintaining high quantitative precession (technical error ~3%) (Lewis, Schommer et al. 2007).

The primary factor that contributes to lengthy 2D experiments is the number of increments collected in the indirect dimension. Time savings are proportional to the number of increments that can be eliminated from the acquisition. Because resolution in the indirect dimension is a

function of the number of increments and the indirect spectral-width, achieving adequate resolution in the minimal amount of time usually involves trimming both the spectral-width and the number of increments. If the indirect spectral-width is cut in half, then the number of indirect points can be halved without affecting resolution. The spectral-width can be reduced well beyond the point where resonances are no longer contained within the spectral window. Resonances with signals outside of the spectral-width will wrap back into the spectral window with a chemical shift that is equal to their original shift plus or minus a multiple of the spectral-width (Fig. 2.7). In more extreme cases, resonances may be wrapped multiple times. If the original chemical shifts and the spectral-width are known, then a heavily wrapped spectrum can be unwrapped by using simple arithmetic. Spectral folding may result in undesirable overlap between analytically important signals, but overlap can often be alleviated with small adjustments of the indirect carrier frequency. As discussed above, off resonance effects play a major role in observed peak intensities. After acceptable settings for the spectral-width, number of increments, and the transmitter offset have been found, it is critical that these values be kept constant across all of the test samples and standards.
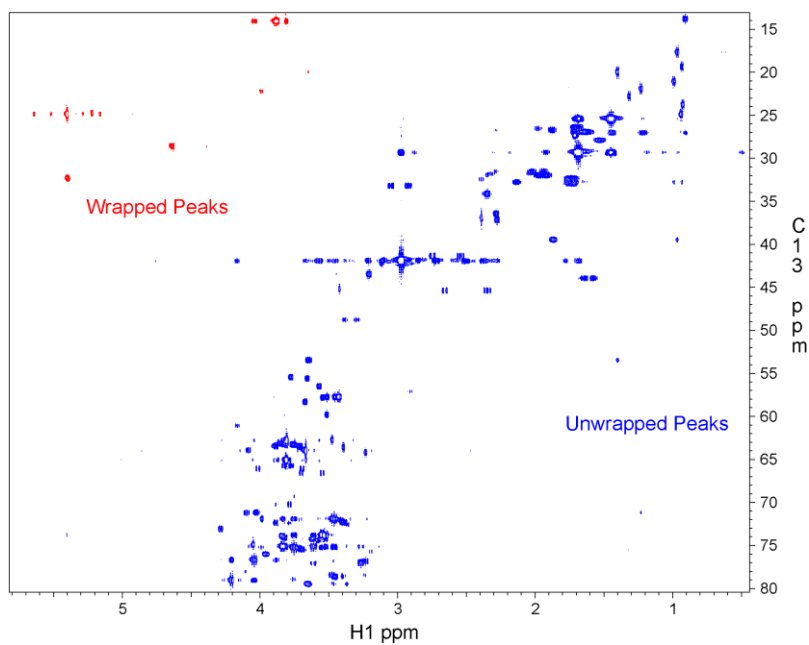
Figure 2.7. Quantitative ^{1}H-^{13}C HSQC NMR spectrum of a synthetic mixture of 26 metabolites. The spectral width in the indirect dimension was reduced to allow for shorter acquisition times. Blue peaks are in their correct locations, whereas red peaks have been wrapped into the top of the spectrum from their normal downfield positions.

2.6.3. Selective NMR methods

Selective NMR is a back to the future idea revived from the early days of protein NMR. Although selective experiments have been largely outdated in protein NMR by modern multidimensional pulse sequences, the technique is appropriate for targeted metabolic studies. Selective NMR works by carefully sculpting the excitation pulse to cover a narrow bandwidth. This, in combination with other sequences such as TOCSY, allows individual spin systems to be isolated from amazingly complex mixtures (Fig. 2.8). Similarly, the technique can be used to selectively remove overpowering resonances from a spectrum. Dan Raftery's group first introduced the idea to the metabolomics community by showing that signals from minor components of honey could be accurately isolated and quantified without being influenced by the large glucose and fructose signals (Sandusky and Raftery 2005). One of the most powerful aspects of the selective TOCSY is that data can be collected very quickly (~1 min per spectrum). For studies requiring accurate quantification of a few metabolites in complex mixtures, this is one of the fastest NMR techniques available.

One must keep a few practical considerations in mind when using selective pulse sequences. Selective experiments require hand tuning of the excitation pulse, and this tuning may change from sample to sample if the target metabolite is subject to chemical shift variation. Secondly, selective experiments lose their time advantage in studies involving multiple metabolites. After the number of target metabolites reaches about ten, then a full 2D ^{1}H-^{1}H TOCSY or ^{1}H-^{13}C HSQC are more efficient and do not require hand manipulation of the pulses between experiments. However, if a study only calls for analysis of a few molecules, then selective TOCSY is one of the best tools for the job.
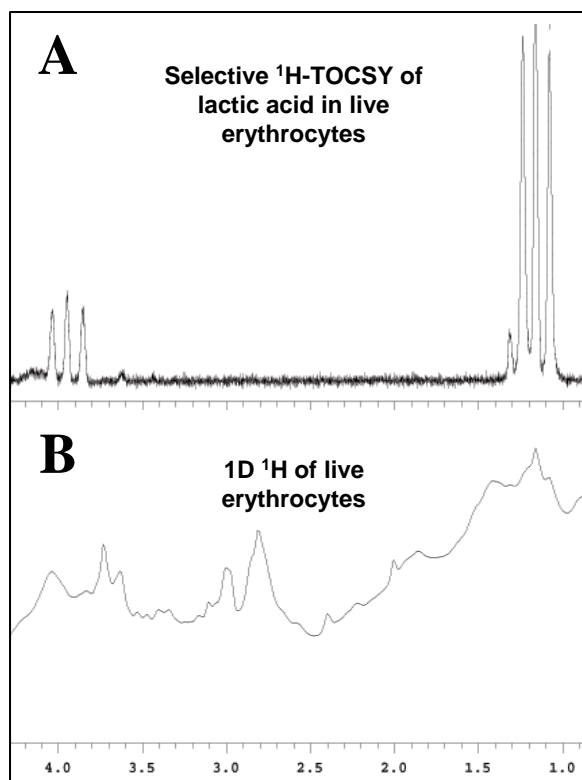
Figure 2.8. (A) Selective 1D ^{1}H-^{1}H TOCSY for lactate in live red blood cells. (B) Standard 1D ^{1}H NMR spectrum of the same sample. Cells were labeled with [U-^{13}C]-glucose, the triplet-like splitting observed in the TOCSY arises from [U-^{12}C]-lactate (center peak) and fully labeled [U-^{13}C]-lactate (two satellite peaks).

2.7. Future prospects

Technological advances of recent years have dramatically increased the efficiency with which metabolites can be identified and accurately quantified. High quality empirical libraries, bioinformatics-based spectral assignment tools, improved analytical software, and practical methods for identifying and quantifying NMR and MS signals in complex spectra have taken the field a step closer to the automation enjoyed by mainstream bioanalytical methods. Despite these advances, bioanalytical metabolomics is still in its early development and is far from capitalizing fully on state-of-the-art NMR and MS technology. A significant proportion of signals in every study go unassigned, bioinformatics tools suffer from unacceptable false discovery rates, quantification requires *ad hoc* correction of quantitatively unreliable data, and specialized expertise is required to collect, analyze, and interpret data. Finding practical solutions to these problems is paramount to bioanalytical metabolomics and is a promising area for future technology development.

CHAPTER 3

New Bioinformatics Resources for Metabolomics

Adapted from:

John L. Markley, Mark E. Anderson, Qiu Cui, Hamid R. Eghbalnia, Ian A. Lewis, Adrian D. Hegeman, Jing Li, Christopher F. Schulte, Michael R. Sussman, William M. Westler, Eldon L. Ulrich, and Zsolt Zolnai. New Bioinformatics Resources for Metabolomics. <u>Pacific Symposium on Biocomputing</u>, World Scientific Press: 157-168 (2007).

and

Qiu Cui, Ian A. Lewis, Adrian D. Hegeman, Mark E. Anderson, Jing Li, Christopher F. Schulte, William M. Westler, Hamid R. Eghbalnia, Michael R. Sussman, and John L. Markley. Metabolite identification via the Madison Metabolomics Consortium Database. *Nature Biotechnology*, 26,162 (2008)

This chapter documents two bioinformatics resources that are essential to bioanalytical metabolomics. Prior to these tools, identifying metabolites required months of data analysis and considerable NMR expertise. These resources made metabolite identification an efficient semi-automated procedure. Developing these tools required the coordinated efforts of many people; my contributions included designing the standardized solution conditions, selection of the metabolite standards, some of the early data collection and sample preparation, assistance with the design of database and bioinformatics tools, and beta testing. I also wrote the two papers published on this topic and conducted the statistical analyses presented in these publications.

3.1. Abstract

We have developed two freely available databases as resources for the metabolomics community. These tools simplify metabolite identification in MS- and NMR-based studies. The first database, a metabolomics extension to the BioMagResBank (BMRB, www.bmrb.wisc.edu), contains peak lists and NMR spectra of over 800 pure compounds.[2] These data allow tentative NMR assignments to be compared with metabolite standards prepared under controlled conditions. The second database, the Madison Metabolomics Consortium Database (MMCD, http://mmcd.nmrfam.wisc.edu), contains extensive bioinformatics tools for identifying metabolites in experimental MS and/or NMR data. We introduce these tools and discuss their utility to metabolomics.

3.2. Introduction

High-throughput metabolic profiling, known as metabolomics (Mendes 2002) or metabonomics (Nicholson, Lindon et al. 1999), has been an active area of research for over 35 years (Pauling, Robinson et al. 1971). Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are the primary analytical techniques used for metabolomics (Fiehn 2002; Nicholson, Connelly et al. 2002). Despite the routine use of these tools, determining the molecular composition of mixtures remains a major analytical challenge. One reason for this difficulty is a lack of publicly available tools for comparing experimental data with the existing literature on the masses and chemical shifts of common metabolites.

To simplify metabolite identification in complex mixtures, we have two databases of biologically relevant small molecules: 1) a metabolomics extension to the existing Biological Magnetic Resonance Data Bank (BioMagResBank, BMRB) and 2) the Madison Metabolomics

---

[2] Data collection efforts are ongoing; this figure is current as of March, 2010.

Consortium Database (MMCD). The BMRB database (http://www.bmrb.wisc.edu) contains experimental NMR data from over 800 pure compounds whereas the MMCD (http://mmcd.nmrfam.wisc.edu) contains a collection of bioinformatics tools for assigning MS and NMR spectra.

3.3. Metabolite Database at the BMRB

Overview. The metabolomics community would clearly benefit from an extensive, freely-accessible spectral library of metabolite standards collected under standardized conditions. Our chief complains with the existing resources are that they either 1) do not contain the original spectral data (Sadtler 1967; Steinbeck, Krause et al. 2003) 2) contain data that were collected under non-standardized conditions (Steinbeck, Krause et al. 2003) or 3) do not make their data freely available (AMIX/SBASE http://bruker-biospin.de). To the best of our knowledge, the Human Metabolome Project (http://www.hmdb.ca/) is the only resource that does not suffer from any of these problems. The deficiency of NMR resources results partially from the prohibitive expense and time expenditure necessary to compile a comprehensive library of biologically relevant small molecules under standardized conditions. Our solution is to provide a well-defined, curated platform that will allow the deposition of data from multiple research groups and free access to all.

Rationale for Metabolomics at BMRB. The BMRB is a logical host for a metabolomics spectral library because of its history as a worldwide repository for biological macromolecule NMR data (Ulrich, Markley et al. 1989; Seavey, Farr et al. 1991; Doreleijers, Mading et al. 2003). The BMRB is a public domain service and is a member of the World Wide Protein Data Bank. Along with its home office in Madison, Wisconsin, the BMRB website has mirror sites in Osaka, Japan and Florence, Italy. BMRB is funded by the National Library of Medicine and its activities are monitored by an international advisory board. BMRB data is well archived with daily onsite tape backups and offsite third party data backup.

Data Collection and Organization. Currently, the BMRB metabolomics archive contains experimental NMR data for more than 800 compounds collected by the Madison Metabolomics Consortium. Entries contain time-domain NMR data, peak lists, processed spectra, and data acquisition and processing files for one-dimensional (^{1}H, ^{13}C, ^{13}C DEPT 90$^\mathrm{o}$, and ^{13}C DEPT 135$^\mathrm{o}$) and two-dimensional (^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC) NMR experiments.

Each BMRB metabolite entry represents a set of experimental and theoretical data reported for a metabolite. Entries are assigned a unique identifier and metadata describing the compound, experimental details, and quantitative data are archived in NMR-STAR formatted text files (Hall 1991; Hall and Spadaccini 1994; Hall and Cook 1995). Data are stored on the BMRB ftp site (ftp://ftp.bmrb.wisc.edu/pub/metabolomics) with directories defined by each compound. Subdirectories for NMR, MS, and literature data are listed for each compound directory. Data for compounds that form racemic mixtures in solution (e.g. glucose) are grouped under a generic compound name.

Presentation and Website Design. The BMRB metabolomics website has been developed to meet needs expressed by many of its users. The layout and usage of the metabolomics web pages have had several public incarnations and will probably undergo more as the site matures and grows. The first page a visitor sees contains a two-paragraph introduction to the field and a collection of Internet links to a few important small molecule sites with a more complete listing of metabolomics websites at the side of the page. The information contained in these websites and databases is complementary to that collected by BMRB. The Standard Compounds page (Fig. 3.1) provides the means for searching for metabolites of interest.

For each compound archived, an individual summary page (Fig. 3.2) is created dynamically from the collection of files located in the standard substance sub-directory associated with that compound. A basic chemical description (nomenclature, formula, synonym list) is provided from

information collected from PubChem (http://www.ncbi.nlm.nih.gov/) and a two-dimensional stick drawing is created. Three-dimensional '.mol' files are generated from the two dimensional '.sdf' files obtained from PubChem, and these are displayed using Jmol. Links are created to one or more PubChem entries and to the corresponding KEGG entry. Synonym information and various nomenclature descriptions such as INChI codes, IUPAC names, and SMILES strings are given.

Figure 3.1. (Above) Metabolomics standards page on the BMRB website and (below) a portion of a metabolite summary page for N-acetyl-D-glucosamine-6-phosphate.

The dynamic user interface at the BMRB allows us to create tools that search through the data or calculate answers according to specific user input. NMR data can be displayed in a variety of ways: as a collection of spectra, as a spectrum along with its peak list, or simply a single spectrum of interest. Links allow the user to access the time-domain or processed data by FTP. The Peak Query tool allows the user to enter a list of peaks in one- or two-dimensional formats with tolerances and retrieve a list of compounds with matching signals.

Metabolites often have a variety of synonyms. The BMRB allows records to be located by common names, INChI codes, IUPAC names, SMILES strings, and various database IDs. The BMRB also supports selective searches. Users wishing to see all entries for molecules containing nitrogen, for example, can search chemical formulas by 'N'. Molecules with similar substructures can be located using INChI or SMILES searches.

Prospects. Over the past year, the BMRB has acquired a collection of high-quality NMR data of metabolite standards and has developed a platform for distributed these data to the public. This resource will continue to expand as additional datasets are deposited. Currently, all of the BMRB data has been contributed by the Madison Metabolomics consortium. However, we are actively soliciting datasets from the community and encourage users to provide feedback on the BMRB interface. Several users have requested that the BMRB data be made available through bulk transactions. We have accommodated this request by collating data from the ftp repository into a collection of tar (tape archive) files that can be easily downloaded.

3.4. Madison Metabolomics Consortium Database (MMCD)

Overview. The task of identifying metabolites present in unfractionated biological samples is fundamentally different from the process of identifying novel natural products. The key distinctions are that the molecular structures of common metabolites are already known, and pure small molecule standards are often commercially available. This means that many of the

time consuming steps required for natural product compound identification can be replaced by bioinformatics. Currently, metabolomics studies rely on traditional natural products-based methods for identifying molecules in complex biological extracts. The Madison Metabolomics Consortium Database (MMCD, http://mmcd.nmrfam.wisc.edu) extends the practical scope of metabolomics studies by replacing labor-intensive metabolite identification with efficient bioinformatics-based assignments.

Data collection and organization. The MMCD contains information on over 20,000 metabolites and other small molecules of biological interest. These molecules, which were chosen from entries in KEGG, BioCyc, BMRB, CHEBI, HMDB, UM-BBD, and PDB, are a collection of primary and secondary metabolites, xenobiotics, and common small molecule contaminants. A total of 459 small molecule entries contain experimental NMR data collected by the Madison Metabolomics Consortium (MMC) and an additional 525 compounds contain links to NMR data collected by the HMDB. The MMC and HMDB NMR data have 234 compounds in common.[3] Although the HMDB and MMC collect data under different conditions (MMC: 99.9% D_2O containing 50 mM phosphate buffer pH 7.4; HMDB: H2O, 50mM phosphate, pH 7.0), the chemical shifts of the two datasets are in good general agreement with an average ^{1}H chemical shift variation of 0.05 ppm and 0.15 ppm variation for ^{13}C chemical shifts.

The MMCD is more than a data repository; it is equipped with a flexible and efficient query system that supports complex queries from any combination of its five basic search engines (text, structure, NMR, mass, and miscellanea). Search results provide users with all of the MMCD information about a molecule and direct links to related records in other public databases.

---

[3] The metabolites numbers in the MMCD and HMDB quoted here reflect 2008 levels.

Text-based searches. The MMCD has an efficient text-based search feature that allows users to locate metabolites by name or by the ID numbers used by other public databases (eg. KEGG and CAS). This search feature supports ambiguous and wildcard matches and will return possible alternatives in cases of misspelled names. An extensive synonyms list supports multiple molecular naming conventions. Queries can be typed into the graphical user interface, or files can be uploaded for batch searching.

Structure-based searches. The MMCD allows records to be located on the basis of molecular formula, average mass, SMILES (Simplified Molecular Input Line Entry System) (Weininger, Weininger et al. 1989) string, INCHI (International Chemical Identifiers; http://www.iupac.org/inchi/) string, or common structure files (eg. .mol and .pdb). Alternatively, structure can be drawn directly into a molecular graphics window. Users can combine as many as six structural criteria in logical and/or/not fashion to further refine the search and can use controllable similarity thresholds to search for substructures, stereoisomers, or related covalent structures.

NMR-based searches. This feature allows users to upload experimental NMR data to the MMCD for metabolite identification. NMR-based searches give users considerable flexibility with regard to the type and quality of data entered. Chemical shifts can be combined with filters that search for complex multinuclear spin topologies. For example, users can specify chemical shift or atom connectivities (e.g., number of hydrogens attached to a carbon atom). Batch-mode searches return probabilistic identifications of metabolites in mixtures on the basis of various types of NMR data: 1D-^{1}H, 1D-^{13}C, 2D-^{1}H-^{13}C HSQC, 2D-^{1}H-^{13}C HMBC, and 2D-HSQC-TOCSY. Peak lists can be typed in manually, or files can be uploaded in the variety of formats used by NMR spectroscopists. NMR searches can use any one of MMCD's three chemical shift databases: experimental (preferred default), empirically predicted from structure (most extensive), or quantum chemical calculated (of interest to theoreticians and useful for quality

control and assignment of experimental data). Search results can be downloaded as a tab-delimited file (Excel type) or viewed directly in the MMCD interface. NMR data and structural information have been seamlessly integrated into the search engine. The search engine makes full use of chemical shift, J-coupling, and structure- related information, such as connectivity (atom neighbors). The search engine automatically handles issues related to differences in NMR field strength by storing chemical shifts, J-couplings, peak intensities, and concentration/intensity ratios in a field-independent manner. The NMR search engine reconstructs these parameters at the field strength of the data submitted by the user. Thus, qualitative analyses can be carried out using data from any NMR filed strength.

As a test of the NMR search engine for qualitative analyses, we used the MMCD to identify metabolites in ^{1}H-^{13}C HSQC spectra of complex mixtures pure compounds comprising, in total, 54 metabolites. These compounds were intentionally selected from the MMCD archives to ensure that sensitivity (correctly identified compounds) and false discovery (incorrectly identified compounds) estimates were restricted to the database's peak matching performance. Under these controlled conditions, automated MMCD compound identification averaged 95% sensitivity and 4% false discovery under the default tolerances.

We are aware that identifying compounds in spectra of biological extracts is complicated by signals from compounds not in the database and by variation in the positions of NMR peak caused by small variations in salt concentration and pH. To evaluate the MMCD performance on biological samples, we analyzed ^{1}H-^{13}C HSQC spectra of *Arabidopsis*, *Medicago* and *Saccharomyces* extracts that had been assigned by hand (Lewis, Schommer et al. 2007). The divergence of the two assignment lists showed MMCD sensitivity to range from 45 to 65% and its false discovery rate ranges from 0 to 18% when applied to spectra of biological extracts. Like the analysis of the standards mixtures, these sensitivity and false discovery estimates were restricted to compounds known to be present in the database and thus reflect the search

engine's capability rather than the extent of the database. The MMCD NMR-based mixture identification tool responds in the expected manner to changes in the default tolerances (the MMCD uses two tolerances, one controls allowable chemical shift variation and the other controls the number peaks from each molecule that must be correctly matched to classify a compound as identified; the default tolerances are $\pm$ 0.05 ppm chemical shift variation for both ^{1}H and ^{13}C and an 80% peak matching threshold). Loosening the default tolerances increases sensitivity and false discovery whereas tightening the tolerances decreases false discovery at the expense of sensitivity (Fig. 3.2). Regardless of the search tolerances, automatic identifications should always be hand-verified by overlaying spectra of pure standards over the experimental spectrum (Lewis, Schommer et al. 2007).
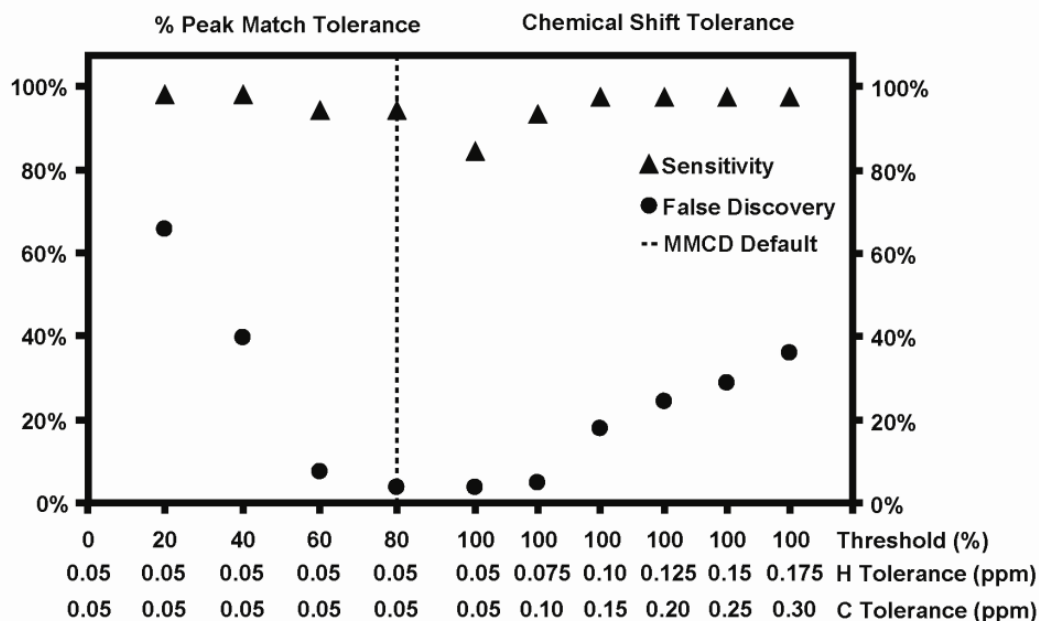
Figure 3.2. Sensitivity and false discovery rates of the MMCD for batch analysis of mixtures using ^{1}H-^{13}C HSQC NMR data as a function of the user-controllable peak matching thresholds. The metabolite mixtures used in this analysis were prepared from pure compounds under the standard MMCD conditions (50 mM phosphate buffer; 99.9% D$_2$O; pH 7.40, glass electrode reading). Sensitivity (number of correct ID / actual composition) and false discovery (number of incorrect IDs / total metabolites returned) rates were determined as a function of the user definable ^{1}H and ^{13}C chemical shift tolerances and peak match thresholds (the peak matching threshold defines the number of peaks, as a percentage, that must be observed in the experimental data for a metabolite to be counted as an positive identification). As expected, tightening the chemical shift and peak matching tolerances decreased false discovery and reduced sensitivity. Conversely, loosening the default thresholds increased both sensitivity and false discovery.
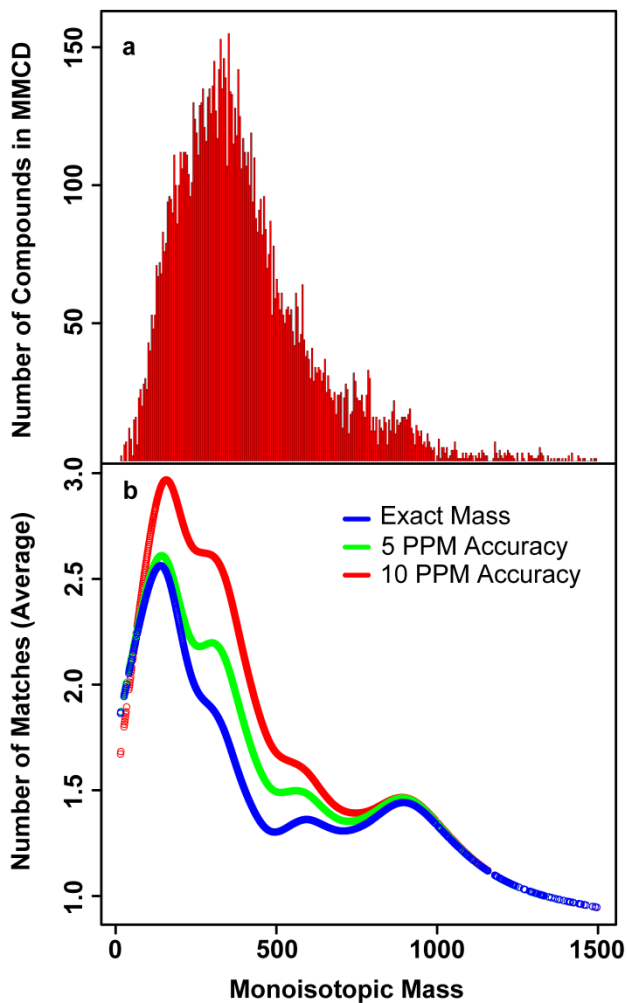
Figure 3.3. (a) Distribution of monoisotopic masses ($N = 19{,}711$) in the MMCD and (b) the average number of metabolites returned from a monoisotopic mass query as a function of mass and mass accuracy at 0 ppm, 5 ppm, and 10 ppm mass accuracy. Results were smoothed to determine the average number of matches across a mass range of 25 ppm.

Mass searches. The MMCD mass search engine is primarily designed for identifying metabolites by exact mass (Fig. 3.3), although the MMCD can also handle LC-MS, and MS/MS data. Users can specify the ionization mode, mass accuracy, carbon and nitrogen isotopic composition, and allow for common adducts. Experimental LC-MS and MS/MS peak lists can be uploaded directly either as flat text files or in JCAMP-DX format for batch queries. MS search results can be viewed directly in the MMCD interface or downloaded as a tab-delimited file for viewing with spreadsheet software.

Miscellaneous searches. A variety of other filters allows users to limit results by biological species, the type of database to be searched, or other criteria. These options allow users to rapidly locate particular sets of records or limit their queries to a preferred subset of records.

In summary, the MMCD is a practical tool for expediting the time consuming steps of identifying and researching small molecules. This freely available resource is compatible with the use of NMR and MS data (singly or in combination) for small molecule identifications as an integral part of high-throughput metabolomics investigations. The ongoing MMCD support is provided by the National Magnetic Resonance Facility at Madison. Users are encouraged to submit data to the BMRB (supported by the National Library of Medicine), which maintains one of the growing data archives that the MMCD relies upon.

CHAPTER 4

Method for Determining Molar Concentrations of Metabolites in Complex Solutions from Two-Dimensional ^{1}H-^{13}C NMR Spectra

Adapted from:

This chapter describes an empirical method I devised for correcting the quantitative defects in multidimensional NMR analyses. When combined with the tools discussed in Chapter 3, this method made comprehensive quantitative studies possible for NMR-based metabolomics. This study also illustrated a practical problem with the bioanalytical approach; the NMR data analysis, which I conducted with the help of three undergraduate students (S.C.S., B.H., and K.A.R), required more than a month to complete. The software available at the time required each spectrum to be assigned individually, which made it difficult to assign resonances consistently across the dataset. I resolved this problem by developing a software package that is better suited to metabolomics resonance assignments (Chapter 5).

4.1. Abstract

One-dimensional (1D) ^{1}H nuclear magnetic resonance (NMR) spectroscopy is used extensively for high-throughput analysis of metabolites in biological fluids and tissue extracts. Typically, such spectra are treated as multivariate statistical objects rather than as collections of quantifiable metabolites. We report here a two-dimensional (2D) ^{1}H-^{13}C NMR strategy (Fast Metabolite Quantification, FMQ by NMR) for identifying and quantifying the ~40 most abundant metabolites in biological samples. To validate this technique, we prepared mixtures of synthetic compounds and extracts from *Arabidopsis thaliana*, *Saccharomyces cerevisiae* and *Medicago sativa*. We show that accurate (technical error 2.7%) molar concentrations can be determined in 12 minutes using our quantitative 2D ^{1}H-^{13}C NMR strategy. In contrast, traditional 1D ^{1}H NMR analysis resulted in 16.2% technical error under nearly ideal conditions. We propose FMQ by NMR as a practical alternative to 1D ^{1}H NMR for metabolomics studies in which 50 mg (extract dry weight) samples can be obtained.

4.2. Introduction

One-dimensional (1D) ^{1}H NMR spectroscopy has been used for decades as an analytical tool for identifying small molecules and measuring their concentrations (Radda and Seeley 1979; Pauli, Jaki et al. 2005). Traditionally, quantitative analysis by NMR has been restricted to relatively simple mixtures with minimal peak overlap. In these applications, 1D ^{1}H NMR is a natural choice, because its peaks scale linearly with concentration and its analytical precision is usually independent of the chemical properties of target molecules. Recently, interest has surged in using NMR for high-throughput analysis of complex biological processes at the metabolic level (Lindon, Nicholson et al. 2000; Lindon, Holmes et al. 2001). These studies, defined as "metabolomics" or "metabonomics", place an emphasis on biomarker discovery or disease classification and are typically centered on unfractionated biological fluids and tissue extracts. 1D ^{1}H NMR spectra of these samples typically contain hundreds of overlapping

resonances (Fig. 4.1) that make traditional NMR-based analytical practices, such as resonance assignment and accurate peak integration, a challenging prospect. As a result, sophisticated statistical tools have been developed to translate spectral data into biologically meaningful information (Lindon, Holmes et al. 2001; Holmes, Cloarec et al. 2006).
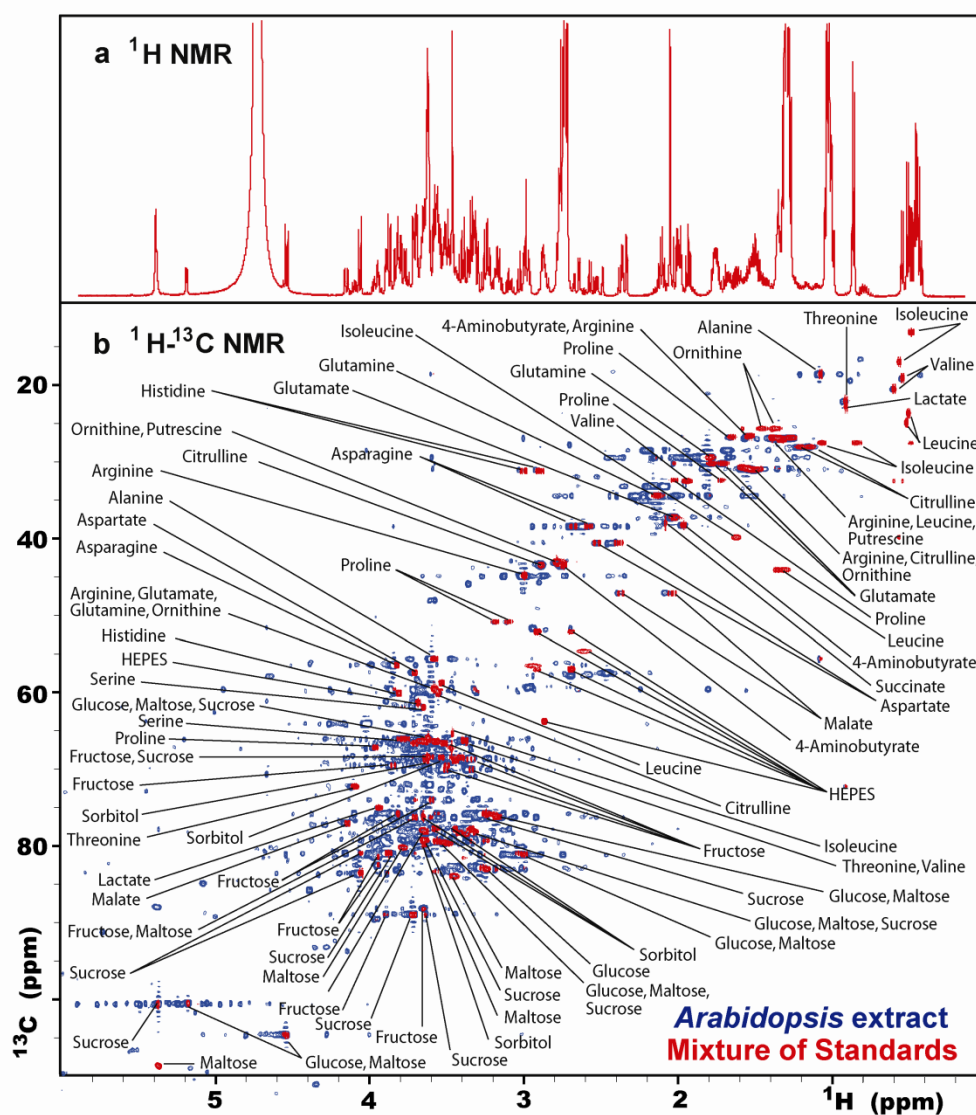
Figure 4.1. (a) One-dimensional ^{1}H NMR spectrum of an equimolar mixture of the 26 small molecule standards listed in Table 4.2. (b) Two-dimensional ^{1}H-^{13}C HSQC NMR spectra of the same synthetic mixture (red) overlaid onto a spectrum of aqueous whole-plant extract from *Arabidopsis thaliana* (blue).

All statistical tools used for analyzing complex spectra face the same fundamental barrier: overlapped peaks do not scale in the discrete linear fashion that typifies well-isolated peaks. They scale as the sum of the total overlapped resonance. Consequently, multivariate and correlation statistics are reporters of overlapped spectral density, not concentrations of specific compounds. Although peak overlap does not interfere with the reproducibility of traditional analyses (Dumas, Maibaum et al. 2006), it does prevent accurate quantification.

Two approaches can be used to overcome this barrier, one mathematical the other experimental. The mathematical approach is to fit overlapped 1D NMR spectra with modeled peaks. This approach has been successfully applied by Weljie and co-workers (Weljie, Newton et al. 2006). The experimental approach is to collect NMR spectra that disperse peaks into two or more dimensions. This allows non-overlapped peak intensities to be measured directly.

Two-dimensional NMR (Fig. 4.1) is a well established technique for reducing peak overlap and has been recognized for over ten years as an excellent tool for metabolomics (Fan 1996; Fan, Lane et al. 2001; Viant 2003; Kikuchi, Shinozaki et al. 2004). Despite this, published applications of multidimensional NMR in the metabolomics literature have been largely restricted to qualitative analyses. One reason for this is that 2D cross-peak intensities (or volumes) are influenced by a greater number of variables (e.g. uneven excitation, non-uniform relaxation, evolution times, mixing times, etc.) than are 1D ^{1}H NMR peaks. This non-uniform behavior makes it difficult to translate peak intensities into metabolite concentrations. A second reason is that 2D NMR spectra usually require more time to collect than 1D ^{1}H spectra. Long NMR acquisition times are impractical for metabolomics studies requiring the analysis of hundreds of samples.

In this paper we introduce a simple experimental protocol, Fast Metabolite Quantification (FMQ) by NMR (Fig. 4.2), for measuring molar concentrations of metabolites in complex

solutions using 2D ^{1}H-^{13}C NMR. This method is accurate and allows 2D experiments to be collected in as little as 12 minutes. Our method requires that all target metabolites be identified and that 2D ^{1}H-^{13}C NMR spectra of standards have been collected under comparable conditions. Metabolite identification has recently become feasible with three new public databases: the metabolomics extension of the Biological Magnetic Resonance Data Bank (BMRB, www.bmrb.wisc.edu) (Seavey, Farr et al. 1991; Markley, Anderson et al. 2007); the Madison Metabolomics Consortium Database (MMCD, http://mmcd.nmrfam.wisc.edu) (Cui, Lewis et al. 2008); and the Human Metabolome Project (HMP, www.metabolomics.ca) (Wishart, Tzur et al. 2007). We describe how to identify metabolites using these tools and demonstrate the process by identifying and quantifying metabolites in *Arabidopsis thaliana*, *Saccharomyces cerevisiae* and *Medicago sativa* tissue extracts. In addition, we estimate the technical error associated with 1D and 2D NMR analyses of complex mixtures. The protocol we describe is a general quantitative procedure that can be applied to any biological system in which 50 mg (extract dry mass) samples can be obtained.
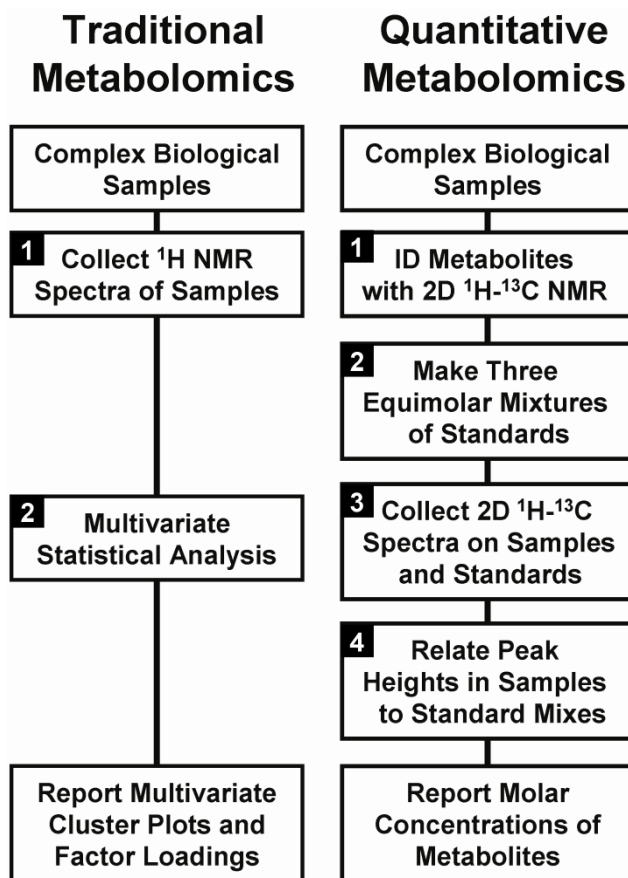
Figure 4.2. (Left) The two-step experimental design used in traditional metabolomics, which typically reports multivariate statistics related to spectra. (Right) Our four-step quantitative metabolomics protocol, "FMQ by NMR", which yields molar concentrations for all identified metabolites.

4.3. Materials and Methods

Experimental Rationale. One of the main purposes of this study was to evaluate 1D ^{1}H NMR and 2D ^{1}H-^{13}C NMR as quantitative tools for metabolic profiling. To do this, we identified ~80% of the NMR-observable metabolites present in *Arabidopsis, Saccharomyces* and *Medicago* extracts. A subset of these compounds was selected to represent an average extract, and several synthetic mixtures of pure compounds were prepared. All of the error estimates reported in this study are based on the synthetic mixtures rather than real biological extracts. This approach allowed us to measure the absolute error associated with the experimental techniques. Although our synthetic samples contained considerably fewer compounds than a real extract, the mixtures were created with biologically realistic concentrations and chemical diversity.

Plant Growth and Tissue Extraction. *Arabidopsis thaliana* seeds were germinated and grown in sterile liquid cultures of MS medium (Murashige 1978). Seedlings were incubated under continuous illumination on a shaker platform. After two weeks of growth, plants were harvested, washed in ddH$_2$O and flash-frozen in liquid nitrogen. Wild type (DS10) *Saccharomyces cerevisiae* were prepared by growing a culture in YPD medium until cells had reached the stationary phase (optical density of 15). Cells suspensions were centrifuged, and the pellet was washed in 20 volumes of phosphate buffered saline (10 mM sodium phosphate, 150 mM NaCl, pH 7.0). Washed cell suspensions were centrifuged, and pellet was washed in 20 volumes of phosphate-buffered saline (10 mM sodium phosphate, 150 mM NaCl, pH 7.4). Washed suspensions were centrifuged, and the remaining cell pellet was flash frozen in liquid nitrogen. *Medicago sativa* sprouts were obtained from a local grocery store. Sprouts were washed in ddH$_2$O and flash frozen in liquid nitrogen.

Frozen *Arabidopsis* seedlings, *Medicago* sprouts, and *Saccharomyces* cells were lyophilized for 48 h and homogenized in a coffee grinder. 400 mg of each dry homogenate was suspended

in 16 mL of boiling ddH_2O and incubated at 100°C in a screw-top 22 mL vial for 15 minutes. Extracts were microfiltered through ddH_2O-washed 3 kDa cutoff spin filters, and the filtrate was lyophilized to a dry powder. Each 400 mg tissue sample yielded approximately 50 mg of dry metabolite extract. The lyophilized extract was suspended in NMR buffer B: D_2O, 5 mM HEPES (4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid), 500 µM NaN_3 and 500 µM DSS (2,2-dimethylsilapentane-5-sulfonic acid) at a volume-to-weight ratio of 8.75 µL buffer B per milligram extract. The resulting solution was titrated with DCl or NaOD as needed to achieve an observed pH of 7.400 ($\pm$ 0.004).

Preparation of Synthetic Samples. Thirty synthetic mixtures were prepared for the error analysis study. Twenty-four of these samples were designated as "test mixtures", and six samples were designated as "concentration references." Both the test mixtures and concentration references contained twenty-six small molecules. Twenty-five of these standards were metabolites selected from the larger list of molecules identified in the three biological extracts (see Quantitative Protocol, Step 1). HEPES was also included in the synthetic samples as an internal concentration reference. All mixtures were prepared from weighed pure standards (Sigma-Aldrich) suspended in NMR buffer B and were titrated to an observed pH of 7.400 (+/- 0.004). Test mixtures contained nineteen metabolites with invariant concentrations (all 5 mM) and seven metabolites with variable concentrations ranging from 5.5 to 29.1 mM. Although each test mixture had a unique metabolite profile, the samples were designed to group into six classes with biologically relevant concentrations and standard deviations (Table 4.1). The six concentration reference samples were prepared with equimolar mixtures of the twenty-six metabolites. These samples contained each metabolite at 2 mM (N=2), 5 mM (N=2), or 10 mM (N=2).

A separate set of biological concentration reference standards was prepared for estimating concentrations in the three tissue extracts (Table 4.2). Biological reference samples had a total

of 52 metabolites split between three groups. The groups were designed to minimize overlap between metabolites signals in 2D ^{1}H-^{13}C NMR spectra. Biological references were prepared in the same manner as described for the synthetic concentration reference samples.

Table 4.1. Concentrations of seven target metabolites in the 24 synthetic mixtures analyzed.*

Sample	Asparagine	Alanine	Glutamine	Proline	Valine	Sucrose	Malate
A1	9.76	5.52	15.74	15.12	11.46	7.21	9.76
A2	7.22	7.21	12.32	12.67	12.92	8.31	8.4
A3	9.36	6.89	11.89	15.32	9.33	6.11	9.77
A4	10.37	7.44	13.33	13.87	12.01	6.88	8.21
B1	21.79	10.17	8.6	10.36	22.68	16.07	29.05
B2	19.58	10.71	7.42	13.72	15.24	16.25	26.84
B3	17.8	14.31	10.74	11.35	17.33	20.41	23.51
B4	19.25	11.49	10.11	12.28	18.3	16.14	25.68
C1	15.32	19.32	16.26	27.21	15.97	20.48	14.2
C2	17.68	16.74	20.78	16.52	17.32	16.53	14.62
C3	16.6	21.31	19.3	18.76	13.11	19.23	12.21
C4	18	18.8	18.9	25.44	12.25	21.08	12.58
D1	6.03	13.74	17.1	6.53	6.23	10.49	21.55
D2	6.81	14.42	15.61	6.87	6.54	13.5	17.94
D3	7.34	16.41	19.84	5.88	6.99	14.11	20.6
D4	5.97	14.31	15.32	7.5	7.66	14.63	21.33
E1	29.94	27.31	25.36	10.06	10.65	8.92	8
E2	23.24	22.48	28.92	8.43	11.53	8.74	5.9
E3	28.57	26.78	20.35	9.2	8.81	9.82	6.73
E4	22.95	22.67	25.43	9.78	8.32	9.76	6.36
F1	14.81	17.42	8.21	17.19	25.72	23.4	17.23
F2	12.17	15.87	5.78	20.55	20.11	26.51	19.03
F3	13.18	18.12	6.7	15.16	22.73	19.99	15.49
F4	13.75	18	6.72	19.19	26.31	23.49	16.32

* Each metabolite mixture contained 26 metabolites (Table 4.2). Nineteen non-target metabolites were present at 5 mM and seven target metabolites were present with variable concentrations. Concentrations were chosen to segregate the samples into four groups (A, B, C, and D) with biologically realistic means and standard deviations.

Table 4.2. Concentrations of metabolites (mM in the NMR tube) determined in extracts of *Arabidopsis, Medicago, Saccharomyces*, and synthetic mixtures of standards by fast 2D ^{1}H-^{13}C HSQC NMR spectroscopy. (OL) indicates that every peak overlapped with other signals, (-) metabolite was not observed, (IS) 5 mM internal standard, and (< 1) too dilute for quantification.

	Arabidopsis	*Medicago*	*Saccharomyce*	Synthetic
Acetate	OL	OL	OL	-
Adenosine	< 1	-	0.5	-
Alanine	6.8	8.8	26.1	5.5 to 27.3
AMP	<1	-	0.3	-
Arginine	10.4	15.3	43.7	5
Asparagine	61.5	229.7	6.9	6 to 30.0
Aspartate	6.1	6.9	4.3	5
Betaine	-	<1	0.3	-
Canavanine	-	8.4	-	-
Carnitine	-	-	0.3	-
Choline	14.7	6.2	2.5	-
Citrate	-	0.04	2.3	-
Citrulline	<1	-	-	5
Cystine	-	-	<1	-
Ethanolamin	4.6	2.9	-	-
Fructose	5.9	32.8	-	5
GABA	12.3	4.7	3.6	5
Galactose	<1	1.1	-	-
Glucose	6	8.2	-	5
Glutamate	9	0.7	96.6	5
Glutamine	37.4	0.9	12.9	5.8 to 28.9
Glycerol	10.4	2.4	35.5	-
Glycine	0.5	3.4	28.6	-
GSH	-	-	1.3	-
GSSG	<1	-	1	-
HEPES	IS	IS	IS	IS
Histidine	<1	9.1	16	5
Homoserine	-	9.1	0.7	-
Isoleucine	2	4.4	6.9	5
Lactate	OL	OL	OL	5
Leucine	1.1	2.8	14	5
Lysine	0.4	3.4	34.3	-
Malate	6.7	3.7	4.9	6 to 29.0
Maltose	0.1	-	-	5
MES	32.6	-	-	-
Methionine	-	-	0.3	-
Myo-Inositol	1.3	1.7	0.2	-
NAD	-	-	0.8	-
Ornithine	-	<1	4	5
Phenylalani	1.2	5.1	2.5	-
Proline	1.9	2.5	5.5	5.9 to 27.2
Putrescine	OL	OL	OL	5
Serine	11.5	18.6	20	5
Sorbitol	-	-	-	5
Succinate	<1	0.04	29	5
Sucrose	<1	-	-	6.1 to 26.5
THPro	-	-	2.7	-
Threonine	10.2	10.3	17	5
Trehalose	-	-	<1	-
Tryptophan	0.6	1.2	-	-
Uridine	<1	<1	<1	-
Valine	4	13.3	10.7	6.2 to 26.3

NMR Spectroscopy. All NMR spectroscopy was carried out at the National Magnetic Resonance Facility in Madison. Spectra were collected on a Varian 600 MHz spectrometer equipped with a triple-resonance (^{1}H, ^{13}C, ^{15}N, ^{2}H lock) cryogenic probe and a sample changer. The probe was tuned, matched, and shimmed by hand for the first sample. All subsequent samples were collected using an automated shimming and data acquisition macro written in house. 1D ^{1}H and 2D ^{1}H-^{13}C HSQC (heteronuclear single-quantum correlation) spectra of each sample were collected. 1D 1H spectra were collected using 90° pulses with four acquisitions, four silent scans, an initial delay of 2 s and an acquisition time of 2 s. Sensitivity enhanced ^{1}H-^{13}C HSQC spectra were collected with four scans, 32 silent scans, an initial delay of 1 s and an acquisition time of 0.3 s with broadband decoupling. The spectral width and number of increments were adjusted to achieve a good compromise between resolution and total acquisition time. Quantitative ^{1}H-^{13}C HSQC spectra were collected in 128 increments using a 70 ppm spectral width in the indirect (^{13}C) dimension. The carbon transmitter offset frequency was tuned to allow all aliphatic resonances to be contained within the spectrum. Our objective was to minimize spectral width while avoiding peak aliasing in the aliphatic region. Aromatic resonances and the anomeric resonances of sugars were allowed to wrap into the top of the spectral window (Fig. 2.7). Each quantitative ^{1}H-^{13}C HSQC spectrum required 12 min to collect. With these spectrometer settings, every molecule in the synthetic mixtures yielded at least one non-overlapped cross peak as did each of the identified metabolites in the biological extracts, with the exception of putrescine, lactate, and acetate. One high-resolution ^{1}H-^{13}C HSQC spectrum was collected for each biological extract. These spectra were acquired with 512 increments, 16 scans and a 13C spectral width of 140 ppm. High-resolution spectra were used to identify metabolites. The larger spectral width helped avoid resonance assignment errors resulting from spectral folding.

Data Processing and Statistical Analysis. All spectra were chemical-shift referenced, phased, Fourier-transformed with a shifted sine bell window function, zero-filled and peak-picked using automated NMRPipe (Delaglio, Grzesiek et al. 1995) processing scripts written in-house. Concentration calculations, regressions, and error analyses were done using automated scripts written in R, a free statistics software package (www.r-project.org). A detailed description of the calculations, annotated R scripts, and all of the raw data are available (see Data availability). Regression analyses report the best fit linear model of calculated concentrations (N=168) vs. the known concentration of each metabolite. Error analyses report the mean absolute difference between calculated metabolite concentrations (N=168) and known concentrations.

4.4. Protocol for FMQ by NMR

Step 1: Identification of Metabolites in Cell Extracts. High-resolution ^{1}H-^{13}C HSQC data were collected from each biological extract, and NMRPipe (Delaglio, Grzesiek et al. 1995) was used to process and peak-pick each spectrum (see NMR spectroscopy). A list of possible metabolite matches was obtained by cross-referencing observed ^{1}H-^{13}C chemical shifts (NMR peak locations) with shifts in the BMRB and MMCD databases. All matches were checked visually by overlaying spectra of pure standards from the BMRB onto the high-resolution ^{1}H-^{13}C HSQC spectrum of each extract. Sparky (freely available from www.cgl.ucsf.edu/home/sparky) was used to prepare the overlaid spectra. All of the assigned metabolites with concentrations greater than ~5 mM were further validated on the basis of long-range ^{1}H-^{13}C couplings observed in the 2D spectra (Fig. 2.5). A total of 52 metabolites met our assignment criteria, and twenty-six of these metabolites were included in the synthetic mixtures used in the error analysis study (Table 4.2).

Step 2: Concentration Reference Samples. Six equimolar mixtures of the identified metabolites (2 mM N=2, 5 mM N=2, and 10 mM N=2) were prepared as concentration reference samples (see Preparation of synthetic samples).

Step 3: Data Collection. All test samples and concentration reference samples were run as a block under identical acquisition parameters (see NMR Spectroscopy). The sample block was run twice to produce two technical replicates for each sample. To minimize experimental bias, a random number generator (www.random.org) was used to determine the sample order. 1D ^{1}H and 2D ^{1}H-^{13}C HSQC spectra were collected sequentially on each sample with custom macros written in-house.

Step 4: Calculation of Concentrations. For the error analysis study, two well-dispersed peaks in the ^{1}H-^{13}C HSQC spectra were selected for each of the seven target metabolites. A standard curve was constructed for each metabolite by regressing absolute peak intensities from the concentration reference samples with their known concentrations (Fig. 4.3). Standard curves were averaged across the technical replicates (N=4, two block replicates and two sample replicates), and the resulting regression coefficients were used to predict metabolite concentrations in the test samples. Concentration estimates were also averaged across technical replicates (N=4, two block replicates, and two peaks were selected from each molecule) to produce a final predicted concentration for each metabolite. Identical procedures were used to estimate concentrations from both 1D and 2D NMR data. Proton chemical shifts used in the 1D ^{1}H analysis were those identified from the positions of 2D ^{1}H-^{13}C HSQC cross peaks. These shifts were hand-verified to ensure that the correct resonance was selected. Some minor adjustments to the HSQC-based chemical shifts ($\pm$ 0.005 ppm) were necessary because of the higher resolution of the 1D ^{1}H spectra. Chemical shift translation was done with custom scripts written in R.

Step 4 (alternative): Normalized Calculation of Concentrations. All of the samples used in this study contained 5mM HEPES as an internal standard. An alternative strategy used for calculating concentrations was to normalize all observed intensities to the average signal from the two dispersed HEPES peaks. Standard curves were then constructed with the HEPES normalized intensities, and concentrations were predicted in the same manner as described above. This method corrects for any changes in the NMR sensitivity between experiments and allows standards to be collected at a different time from the test samples. Metabolite concentration estimates for the biological extracts were carried out in this manner, and all well-dispersed peaks were used in the calculations. Chemical shifts and raw intensities of peaks used for quantification are provided as text files in the supplementary materials.
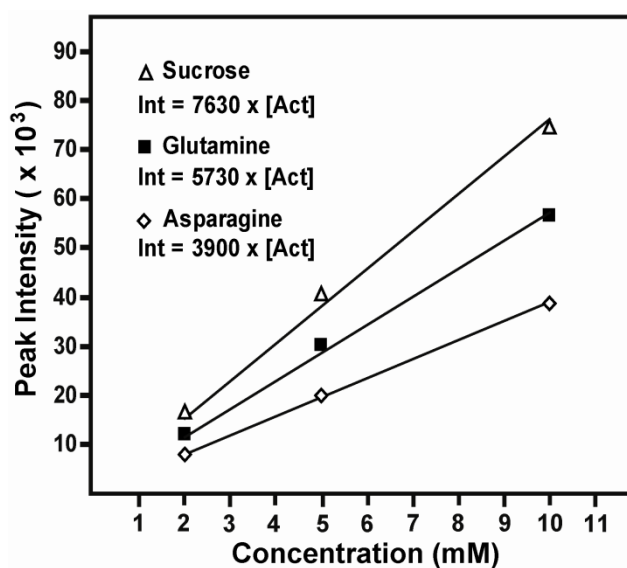
Figure 4.3. Two-dimensional ^{1}H-^{13}C HSQC NMR peak intensities for three metabolites in the concentration reference samples plotted as a function of known concentration. The concentration of metabolites in the test samples were calculated from the best fit regression lines of the concentration reference samples.

4.5. Results

   Regression and Error Analyses. One hundred sixty-eight NMR-based concentration estimates were made from NMR spectra of the 24 synthetic mixtures. Each estimate was regressed against the known concentration of the metabolite (Fig. 4.4). Parallel regressions of 2D ^{1}H-^{13}C HSQC and 1D ^{1}H NMR measurements indicated that the 1D estimates were considerably noisier than the 2D estimates (r^2 = 0.88 versus r^2 = 0.99). In addition, regression slopes indicated that the 2D ^{1}H-^{13}C HSQC estimates were precise (slope = 0.97; theoretical slope = 1), whereas the 1D ^{1}H-based data underestimated concentration (slope = 0.71; theoretical slope = 1). Error estimates calculated from the divergence of observed from actual concentrations indicated that the HSQC-based method averaged 2.7 % error with a maximum of 10.3 %, while the 1D ^{1}H analysis averaged 16.2 % error with a maximum of 44.5 % (Fig. 4.5). This error translates to an average (root mean square) accuracy of 0.6 mM for the 2D and 3.5 mM accuracy for 1D estimates.
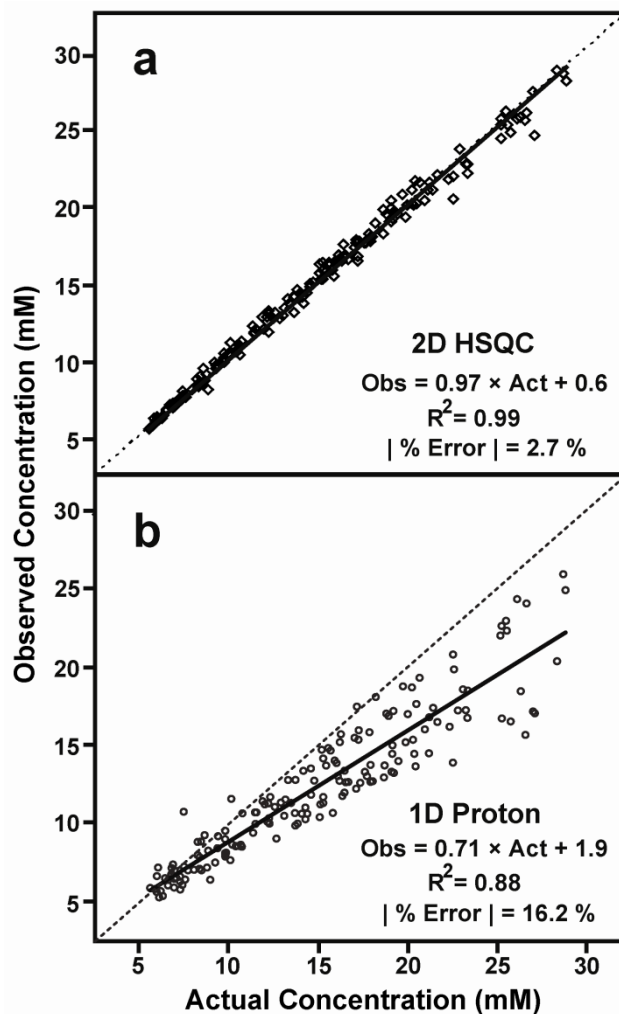
Figure 4.4. (a) Concentration estimates (N = 168) based on two-dimensional ^{1}H-^{13}C HSQC and (b) one-dimensional ^{1}H NMR data. Estimates are plotted as a function of the known concentration of metabolites in synthetic mixtures. Dotted lines indicate the ideal regression (slope = 1) and the solid lines indicate the best fit regression.
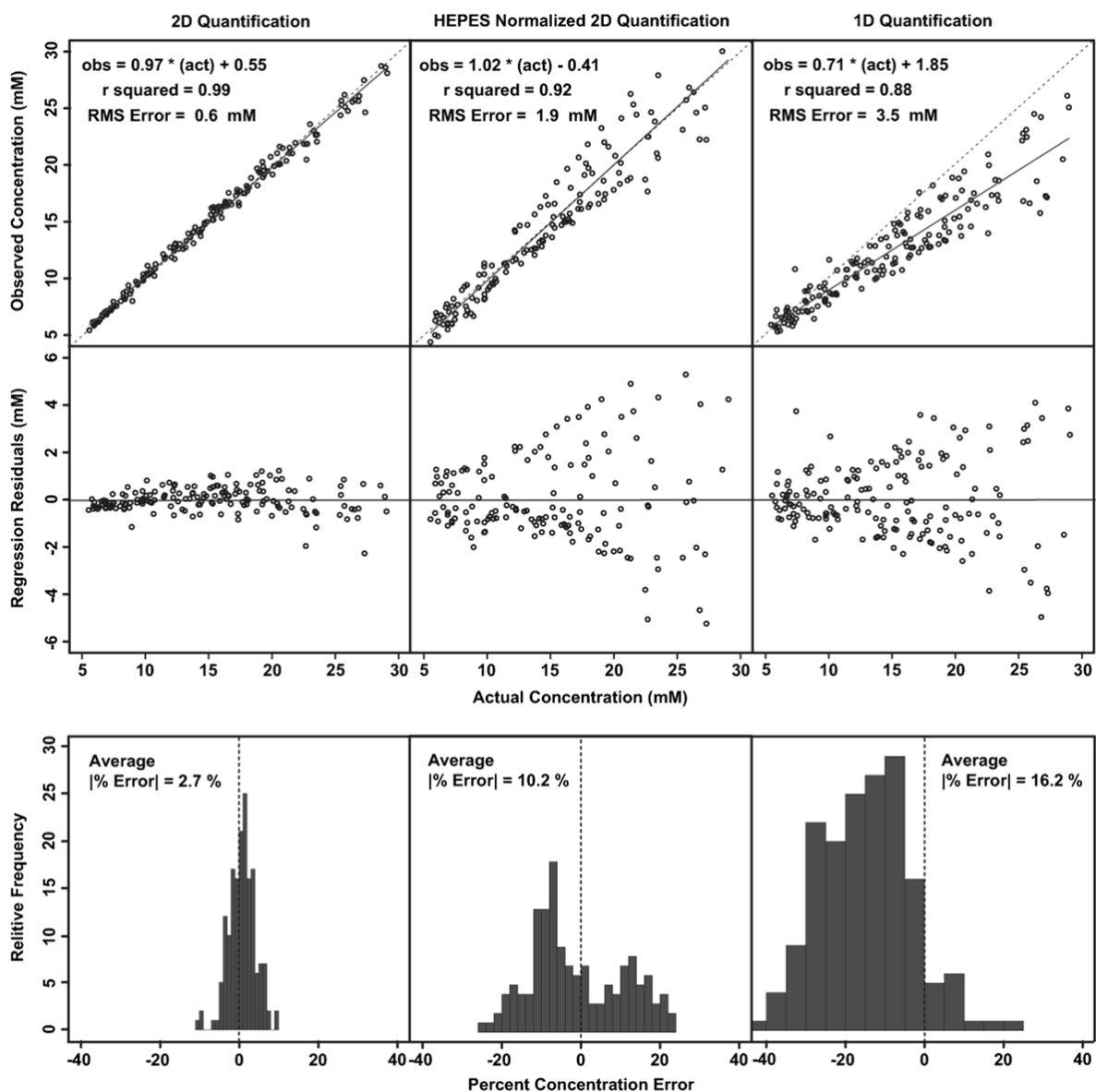
Figure 4.5. Quantitative performance of 2D ^{1}H-^{13}C NMR with (center panels) and without (left panels) HEPES normalization versus 1D ^{1}H NMR (right panels). The top row shows observed vs. actual concentrations of 168 metabolites in complex solutions of standards, the middle row shows linear regression residuals, and the bottom row shows the distribution of error. Solid lines indicate the best fit linear regression curve and dotted lines indicate the theoretical ideal (slope = 1) regression. RMS error reports the average divergence of observed from actual concentrations; average error is reported as the mean absolute percent difference between observed and actual concentrations.

Quantification of biological extracts. Fifty metabolites (excluding HEPES and MES) were identified in high resolution spectra of the three biological extracts. Forty of these were observed in *Arabidopsis*, thirty-seven were seen in alfalfa sprouts, and forty-one were present in yeast extracts. These metabolites represented, respectively: 81%, 81%, and 86% of the observed peaks in the *Arabidopsis*, alfalfa, and yeast extracts (Table 4.3). Of the 50 observed metabolites, 41 (82%) could be quantified in automated fashion from 12-min 2D ^{1}H-^{13}C data acquisitions. Observed concentrations, as measured in the NMR tube, ranged from 230 mM to as little as 40 μM (Table 4.2). Metabolite concentrations below 1 mM should be treated with caution because they are below the limit of our measured error. Half of the metabolites seen in high resolution ^{1}H-^{13}C NMR had concentrations lower than 4 mM and metabolites observed in the three tissues were distributed proportionally to the inverse of concentration (Fig. 4.6).
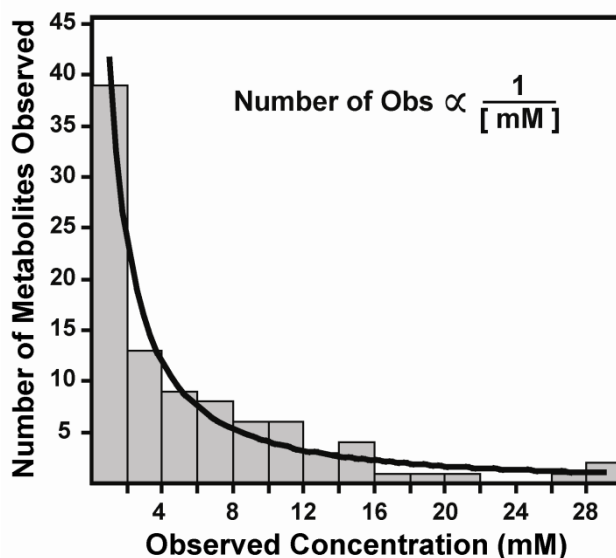
Figure 4.6. Number of observed metabolites in *Arabidopsis*, *Medicago*, and *Saccharomyces* extracts as a function of observed concentration (mM in the NMR tube; N=94). Bars indicate the total number of metabolite occurrences in the three extracts within consecutive 2 mM windows, and the regression line indicates the best-fit power regression ($y = 42 \times x^{-1.36}$, $r^2 = 0.86$). The internal standard (HEPES; N = 3), overlapped metabolites (acetate, lactate, and putrescine; N = 9) and nine metabolites with concentrations over 30 mM were excluded from the analysis. Metabolites observed in high resolution ^{1}H-^{13}C HSQC spectra that were too dilute for FMQ by NMR (N = 13) were included with the quantifiable metabolites in the 0-2 mM bar. A complete list of all observed concentrations is provided in Table 4.2.

Factors influencing quantification. It should be noted that most metabolite concentrations were predicted well beyond the range of the standard intensity versus concentration curves. In an earlier trial of this experiment, we found that higher standard concentrations (>10 mM for each metabolite) influenced the sensitivity of the probe and produced nonlinear calibration curves. When included in an analysis, the nonlinear standards caused a systematic underestimate of concentration and greater technical error. It should also be noted that the standards data must be collected at multiple concentrations for accurate quantification. Multidimensional peak intensities are influenced by a range of variables (uneven excitation, multiple relaxation pathways, etc.) that change the scaling factors for individual peaks from different metabolites (Fig. 4.3). Our data show that calculating concentrations from three standard concentrations (2 mM, 5 mM, and 10 mM) is sufficient for estimating concentrations between 1 mM and 30 mM.

We recognize that quantitative estimates are influenced by any variable that affects spectrometer sensitivity (primarily salt concentration and NMR line shape). HEPES (5 mM) was included in all the samples to act as an internal control for these variables. HEPES is a convenient choice for 2D ^{1}H-^{13}C NMR studies, because it is a pH indicator and has several well dispersed peaks (Fig. 4.1). An additional benefit of HEPES normalization is that it allows standard curves generated for one study to be used in others, provided that all of the NMR instrument settings are identical. However, we found that HEPES normalization in the synthetic samples led to greater (10.2%) technical error (Fig. 4.5). In our experience, accurate quantification (<3% error) requires measurements of absolute peak intensities referenced to concentration standards collected at the same time as the test samples.

Sample Size Requirements for Fast Quantification. Fast 2D ^{1}H-^{13}C NMR quantification requires a substantial amount of starting material. This study required 50 mg of metabolite extract to generate signals that could be quantified in 12 minutes. The sample requirement can

be reduced to 25 mg by using 5 mm susceptibility matched NMR tubes (Shigemi), which require smaller volumes. Smaller amounts of starting material will result in loss of information about metabolites present below the detection threshold. Longer NMR data collection times can partially compensate for lower concentrations, but this approach is limited because NMR sensitivity scales with the square-root of the number of scans.

4.6. Discussion

Our data indicate that 2D ^{1}H-^{13}C HSQC is superior to 1D ^{1}H NMR for quantitative analyses of complex mixtures. Although lower error estimates (~1%) have been reported for 1D ^{1}H NMR analyses (Pinciroli, Biancardi et al. 2004; Burton, Quilliam et al. 2005; Malz and Jancke 2005), these studies have been limited to well dispersed peaks. One-dimensional ^{1}H NMR spectra of unfractionated biological extracts contain hundreds of overlapping resonances. Our analysis indicates that peak overlap considerably increases technical error.

We believe that the error estimates reported here represent the practical limit of quantitative precision for 1D ^{1}H NMR based metabolomics. Our synthetic mixtures included only 26 metabolites, and we had the luxury of knowing the exact chemical shifts of every molecule from 2D ^{1}H-^{13}C HSQC spectra. NMR-based analyses of real biological samples must contend with >50 observable compounds with ambiguous resonance assignments and larger variations in chemical shift than our synthetic mixtures. Although these present serious obstacles for 1D ^{1}H NMR analysis, they can be overcome using the 2D ^{1}H-^{13}C NMR protocol we introduce here.

FMQ by NMR requires that target metabolites be identified prior to quantitative analysis. Although the identification of metabolites is laborious, it enables well controlled quantitative analyses to be based on intensities of minimally overlapping peaks. Our experimental method and guidelines for reducing NMR acquisition times should make quantitative metabolite analyses feasible for biological studies in which 50 mg of extract can be obtained.

CHAPTER 5

rNMR: open source software for identifying and quantifying metabolites in NMR spectra

Adapted from:

The original incarnation of rNMR was a basic software tool I developed to serve as an interface between the MMCD, BMRB, and an existing NMR software package (Sparky). The original software generated a preliminary assignment table from standards and uploaded these data to Sparky for visualization. However, the positions of the NMR signals differ slightly from spectrum-to-spectrum. Despite considerable efforts, and collaborations with highly competent mathematicians (Dr. Arash Bahrami and Dr. Hamid Eghbalnia), I was unable to devise a reliable algorithm for matching peak lists derived from one spectrum to signals in other spectra. My solution was to discard the traditional peak assignment approach and define NMR signals by user-manipulatable subsections of spectra. This strategy allows metabolite signals from hundreds of spectra to be visualized simultaneously and eliminates problems associated with spectrum-to-spectrum variation in peak positions. Because none of the existing NMR software supported this type of analysis, it was necessary to build a new program from scratch. I wrote rNMR with the help of an undergraduate student (Seth C. Schommer). I coded the underling bioinformatics, data processing, and graphics functions, whereas Seth coded the graphical user interfaces. In the first three months after the rNMR manuscript was published, the program was downloaded 225 times by spectroscopists in 35 countries.

5.1. Abstract

Despite the extensive use of NMR for metabolomics, no publicly available tools have been designed for identifying and quantifying metabolites across multiple spectra. We introduce here a new open source software tool, rNMR, that provides a simple graphics-based method for visualizing, identifying, and quantifying metabolites across multiple one- or two-dimensional NMR spectra. rNMR differs from existing software tools for NMR spectroscopy in that analyses are based on regions of interest (ROIs) rather than peak lists. ROIs contain all of the underlying NMR data within user-defined chemical shift ranges. ROIs can be inspected visually, and they support robust quantification of NMR signals. ROI-based analyses support simultaneous views of metabolite signals from up to hundreds of spectra, and ROI boundaries can be adjusted dynamically to ensure that signals corresponding to assigned atoms are analyzed consistently throughout the dataset. We describe how rNMR greatly reduces the time required for robust bioanalytical analysis of complex NMR data. An rNMR analysis yields a compact and transparent way of archiving the results from a metabolomics study so that it can be examined and evaluated by others. The rNMR website at http://rnmr.nmrfam.wisc.edu offers downloadable versions of rNMR for Windows, Macintosh, and Linux platforms along with extensive help documentation, instructional videos, and sample data.

5.2. Introduction

Nuclear magnetic resonance (NMR) spectra of biological extracts contain thousands of overlapping signals from many molecules. Translating this complex spectral data into concentrations of individual metabolites is one of the most significant challenges facing modern metabolomics. Recently, methods have been developed for identifying (Markley, Anderson et al. 2007; Wishart, Tzur et al. 2007; Cui, Lewis et al. 2008; Xia, Bjorndahl et al. 2008), and accurately quantifying (Lewis, Schommer et al. 2007), metabolites using multidimensional NMR and databases of metabolite standards. Although these methods are effective in small scale

studies, the significant time required for identifying metabolites, assigning individual NMR signals, and quantifying resonances has made larger scale applications of these tools impractical.

The main difficulties in translating quantitative metabolomics to larger scale studies result from the time consuming nature of traditional NMR resonance assignment. Metabolomics studies often involve fifty or more metabolites, hundreds of samples, and thousands of individual resonance assignments. Moreover, the positions (chemical shifts) of NMR signals in complex biological extracts are subject to unpredictable variation caused by small differences in solution chemistry between samples (Cloarec, Dumas et al. 2005). The complexity of the NMR spectra and inherent variations in chemical shifts make automated peak assignment error prone when assignments are transferred from one spectrum to another (Fig. 5.1A). If peak matching algorithms are allowed to be flexible to chemical shift variation, then NMR peaks can be improperly matched to neighboring signals; if tolerances are too rigid, then signals can be missed completely. As a result, reliable metabolite assignments ultimately require visual inspection of the raw NMR data. Currently, none of the publicly available software tools allow NMR signals from individual metabolites to be easily compared across multiple spectra.

To make quantitative NMR-based metabolomics more feasible in large scale studies, we have developed a simple, graphics-based method for comparing resonance assignments across multiple spectra. Our solution is based on the concept of a region of interest (ROI) (Fig. 5.1B). ROIs are dynamic user-defined subsections of a spectrum that can be actively moved or resized to enclose an NMR signal. In contrast to peak lists, which are static summaries containing limited information, ROIs contain all of the underlying NMR data within the ROI boundaries and can be rapidly inspected. We have implemented this approach in the software tool (rNMR) described here.
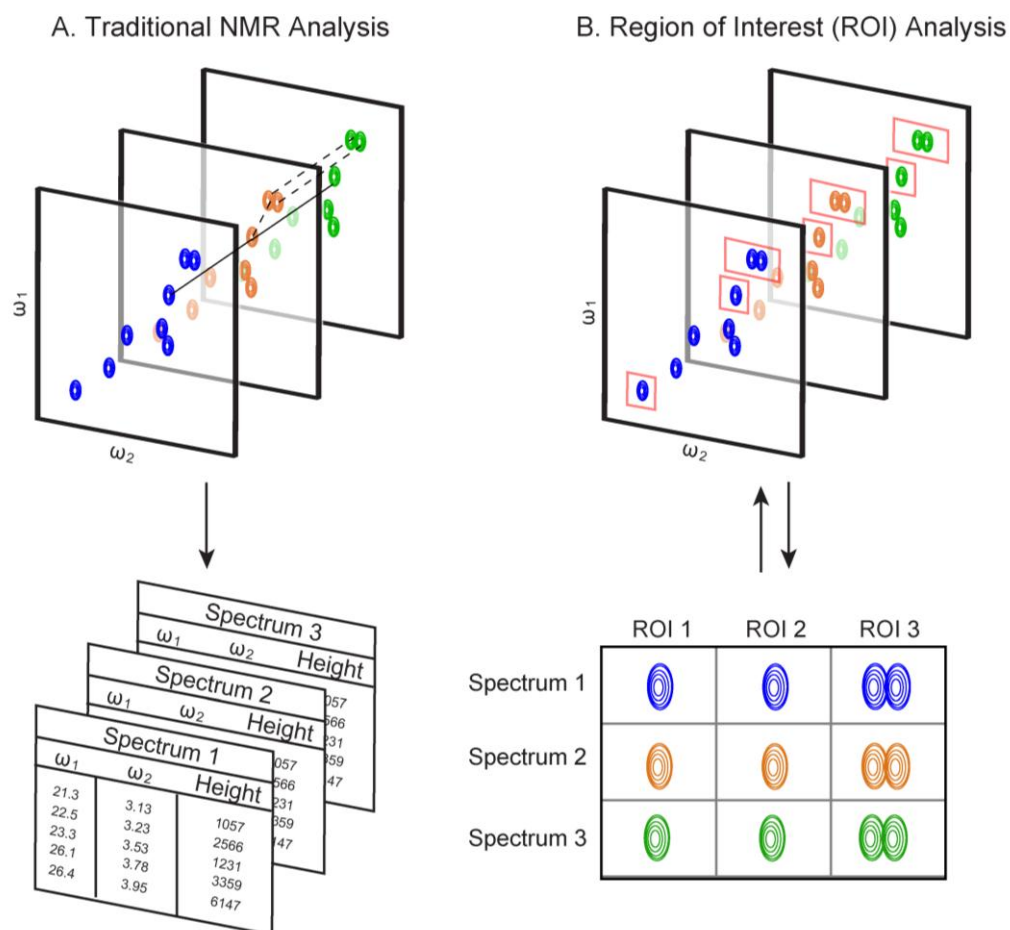
Figure 5.1. Traditional (A) versus region of interest (ROI) based (B) analysis of two-dimensional NMR data. Traditional approaches for matching signals across multiple spectra are prone to error because of chemical shift variation. The ROI-based approach used by rNMR avoids this problem by allowing users to view all of the NMR data related to a resonance assignment and to dynamically resize or move ROIs so that only the target signals are included in an analysis.

5.3. Results and Discussion

Program design objectives. rNMR was designed with four major objectives: 1) to simplify analyses of multiple NMR spectra, 2) to provide a transparent mechanism for connecting quantitative summaries with the underlying NMR data, 3) to provide a customizable framework for developing new NMR software tools and 4) to create a user friendly program for analyzing NMR spectra. We developed rNMR as an add-on package for the R statistical software environment (freely available from www.r-project.org) because R is inherently suited to these objectives. Programs written in R provide direct access to the code and data tables. Users can insert custom functions, view and modify the data, and redirect existing functions for other purposes. These manipulations can be performed at any time within the main R console. Furthermore, R is supported by extensive public libraries for mathematics, statistics, and graphics. These tools can be easily integrated into existing rNMR functions or can be applied ad hoc as the need arises. Any modifications can be readily shared with the community because rNMR's licensing (GPLv3; www.gnu.org) gives users the freedom to modify and redistribute the program.

Batch manipulations. To simplify analyses of multiple one- or two-dimensional NMR spectra, all of rNMR's basic tools were designed to support batch operations. Basic rNMR tools include functions for: overlaying spectra, displaying slices and projections of two-dimensional spectra, adjusting chemical shift referencing, peak picking, graphics settings, and a variety of plotting methods (Fig. 5.2). These functions can be applied in batch to any of the open spectra via point-and-click graphical user interfaces or line commands. Moreover, settings designated for one spectrum can be transferred directly to other spectra because rNMR's functions are designed to operate independently of the NMR acquisition parameters. Peak picking thresholds and contour levels, for example, are defined by standard deviations above the thermal noise.
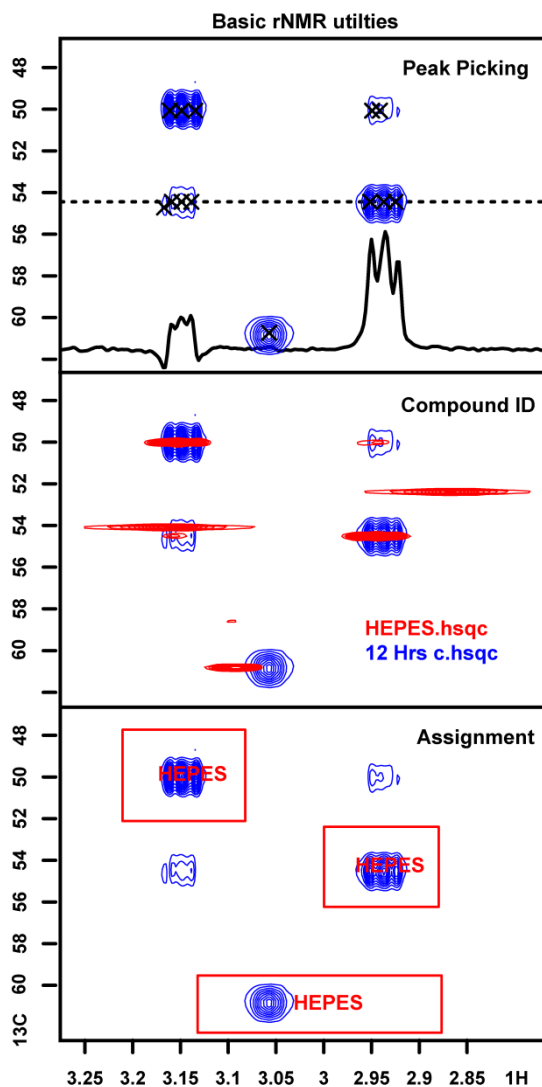
Figure 5.2. The rNMR main plotting window illustrating several basic features of the program: peak picking, one-dimensional slice visualization, spectral overlay, and ROI-based assignment. Metabolites can be identified by submitting peak lists generated by rNMR to the Madison Metabolomics Consortium Database (http://mmcd.nmrfam.wisc.edu). These automated identifications can be verified by overlaying spectral standards available from the BioMagResBank (www.bmrb.wisc.edu). Resonance assignments are thus based on ROIs enclosing each target resonance. Once ROIs have been generated, data from multiple spectra can be visualized simultaneously (Fig. 5.3) and quantified.

ROI-based tools. The ROI-based tools included in rNMR provide a simple mechanism for visualizing NMR data across multiple spectra. ROIs can be imported from a tab delimited file, designated automatically, or generated manually by clicking on a desired region of a spectrum. Once ROIs have been created (Fig. 5.2), data can be extracted from hundreds of files and displayed side-by-side in a multiple-file graphics window (Fig. 5.3). This strategy allows gigabytes of raw NMR data to be visually inspected in a few minutes and allows assignment errors detected at any stage of an analysis to be corrected by simply adjusting the ROI boundaries. Because rNMR requires assigned resonances to fall within a defined chemical shift range, and because each spectrum is subjected to the same assignment criteria, rNMR enforces consistent assignments across samples while maintaining flexibility for variation in chemical shift (Fig. 5.3).

In addition to simplifying resonance assignments, ROI-based analyses provide a transparent link between quantitative analyses and the underlying NMR data. Quantification is based directly on the NMR data displayed in the ROI windows. The NMR signal behind any data point can be visually inspected by simply clicking on the appropriate ROI. Because all rNMR analyses are regenerated on-the-fly from raw NMR data and a table containing the chemical shift boundaries of each ROI, any rNMR analysis can be duplicated and rigorously evaluated by other researchers. To illustrate this point, we have provided (at http://rnmr.nmrfam.wisc.edu) the NMR data, ROI table, and instructions detailing the three-step process required to reproduce Fig. 5.3.
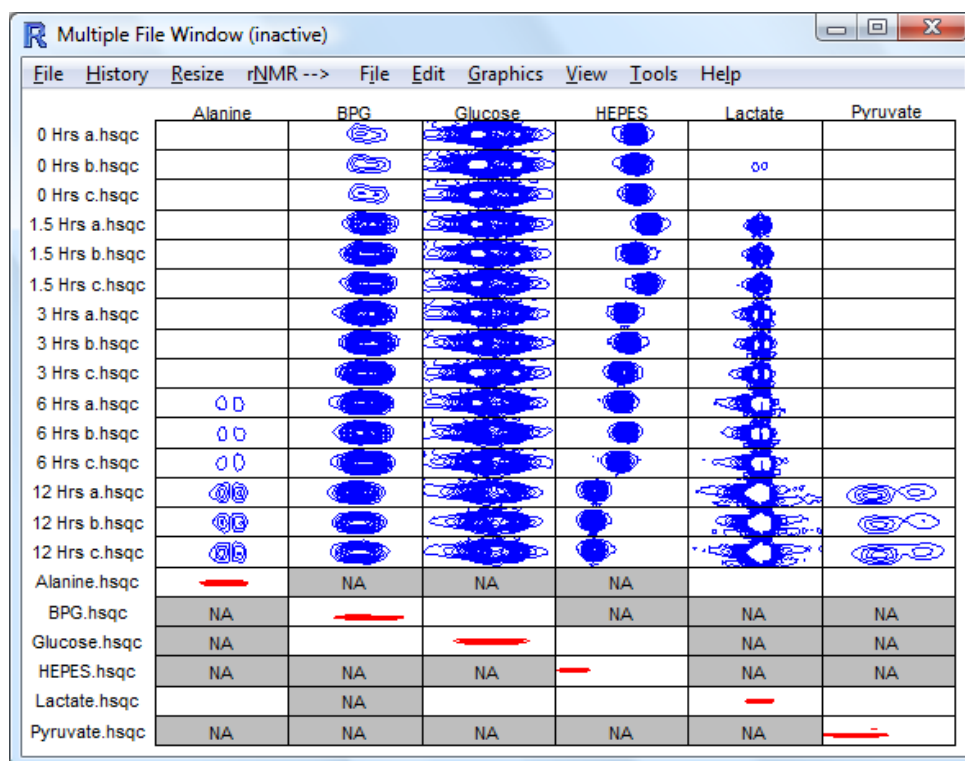
Figure 5.3. A screen shot from the rNMR multiple file window showing ROIs containing 2D NMR signals (blue) from six molecules observed in fifteen spectra of red blood cell extracts incubated with [U-13C]-glucose. Spectra correspond to aliquots withdrawn over a 12-h time course. The red peaks are from small molecule standards corresponding to each of the observed metabolites. HEPES, which was included in the incubation medium, serves as a monitor of pH; its chemical shift can be used to determine the pH of each aliquot. If an ROI falls outside the chemical shift range of a spectrum, rNMR displays a gray box with the designation 'NA'.

rNMR user interface. One of our major objectives in designing rNMR was to provide a user friendly interface for new users while providing more advanced users an easy mechanism for modifying the program. For new users, we have created a suite of point-and-click graphical tools that are accessible from drop-down menus. Although these tools provide direct accesses to all of rNMR's features, the main objective in their design was to provide a simple interface for interacting with data. Wherever possible, we have combined or eliminated superfluous features to simplify the user interface and to minimize the number of steps required to conduct quantitative analyses. Much of this economy has been made possible by rNMR's "what you see is what you get" philosophy. Quantitative analyses and peak picking thresholds, for example, are based directly on the data displayed to the user. This approach has the advantages of both eliminating redundancy and making quantitative procedures more transparent.

To facilitate customization of rNMR, the program was designed with a pseudo object-oriented framework in which all of the plotting and quantitative functions act on a series of user-modifiable lists and tables. All of rNMR's graphics are regenerated on-the-fly from these objects. This strategy allows custom functions be seamlessly integrated into the existing software. In addition, all of rNMR's features are accessible via line command and can be strung together to create standardized processing scripts. We have also provided a suite of utility functions that simplify customization of the program. These tools include functions for manipulating graphics settings, for creating custom dialog boxes, and for managing windows.

Support for rNMR. We designed rNMR initially to solve practical problems encountered in our own research. In expanding this software for more general metabolomics applications, we have solicited feedback from rNMR beta testers around the world. The current release of rNMR has undergone more than six months of rigorous beta testing, and several of the features currently implemented in rNMR were suggested by outside users. We have structured rNMR to be a community maintained program, and we encourage others to modify it to meet their needs.

We plan to continue supporting our version of the software and are actively expanding rNMR's capabilities with input from the user community.

5.4. Conclutions

rNMR is an open source software tool that solves three critical problems facing modern bioanalytical metabolomics: the difficulty in analyzing one- and two-dimensional NMR data from studies involving large numbers of spectra, the combinatorial problems associated with automated resonance assignments, and the difficulty in evaluating and documenting NMR-based analyses. The rNMR software program, in combination with existing tools for identifying (Markley, Anderson et al. 2007; Wishart, Tzur et al. 2007; Cui, Lewis et al. 2008; Xia, Bjorndahl et al. 2008) and quantifying (Lewis, Schommer et al. 2007) metabolites in complex biological extracts, offers a robust platform for extending quantitative bioanalytical metabolomics to large scale studies. Versions of rNMR for Windows, Macintosh, and Linux, plus extensive help documentation, instructional videos, and sample data (including all of the data used to create Fig. 5.2 and Fig. 5.3), are freely available from http://rnmr.nmrfam.wisc.edu.

CHAPTER 6

Method for Determining Carbon-13 Isotopic Enrichment of Metabolites in Complex Solutions

Adapted from:

Ian A. Lewis, Ryan H. Karsten, Mark E. Norton, Marco Tonelli, William M. Westler and John L. Markley. NMR Pulse Sequence for Measuring Carbon-13 Isotopic Enrichment of Metabolites in Complex Solutions. (submitted)

The NMR pulse sequence and analytical strategy presented here were originally designed as a method for quantifying metabolites that did not rely on empirical correction. My strategy was to use isotopic enrichment ratios between unlabeled extracts and enriched standards to derive concentrations. I was interested in this approach because most of the quantitative problems in multidimensional NMR pulse programs affect labeled and unlabeled versions of the same molecule equivalently. Thus, isotopic ratios normalize many sources of quantitative error. Unfortunately, the few NMR phenomena that are not canceled by this approach have a major influence on the observed signal intensities. Although I later devised a method for controlling the biggest contributor to isotope-related quantitative problems (Chapter 7), accurate quantification still requires empirical calibration. Despite this setback, the method I present here is useful for measuring isotopic enrichment and allows metabolomics studies to capitalize on isotope-based analytical strategies. I conducted this study with the help of an undergraduate student (R.H.K.) and designed the pulse program with the help of NMRFAM scientists (W.M.W and M.T.). The biological samples used in this chapter were contributed by M.E.N.

6.1 Abstract

Isotope-based methods are commonly used for metabolic flux analysis and metabolite quantification in biological extracts. Nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical tool for these studies because NMR can unambiguously identify compounds and accurately measure ^{13}C enrichment. We have developed a new pulse sequence, Difference Edited TOtal Correlation Spectroscopy (DETOCS), that filters two-dimensional ^{1}H-^{1}H NMR spectra from ^{12}C- and ^{13}C-containing molecules into separate, quantitatively equivalent spectra. DETOCS spectra of labeled and unlabeled molecules are directly comparable and can be assigned using existing bioinformatics tools. In this study, we evaluate DETOCS using synthetic mixtures of standards and extracts from *Escherichia coli*. We show that DETOCS has low technical error (6.6% for metabolites ranging from 0.34 to 6.2 mM) and can detect molecules at concentrations less than 10 μM. We propose DETOCS as a practical NMR strategy for metabolic flux analysis, isotope dilution experiments, and other methods that rely on carbon-13 labeling.

6.2. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the primary analytical tools used for investigating metabolites in complex biological extracts. Recent advances in NMR techniques (Lewis, Schommer et al. 2007), data analysis software (Xia, Bjorndahl et al. 2008; Lewis, Schommer et al. 2009), and spectroscopic libraries of metabolite standards (Markley, Anderson et al. 2007; Wishart, Tzur et al. 2007), have considerably increased the number of metabolites that can be identified and quantified in routine studies. Currently, more than 80% of the signals observed in ^{1}H-^{13}C spectra of biological extracts can be assigned (Lewis, Schommer et al. 2007), and hundreds of spectra can be analyzed in a few hours (Lewis, Schommer et al. 2009). However, these tools are primarily constrained to steady-state analyses of metabolites.

A wide variety of isotope-based methods have been developed for tracing metabolic pathways (Vallino and Stephanopoulos 1993), measuring reaction kinetics (Northrop 1975), and quantifying molecules after extensive sample handling (Rittenberg and Foster 1940). These techniques are of obvious utility to metabolomics, but have not been widely applied to NMR-based studies because of complications related to signal overlap. NMR spectra of unenriched biological extracts can contain thousands of ^{1}H resonances; spectra of ^{13}C enriched extracts are further complicated by ^{1}H-^{13}C J-couplings. In isotope-based metabolomics studies, which involve complex mixtures of both ^{13}C labeled and unlabeled metabolites, NMR spectra are too heavily overlapped to support comprehensive quantitative analyses. As a result, studies that have employed isotope-based methods have been restricted to the subset of metabolites with well isolated resonances (Carvalho, Jeffrey et al. 1998; Lane and Fan 2007; Massou, Nicolas et al. 2007; Massou, Nicolas et al. 2007).

To make isotope-based techniques more accessible to comprehensive metabolic analyses, we have developed a new homonuclear ^{1}H NMR experiment that separates signals from ^{12}C- and ^{13}C-containing molecules into distinct, quantitatively equivalent spectra. This pulse sequence, Difference Edited TOtal Correlation Spectroscopy (DETOCS), allows metabolomics studies to capitalize on isotope-based methods without increasing the complexity of spectra. DETOCS is based on the established TOCSY pulse sequence (Braunschweiler and Ernst 1983; Bax and Davis 1985; Shaka, Lee et al. 1988), which is desirable for metabolomics (Morvan, Demidem et al. 2003; Sandusky and Raftery 2005; Massou, Nicolas et al. 2007; Swanson, Keshari et al. 2008; Zhang, Bruschweiler-Li et al. 2008) because of its high sensitivity and compatibility with existing spectroscopic libraries and bioinformatics tools (Markley, Anderson et al. 2007; Wishart, Tzur et al. 2007; Cui, Lewis et al. 2008; Xia, Bjorndahl et al. 2008; Lewis, Schommer et al. 2009). We introduce here the DETOCS pulse sequence and evaluate its performance under both controlled and biologically relevant conditions. We show that DETOCS

is a reliable quantitative tool, and we foresee its potential utility in metabolic flux analysis, isotope dilution experiments, and other methods that rely on isotopic labeling.

6.3. Materials and Methods

DETOCS Pulse Sequence. DETOCS data are collected as two ^{1}H-^{1}H DIPSI-TOCSY spectra (Braunschweiler and Ernst 1983; Bax and Davis 1985; Shaka, Lee et al. 1988), one specific to ^{12}C and the other isotopically non-specific (^{12}C+^{13}C). Signals from ^{13}C-bound protons are removed from ^{12}C-specific spectra by the use of $90°(^1H)–1/4^1J_{HC}–180°(^1H,^{13}C)–1/4^1J_{HC}–90°(^{13}C)$-PFG filter elements. These low-pass J filters convert antiphase ^{13}C-bound ^{1}H magnetization to undetectable double- and zero-quantum coherence (Kogler, Sorensen et al. 1983). DETOCS employs three consecutive filtering elements with $1/4^1J_{HC}$ delays tuned to aromatic (170 Hz), aliphatic (140 Hz), and anomeric (125 Hz), coupling constants (Fig. 6.1). Isotopically non-specific spectra are collected using the same pulse sequence, but with the carbon transmitters tuned 300 ppm off-resonance to prevent the filtering elements from functioning. This approach ensures that the complementary spectra are collected under identical conditions. ^{12}C+^{13}C and ^{12}C data are collected in interleaved scans, and both utilize carbon decoupling to minimize spectral complexity. ^{13}C-specific data are derived as a difference spectrum from the non-specific (^{12}C+^{13}C) and ^{12}C-specific spectra.
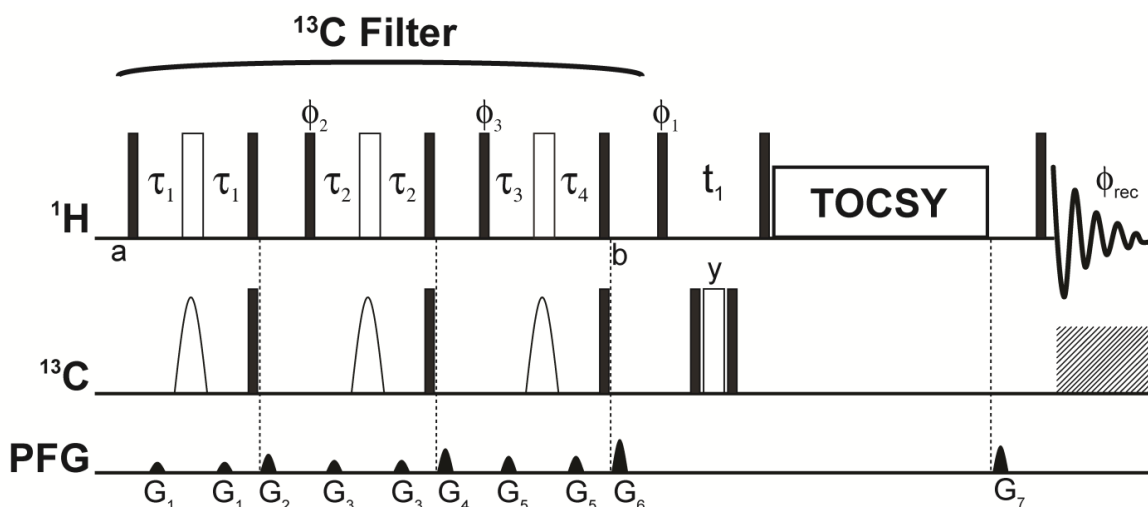
Figure 6.1. Pulse program for acquiring DETOCS. Filled and open bars indicate 90° and 180° hard pulses, respectively. Shaped ^{13}C pulses denote 1 ms (200 ppm) hyperbolic secant inversions. Isotope-specific and non-specific data are collected on alternate scans; filtered data are collected with the ^{13}C carrier set on resonance (70 ppm), and isotopically non-specific data are collected with the ^{13}C carrier set off resonance (300 ppm) for pulses between points **a** and **b**. The τ_1, τ_2 and τ_3 delays are set to ¼ $^1J_{HC}$ with the filtering elements optimized for 125 Hz, 140 Hz and 170 Hz coupling constants. All ^{13}C pulses are placed on resonance after point **b** for ^{13}C decoupling; a composite pulse (90x-180y-90x) and a WURST CPD sequence (140 ppm) are used for ^{13}C decoupling during the t_1 and t_2 evolution periods, respectively. Isotropic mixing is achieved via a 80 ms DIPSI-2 sequence with a bandwidth of 4 kHz. Pulses are applied along the x-axis unless otherwise indicated. Phase cycling for $\phi_1 = x,-x$; $\phi_2 = x,x,-x,-x$; $\phi_3 = x,x,x,x,-x,-x,-x,-x$; $\phi_{rec} = x,-x,-x,x,-x,x,x,-x$. Indirect quadrature detection requires a second FID for each increment, with the phase cycling for ϕ_1 set to y and −y; states-TPPI requires the ϕ_1 and ϕ_{rec} phases to alternate 180˚ every other increment. Pulse field gradients (PFG) employ shaped z-axis WURST pulses applied for 1.0 ms. Gradients G_1, G_3 and G_5 select magnetization inverted by the 180° ^{1}H pulses and are varied between 2 and 5 G/cm. Gradients G_2, G_4, G_6 and G_7 destroy xy-plane magnetization and are set to 7, 9, 11 and 5 G/cm, respectively.

Preparation of Synthetic Mixtures. The quantitative efficacy of DETOCS was evaluated by analyzing synthetic mixtures containing both ^{13}C labeled and unlabeled standards prepared at known concentrations. Synthetic mixtures (N = 24) contained thirty biologically relevant compounds, six of which included both ^{13}C labeled and unlabeled species. These differentially labeled compounds were alanine, fructose, glucose, glutamate, glutamine, and lactate. All unlabeled molecules were prepared at 2 mM, and the concentrations of ^{13}C compounds ranged from 130 µM to 6.5 mM (Table 6.1). All synthetic samples were prepared in D_2O containing 300 µM NaN_3 and 300 µM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Samples were titrated with DCl/NaOD as needed to an observed (glass electrode) pH of 7.400 ± 0.004.

Table 6.1. Concentrations of [U-^{13}C] metabolite standards in synthetic metabolite mixtures. Each mixture contained 30 unlabeled metabolites** prepared at 2 mM. Synthetic samples used for the companion publication (Lewis, Karsten et al. 2010) were prepared as a 1:1 dilution of each mixtures with a 2 mM Fe(III)-EDTA solution.

	Alanine	Fructose	Glucose	Glutamate	Glutamine	Lactate
1	2.15	5.76	3.71	4.96	1.59	1.87
2	2.69	5.22	1.06	3.22	3.31	4.54
3	1.48	4.96	0.27*	5.50	4.77	3.07
4	0.54	2.95	3.05	1.74	5.30	6.41*
5	4.85	4.29*	0.27	2.95	1.59	6.14
6	2.02	2.01	1.86	6.04*	4.90	3.20
7	0.40*	5.63	4.11	4.29	4.24	1.33
8	3.37	5.63	3.05*	4.29	0.79	2.94*
9	0.67	3.62	6.10	1.61	4.10	3.87
10	2.42	4.42	4.51	4.96	1.59	2.14
11	5.93	6.03	2.52	2.28	1.06	2.27
12	4.98	0.67	5.57	4.96	2.12	1.74
13	2.15	4.82	3.45	1.88	3.97	3.74
14	6.33	1.88	3.31	0.54	5.56	2.40
15	6.46	1.88	2.25	4.43	2.65	2.40
16	2.69*	0.13*	4.11	5.10	3.84	4.14
17	4.71	4.42	3.18	1.21	6.35*	0.13*
18	6.46*	1.74	1.06	3.89	2.52*	4.41
19	4.04	4.69	4.51	2.15	1.59	3.07
20	4.58	6.03*	1.06	2.82*	3.04	2.54
21	5.12	4.96	3.84	0.27*	0.26*	5.61
22	2.42	5.63	2.78	1.74	2.38	5.07
23	3.77	1.34	6.36*	5.90	0.53	2.14
24	3.10	2.95	4.77	2.15	2.78	4.27

* Signals from these metabolites were excluded from the regression analysis (Fig. 6.3) because they were used as standards for deriving the empirical correction factors.

** Unlabeled metabolites included acetate, adenosine, alanine, arginine, asparagine, aspartate, betaine, choline, citrate, fructose, 4-aminobutyrate, glucose, glutamate, glutamine, glycerol, glycine, HEPES, histidine, isoleucine, lactate, leucine, malate, ornithine, phenylalanine, proline, serine, succinate, trans-4-hydroxyproline, threonine, valine.

***In Vivo* Isotopic Labeling Strategy.** DETOCS-based quantification requires isotopically labeled standards, but it is impractical to purchase the number of enriched compounds required for comprehensive metabolic analyses. For the osmotic stress study (see below), [U-^{13}C] standards were produced *in vivo* from *E. coli* incubated in [U-^{13}C] glucose under the control condition. To quantify metabolites in the labeled extracts, a second set of unlabeled cultures was grown from the same starter in a medium containing unlabeled glucose. Metabolites in the unlabeled cultures were quantified using established methods (Lewis, Schommer et al. 2007) and these levels were used as a benchmark for the labeled extracts. Regression coefficients used for quantifying enriched metabolites (see below) were derived from solutions of enriched extracts prepared at three dilutions (1:0, 1:1, 1:3) in D$_2$O.

Osmotic Stress Study. As a biological application of DETOCS, we investigated the metabolic response of *E. coli* incubated under varying levels of osmotic stress. Cultures (500 mL each; N = 4 per condition) of *E. coli* (MG1655) were grown in M9 media containing 22.8 mM glucose and an additional 0 (control), 150, 300, or 500 mM NaCl. To demonstrate osmolarity dependence of observed metabolic responses, additional cultures (N=4) were prepared in M9 with 500 mM NaCl and 10 mM glycine betaine, an established osmoprotectant (Cayley and Record 2003). Isotopically-enriched cultures used for metabolite standards (250 mL each; N=20) were grown in M9 medium containing 22.1 mM [U-^{13}C] glucose (Cambridge Isotope Laboratories). All cultures were incubated on a shaking platform at 37 °C. When cultures reached an optical density of 0.80 (660 nm) they were transferred to an ice-water bath to minimize metabolic activity. All subsequent sample preparation was conducted in a 4 °C cold room. Chilled samples (500 mL for natural abundance, 250 mL for ^{13}C cultures) were centrifuged (18,500 $\times$ g), and the supernatant was discarded. Pellets were washed with 10 mL of glucose-free M9 media osmotically adjusted with NaCl to match each of the culture

conditions. Washed pellets were re-centrifuged, and the resulting pellets were flash frozen in liquid N_2.

Preparation of Biological Extracts. Frozen *E. coli* pellets were resuspended in 16 mL of boiling water containing 250 µM MES. Samples were incubated in sealed reaction vials in a boiling water bath for 7.5 min, then centrifuged to remove cellular debris (8,000 × g). The pellet was discarded, and extracts were microfiltered (Vivaspin 20; 3000 MWCO) to remove high molecular weight components. Filtrates were lyophilized and redissolved in 800 µL D_2O with 300 µM NaN_3 and 300 µM DSS. Samples were titrated with DCl/NaOD as needed to an observed pH of 7.400 (± 0.004). All ^{13}C labeled samples were pooled to create a single ^{13}C enriched library. Labeled standards were mixed with unlabeled samples (300 µL each) and transferred to 5 mm NMR tubes (Wilmad) for spectroscopic analysis.

NMR Spectroscopy. All NMR spectroscopy was conducted at the National Magnetic Resonance Facility at Madison. Two-dimensional ^1H-^1H DETOCS spectra were collected on a 600 MHz Varian spectrometer equipped with a cryogenic probe. Spectra were collected with a 1.5 s repetition delay, 64 steady-state transits, 2 acquisition transits, 128 increments, and an acquisition time of 0.5 s (7530 points). Sweep widths for the direct and indirect dimensions were 7500 Hz and 5400 Hz, respectively. ^{12}C+^{13}C and ^{12}C-specific spectra were collected concurrently using interleaved scans. Time-domain data were Fourier-transformed with a shifted sine bell window function, zero-filled, phased, and referenced to DSS using automated NMRPipe (Delaglio, Grzesiek et al. 1995) macros written in-house. ^{13}C-specific difference spectra were derived from the ^{12}C+^{13}C and ^{12}C-specific data using custom NMRPipe macros.

Metabolite Quantification. All NMR data analyses were performed using the rNMR software package (Lewis, Schommer et al. 2009). Metabolite signals in DETOCS spectra were assigned using previously established methods (Lewis, Schommer et al. 2007). Briefly,

metabolites were identified by submitting peak lists to the Madison Metabolomics Consortium Database (Cui, Lewis et al. 2008); possible metabolite matches were verified by visually inspecting overlaid spectral standards from the BioMagResBank (Markley, Anderson et al. 2007). Dispersed resonances were selected for each of the assigned metabolites, and the peak intensities of these signals were measured.

Deriving accurate metabolite concentrations from multidimensional NMR signal intensities requires the use of calibration coefficients (Lewis, Schommer et al. 2007). These isotope- and resonance-specific coefficients (m_{12}, m_{13}) are the linear regression slopes describing concentration as a function of intensity (Δ concentration / Δ intensity) for metabolite signals in standards prepared at three concentrations (2, 5, and 10 mM for unlabeled *E. coli* standards; 1:0, 1:1, and 1:3 dilutions of [U-^{13}C] *E. coli* extracts; minimum, median, and maximum concentrations for synthetic samples). Metabolite concentrations were calculated via three methods (eq. 1-3). The established Fast Metabolite Quantification (FMQ) method (Lewis, Schommer et al. 2007), which was used as a benchmark for *E. coli* study, determines the total concentration (U_{tot}) of unlabeled (U_{12}) and natural abundance ^{13}C (U_{na}) resonances from signal intensities (I_{12}) observed in ^{12}C-specific DETOCS spectra (eq. 1).

$$U_{tot} = U_{12} + U_{na} = I_{12} \times m_{12} \tag{1}$$

Metabolites were also quantified using the calibrated DETOCS method (eq. 2-3), which relates unenriched signals (I_{12}) to the corresponding resonances from [U-^{13}C] labeled molecules (I_{13}). The calibrated DETOCS approach can be used to either (eq. 2) calculate concentrations of enriched molecules (U_{13}) relative to known concentrations of unenriched standards (S_{12}) or (eq. 3) calculate concentrations of unenriched molecules (U_{tot}) relative to concentrations of [U-^{13}C] standards (S_{13}). The equations used for calibrated DETOCS calculations differ between S_{12}- and S_{13}-standarized experiments to account for the natural abundance levels of ^{13}C (1.1%) in

unenriched molecules. In this study, the S_{12} protocol (eq. 2) was used for calculating metabolite concentrations in synthetic mixtures whereas the S_{13} protocol (eq. 3) was used in the *E. coli* study.

$$U_{13} = S_{12} \times \left\{ \frac{I_{13} \times m_{13}}{I_{12} \times m_{12}} - 0.011 \right\} \tag{2}$$

$$U_{tot} = S_{13} \times \left\{ \frac{I_{12} \times m_{12}}{I_{13} \times m_{13} - (I_{12} \times m_{12} \times 0.011)} \right\} \tag{3}$$

6.4. Results

The efficacy of DETOCS-based analyses was investigated using complex mixtures of standards prepared with [U-^{13}C] labeled and unlabeled compounds (Table 6.1). As expected, DETOCS was effective for separating resonances on the basis of isotopic enrichment (Fig. 6.2). We had hoped that quantification based on ^{13}C/^{12}C ratios for individual resonances would normalize the systematic quantitative defects inherent to multidimensional NMR (Lewis, Schommer et al. 2007), and eliminate the need for external signal calibration. Although enrichment-based quantification eliminated several major sources of systematic error, differential T_1 relaxation and other isotope-related phenomena prevented accurate quantification from the uncalibrated DETOCS signals. The various sources of quantitative error and a calibration-independent method for controlling differential T_1 relaxation are presented in a companion paper (Lewis, Karsten et al. 2010).
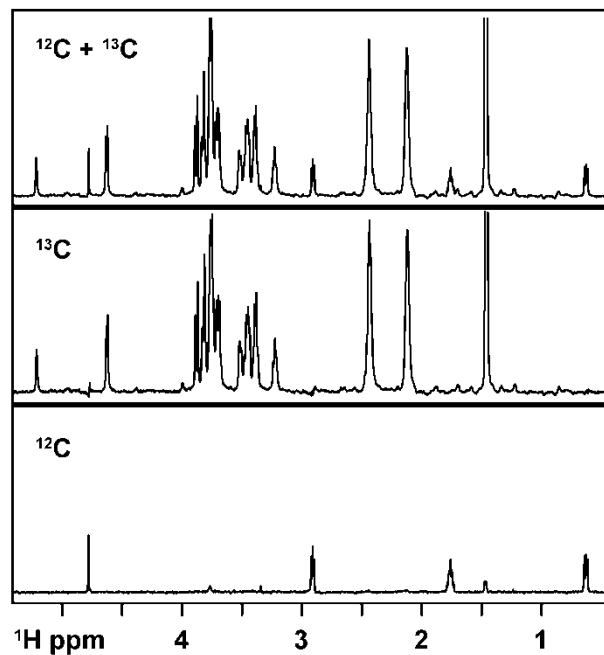
Figure 6.2. Isotopically non-specific (^{12}C+^{13}C), difference edited (^{13}C), and isotope filtered (^{12}C) 1D ^{1}H-^{1}H DETOCS spectra of a mixture of unlabeled and uniformly ^{13}C-labeled molecules. Signals in the ^{13}C spectrum correspond to [U-^{13}C]-glucose, [U-^{13}C]-glutamine, and [U-^{13}C]-alanine, whereas signals in the ^{12}C spectrum correspond to unenriched DSS.

We have previously reported a general method for correcting systematic quantitative defects in NMR pulse sequences using empirically-determined calibration coefficients (Lewis, Schommer et al. 2007). DETOCS-derived isotopic ratios were calibrated in this manner using standards prepared at multiple dilutions. As expected, the corrected $^{13}C/^{12}C$ values were quantitatively reliable; regression of 121 DETOCS-derived concentrations versus known levels yielded a slope of 0.96 (ideal slope of 1) with $R^2 = 0.98$. This translates to 6.6% average technical error for concentrations ranging from 0.34 to 6.2 mM (Fig. 6.3).

To evaluate DETOCS in the context of a biological study, *E. coli* cultures were prepared under varying levels of osmotic stress. Unlabeled extracts were mixed with a ^{13}C metabolite library, and compounds were quantified using the FMQ (Lewis, Schommer et al. 2007) and corrected DETOCS protocols. FMQ-based values were calculated from signals observed in ^{12}C-specific DETOCS spectra, whereas DETOCS-based values were calculated from $^{12}C/^{13}C$ ratios. Regression of concentrations (N=339; 0.009 to 11 mM) determined by DETOCS versus FMQ showed a slope of 1.02 and an R^2 of 0.99 (Fig. 6.4). Concentration-dependent error was observed, with values under 100 µM showing a substantial increase in error. These data indicate that the sensitivity limit of DETOCS is less than 10 µM, but the limit for reliable quantification is 100 µM.
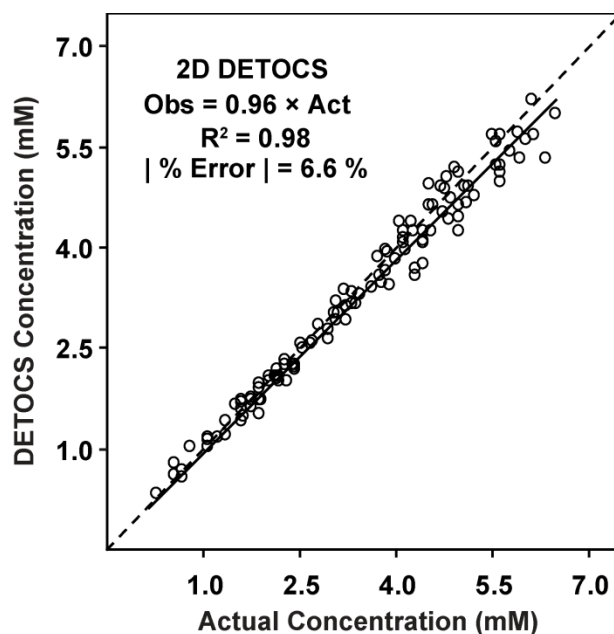
Figure 6.3. Metabolite concentrations (N = 121) measured by calibrated 2D ^{1}H-^{1}H DETOCS vs. known values. The dotted line indicates the ideal regression line (slope = 1). The data represent 24 synthetic mixtures; each mixture contained 30 metabolite standards, 6 of which were supplemented with a [U-^{13}C]-labeled version. DETOCS values were calculated from the observed ^{13}C/^{12}C ratio and were corrected for systematic quantitative errors using empirically-derived coefficients. The error estimate reflects the average absolute error for all DETOCS-derived metabolite concentrations. The compositions of the synthetic mixtures are listed in Table 6.1.
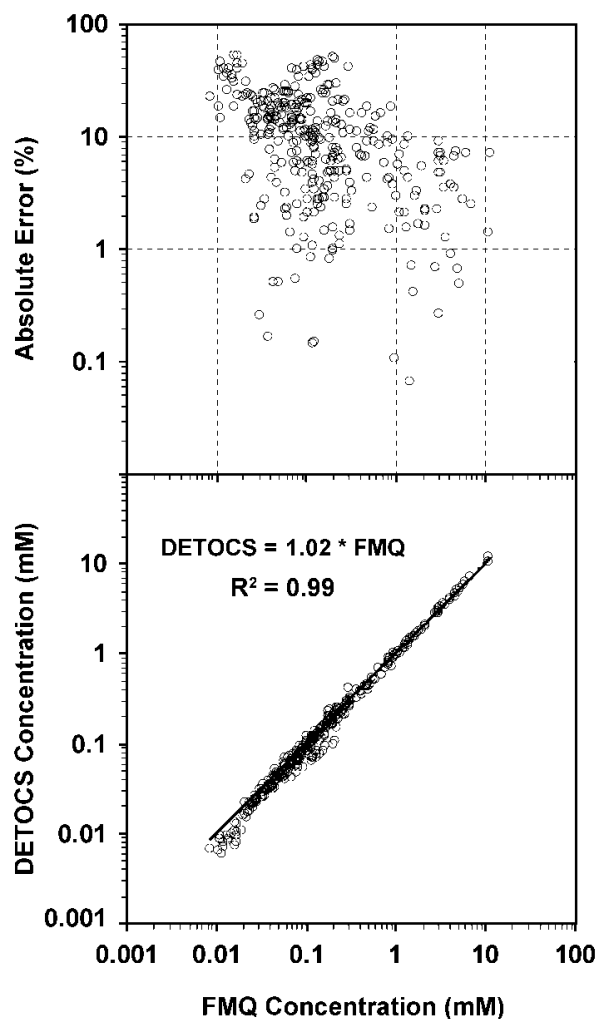
Figure 6.4. Metabolite concentrations (N = 339) and quantitative errors observed in 20 *E. coli* extracts as measured by calibrated DETOCS vs. the established FMQ protocol. Each extract was mixed 1 to 1 with a ^{13}C-enriched *E. coli* extract. FMQ values were based on the intensities of ^{12}C-specific spectra where as DETOCS values were derived from ^{12}C/^{13}C ratios observed for metabolites in each mixture. Quantitative errors observed in these data were concentration dependent; metabolites with concentrations ranging from 0.010 to 0.1, 0.1 to 0.5, 0.5 to 1, and 1 to 11 mM showed average absolute errors of 17%, 13%, 8%, and 4%, respectively. A total of 24 compounds were quantified in this study; glucose, glutamine, isoleucine, MES, phenylalanine, and trehalose were excluded from the figure because these compounds were not observable in DETOCS spectra of the ^{13}C-enriched library.

DETOCS-based analyses of the *E. coli* extracts relied on isotopically enriched standards produced *in vivo*. These standards were quantified under the assumption that steady-state metabolite levels observed in labeled and unlabeled extracts are equivalent. The strong correlation observed between FMQ (Lewis, Schommer et al. 2007) and DETOCS values verifies this assumption. If steady-state levels differed between labeled and unlabeled samples, then compound-specific systematic error would have been observed.

One limitation of DETOCS evident from the *E. coli* study is that reliable quantification requires both the labeled and unlabeled molecules to be present at concentrations above the detection limit. Six of the 24 metabolites quantified by FMQ were not observable in DETOCS spectra of the enriched extracts. Four of these compounds were present at concentrations near the limit of detection for FMQ and were unobservable in enriched extracts because of the faster R_1 relaxation of ^{13}C labeled molecules. The other two compounds, glucose and trehalose, were absent from the enriched library because these metabolites are only accumulated under high-salt conditions. The main biological findings of this study were salt-dependent increases in trehalose and glucose, diminished putrescine levels at high salt, and the reversal of these phenotypes by glycine betaine (Fig. 6.5). These findings, which had been previously reported (Munro, Morgan et al. 1972; Measures 1975; Lewis, Cayley et al. 1990; Record, Courtenay et al. 1998), were not quantifiable because of our design of the *in vivo* labeling conditions. This limitation could be mitigated in future studies by including samples from each condition as standards.
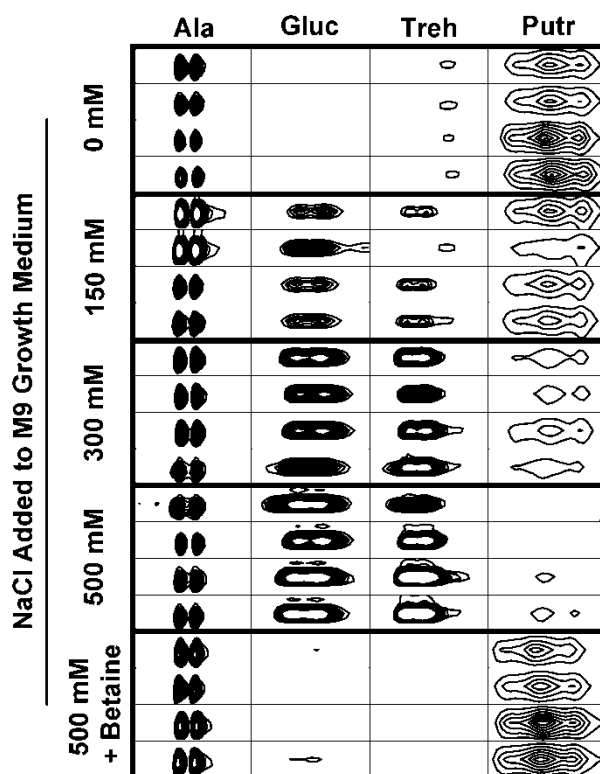
Figure 6.5. Osmolarity-dependent alterations in metabolites observed in 2D ^{1}H-^{1}H DETOCS spectra (^{12}C-specific) of *E. coli* extracts. Each box shows a metabolite-specific region of an NMR spectrum; columns denote the four selected compounds; rows denote the five conditions (N = 4 each) used for *E. coli* cultures. All spectra are shown with the minimum and maximum contour thresholds set to 25 and 55 standard deviations above the thermal noise, respectively. Steady-state concentrations of glucose (Gluc) and trehalose (Treh) are elevated under osmotic stress whereas putrescine (Putr) levels are diminished. Samples incubated in 500 mM NaCl with 10 mM glycine betaine (last four rows) show a metabolic phenotype similar to cultures grown in standard M9 medium (first four rows). Alanine (Ala) signals are shown as a reference. A complete listing of all of the metabolites quantified in this study can be found in Table 6.2.

Table 6.2. Metabolites observed in *E. coli* cultures grown under various levels of salt stress. Samples (N = 4 per condition) were cultured in M9 medium containing an additional 0, 150, 300, or 500 mM NaCl or 500 mM NaCl plus 10 mM glycine betaine. Metabolite concentrations are shown as millimolar levels observed in the NMR tubes as calculated by the FMQ and corrected DETOCS protocols, respectively.

	0 mM NaCl				150 mM NaCl			
	A	B	C	D	A	B	C	D
Acetate	0.07, 0.07	0.06, 0.06	0.11, 0.10	0.10, 0.10	0.17, 0.15	0.18, 0.18	0.17, 0.16	0.20, 0.20
Adenosine	0.17, 0.13	0.20, 0.16	0.21, 0.22	0.21, 0.20	0.19, 0.16	0.25, 0.23	0.18, 0.20	0.18, 0.18
Alanine	3.03, 2.89	2.10, 2.06	1.61, 1.57	1.33, 1.35	10.6, 10.5	11.0, 11.8	2.99, 3.17	3.17, 3.32
AMP	0.06, 0.06	0.06, 0.06	0.08, 0.07	0.08, 0.07	0.10, 0.08	0.16, 0.15	0.06, 0.05	0.07, 0.06
Aspartate	0.05, 0.04	0.08, 0.05	0.06, 0.06	0.06, 0.05	0.07, 0.06	0.15, 0.12	0.07, 0.06	0.08, 0.06
Glucose	0.05, 0.02*	0.08, 0.03*	0.07, 0.02*	0.08, 0.03*	0.39, 0.17*	0.52, 0.24*	0.39, 0.19*	0.40, 0.18*
Glutamate	0.83, 0.82	0.82, 0.79	2.12, 2.09	2.08, 2.03	2.99, 2.78	3.05, 3.04	4.56, 4.40	4.93, 4.89
Glutamine	0.05, 0.01*	0.06, 0.02*	0.06, 0.01*	0.08, 0.02*	0.10, 0.03*	0.10, 0.04*	0.12, 0.04*	0.10, 0.04*
Glutathione	0.10, 0.08	0.09, 0.08	0.17, 0.17	0.14, 0.13	0.20, 0.18	0.42, 0.36	0.21, 0.19	0.23, 0.18
Histidine	0.02, 0.01	0.02, 0.01	0.02, 0.01	0.01, 0.01	0.02, 0.01	0.03, 0.02	0.01, 0.01	0.01, 0.01
Isoleucine	0.05, 0.03*	0.05, 0.03*	0.06, 0.03*	0.05, 0.02*	0.05, 0.03*	0.08, 0.05*	0.05, 0.04*	0.06, 0.04*
Leucine	0.11, 0.10	0.10, 0.09	0.12, 0.11	0.12, 0.11	0.26, 0.21	0.23, 0.21	0.12, 0.11	0.11, 0.10
Lysine	0.11, 0.07	0.12, 0.07	0.14, 0.08	0.15, 0.08	0.16, 0.12	0.22, 0.15	0.21, 0.11	0.20, 0.10
MES	3.86, 2.03*	3.78, 2.05*	4.18, 2.09*	4.22, 1.92*	4.17, 2.04*	3.78, 2.23*	4.19, 1.97*	4.08, 2.19*
Methionine	0.06, 0.04*	0.06, 0.03*	0.08, 0.04*	0.09, 0.05*	0.08, 0.04*	0.09, 0.04*	0.10, 0.06*	0.10, 0.04*
Phenylalanine	0.03, 0.02	0.02, 0.02	0.04, 0.03	0.03, 0.03	0.04, 0.03	0.04, 0.04	0.02, 0.02	0.02, 0.02
Proline	0.18, 0.04*	0.18, 0.04*	0.10, 0.02*	0.09, 0.02*	0.22, 0.04*	0.20, 0.03*	0.10, 0.03*	0.08, 0.02*
Putrescine	0.19, 0.24	0.18, 0.23	0.31, 0.30	0.28, 0.26	0.19, 0.18	0.08, 0.49	0.15, 0.15	0.16, 0.15
Succinate	0.81, 0.74	0.75, 0.7	1.24, 1.17	1.04, 0.98	1.38, 1.24	1.97, 1.86	1.36, 1.30	1.55, 1.56
Threonine	0.07, 0.06	0.04, 0.03	0.05, 0.05	0.05, 0.04	0.09, 0.07	0.12, 0.10	0.06, 0.04	0.05, 0.05
Trehalose	0.02, 0.01*	0.03, 0.01*	0.03, 0.01*	0.03, 0.01*	0.07, 0.03*	0.04, 0.01*	0.11, 0.05*	0.10, 0.04*
Tyramine	0.05, 0.04	0.03, 0.02	0.04, 0.03	0.04, 0.03	0.05, 0.04	0.05, 0.04	0.04, 0.03	0.04, 0.03
Uracil	0.53, 0.46	0.47, 0.43	0.57, 0.51	0.60, 0.51	0.89, 0.73	0.69, 0.57	0.50, 0.44	0.55, 0.53
Valine	0.11, 0.09	0.07, 0.06	0.09, 0.08	0.07, 0.06	0.27, 0.22	0.29, 0.28	0.09, 0.08	0.09, 0.08

* Glucose, isoleucine, MES, methionine, proline, and trehalose, levels calculated by the DETOCS quantification protocol are unreliable because signals from these molecules were below the limit of detection in ^{13}C-specific DETOCS spectra.

Table 6.2 (continued).

	300 mM NaCl				500 mM NaCl			
	A	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Acetate	0.20, 0.20	0.56, 0.51	0.32, 0.31	0.16, 0.17	0.23, 0.24	0.10, 0.10	0.18, 0.20	0.28, 0.30
Adenosine	0.08, 0.07	0.08, 0.08	0.10, 0.09	0.10, 0.12	0.06, 0.07	0.08, 0.08	0.08, 0.08	0.09, 0.09
Alanine	4.04, 3.89	3.57, 3.53	4.05, 4.08	2.97, 3.24	2.78, 2.8	2.14, 2.11	3.16, 3.37	3.45, 3.58
AMP	0.06, 0.04	0.06, 0.05	0.09, 0.08	0.05, 0.05	0.04, 0.04	0.03, 0.02	0.05, 0.05	0.07, 0.06
Aspartate	0.03, 0.02	0.03, 0.02	0.03, 0.03	0.04, 0.03	0.03, 0.02	0.02, 0.01	0.04, 0.04	0.03, 0.03
Glucose	2.57, 0.81*	2.29, 0.75*	2.19, 0.86*	1.72, 0.70*	3.13, 0.84*	3.03, 0.85*	3.81, 1.04*	4.31, 1.04*
Glutamate	1.48, 1.46	0.47, 0.45	3.13, 3.35	3.04, 3.26	0.65, 0.71	0.95, 0.95	4.50, 4.79	4.71, 5.06
Glutamine	0.12, 0.03*	0.07, 0.02*	0.10, 0.04*	0.08, 0.04*	0.10, 0.03*	0.15, 0.04*	0.18, 0.07*	0.15, 0.07*
Glutathione	0.08, 0.06	0.06, 0.05	0.20, 0.20	0.14, 0.14	0.07, 0.06	0.08, 0.06	0.24, 0.24	0.20, 0.18
Histidine	0.02, 0.01	0.01, 0.01	0.02, 0.01	0.01, 0.01	0.02, 0.01	0.01, 0.01	0.01, 0.01	0.01, 0.01
Isoleucine	0.04, 0.02*	0.04, 0.02*	0.06, 0.04*	0.05, 0.03*	0.05, 0.03*	0.03, 0.01*	0.05, 0.03*	0.06, 0.03*
Leucine	0.13, 0.12	0.13, 0.11	0.12, 0.11	0.10, 0.09	0.10, 0.11	0.07, 0.07	0.12, 0.12	0.13, 0.13
Lysine	0.13, 0.07	0.12, 0.07	0.16, 0.10	0.10, 0.08	0.09, 0.05	0.11, 0.07	0.07, 0.06	0.12, 0.07
MES	3.97, 1.83*	4.01, 1.68*	4.02, 1.99*	3.82, 1.96*	3.33, 1.86*	3.12, 2.07*	3.81, 2.14*	3.96, 2.02*
Methionine	0.06, 0.03*	0.05, 0.02*	0.08, 0.04*	0.06, 0.02*	0.03, 0.01*	0.04, 0.02*	0.05, 0.02*	0.05, 0.02*
Phenylalanine	0.02, 0.02	0.03, 0.02	0.03, 0.03	0.02, 0.02	0.03, 0.02	0.01, 0.01	0.02, 0.02	0.03, 0.03
Proline	0.20, 0.04*	0.21, 0.04*	0.12, 0.02*	0.09, 0.02*	0.19, 0.04*	0.18, 0.03*	0.09, 0.02*	0.09, 0.01*
Putrescine	0.08, 0.08	0.08, 0.08	0.12, 0.12	0.09, 0.10	0.07, 0.07	0.05, 0.05	0.06, 0.07	0.07, 0.08
Succinate	1.11, 1.03	1.26, 1.15	1.74, 1.68	0.94, 1.03	0.22, 0.23	0.11, 0.10	0.85, 0.89	1.09, 1.07
Threonine	0.05, 0.04	0.06, 0.05	0.06, 0.05	0.04, 0.03	0.06, 0.05	0.04, 0.03	0.06, 0.04	0.07, 0.04
Trehalose	0.83, 0.30*	0.49, 0.16*	0.71, 0.26*	0.99, 0.28*	0.42, 0.16*	1.09, 0.32*	2.41, 0.46*	2.45, 0.36*
Tyramine	0.04, 0.03	0.04, 0.03	0.04, 0.04	0.03, 0.03	0.03, 0.02	0.02, 0.02	0.03, 0.02	0.03, 0.03
Uracil	0.30, 0.27	0.31, 0.26	0.29, 0.28	0.31, 0.31	0.32, 0.32	0.25, 0.22	0.24, 0.24	0.26, 0.24
Valine	0.15, 0.13	0.14, 0.13	0.13, 0.12	0.10, 0.11	0.13, 0.14	0.07, 0.06	0.12, 0.12	0.12, 0.12

* Glucose, isoleucine, MES, methionine, proline, and trehalose, levels calculated by the DETOCS quantification protocol are unreliable because signals from these molecules were below the limit of detection in ^{13}C-specific DETOCS spectra.

Table 6.2 (continued).

	500 mM + 10 mM Betaine			
	A	**B**	**C**	**D**
Acetate	0.09, 0.09	0.14, 0.15	0.13, 0.14	0.08, 0.08
Adenosine	0.12, 0.12	0.15, 0.11	0.12, 0.12	0.14, 0.15
Alanine	5.64, 5.80	5.20, 5.23	5.96, 6.39	6.86, 7.03
AMP	0.04, 0.04	0.04, 0.04	0.05, 0.05	0.05, 0.06
Aspartate	0.04, 0.04	0.03, 0.02	0.05, 0.04	0.12, 0.07
Glucose	0.10, 0.05*	0.18, 0.05*	0.11, 0.04*	0.08, 0.05*
Glutamate	1.26, 1.28	0.81, 0.89	2.88, 2.94	3.40, 3.60
Glutamine	0.06, 0.02*	0.10, 0.03*	0.09, 0.03*	0.13, 0.04*
Glutathione	0.12, 0.12	0.20, 0.21	0.19, 0.18	0.37, 0.33
Histidine	0.01, 0.01	0.02, 0.02	0.02, 0.01	0.02, 0.01
Isoleucine	0.06, 0.03*	0.08, 0.04*	0.07, 0.03*	0.05, 0.03*
Leucine	0.11, 0.12	0.12, 0.11	0.12, 0.11	0.12, 0.11
Lysine	0.08, 0.05	0.13, 0.07	0.13, 0.08	0.08, 0.06
MES	3.67, 2.13*	3.52, 1.89*	3.88, 2.27*	3.94, 1.93*
Methionine	0.06, 0.02*	0.05, 0.03*	0.10, 0.05*	0.08, 0.03*
Phenylalanine	0.02, 0.02	0.03, 0.03	0.03, 0.03	0.03, 0.02
Proline	0.40, 0.02*	0.34, 0.01*	0.41, 0.02*	0.35, 0.01*
Putrescine	0.21, 0.22	0.22, 0.25	0.36, 0.39	0.29, 0.41
Succinate	0.93, 0.90	0.99, 0.96	1.46, 1.46	1.79, 1.76
Threonine	0.08, 0.07	0.11, 0.09	0.13, 0.14	0.15, 0.13
Trehalose	0.04, 0.01*	0.10, 0.01*	0.08, 0.01*	0.03, 0.01*
Tyramine	0.04, 0.03	0.04, 0.03	0.04, 0.03	0.04, 0.04
Uracil	0.40, 0.38	0.46, 0.43	0.42, 0.39	0.49, 0.40
Valine	0.23, 0.25	0.20, 0.17	0.19, 0.18	0.22, 0.19

* Glucose, isoleucine, MES, methionine, proline, and trehalose, levels calculated by the DETOCS quantification protocol are unreliable because signals from these molecules were below the limit of detection in ^{13}C-specific DETOCS spectra.

6.5. Discussion

As expected, DETOCS was effective in separating signals from labeled and unlabeled molecules. Although accurate quantification required calibration, the corrected signals had sufficiently low error to be useful for metabolomics studies. We foresee the primary applications of DETOCS to be metabolic flux analysis, isotope dilution, and metabolite quantification using standards produced *in vivo*. The *in vivo* labeling strategy used in this study has obvious applications beyond the limited case presented here. Enriched metabolites produced in cost-effective systems could be used to quantify extracts from unrelated species. The ^{13}C labeled metabolite library generated for this study, for example, could be used to quantify metabolites in human serum. The main limitation to this approach is that quantification requires metabolites to be present in both library and test samples. However, a more comprehensive library could be constructed using a variety of free-living organisms, or under a range of conditions appropriate to a particular study.

The DETOCS quantification protocol is an NMR adaptation of the traditional radioisotope dilution method for measuring metabolite concentrations (Rittenberg and Foster 1940). Isotope dilution methods are advantageous because they can correct sample-to-sample variations in extraction efficiency, chromatography, and other sample handling errors if standards are added prior to any manipulation of the samples. Although extracts were analyzed as a complex solution in this study, the DETOCS quantification protocol presented here allows samples to be fractionated and concentrated without affecting quantification.

Capitalizing on enriched libraries depends on our ability to quantify the labeled metabolites. DETOCS provides a simple mechanism for achieving this; labeled compounds can be identified using existing bioinformatics tools and quantified using unlabeled standards. In summary, DETOCS is a convenient tool for differentiating between molecules on the basis of enrichment and provides the mean for extending metabolomics into comprehensive isotope-based studies.

CHAPTER 7

Method for Controlling Differential T_1 Relaxation in NMR Analyses of Metabolites in Complex Solutions

Adapted from:

Ian A. Lewis, Ryan H. Karsten, William M. Westler and John L. Markley. Method for Controlling Differential T_1 Relaxation in NMR Analyses of Metabolites in Complex Solutions. (submitted)

The strategy presented in this chapter was my second attempt at devising a method for quantifying metabolites that did not rely on external signal calibration. The theory was to couple the DETOCS quantification protocol (Chapter 6) with a method for standardizing T_1 relaxation times of NMR signals. This combined strategy was sufficient for semi-quantitative analyses, and is currently the most robust calibration-independent method for quantifying signals in NMR spectra. The most exciting aspect of this chapter is the potential application beyond the limited scope of isotopic enrichment studies. Relaxation control could allow unlabeled standards to be used for calibrating enriched extracts via my established quantification protocol (Chapter 4). This potential application of relaxation control is an ongoing area of research in the Markley laboratory. I conducted this study with the help of an undergraduate (R.H.K) with guidance from the NMRFAM (W.M.W).

7.1. Abstract

Nuclear magnetic resonance (NMR) spectroscopy is frequently used for measuring carbon-13 enrichment of metabolites in complex biological extracts. Enrichment measurements are complicated by differential longitudinal (T_1) relaxation times of NMR resonances from labeled and unlabeled molecules. Existing methods for controlling differential relaxation include long recycle delays and empirical correction, both of which are impractical in multidimensional NMR analyses of complex solutions. We propose the use of paramagnetic relaxation agents as a practical strategy for equalizing T_1 relaxation and reducing error in isotope-based NMR analyses. We show that adding 1 mM Fe(III)-EDTA to samples is sufficient for controlling differential relaxation to within 3% and that quantitative error is considerably improved in samples containing Fe(III)-EDTA. We foresee the primary applications of this strategy to be calibration-free estimates of isotopic enrichment and accurate quantification of ^{13}C labeled metabolites.

7.2. Results and Discussion

Nuclear magnetic resonance (NMR) spectroscopy is frequently used for metabolic flux analysis, isotope dilution experiments, and other studies requiring accurate measures of carbon-13 enrichment in metabolites. Isotopic enrichments are generally derived from the relative intensities of ^{12}C- versus ^{13}C-bound protons observed in one-dimensional (1D) ^{1}H or two-dimensional (2D) ^{1}H-^{1}H NMR spectra. Accurate quantification of isotopic enrichment is complicated by the differential longitudinal (T_1) relaxation times of labeled and unlabeled molecules. The anomeric ^{1}H resonance of glucose, for example, relaxes 2.3 times faster when bound to ^{13}C as opposed to ^{12}C, whereas the methyl signal of alanine is relatively unaffected by ^{13}C enrichment (Table 7.1). These variations in T_1 times represent a serious practical challenge to comprehensive isotopic analyses in complex biological extracts.

Table 7.1. Longitudinal (T_1) relaxation times of protons in a mixture containing unlabeled and uniformly ^{13}C-labeled glucose and alanine. [1]H relaxation times (in seconds) for the anomeric proton of glucose and the methyl protons of alanine were measured by a 1D inversion-recovery experiment. Samples were prepared with 1, 100, or 1000 μM Fe(III)-EDTA. Ratios in the table indicate the level of differential relaxation between protons bound to ^{12}C and [13]C. The resonances selected for this table are representative of the range of T_1 values observed in biological extracts. Inversion-recovery spectra of the entire mixture are shown in Fig. 7.1.

Fe(III)-EDTA	Glucose			Alanine		
	[12]C	[13]C	[12]C/[13]C	[12]C	[13]C	[12]C/[13]C
1 μM	3.40	1.46	2.32	1.71	1.41	1.21
100 μM	1.93	1.03	1.88	1.39	1.22	1.14
1000 μM	0.44	0.43	1.02	0.46	0.44	1.03

The traditional method for controlling differential T_1 relaxation is to collect data with a long recycle delay (Fan and Lane 2008). Quantitative NMR methods recommend a delay of five times the longest T_1 (99.3% signal recovery) (Evilia 2001). Using this strategy, a single 2D ^{1}H-^{1}H NMR spectrum requires over 8.5 h to collect (assuming 4 transits per increment, 256 increments, and states-TPPI quadrature detection) due to the long T_1 values of many metabolite signals (3 s; Fig. 7.1). An alternative strategy is to correct differential T_1 relaxation by applying empirically-determined correction coefficients (Lewis, Karsten et al. 2010). Correction factors can be derived from signals observed in standards prepared at multiple concentrations (Lewis, Schommer et al. 2007) or by measuring T_1 relaxation times for each resonance using inversion-recovery experiments (Vold, Waugh et al. 1968). Although empirical correction methods are effective (Lewis, Schommer et al. 2007; Lewis, Karsten et al. 2010), they require additional experimental steps and external calibration of the observed NMR signals. Moreover, standards-based calibrations are limited by the expense and unavailability of enriched standards, and comprehensive T_1 measurements in complex mixtures require lengthy multidimensional inversion-recovery experiments (~16 h using the previous assumptions).

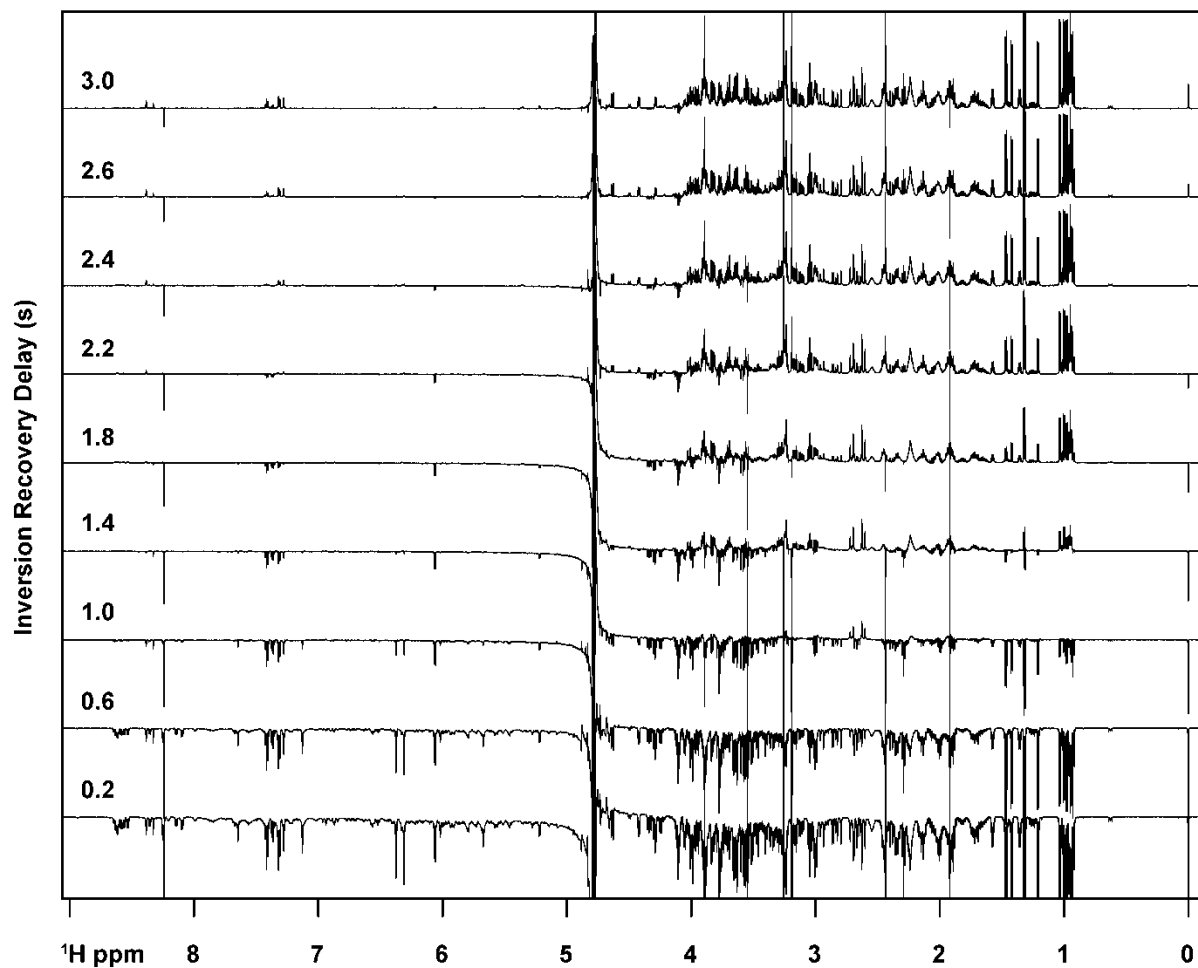**Figure 7.1.** Longitudinal relaxation of 30 metabolites in a synthetic mixture of standards. Relaxation times (T_1) were measured by inversion-recovery. The relaxation delay was incremented from 0 to 3 s in 0.2 s increments (every other increment is shown here). Relaxation times can be approximated by $T_1 = \tau / \ln(2)$, where τ is the inversion-recovery delay at the zero-point crossing for a resonance.

We propose an alternative method for T_1 control in metabolomics studies that utilizes a paramagnetic agent to standardize relaxation. Paramagnetic atoms, such as Fe(III), act as relaxation sinks and can decrease T_1 times sufficiently to minimize relaxation differences between labeled and unlabeled molecules (Fig. 7.2). Paramagnetic T_1 equalization is advantageous because it does not require external signal calibration or additional empirical data. Although paramagnetic agents are used to shorten acquisition times of 1D ^{13}C NMR spectra (Caytan, Remaud et al. 2007), they have not been applied to [1]H-based studies of metabolites in complex solutions. Paramagnetic species (either present in biological samples or introduced during sample preparation) generally are viewed as an impediment to metabolomics because they can increase quantitative error. However, sample-to-sample variation can be tightly controlled when paramagnetic species are intentionally added at concentrations far greater than natural levels.

**Figure 7.2.** Inversion-recovery curves for ^{1}H NMR signals from mixtures of unlabeled and ^{13}C-labeled metabolites containing variable concentrations (1, 100 and 1000 µM) of Fe(III)-EDTA. The methyl alanine and anomeric glucose resonances shown here are representative of the range of T_1 times observed in biological extracts. T_1 values calculated from these curves are presented in Table 1.
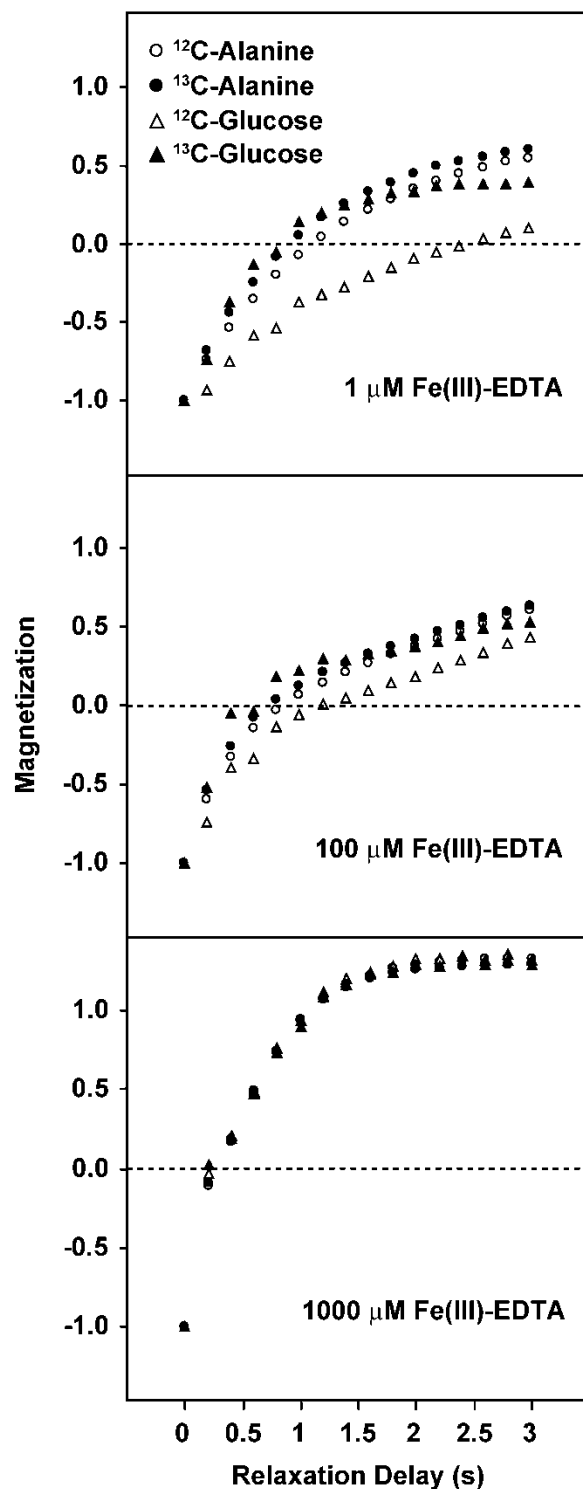
We prepared synthetic mixtures (N=23) containing thirty metabolites to evaluate the efficacy and feasibility of paramagnetic T_1 equalization in isotope-based metabolomics. Six of the metabolites were prepared with both [U-^{13}C] and unlabeled species at ^{13}C to ^{12}C ratios ranging from 0.067 to 3.24. Synthetic mixtures were split into three aliquots and diluted with an equal volume of solution containing either 1, 100, or 1000 µM Fe(III)-EDTA (final concentration). EDTA was selected as a chelator because its high affinity for ferric iron ($K_a \sim 10^{25}$ at neutral pH) (Fischer and Peters 1968) prevents Fe(III) from binding to metabolites in solution. All synthetic samples were prepared in D_2O containing 300 µM NaN_3 and 300 µM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) and were titrated to an observed pH of 7.400 ± 0.004. T_1 relaxation times were measured by inversion-recovery (Fig. 7.1), and isotopic enrichments were measured by 2D ^{1}H-^{1}H Difference Edited TOtal Correlation Spectroscopy (DETOCS) (Lewis, Karsten et al. 2010). The rNMR software package (Lewis, Schommer et al. 2009) was used in analyzing and displaying the NMR data. A detailed description of the synthetic mixtures, data processing methods, and the DETOCS pulse program are available in a companion publication (Lewis, Karsten et al. 2010).

We predicted that Fe(III)-EDTA would accelerate relaxation sufficiently to eliminate resonance-to-resonance variation in T_1 values. Relaxation times observed by inversion-recovery supported this prediction; the range of T_1 values in synthetic metabolite mixtures was 3 s in the presence of 1 µM Fe(III)-EDTA and only 0.03 s in the presence of 1000 µM Fe(III)-EDTA (Fig. 7.2, Table 7.1). To determine the effects of T_1 equalization on quantification of isotopic enrichment, we analyzed 288 cross-peaks from 2D ^{1}H-^{1}H DETOCS spectra of differentially enriched metabolite standards. Linear regression of observed versus actual isotopic enrichments (Fig. 7.3) yielded $R^2 = 0.59$ from iron-free samples (Fig. 7.3B) and $R^2 = 0.76$ (Fig. 7.3C) from samples containing 1000 µM Fe(III)-EDTA. These findings indicate that uncalibrated enrichment estimates can be improved by the addition of a paramagnetic agent.

Figure 7.3. Experimental versus actual ^{13}C isotopic enrichment of carbon atoms in a mixture of metabolites of known $^{13}C/^{12}C$ isotopic composition. Experimental values were determined in three different ways using ^{1}H-detection. (a) ^{1}H-^{13}C peak heights were normalized to that of an internal standard. (b) ^{1}H-^{13}C / ^{1}H-^{12}C peak height ratios were determined directly from (uncalibrated) DETOCS data. (c) ^{1}H-^{13}C / ^{1}H-^{12}C peak height ratios were determined directly from (uncalibrated) DETOCS data from samples containing 1 mM Fe(III)-EDTA. The error reduction observed in plot b, as compared to plot a, reflects the normalization of relaxation-independent phenomena related to the NMR pulse program (e.g. off-resonance effects, line shape, peak multiplicity, and differential isotropic mixing) achieved by taking ^{1}H-^{13}C / ^{1}H-^{12}C peak ratios. The error reduction observed in plot c, as compared to plot b, stems from the paramagnetic equalization of T_1 relaxation. The non-unit slope observed in plot c is due to several factors including: residual differences in T_1 relaxation of ^{1}H bound to ^{13}C vs.^{12}C- (~3%), decoupling artifacts affecting ^{1}H bound to ^{13}C but not ^{12}C (~10%), mismatch between ^{1}H-^{13}C J-couplings and the isotope filtering pulses used by DETOCS, and differential signal recovery from the isotropic mixing sequence due to ^{1}H-^{13}C coupling.
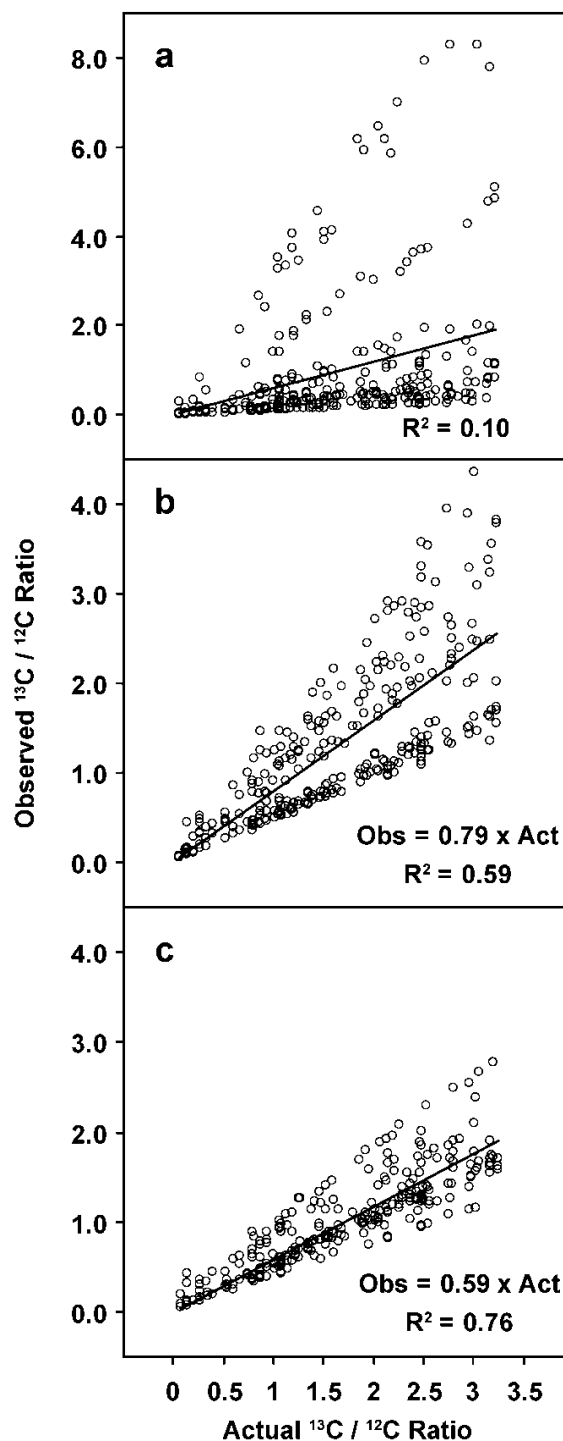
Although the paramagnetic strategy presented here eliminates differential T_1 relaxation, it does not control other sources of quantitative error inherent to multidimensional NMR (e.g. off resonance effects, mismatch of delays to J-couplings, decoupling artifacts, and solution conditions). We have previously reported that the intensities of non-overlapped signals in 2D NMR spectra are linearly proportional to concentration and that the regression slopes describing this proportionality (m_{res}) vary considerably between resonances (Lewis, Schommer et al. 2007). However, many of the relaxation-independent factors affecting m_{res} are normalized in isotopic enrichment studies because quantification is based on $^{13}C/^{12}C$ ratios of individual resonances within a spectrum (Fig. 7.3A versus 7.3B). The use of Fe(III)-EDTA, in combination with quantification based on $^{13}C/^{12}C$ ratios, can control experimental error sufficiently to allow for semi-quantitative analyses without any external calibration of observed NMR signals (Fig. 7.3C).

Calibration-independent enrichment measurement is a plausible application of paramagnetic T_1 equalization, but we foresee its primary utility as a practical tool for accurate quantification of labeled metabolites. Currently, variations in m_{res} are corrected using empirically-determined regression slopes of standards collected at multiple concentrations (Lewis, Schommer et al. 2007; Lewis, Karsten et al. 2010). The expense and unavailability of enriched standards makes this strategy impractical in metabolomics studies. Paramagnetic equalization of differential T_1 relaxation would allow unlabeled standards to be used for measuring m_{res} and would extend the scope of quantitative analyses to ^{13}C labeled metabolites.

CHAPTER 8

Role of Band 3 in regulating metabolic flux of red blood cells

Adapted from:

This study was conducted in collaboration with Dr. Philip S. Low's laboratory (Purdue University). Several years ago, I served as Dr. Low's student host during one of his visits to our department. Dr. Low has a productive research program related to the structure and organization of erythrocyte membranes and has accumulated considerable indirect evidence supporting an unusual regulatory mechanism in erythrocyte metabolism. At the time of his visit, Dr. Low had recently developed two model systems (a knockout mouse and a chemical treatment) for testing his regulatory mechanism and was looking for a tool for quantifying metabolic flux. I was familiar with NMR-based metabolic flux analysis and suggested we adapt our new bioanalytical metabolomics tools for the study. The result was a fruitful collaboration that drove my technology development efforts and yielded an interesting biological study (presented here). I performed the erythrocyte studies and sample preparation at Purdue University with help from a senior scientist (M.E.C.) in Dr. Low's laboratory. I conducted the NMR work and data analysis at the NMRFAM.

8.1. Abstract

Deoxygenation elevates glycolytic flux and lowers pentose phosphate pathway (PPP) activity in mammalian erythrocytes. The membrane anion transport protein (band 3 or AE1) is thought to facilitate this process by binding glycolytic enzymes (GE) and inhibiting their activity in an oxygen-dependent manner. However, this regulatory mechanism has not been demonstrated under physiological conditions. In this study, we introduce a new ^{1}H-^{13}C nuclear magnetic resonance (NMR) technique for measuring metabolic fluxes in intact cells. The role of band 3 in mediating the oxygenated/deoxygenated metabolic transition was examined by treating cells with pervanadate, a reagent that prevents the GE-band 3 complex from forming. We report that pervanadate suppresses oxygen-dependent changes in glycolytic and PPP fluxes. Moreover, these metabolic alterations were not attributable to modulation of bisphosphoglycerate mutase, direct inhibition of GEs by pervanadate, or oxidation, which are the major side effects of pervanadate treatment. These data provide the first direct evidence supporting the role of band 3 in mediating oxygen-regulated metabolic transitions.

8.2. Introduction

Human erythrocytes circulate between the lungs and peripheral tissues approximately every minute, exposing the cells to rapidly changing metabolic demands (Uthman 1998). In the lungs, where O_2 partial pressures are high, erythrocytes are exposed to oxidative stresses that must be controlled by accelerated production of reducing equivalents derived from the pentose phosphate pathway (PPP). In the peripheral tissues, where O_2 pressures are low, erythrocytes must pass through capillaries much smaller than their own diameters (Potter and Groom 1983), causing the cells to distort as they flow from the arterioles to the postcapillary venules. The resulting mechanical stresses induce cation leaks (Johnson 1994) that create an enhanced demand for glycolytically-derived ATP to restore intracellular ion balances. Several layers of metabolic regulation allow erythrocytes to match ATP, NADH, and NADPH production with fluctuations in oxidative stress and ion leakage. One of these mechanisms is oxygen-dependent control of glycolytic and PPP activity.

Historically, oxygen-dependent metabolic regulation has been attributed to alterations in pH, classical allosteric mechanisms, and demand for ATP (Hamasaki, Asakura et al. 1970; Rapoprot, Berger et al. 1976; Mulquiney, Bubb et al. 1999; Mulquiney and Kuchel 1999; Mulquiney and Kuchel 1999). However, experimental evidence indicates that these metabolic regulators operate independently from oxygen tension (Jensen 2004), suggesting that the models are incomplete. An alternative mechanism, which is supported by a growing body of evidence, argues that erythrocyte metabolism is regulated through the oxygen-dependent assembly of glycolytic enzymes (GEs) into inhibitory complexes on the membrane anion exchange protein (band 3 or AE1). Although several lines of indirect evidence support this hypothesis, the GE-band 3 model has not been demonstrated in intact cells.

Early evidence for band 3-dependent glycolytic regulation was derived from *in vitro* binding assays, which showed that purified fragments of the membrane anion exchange protein (band 3

or AE1) interact with several GEs (Murthy, Liu et al. 1981; Tsai, Murthy et al. 1982; Jenkins, Kezdy et al. 1985; Rogalski, Steck et al. 1989). Although debate subsequently emerged over the *in vivo* significance of these data (Maretzki, Reimann et al. 1989), it is now clear that phosphofructokinase (PFK), aldolase, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) bind to band 3 in intact cells (Campanella, Chu et al. 2005). A variety of band 3-specific perturbations have been shown to disrupt the GE-band 3 interaction; GEs are readily displaced from the membrane by antibodies to the amino terminus of band 3 and by phosphorylation of band 3's tyrosine residues (Campanella, Chu et al. 2005). Furthermore, peptide fragments of band 3's GE binding site compete with native band 3 for GEs in resealed cells (Campanella, Chu et al. 2008), and transgenic mice lacking band 3 exhibit no membrane-associated GEs (Campanella, Chu et al. 2008).

Several indirect lines of evidence suggest that GE-band 3 interactions act as an oxygen-dependent metabolic regulator. Whereas oxyHb has no affinity for band 3, deoxyHb binds to the GE site on band 3 with high affinity (Chu, Breite et al. 2008). Consequently, competition between deoxyHb and GEs causes GEs to be released from membrane upon deoxygenation (Campanella, Chu et al. 2005). Deoxygenation also triggers higher glycolytic flux and reduced PPP activity in intact erythrocytes (Murphy 1960; Schrader, Eskey et al. 1993; Messana, Orlando et al. 1996; Delgado, Castro et al. 2004). This observation is significant in that purified fragments of band 3's cytoplasmic tail reduce the *in vitro* catalytic activity of GEs (Murthy, Liu et al. 1981; Tsai, Murthy et al. 1982; Jenkins, Kezdy et al. 1985). Together, these studies suggest that the lower glycolytic fluxes observed in oxygenated erythrocytes result from catalytic inhibition of GEs via the GE-band 3 interaction, whereas elevated glycolytic fluxes of deoxygenated erythrocytes result from deoxyHb displacing GEs from their inhibitory site on band 3.

Although the GE-band 3 model is commonly used to explain oxygen-dependent metabolic phenomena in RBCs (De Rosa, Carelli Alinovi et al. 2008), the model has not been convincingly demonstrated in intact cells. To date, only two publications have tested the band 3 model directly. The first, by Messana et al., showed that red cells treated with DIDS, a covalent inhibitor of band 3-mediated anion transport, exhibit altered pentose phosphate pathway activity (Messana, Orlando et al. 1996). However, Messana's study did not demonstrate any effect of DIDS on GE-band 3 interactions nor did it control for the effects of DIDS on intracellular pH, membrane skeletal interactions (Van Dort, Moriyama et al. 1998), or ion concentrations. The second paper, by Kinoshita et al., showed that computer models of red cell metabolism are consistent with band 3-mediated metabolic regulation (Kinoshita, Tsukada et al. 2007), but their paper provided limited experimental evidence to support their computations.

In this study, we introduce a new ^{1}H-^{13}C nuclear magnetic resonance (NMR) technique for quantifying metabolic pathway flux in intact cells. We examine the metabolic consequences of disrupting the GE-band 3 interaction in intact cells by stimulating tyrosine phosphorylation on band 3 with pervanadate. Pervanadate disrupts the glycolytic enzyme complex by inducing phosphorylation of the two tyrosines (Y8 and Y21) that are located within the GE binding site on band 3 (Low, Allen et al. 1987; Yannoukakos, Meyer et al. 1991). We report that oxygenated RBCs treated with pervanadate have increased glycolytic flux, reduced pentose shunt activity, and are metabolically unresponsive to deoxygenation. These findings are consistent with the band 3-mediated metabolic regulatory model and suggest that the GE-band 3 complex plays a direct role in regulating glycolytic and pentose shunt fluxes of intact erythrocytes.

8.3. Results

We hypothesized that pervanadate-induced disruption of the GE-band 3 complex would stimulate glycolytic flux in oxygenated RBCs and prevent cells from responding metabolically to

changes in oxygen tension. To test this hypothesis, we incubated pervanadate-treated erythrocyte suspensions in isotopically enriched glucose and measured the concentrations of isotopically enriched metabolites over a 12 h time course. Oxygen-dependent metabolic regulation was monitored by conducting the isotopic labeling experiments under both oxygenated and deoxygenated conditions. Concentrations of ^{13}C labeled metabolites were measured by ^{1}H and ^{1}H-^{13}C NMR (Fig. 8.1), and metabolic fluxes were calculated by regressing observed concentrations as a function of incubation time (Fig. 8.2 and Table 8.1).
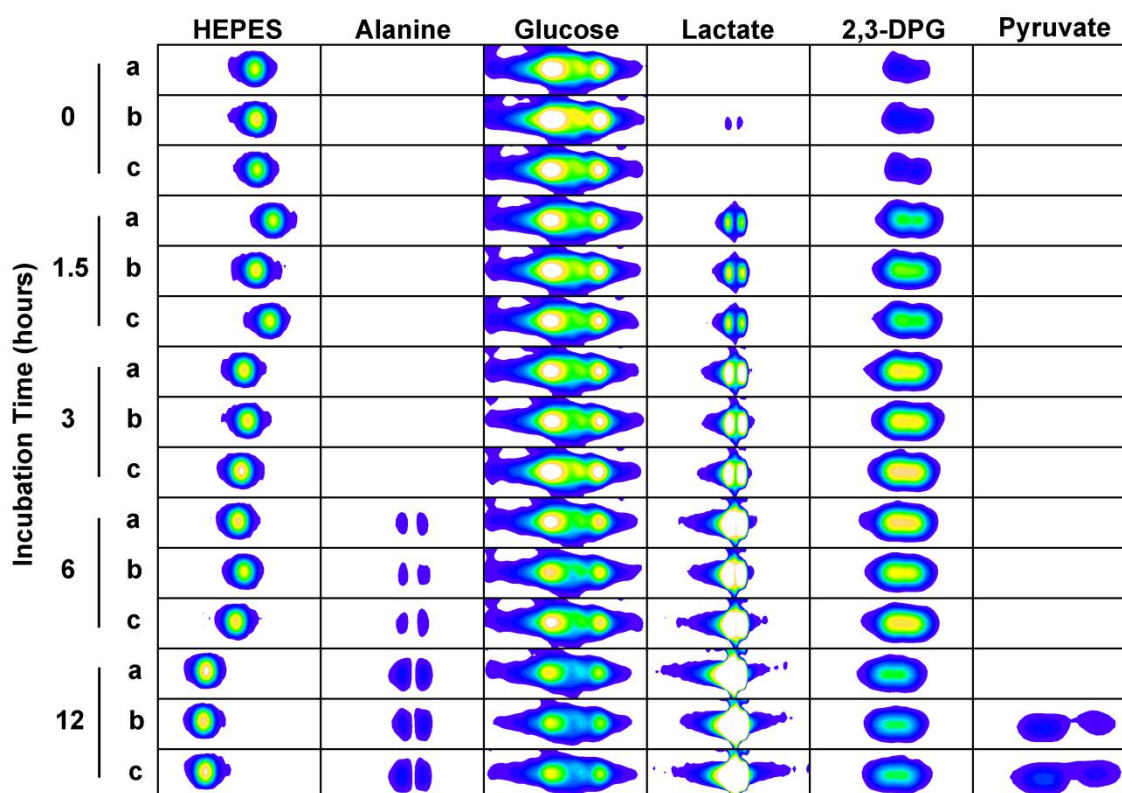
Figure 8.1. Two-dimensional ^{1}H-^{13}C HSQC NMR spectra of three (a, b, c) erythrocyte samples. Each box shows an NMR cross-peak of a metabolite observed in untreated samples incubated in U-^{13}C glucose over a 12 h time course. The minimum and maximum cross peak contours shown in each spectrum depict signal intensities between six and twenty standard deviations above the thermal noise threshold.
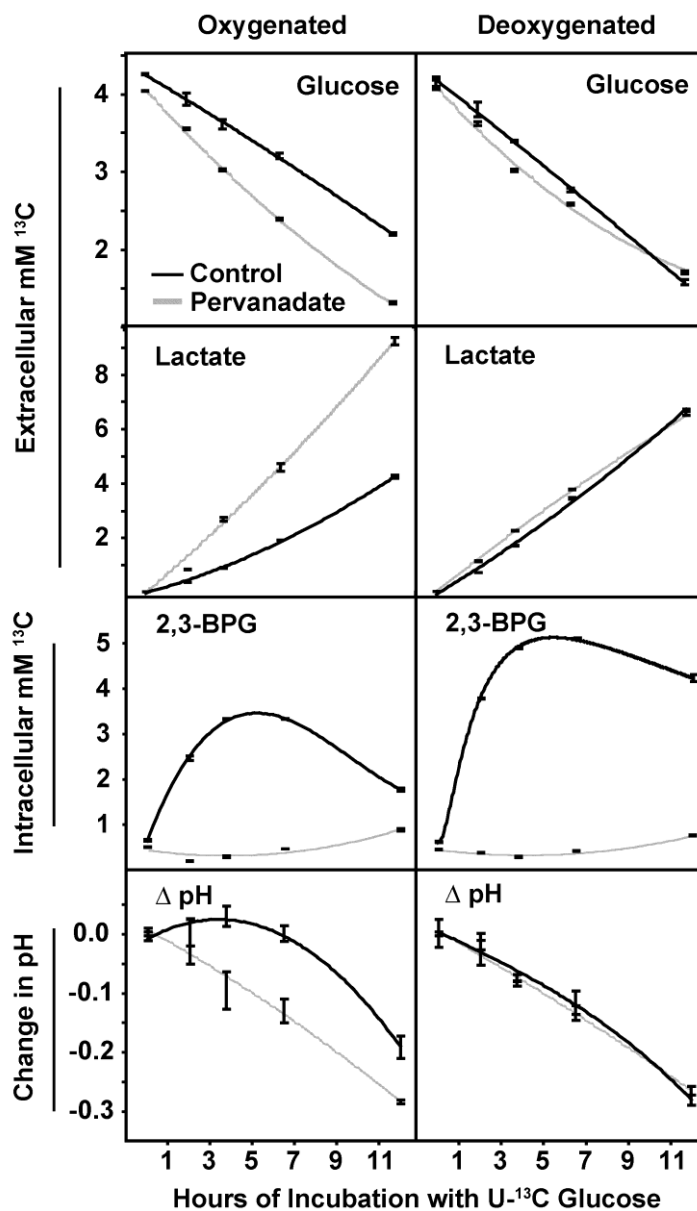
Figure 8.2. Concentrations of ^{13}C labeled metabolites and pH observed in erythrocyte extracts. RBC suspensions (20% hematocrit) were incubated with U-^{13}C-glucose for 12 h and metabolite concentrations were measured by ^{1}H-^{13}C NMR. Black lines indicate untreated RBCs, gray lines indicate pervanadate-treated RBCs, and error bars show standard error.

Table 8.1. Metabolic fluxes in isotopically labeled red cells measured by ^{1}H NMR and ^{1}H-^{13}C NMR.[a]

|	Metabolite	Control O_2	Control Argon	Pervanadate O_2	Pervanadate Argon	Methylene Blue (O_2)
^{1}H NMR	Glucose	0.62 ± 0.04	0.89 ± 0.05	1.12 ± 0.08	0.86 ± 0.02	0.67 ± 0.37
	Lactate	1.55 ± 0.07	2.14 ± 0.02	2.9 ± 0.03	2.09 ± 0.05	1.45 ± 0.05
	PC[‡]	0.06 ± 0.01	0.03 ± 0.005	0.03 ± 0.0001	0.02 ± 0.01	0.21 ± 0.03
^{1}H-^{13}C NMR	Glucose	0.86 ± 0.03	1.09 ± 0.06	1.14 ± 0.02	0.99 ± 0.03	
	2,3-DPG	0.91 ± 0.04	1.58 ± 0.01	*	*	
	Pyruvate[†]	0.1 ± 0.01	0.03 ± 0.003	0.13 ± 0.01	0.1 ± 0.03	
	Lactate	1.81 ± 0.04	2.84 ± 0.02	3.28 ± 0.02	2.53 ± 0.04	
	Alanine[†]	0.02 ± 0.0004	0.02 ± 0.001	0.06 ± 0.01	0.03 ± 0.002	
	Δ pH	-0.08 ± 0.02	-0.12 ± 0.02	-0.12 ± 0.01	-0.11 ± 0.003	

[a] Data are expressed as mean ± S.D. in μmol / hour per ml RBC. Pentose shunt and general flux studies were conducted on independent pools of blood. In pentose activity studies, samples were labeled with 2C-^{13}C glucose and were analyzed by ^{1}H NMR. In general flux studies, samples were labeled with U-^{13}C glucose and were analyzed by ^{1}H-^{13}C NMR.

[‡] Pentose activity (PC) expresses the fraction of incoming glucose used to produce pentose-derived glyceraldehyde-3-phosphate; metabolic fluxes for various points in the pentose shunt can be calculated by multiplying PC by the glucose uptake rate and adjusting for the appropriate stochimetry

* No 2,3-BPG production was observed in pervanadate-treated RBCs

[†] Rates of pyruvate and alanine production include extracellular accumulation of these metabolites

Untreated erythrocytes were used to validate our analytical technique and served as a negative control for pervanadate treatment. As has been shown previously (Schrader, Eskey et al. 1993), glycolytic fluxes of untreated controls differed significantly between oxygenated and deoxygenated conditions. Deoxygenated controls showed higher rates of glucose uptake, 2,3-BPG flux, lactate production, and pH change than oxygenated controls (Fig. 8.2). Overall, glycolytic fluxes of deoxygenated controls were 41% ($p < 0.001$) higher than oxygenated samples. This effect was replicated in two independent pools of blood that were collected and analyzed on different dates. Although some variation in glycolytic fluxes was observed between the two studies, observed kinetics were consistent within the range of values reported in similar studies (Ninfali, Accorsi et al. 1983; Schrader, Eskey et al. 1993; Messana, Orlando et al. 1996) (Table 8.1).

In contrast to the normal metabolic responses observed in untreated controls, pervanadate-treated samples showed significant alterations in oxygen-dependent metabolic regulation. Under oxygenated conditions, pervanadate-treated samples showed higher rates of glucose uptake, lactate production and pH change than untreated controls (Table 8.1). On average, glycolytic fluxes of pervanadate-treated samples were 45% higher ($p < 0.001$) than their corresponding oxygenated controls. In contrast, rates of glucose uptake and lactate production observed in deoxygenated pervanadate-treated samples did not differ significantly from deoxygenated controls (Fig. 8.3). The 45% increase in glycolytic flux of oxygenated pervanadate-treated samples, in combination with the negligible alterations to deoxygenated flux, reversed the normal oxygen-dependent metabolic response of red cells. Whereas untreated controls showed higher glycolytic activity under deoxygenated conditions, pervanadate-treated samples were more glycolytically active under oxygenated conditions (Fig. 8.3).
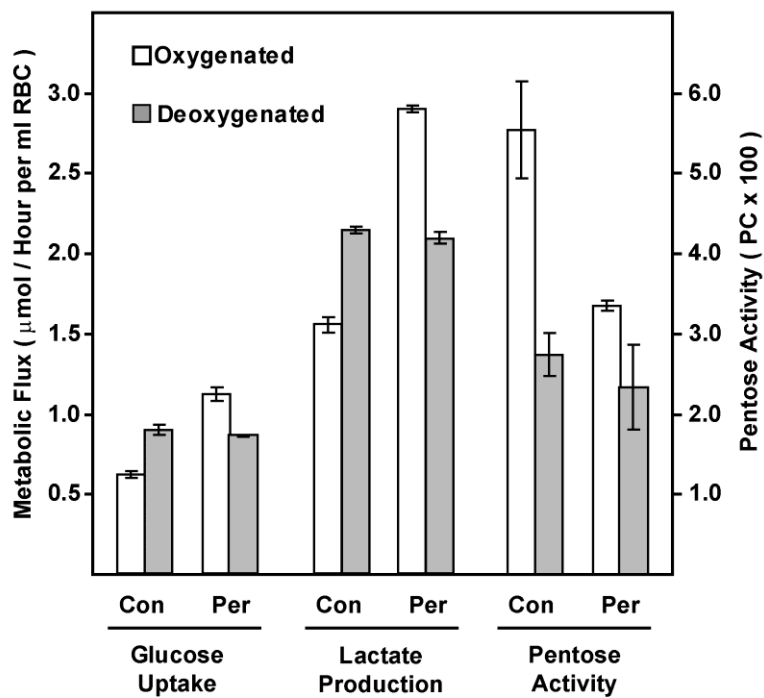
Figure 8.3. Rates of glucose consumption, lactate production, and pentose shunt activity (PC) observed in untreated (Con) and pervanadate-treated (Per) RBCs. Light bars indicate oxygenated samples, dark bars indicate deoxygenated samples and error bars show standard error.

Band 3-dependent metabolic regulation has primarily been studied in the context of glycolysis. However, glycolytic inhibition by the GE-band 3 complex could stimulate PPP flux by making more substrate available to the pentose shunt (Thorburn and Kuchel 1985; Messana, Orlando et al. 1996). To determine the role of the GE-band 3 complex in regulating PPP flux, cell suspensions were incubated with 2-^{13}C-glucose for 12 h. PPP fluxes were calculated from the positional isotopic enrichment of lactate observed in ^{1}H NMR spectra of cell extracts. In accordance with previous studies (Schrader, Eskey et al. 1993; Messana, Orlando et al. 1996; Delgado, Castro et al. 2004), pentose shunt flux accounted for 6% of total glucose consumption in oxygenated controls, but only 3% in deoxygenated controls (51% decrease; $p = 0.013$; Table 8.1). As expected, methylene blue, a traditional positive control for pentose shunt stimulation, increased pentose shunt activity to 21% of total incoming glucose ($p < 0.001$ relative to untreated samples).

We predicted that pervanadate-induced disruption of the GE-band 3 complex would shift metabolic flux towards glycolysis and thus diminish PPP flux under oxygenated conditions. Positional isotopic enrichment data supported this prediction (Fig. 8.3). Under oxygenated conditions, pervanadate decreased pentose shunt activity by 66% ($p = 0.022$). Similar to the pattern observed in glycolytic fluxes, pervanadate-induced metabolic alterations were most pronounced under oxygenated conditions. Whereas deoxygenation induced a significant ($p = 0.013$) reduction in the PPP activity of control samples, deoxygenation led to no significant change in the PPP activity ($p = 0.13$) of pervanadate treated samples.

In addition to the metabolic alterations predicted in the GE-band 3 model, pervanadate elicited two detectable side effects: a complete absence of 2,3-BPG and elevated production of pyruvate and alanine. Both pervanadate-induced disappearance of 2,3-BPG and elevated pyruvate production have been previously attributed to phosphatase activity by bisphosphoglycerate mutase (Ninfali, Accorsi et al. 1983). Elevated alanine production has not

been reported in conjunction with pervanadate treatment, but pyruvate and alanine are readily interconverted via aminotransferases (Manuel y Keenoy, Conget et al. 1991). Isotopically enriched pyruvate and alanine were detected in untreated erythrocytes, although observed biosynthetic rates were considerably lower in untreated controls than in pervanadate-treated samples (Table 8.1). NMR analysis of the incubation medium indicated that these metabolites were accumulating extracellularly (Fig. 8.4). When expressed as a fraction of total carbon output, pyruvate and alanine accounted for 6% and 2% in oxygenated, and deoxygenated controls, respectively. In contrast, the glycolytic side products accounted for 5% of the total carbon output in pervanadate-treated samples, regardless of their oxygenation state.

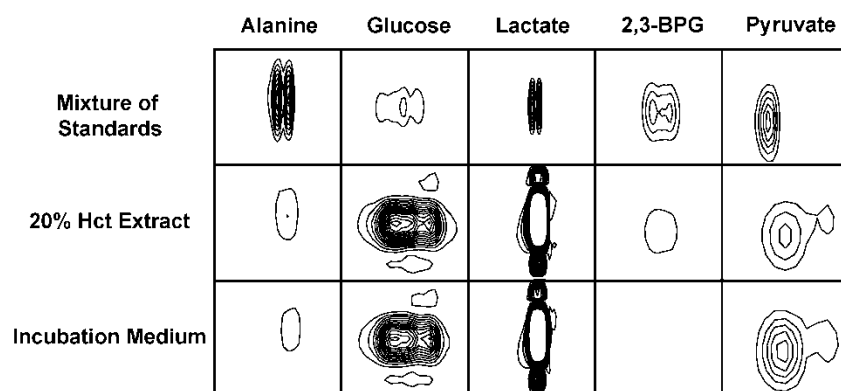**Figure 8.4.** Two-dimensional ^{1}H-^{13}C NMR spectra of erythrocyte extracts and standards. Boxes show the cross-peaks from alanine, glucose, lactate, 2,3-DPG, and pyruvate observed in a mixture of standards, extracts from a 20% suspension of human RBCs after a 12 h incubation with U-^{13}C-glucose, and the supernatant from a centrifuged 20% suspension of human RBCs after a 12 h incubation in U-^{13}C-glucose.

Enzymatic activity by NADH-dependent methemoglobin reductase is a band 3-independent effect that could contribute to increased glycolytic fluxes observed in pervanadate-treated samples. The mechanism for oxidation of oxyhemoglobin to methemoglobin involves an electron transfer between the heme iron and the bound O_2 (Kosaka, Tyuma et al. 1983). CO can be used to inhibit the formation of methemoglobin and thereby suppress demand for NADH-dependent methemoglobin reductase. To control for effects on methemoglobin reductase activity, rates of glucose consumption in pervanadate-treated and untreated samples were measured in washed erythrocytes incubated in CO, air, and argon. Samples incubated in carbon monoxide showed no significant differences in glucose consumption when compared to samples incubated in air (Fig. 8.5). As expected, glycolytic rates for samples incubated in air and argon matched our previously observed rates (Fig. 8.3 versus Fig. 8.5). Therefore, methemoglobin reduction cannot account for the metabolic effects observed in pervanadate-treated samples.
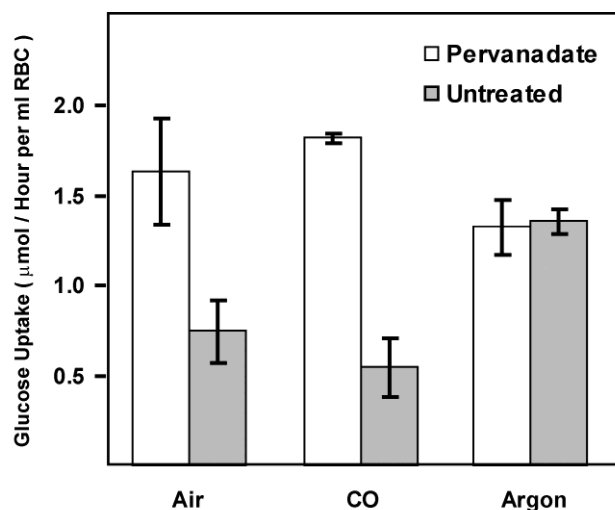
Figure 8.5. Inhibition of methemoglobin formation**.** Rates of glucose uptake from the incubation medium were measured in pervanadate-treated and untreated erythrocytes incubated in air, carbon monoxide, and argon. Light bars indicate pervanadate-treated samples, dark bars indicate untreated samples and error bars show standard deviation. Samples were prepared as described previously (see Preparation of RBCs for metabolic studies) and were incubated at 37° C in isotonic HEPES solution (Table 8.2) containing 5 mM unlabeled glucose (Sigma G8270). Aliquots (1 ml; N=2) were harvested from each sample after 0, 3 and 6 hours of incubation. The incubation medium was separated from the erythrocytes by centrifugation (2.5 min at 2,500 x g, 4° C) and metabolites were extracted (see Sample extraction and preparation of standards). Glucose concentrations were measured using a commercial enzymatic assay (Fisher TR15103) and rates of glucose uptake were estimated by linear regression.

8.4. Discussion

Considerable indirect evidence has accumulated in support of the GE-band 3 model (Low, Rathinavelu et al. 1993; Weber, Voelter et al. 2004). However, the practical challenge of disrupting the GE-band 3 complex and measuring metabolic fluxes in intact cells has prevented the GE-band 3 model from being convincingly demonstrated. In this study, we took advantage of a new ^{1}H-^{13}C NMR strategy for measuring metabolic fluxes in intact cells (Fig. 8.1) in order to investigate the metabolic consequences of disrupting the GE-band 3 complex under physiologically relevant conditions.

The GE-band 3 model (Fig. 8.6) predicts that pervanadate-induced disruption of the GE-band 3 complex will release catalytic inhibition of the GEs and thereby stimulate glycolytic flux and reduce PPP flux. Indeed, oxygenated pervanadate-treated samples had 45% higher glycolytic fluxes and 66% lower pentose shunt fluxes than controls. Moreover, pervanadate-induced metabolic alterations were specific to oxygenated samples: metabolic fluxes of deoxygenated samples did not differ significantly between controls and pervanadate-treated samples. This result was expected because the GE-band 3 complex forms only under oxygenated conditions. The asymmetric effects of pervanadate on oxygenated, but not deoxygenated, samples resulted in an oxygen-insensitive metabolic condition. In contrast, untreated controls were highly responsive to alterations in oxygen-tension.
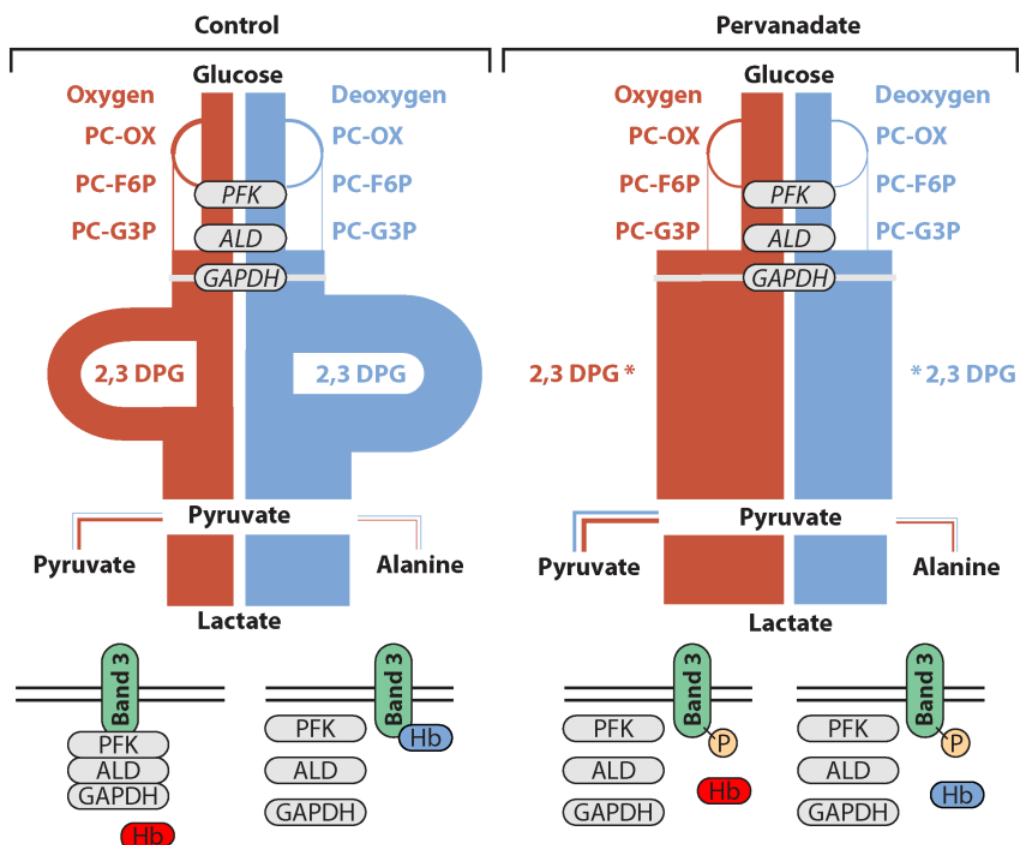
Figure 8.6. Metabolic fluxes in RBCs as measured by ^{1}H and ^{1}H-^{13}C NMR. Line widths are proportional to measured fluxes where red lines indicate oxygenated and blue lines indicate deoxygenated conditions. Metabolites used to estimate flux are shown in bold and the status of the GE-band 3 complex is denoted below each condition. *No 2,3-BPG production was observed in pervanadate-treated RBCs. Abbreviations: PC-OX, flux through the oxidative branch of the pentose phosphate pathway; PC-F6P, pentose shunt derived fructose 6 phosphate; PC-G3P, pentose shunt derived glyceraldehyde-3 phosphate.

Although both the glycolytic and PPP flux data support the GE-band 3 model, pervanadate has several known side effects that must be addressed. Pervanadate and orthovanadate inhibit several glycolytic enzymes (Simons 1979). However, our data indicate that that the pervanadate concentrations used in this study were insufficient to influence glycolytic and pentose shunt fluxes directly. Direct inhibition of GEs by pervanadate lowers glycolytic flux (Benabe, Echegoyen et al. 1987), whereas we observed increased glycolytic flux in oxygenated samples treated with pervanadate. In addition, direct inhibition of GEs by pervanadate is an oxygen-independent phenomenon, whereas the pervanadate-induced metabolic alterations observed in this study selectively affected oxygenated erythrocytes. Moreover, the pervanadate concentrations used in this study have been shown to have little effect on GE activities *in vitro* (Climent, Bartrons et al. 1981). Thus, we conclude that any direct inhibitory effects of pervanadate in this study were secondary to pervanadate-induced disruption of the GE-band 3 complex.

The complete absence of 2,3-BPG and the elevated pyruvate/alanine production we observed in pervanadate-treated samples are band 3-independent effects of pervanadate that have been reported previously (Ninfali, Accorsi et al. 1983). The disappearance 2,3-BPG and elevated pyruvate production are both thought to originate from pervanadate-induced stimulation of the phosphatase activity of bisphosphoglycerate mutase (Ninfali, Accorsi et al. 1983). In contrast to previous reports, which have implicated 2,3-BPG in glycolytic regulation (Duhm 1975), our data show no correlation between 2,3-BPG levels and glycolytic fluxes. Under deoxygenated conditions, we observed nearly identical fluxes between untreated controls and pervanadate-treated samples. This finding indicates that pervanadate-induced alterations in 2,3-BPG concentrations did not significantly alter the observed metabolic fluxes.

Another side effect of pervanadate that could influence glycolytic and pentose shunt fluxes relates to its activity as an oxidizing agent. Two of the most conspicuous metabolic effects of

oxidation in erythrocytes are increased PPP flux and elevated NADH-dependent methemoglobin reductase activity. However, we observed a 66% decrease in PPP activity of pervanadate-treated samples and incubating samples in carbon monoxide, an inhibitor of methemoglobin formation, had no significant effect on glycolytic activity (Fig. 8.5). Although the elevated pyruvate/alanine production we observed in pervanadate-treated samples could be attributed to methemoglobin reductase activity, pervanadate-induced pyruvate production was only 5% of the total carbon output and has been previously attributed to 2,3-BPG degradation (Ninfali, Accorsi et al. 1983). The low production of pyruvate/alanine relative to total carbon output, negligible effects of carbon monoxide, and lack of a substantial PPP response argue that pervanadate-induced oxidation was relatively minor in this study.

In summary, our findings provide direct evidence for *in vivo* regulation of oxygen-dependent metabolic flux via the formation and dissociation of the GE-band 3 complex. Pervanadate-induced disruption of the GE-band 3 interaction in intact cells elicited glycolytic and pentose shunt fluxes similar to those found in deoxygenated erythrocytes. Samples that were treated with pervanadate were unable to respond to changes in oxygen tension whereas untreated controls were highly responsive. Although pervanadate has several known metabolic side effects, these side effects do not appear to be responsible for our results. In contrast, all of the observed glycolytic and pentose shunt flux data were consistent with the GE-band 3 model. These findings suggest that the GE-band 3 complex plays a direct role in regulating oxygen-dependent changes in glycolytic and pentose shunt fluxes of intact erythrocytes.

8.5. Materials and Methods

Preparation of RBCs for metabolic studies. Fresh blood was collected by venipuncture from healthy human volunteers (N=6) and erythrocytes from each donor were isolated and washed three times in isotonic HEPES buffer (Table 8.2). Washed erythrocytes were pooled

and resuspended at 20% hematocrit in HEPES buffer containing 20 IU/mL Penicillin/Streptomycin, after which separate aliquots of the suspension were treated as follows. To stimulate phosphorylation of band 3, erythrocytes were incubated at 37°C for 30 minutes with 0.5 mM pervanadate (premixed final concentration of hydrogen peroxide 1.5mM and sodium orthovanadate 0.5mM) via established methods (Harrison, Rathinavelu et al. 1991). For maximal stimulation of pentose phosphate shunt activity, a separate aliquot of cells was similarly treated with 6.7 µM methylene blue (Murphy 1960; Schrader, Eskey et al. 1993; Messana, Orlando et al. 1996; Delgado, Castro et al. 2004). For measurement of metabolism in control cells, a third suspension was treated with the same volume of HEPES buffer and then incubated similarly.

Table 8.2. Composition of buffer used for washing and incubating erythrocytes

Compound	[mM]
Glucose*	5
HEPES	25
NaH_2PO_4	1
NaCl	115
KCl	10
$MgCl_2$	2
Pen-Strep**	20 IU/ml

* Glucose was excluded from the wash buffer

** 20 IU/ml Penicillin/Streptomycin was included in the incubation medium to prevent microbial growth during the isotopic labeling experiments

Deoxygenated RBCs were prepared by placing small volumes of RBC suspensions in high volume containers (e.g. 20 mL of suspension in a 250 mL plastic bottle). Containers were fitted with rubber septa, placed on their sides, and gently agitated on a rocking table to maximize the exposed surface area of the suspensions. The agitating suspensions were deoxygenated for two hours under a continuous stream of humidified argon. All red cell preparations, except the 30 minute incubation required for pervanadate-induced phosphorylation of band 3, were conducted in a 4°C cold room to minimize metabolic activity.

Isotopic labeling. We employed two isotope-based strategies for measuring metabolic flux. PPP activity was calculated from the steady-state positional isotopic enrichment of lactate produced by cells incubated with 5 mM 2-^{13}C-glucose (Isotec 310794). Although this strategy is a well established method for calculating pentose shunt activity (Schrader, Eskey et al. 1993; Delgado, Castro et al. 2004), the resulting mixture of isotopomers complicates flux analysis in other pathways. For general metabolic analyses, including measurements of glucose uptake, lactate production, and flux through the 2,3-BPG shunt, we incubated cells with 5mM U-^{13}C-glucose (Isotec 389374) and monitored the resulting concentration of ^{13}C labeled metabolites over time. For general flux analysis, washed RBC pools containing 5 mM U-^{13}C-glucose were split into twelve samples consisting of: control (oxygenated n = 3; deoxygenated n = 3) and pervanadate-treated samples (oxygenated n = 3; deoxygenated n = 3). At time zero, RBC suspensions were transferred from 4°C to a 37 °C warm room. Aliquots (1 mL) of each sample were then harvested after 0, 1.5, 3, 6 and 12 h of incubation. Each aliquot was flash frozen in liquid nitrogen and stored at -80°C until sample extraction. Sample preparation methods for PPP activity experiments were the same as those used for the general flux studies with the exceptions that cells were incubated with 2-^{13}C-glucose and methylene blue treated RBC suspensions (n =3) were added as an additional control.

Sample extraction and preparation of standards. Isotopically labeled samples were suspended in a boiling water bath for 7.5 minutes to lyse cells and halt enzymatic activity. Boiled lysates were spun at 16,000 g to pellet cellular debris, and an 800 μL aliquot of each supernatant was dried in a SpeedVac Concentrator (Thermo Scientific). Each resulting metabolite residue was suspended in 800 μL of NMR analytical solution A (D_2O containing 500 µM NaN_3 and 500 µM DSS). Concentration reference solutions containing equimolar (at 2 mM, 5 mM, and 10 mM) glucose, lactate, 2,3-BPG, pyruvate, and alanine were prepared from weighed pure standards (Sigma) dissolved in HEPES buffered saline (Table 8.2) prepared in D_2O. Each of the three reference solutions was titrated to an observed glass electrode pH of 7.40 by adding deuterated acid (DCl) or base (NaOD) as needed.

NMR spectroscopy. NMR spectroscopy was conducted at the National Magnetic Resonance Facility at Madison. One-dimensional ^{1}H and two-dimensional ^{1}H-^{13}C HSQC spectra were collected on a Varian 600 MHz spectrometer equipped with a cryogenic probe. The pulse sequence used to collect ^{1}H NMR spectra consisted of a 1 s initial delay followed by a 90° pulse, and a 2 s acquisition time (32,000 data points). Each ^{1}H spectrum was collected in four transients with four steady-state transients. Sensitivity enhanced ^{1}H-^{13}C HSQC spectra were collected in four transients with 16 steady-state transients, 256 increments, a carbon spectral width of 70 ppm, an acquisition time of 300 ms (3,000 data points), and an initial delay of 1 s. One- and two-dimensional time-domain data sets were Fourier transformed, phased, and chemical shift referenced to DSS. HSQC spectra were processed with a shifted exponential sine bell widow function in both the direct and indirect dimensions. All data processing was done automatically using custom nmrDraw (Delaglio, Grzesiek et al. 1995) scripts written in-house.

Assigning metabolites and calculating concentrations. We have recently developed methods for identifying (Cui, Lewis et al. 2008) and accurately quantifying (Lewis, Schommer et al. 2007) metabolites in complex solutions using multidimensional ^{1}H-^{13}C NMR. These methods

were developed to mitigate the significant disadvantages associated with 1D NMR analyses of unfractionated biological extracts. Historically, metabolic flux analyses have primarily relied on 1D NMR (McIntyre, Thorburn et al. 1989; Berthon, Bubb et al. 1993; Schrader, Eskey et al. 1993; Shulman and Rothman 2001; Delgado, Castro et al. 2004). The use of this technique has limited both the number of metabolites that can be monitored and the quantitative reliability of experiments (Lewis, Schommer et al. 2007). The methods presented here are a metabolic flux adaptation of our Fast Metabolite Quantification (FMQ) strategy (Lewis, Schommer et al. 2007) and can be adapted to any ^{13}C-based metabolic flux study.

Metabolites observed in ^{1}H-^{13}C NMR spectra were identified by comparing peak lists from experimental spectra to peak lists from spectra of model compounds collected by the Madison Metabolomics Consortium (Cui, Lewis et al. 2008). All assignments were verified by overlaying spectra of pure standards over the spectrum of a representative extract. Peak intensities (area for 1D spectra; peak height for 2D spectra) of dispersed signals from each metabolite (Fig. 8.1) were normalized to the average HEPES signal to minimize error associated with sample handling. NMR spectra of three equimolar mixtures of pure standards (2, 5 and 10 mM for each metabolite) were used to generate calibration curves for relating observed metabolite intensities to molar concentrations. Metabolite signals observed in the RBC extracts were then quantified using the regression coefficients from the calibration curves.

Determination of sample pH by NMR. The ^{1}H NMR chemical shifts of several HEPES signals are sensitive to pH and thus can be used as an internal pH indicator. Chemical shifts of HEPES were measured in each sample by ^{1}H-^{13}C HSQC (Fig. 8.1) and were converted to pH by fitting a titration curve to the observed shifts (Fig. 2.1). The HEPES titration curve was generated from samples of the HEPES buffered saline solution (Table 8.2) prepared in D$_2$O and hand titrated to uncorrected glass electrode readings ranging from 5.5 to 10.5. NMR spectra were collected at the National Magnetic Resonance Facility at Madison.

Calculating Pentose Shunt Flux. Several publications discuss the methods for calculating pentose shunt flux from differential isotopic enrichment in the C2 and C3 positions of lactate following a pulse with 2-^{13}C-glucose (Schrader, Eskey et al. 1993; Delgado, Castro et al. 2004). Briefly, isotopic enrichment in the C3 position of lactate is indicative of pentose shunt activity, whereas C2 enrichment is indicative of glycolytic flux. The fraction of incoming glucose used to produce pentose-derived glyceraldehyde-3-phosphate (PC) can be calculated from ^{13}C enrichment in the C3 and C2 positions of lactate using a model initially developed for radioisotope analyses (Katz and Wood 1960) that has since been modified for NMR (Schrader, Eskey et al. 1993; Delgado, Castro et al. 2004):

$$\frac{C3_{lactate}}{C2_{lactate}} = \frac{2PC}{1+2PC}$$

Metabolic flux through various parts of the PPP can then be calculated by adjusting PC for stoichiometry and glucose uptake; flux through the oxidative branch of the pentose shunt = (3 × PC × glucose consumption), pentose shunt derived fructose-6-phosphate = (2 × PC × glucose consumption), pentose shunt derived glyceraldehyde-3-phosphate = (PC × glucose consumption), and glycolytic flux = ([1-PC] × glucose consumption).

Our method for determining ^{13}C enrichment differs slightly from that presented by Delgado et al. (Delgado, Castro et al. 2004). As in Delgado's publication, we measured isotopic enrichment from ^{13}C coupling detected by ^{1}H NMR. However, Delgado's group measured enrichment in both the C2 and C3 positions of lactate from the methyl signal of lactate. This approach requires curve fitting to deconvolve the overlapped signals of unlabeled and 2-^{13}C lactate. To avoid curve fitting, we measured C2 enrichment directly from the dispersed methine signal of lactate. All other calculations, including the correction for naturally occurring ^{13}C, were carried out as described by Delgado et al. (Delgado, Castro et al. 2004).

Regression analyses and statistics. Rates of glucose consumption, lactate production and flux through 2,3-BPG were based on linear regression of U-^{13}C metabolite concentrations as a function of time. Glucose consumption and lactate production estimates were based on all of the time points. Isotopic labeling kinetics in 2,3-BPG are nonlinear because 2,3-BPG is a pathway intermediate and because steady state concentrations of 2,3-BPG are affected by changes in pH (Mulquiney, Bubb et al. 1999). Consequently, 2,3-BPG flux estimates were based on the initial labeling rates observed between the 0 and 1.5 h time points. Standard deviations of the labeling kinetics reflect variation of regression coefficients between sample replicates. All of the *p* values presented in this manuscript are derived from a two-tailed equal variance *t*-test.

CHAPTER 9

Role of Band 3 in regulating metabolic flux of reticulocytes

Adapted from:

Ian A Lewis, M. Estela Campanella, Nicholas O. Markham, Lisa J Garrett-Beal, David M. Bodine, John L. Markley and Philip S. Low. Role of Band 3 in regulating metabolic flux of reticulocytes. Manuscript to be submitted.

This study, which was conducted in collaboration with Dr. Philip S. Low's laboratory (Purdue University), was follow-up to my previous investigation of metabolic regulation in red blood cells (Chapter 8). Dr. Low's collaborator at the NIH mouse facility (D.M.B) produced a band 3 knockout mouse to test the band 3 metabolic regulatory model in an *in vivo* system. Red cells from the knockout mice are very fragile. Consequently, these mice are anemic and have a high reticulocyte count. We assumed that oxygen-dependent metabolic regulation in reticulocytes would be controlled via the same GE-band 3 mechanism we observed in erythrocytes. We were surprised to discover a robust band 3-independent regulatory mechanism in the knockouts. I conducted the red cell labeling work with assistance from a senior scientist (M.E.C.) in the Low laboratory. I conducted NMR spectroscopy and data analysis at the NMRFAM.

9.1. Abstract

Deoxygenated hemoglobin displaces rate-controlling glycolytic enzymes (GE) from a binding site on the anion exchange protein (AE1 or band 3) of erythrocytes. This process is thought to mediate oxygen-dependent alterations in erythrocyte metabolism. Although the GE-band 3 model is well established in mature erythrocytes, the model has not been studied in reticulocytes. We have recently produced a band 3 knockout mouse to test the GE-band 3 regulatory mechanism in an *in vivo* model. Based on previous studies of erythrocyte metabolic regulation, we predicted that band 3 knockout red cells, which are 95% reticulocytes, would be unable to respond metabolically to changes in oxygen tension. We tested this prediction by measuring the glycolytic and pentose phosphate pathway fluxes of reticulocytes under oxygenated and deoxygenated conditions. Surprisingly, the metabolism of band 3 -/- reticulocytes was highly responsive to oxygen tension and showed oxygen-dependent metabolic alterations equivalent to those observed in wild type reticulocytes. These data indicate a robust band 3-independent mechanism for controlling metabolic flux in reticulocytes.

9.2. Introduction

Mammalian erythrocytes control metabolic flux through glycolysis and the pentose phosphate pathway (PPP) in an oxygen-dependent manner. Deoxygenation stimulates glycolytic flux and decreases PPP activity whereas oxygenated conditions suppress glycolytic flux and increase PPP flux. These oxygen-dependent metabolic alterations have historically been attributed to changes in pH, classical allosteric mechanisms, and demand for ATP (Hamasaki, Asakura et al. 1970; Rapoprot, Berger et al. 1976; Mulquiney, Bubb et al. 1999; Mulquiney and Kuchel 1999; Mulquiney and Kuchel 1999). However, a growing body of literature indicates that oxygen-related metabolic phenomena are regulated through an unusual

mechanism involving competition between glycolytic enzymes (GEs) and deoxyhemoglobin (deoxyHb) for a binding site on the amino-terminus of anion exchange protein (AE1 or band 3).

The GE-band 3 metabolic regulatory model is based on several lines of evidence including: 1) deoxyHb, but not oxyHb, binds to band 3 with high affinity (Chu, Breite et al. 2008), 2) several GEs bind to the same site on band 3 as deoxyHb (Chu, Breite et al. 2008), 3) band 3-associated GEs have reduced enzymatic activity (Murthy, Liu et al. 1981; Tsai, Murthy et al. 1982; Jenkins, Kezdy et al. 1985), 4) deoxygenation displaces GEs from their band 3 binding site (Campanella, Chu et al. 2005), 5) deoxygenation stimulates glycolytic flux (Murphy 1960; Schrader, Eskey et al. 1993; Messana, Orlando et al. 1996; Delgado, Castro et al. 2004) and 6) a variety of band 3-specific perturbations alter the enzymatic activity of GEs (Murthy, Liu et al. 1981; Tsai, Murthy et al. 1982; Jenkins, Kezdy et al. 1985). In addition, we have recently shown that disrupting the GE-band 3 complex in intact erythrocytes produces a pseudo-deoxygenated metabolic condition that is unresponsive to changes in oxygen-tension (Lewis, Campanella et al. 2009). Together, these studies provide strong evidence for a metabolic regulatory model in which deoxyHb displaces GEs from their inhibitory site on band 3 and thereby stimulates glycolytic flux under deoxygenated conditions.

The GE-band 3 model satisfies a wide variety of empirical data and explains how erythrocytes can adjust metabolic flux in response to oxygen tension. However, the GE-band 3 model has only been studied in the context of mature erythrocytes. Reticulocytes, like mature erythrocytes, alter glycolytic and PPP flux in response to deoxygenation (Ghosh and Sloviter 1973). Unlike erythrocytes, however, reticulocytes must conduct a variety of other maturation-related metabolic processes, including new protein synthesis, fatty acid biosynthesis, and oxidative phosphorylation (Ghosh and Sloviter 1973; Siems, Muller et al. 1982; Kim, Koury et al. 1991; LePetitThevenin, Nobili et al. 1997). How reticulocytes balance their diverse biosynthetic functions with oxygen-dependent metabolic alterations remains a mystery.

In this study, we examine the role of the GE-band 3 complex in regulating oxygen-dependent metabolism of reticulocytes. Because reticulocytes form a GE-band 3 complex similar to the one observed in mature erythrocytes (Campanella, Chu et al. 2008), we hypothesized that *in vivo* disruption of the GE-band 3 complex would severely alter oxygen-dependent metabolic regulation in reticulocytes. To test this hypothesis, we generated a band 3 knockout (-/-) mouse line and measured the oxygen-dependent shift in glycolytic and PPP fluxes using ^{1}H and ^{1}H-^{13}C NMR. To our surprise, band 3 -/- reticulocytes did not show any significant metabolic abnormalities when compared with wild type (band 3 +/+) reticulocytes. Band 3 -/- reticulocytes, like the wild type controls, responded to deoxygenation by increasing glycolytic flux and decreasing PPP activity. These findings provide clear evidence that reticulocytes can regulate oxygen-dependent metabolism via a band 3 independent mechanism.

9.3. Materials and Methods

Generating band 3 knockout mice. All of the mice used in this study were bred, housed and manipulated using an NHGRI approved Animal Study protocol (G-04-2). Band 3 knockout mice were generated through homologous recombination of embryonic stem cells. A 3227 base pair 5′ arm from the mouse Slc4a1 locus (coordinates chr.11: 102,472,363 to 102,469,136) that extends into exon 4 was fused to a PGK-Neo gene flanked by loxP sites and a 5232 3' arm (coordinates chr.11: 102,469,136 to 102,464,304) in a modified pPNT vector (Tybulewicz, Crawford et al. 1991). The 3′ arm contains a 33 base pair deletion that removes a beta-hairpin loop in exon 7 (coordinates chr.11: 102,467,497 to 102,467,529). The 11-amino acid beta-hairpin loop in the cytoplasmic domain of band 3 is responsible for ankyrin binding in mouse erythrocytes. The targeting vector was transfected into TC1 ES cells and recombinant clones were isolated using established methods (Stefanovic, Markham et al. 2007). Approximately 12 cells from two clones were injected into C57BL/6J (B6) blastocysts to generate chimeric mice.

Chimeras were bred to Black Swiss female mice and germ line transmission was demonstrated by the presence of agouti pups in the F1 litters. PCR analysis of DNA from tail biopsies of F1 animals was used to identify animals heterozygous for the targeted PGK-Neo insertion. Heterozygotes (-/+) were intercrossed to generate an F2 generation. Homozygous (-/-) transgenic mice were easily detected at birth by their extreme pallor. PCR analysis of F2 mice confirmed that these animals were homozygous for the recombinant locus. Genetic heterogeneity in the colony was maintained by alternate backcrossing with 129 and Black Swiss mice. Complete Blood Counts were determined using a Coulter Counter and osmotic fragility was determined using established methods (Stefanovic, Markham et al. 2007). Reticulocyte counts were determined manually by staining of peripheral blood smears with Crystal Violet.

RNA analysis. RNA was extracted from the bone marrow, spleen and kidney using Trizol Reagent according to the manufacturer's instructions. Relative levels of band 3 expression (forward, 5′TCGAGGATCAGATTCGGCCT3′; reverse, 5′GAGGGTCCTTCTGAGCCCC3′) and beta-2 microglobulin (forward, 5′TGCTATCCAGAAAACCCCTC3′; reverse, 5′GTCATGCTTAACTCTGCAGG3′) were compared using the Superscript III Platinum Two-Step qRT-PCR kit with Syber Green dye detection on an ABI Prism 7700 Sequence Detector (Applied Biosystems, Foster City, CA).

Generating anemic wild type mice. Direct comparisons between red cells isolated from band 3 -/- and band 3 +/+ mice are misleading because over 95% of band 3 -/- red cells are reticulocytes. To generate a more meaningful control group, twenty reticulocyte-enriched wild type mice were produced via biweekly bleeding of the animals over the course of one month. Red cells isolated from these animals had reticulocyte counts of 75%.

Preparation of RBCs. Whole blood was collected from band 3 knockout mice (-/-; *N* = 20), anemic wild type mice (anemic, *N* = 20), and wild type mice (+/+; *N* = 10) at the NIH mouse

facility (Bethesda, MD) and was shipped on ice overnight to Purdue University for isotopic labeling experiments. Blood from individual animals was combined within groups to yield three main pools (-/-, anemic, and +/+) from which all subsequent samples were derived. RBCs were isolated by low speed centrifugation (500 × g for 30 minutes) and the buffy coat was discarded. Low centrifuge speeds were used to prevent lysis of the fragile band 3 -/- cells. RBCs were washed three times in 20 volumes of 330 milliosmol HEPES buffered saline (Table 9.1) and resuspended at 5% hematocrit. Isotopic labeling experiments required the low hematocrit suspensions to ensure a sufficient glucose pool for the metabolically active reticulocytes.

Table 9.1. Composition of buffer used for washing and incubating red blood cells. *Glucose was excluded from the wash buffer.

Compound	mM
Glucose*	5
HEPES	25
NaH_2PO_4	1
NaCl	130
KCl	10
CaCl	0
$MgCl_2$	2

Deoxygenated RBCs were prepared by placing samples in oversized containers (e.g. 20 ml sample in a 250 ml container). Containers were fitted with airtight rubber septa and were placed on their side on a gently agitating rocking table to maximize the exposed surface area of the RBC suspensions. Samples were deoxygenated for two hours under a continuous stream of humidified argon. All RBC preparations were conducted in a 4º C cold room to minimize metabolic activity.

Isotopic labeling. Washed RBC suspensions containing 5 mM 2C-^{13}C glucose (Isotec 310794) were split into sixteen samples consisting of: wild type mice (oxygenated $n = 3$; deoxygenated $n = 3$), band 3 knockout mice (oxygenated $n = 2$; deoxygenated $n = 2$), and anemic wild type mice (oxygenated $n = 3$; deoxygenated $n = 3$). After all preparations were complete, RBC suspensions were transferred simultaneously to a 37° C warm room. Aliquots (1 ml) of each sample were harvested after 0, 1.5, 3, 6 and 12 h of incubation. Each aliquot was flash frozen in liquid nitrogen and stored at -80° C until sample extraction.

Sample extraction and preparation of standards. Isotopically labeled samples were suspended in a boiling water bath for 7.5 min to lyse cells and halt enzymatic activity. Boiled lysates were spun at 16,000 $\times$ g to pellet cellular debris, and an 800 μL aliquot of each supernatant was dried in a SpeedVac Concentrator (Thermo Scientific). Each resulting metabolite residue was suspended in 800 μL of NMR analytical solution (D$_2$O containing 500 μM NaN$_3$ and 500 μM DSS). Concentration reference solutions containing equimolar (at 2 mM, 5 mM, and 10 mM) glucose, lactate, 2,3-BPG, pyruvate, and alanine were prepared from weighed pure standards (Sigma) dissolved in HEPES buffered saline (Table 9.1) prepared in D$_2$O. Each of the three reference solutions were titrated to an observed glass electrode pH of 7.40 by adding deuterated acid (DCl) or base (NaOD) as needed.

NMR spectroscopy. NMR spectroscopy was conducted at the National Magnetic Resonance Facility at Madison. One-dimensional ^{1}H and two-dimensional ^{1}H-^{13}C HSQC spectra were collected on a Varian 600 MHz spectrometer equipped with a cryogenic probe. The pulse sequence used to collect ^{1}H NMR spectra consisted of a 1 s initial delay followed by a 90° pulse, and a 2 s acquisition period (32,000 data points). Each ^{1}H spectrum was collected in four transits with four steady-state transits. Sensitivity enhanced ^{1}H-^{13}C HSQC spectra were collected in four transits with 16 steady-state transits, 256 increments, a carbon spectral width of 70 ppm, an acquisition time of 300 ms (3,000 data points), and an initial delay of 1 s. High

resolution ^{1}H-^{13}C HSQC collected with four transits, 2000 increments, and a carbon spectral width of 140 ppm were also acquired for detection of low abundance metabolites and qualitative analysis of ^{13}C-^{13}C coupling. One- and two-dimensional time-domain data were Fourier transformed, phased, and chemical shift referenced to DSS. HSQC spectra were processed with a shifted exponential sine bell widow function in both the direct and indirect dimensions. All data processing was done automatically using custom NMRPipe (Delaglio, Grzesiek et al. 1995) scripts written in-house.

Metabolic flux analysis. Our methods for identifying and quantifying metabolites have been described in detail elsewhere (Lewis, Schommer et al. 2007; Lewis, Campanella et al. 2009; Lewis, Schommer et al. 2009; Lewis, Shortreed et al. 2010). Briefly, metabolites were identified using the Madison Metabolomics Consortium Database (Cui, Lewis et al. 2008), assignments in spectra of extracts were visually inspected by overlaying spectra of standards from the BioMagResBank (Markley, Anderson et al. 2007), and observed signal intensities were translated to concentrations using empirical calibration coefficients (Lewis, Schommer et al. 2007). All NMR data analyses were performed using the rNMR software package (Lewis, Schommer et al. 2009).

Sample limitation prevented us from conducting separate isotopic labeling experiments with 2-^{13}C glucose and U-^{13}C glucose. As a result, all fluxes reported in this study were derived from RBC suspensions incubated in 2-^{13}C glucose. A disadvantage of this approach is that PPP activity produces a mix of isotopomers in the downstream metabolites. To avoid complications associated with accounting for these isotopomers, glucose and lactate concentrations were quantified from their dispersed 1D ^{1}H signals. Total lactate concentrations reported reflect all of the observed isotopomers. Concentrations of 2,3-BPG, which does not have a dispersed 1D ^{1}H signal in RBC extracts, were based on observed 2-^{13}C 2,3-BPG concentrations observed in ^{1}H-

^{13}C HSQC spectra. Observed 2,3-BPG concentrations were corrected for glycolytic stochimetry but not for scrambling of the labeling pattern due to PPP activity.

Metabolic fluxes were calculated by regressing observed concentrations of metabolites as a function of time. Rates of glucose uptake and lactate production were estimated using linear regression of concentrations across all of the time points. Rates of 2,3-BPG flux, which are non-linear because 2,3-BPG is a pathway intermediate and is affected by pH, were based on initial labeling rates observed between the 0 and 1.5 h time points. Pentose phosphate pathway (PPP) flux was calculated from steady-state positional isotopic enrichment in lactate observed in samples incubated for 12 h in 2-^{13}C glucose. Our procedure for calculating PPP flux has been described elsewhere (Lewis, Campanella et al. 2009). Briefly, concentrations of 2-^{13}C lactate and 3-^{13}C lactate were determined by 1D ^{1}H NMR and corrected for naturally occurring ^{13}C enrichment. Pentose shunt activity (PC), defined as the fraction of incoming glucose resulting in PPP-derived glyceraldehyde-3-phosphate, was calculated from ^{13}C enrichment in the C3 and C2 positions of lactate using an equation developed by others (Katz and Wood 1960; Schrader, Eskey et al. 1993; Delgado, Castro et al. 2004):

$$\frac{C3_{lactate}}{C2_{lactate}} = \frac{2PC}{1+2PC}$$

The ^{1}H chemical shifts of HEPES observed in ^{1}H-^{13}C NMR spectra were used to measure sample pH. Chemical shifts were converted to pH using a titration curve generated from NMR spectra of HEPES buffered saline in D_2O that had been hand titrated to pHs (uncorrected glass electrode readings) ranging from 5.5 to 10.5. Standard deviations reported in this study reflect variations in regression coefficients between independent pool replicates. All of the reported p values were calculated by two-tailed equal variance t-test.

9.4. Results

Phenotypes of band 3 knockout mice. Transgenic mice were generated with a recombinant *Slc4a1* gene that contains a neomycin cassette in intron 4 and a 33 base pair deletion in exon 7 (-). On a mixed 129/Black Swiss background, mice homozygous for the transgene were born in a Mendelian ratio (17/74 = 22.9%) and more than 80% of homozygous pups survived to weaning. Similar to previous observations (Peters, Shivdasani et al. 1996), inbreeding or backcrossing led to increased mortality of homozygous pups. Despite having a nearly intact band 3 gene, homozygous (-/-) mice had a phenotype identical to previous homozygous band 3 null mice (Southgate, Chishti et al. 1996; Peters, Swearingen et al. 2004). Homozygous (-/-) mice had massive splenomegaly, severely reduced red cell number and hematocrit, elevated reticulocyte count, increased number of nucleated red cells and a severe osmotic fragility (Table 9.2). In contrast, heterozygous (+/-) mice had normal hematological values with the exception of a mild elevation in the reticulocyte count and slight osmotic fragility.

Quantitative RT-PCR analysis of the bone marrow and spleen from homozygous (-/-) mice showed a 12.5-fold reduction in band 3 mRNA ($p<0.001$, Table 9.2). Since the percentage of Ter119+ erythroid cells in marrow and spleen of homozygous (-/-) mice is more than four times greater than normal mice, we estimate that level of mRNA generated from the recombinant *Slc4a1* allele is approximately 3% of the wild type allele. We attribute the reduced expression of the recombinant *Slc4a1* allele to the intronic neomycin cassette, which has been shown to interfere with transcription or mRNA processing at other loci (Meyers, Lewandoski et al. 1998; Nagy, Moens et al. 1998).

Table 9.2. Hematological and Q-RT-PCR expression analysis of wild type C57BL6 (+/+), band 3 knockout (homozygous; -/-) and heterozygous (+/-) mice. Data are presented as the mean ± S.E.

	+/+	+/-	-/-
RBC (10^6/ L)	10.3 ± 0.2	10.1 ± 0.8	3.7 ± 0.2*
Hematocrit (%)	52.8 ± 1.7	51.5 ± 1.5	27.0 ± 4.6*
Hemoglobin (g/dL)	15.9 ± 0.2	15.9 ± 0.3	6.5 ± 0.3*
Mean Corpuscular Volume (fL)	51.1 ± 0.3	51.0 ± 0.4	64.8 ± 1.1*
Reticulocyte (%)	2.2 ± 1.2	6.3 ± 2.0*	95.2 ± 2.9*
RBC life span (days)	60	57	2
Nucleated RBC (10^3/ L)	0	0	69 ± 13*
Platelet	1051 ± 105	851 ± 72	620 ± 50
Band 3 mRNA (Bone Marrow)	1	0.67 ± 0.29	0.08 ± 0.007**
Band 3 mRNA (Spleen)	0.05 ± 0.02	0.02 ± 0.01	0.008 ± 0.004**

* $p < 0.001$; ** $p < 0.01$

Metabolism of band 3 knockout reticulocytes. We generated the band 3 (-/-) mouse to investigate the metabolic effects of *in vivo* disruption of the GE-band 3 complex. To control for the high (95%) reticulocyte count of the band 3 knockouts, band 3 -/- samples were matched with samples from healthy wild type mice (band 3 +/+) with an elevated (75%) reticulocyte count. Samples of mature band 3 +/+ erythrocytes were also included to illustrate differences between reticulocyte and erythrocyte metabolism.

Based on the extensive evidence supporting the GE-band 3 model (Murthy, Liu et al. 1981; Tsai, Murthy et al. 1982; Jenkins, Kezdy et al. 1985; Campanella, Chu et al. 2005; Chu, Breite et al. 2008; Lewis, Campanella et al. 2009), we predicted that band 3 -/- reticulocytes would show profound alterations in their oxygen-dependent metabolic regulation. Specifically, we predicted knockouts would have elevated glycolytic flux and reduced PPP activity under oxygenated conditions but show little difference from wild type controls under deoxygenated conditions. These predictions were based on the results of our previous study of GE-band 3 disruption in erythrocytes (Lewis, Campanella et al. 2009).

To our surprise, band 3 -/- reticulocytes did not show any evidence of altered metabolic regulation. Both band 3 -/- and band 3 +/+ reticulocytes responded to deoxygenation by increasing glycolytic flux 60% (band 3 -/-, $p = 0.011$; band 3 +/+, $p < 0.001$) and decreasing pentose shunt activity (PC) from 2-3% to 1% (band 3 -/-, $p = 0.017$; band 3 +/+, $p < 0.001$). Similarly, deoxygenation stimulated a 2.1 fold increase ($p < 0.001$) in 2,3-BPG flux of band 3 -/- reticulocytes and a 1.5 fold increase ($p = 0.01$) in band 3 +/+ reticulocytes (Fig. 9.1, 9.2, and Table 9.3). Although the overall glycolytic activity of band 3 -/- samples was 51% higher than band 3 +/+ reticulocytes, this difference can be attributed to the disparity in reticulocyte count between the band 3 -/- (95%) and band 3 +/+ (75%) samples.
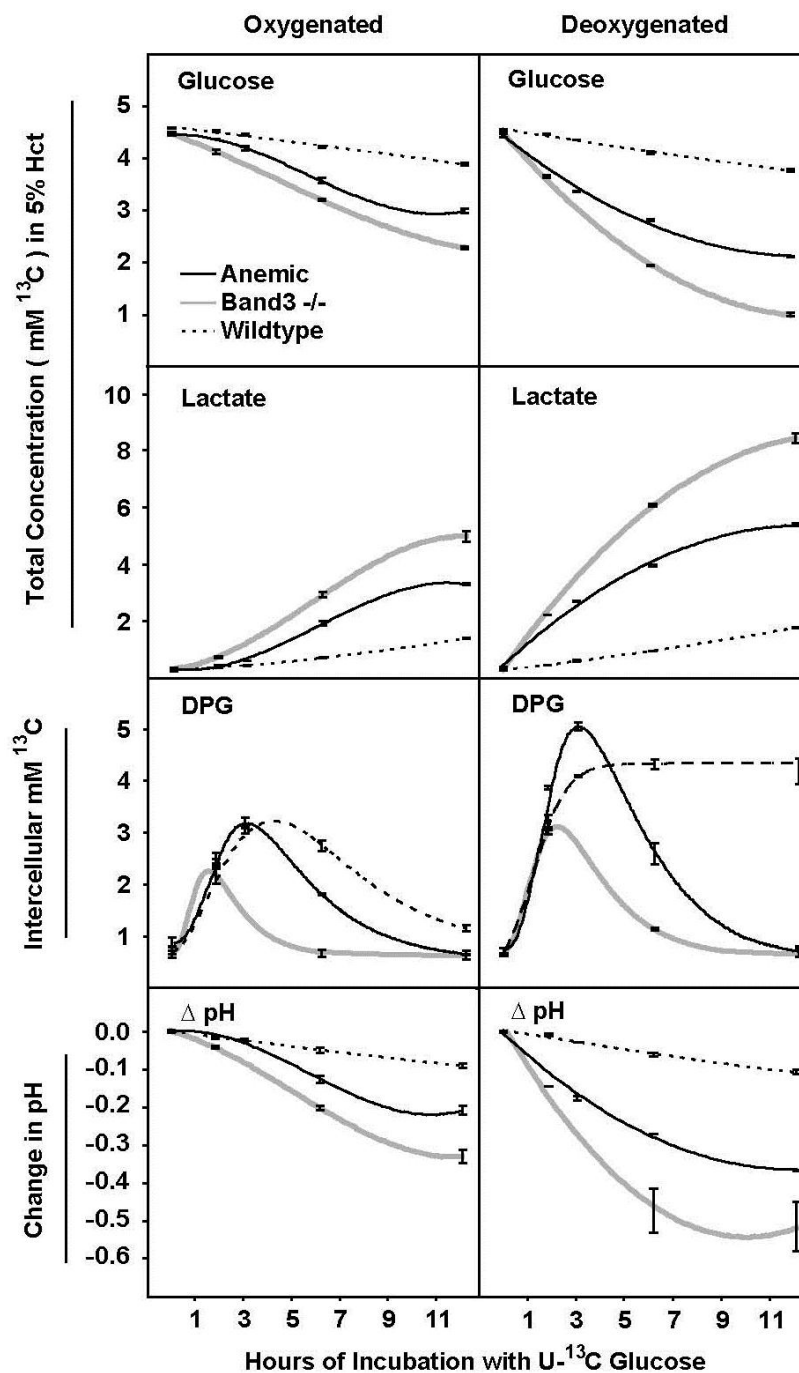
Figure 9.1. Concentrations of U-^{13}C labeled metabolites and pH in erythrocytes and reticulocytes incubated with U-^{13}C-glucose as measured by NMR. Dotted lines indicate wild type (band 3 +/+) erythrocytes, black lines indicate reticulocytes from wild type anemic mice, and gray lines indicate band 3 -/- reticulocytes. Error bars indicate standard error.
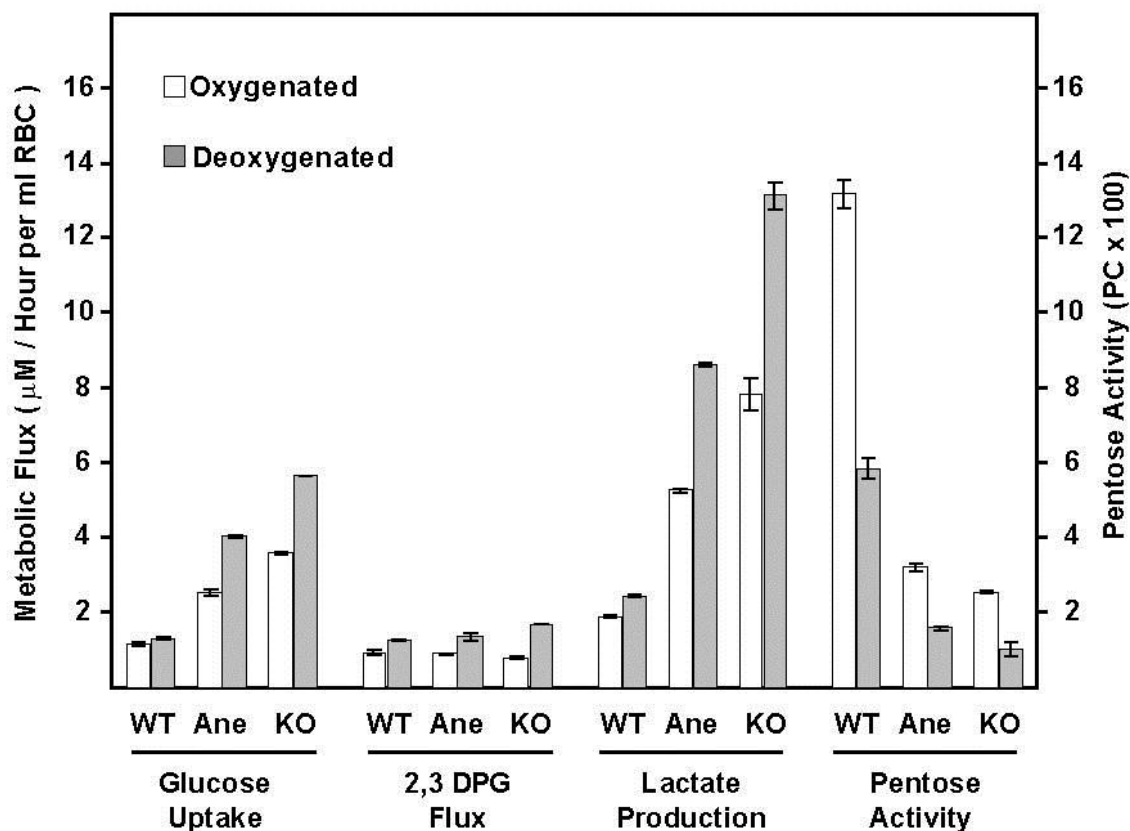
Figure 9.2. Rates of glucose consumption, lactate production, 2,3-BPG production, and pentose shunt activity (PC) observed in erythrocytes and reticulocytes. Light bars indicate oxygenated samples, dark bars indicate deoxygenated samples and error bars show standard error. Abbreviations: WT, wild type (band 3 +/+) erythrocytes; Ane, wild type (band 3 +/+) reticulocytes; KO, knockout (band 3 -/-) reticulocytes.

Table 9.3. Metabolic fluxes in isotopically labeled red cells measured by NMR[a]

	Erythrocyte (+/+) O_2	Erythrocyte (+/+) Argon	Reticulocyte (+/+) O_2	Reticulocyte (+/+) Argon	Reticulocyte (-/-) O_2	Reticulocyte (-/-) Ar
Glucose	1.17 ± 0.04	1.31 ± 0.02	2.54 ± 0.08	4.03 ± 0.04	3.59 ± 0.03	5.65 ± 0.01
Lactate	1.89 ± 0.06	2.44 ± 0.06	5.26 ± 0.06	8.63 ± 0.06	7.84 ± 0.42	13.1 ± 0.37
PC[‡]	0.13 ± 0.004	0.06 ± 0.003	0.03 ± 0.001	0.02 ± 0.001	0.03 ± 0.001	0.01 ± 0.002
2,3-BPG	0.91 ± 0.06	1.26 ± 0.03	0.89 ± 0.02	1.35 ± 0.1	0.78 ± 0.02	1.67 ± 0.02
Alanine	0.02 ± 0.003	0.01 ± 0.01	0.04 ± 0.01	0.07 ± 0.01	0.08 ± 0.01	0.09 ± 0.01
Δ pH	-0.15 ± 0.01	-0.18 ± 0.01	-0.37 ± 0.02	-0.63 ± 0.01	-0.55 ± 0.05	-0.86 ± 0.15

[a] Data are expressed as mean ± S.E.; metabolite fluxes are in units of µM / hour per ml RBC and Δ pH are presented as Δ pH / hour per ml RBC.

[‡] Pentose activity (PC) expresses the fraction of incoming glucose used to produce pentose-derived glyceraldehyde-3-phosphate; metabolic fluxes for various points in the pentose shunt can be calculated by multiplying PC by the glucose uptake rate and adjusting for the appropriate stoichiometry

The only striking metabolic differences observed in this study were between reticulocytes and erythrocytes. Reticulocyte-enriched band 3 +/+ samples had, on average, four fold higher rates of glucose consumption, and 28% lower pentose shunt activities than comparable erythrocytes (Table 9.3). As expected, high resolution ^{1}H-^{13}C HSQC spectra showed evidence for tricarboxylic acid cycle (TCA) activity in reticulocytes but not in erythrocytes (Fig. 9.3). However, concentrations of the two TCA-derived metabolites observed in reticulocyte extracts, succinate and glutamate (glutamate is synthesized from α–ketoglutarate), were below our limit for quantification. Interestingly, 2,3-BPG fluxes were relatively consistent between erythrocytes and reticulocytes despite the four fold difference in glycolytic fluxes (Table 9.3).
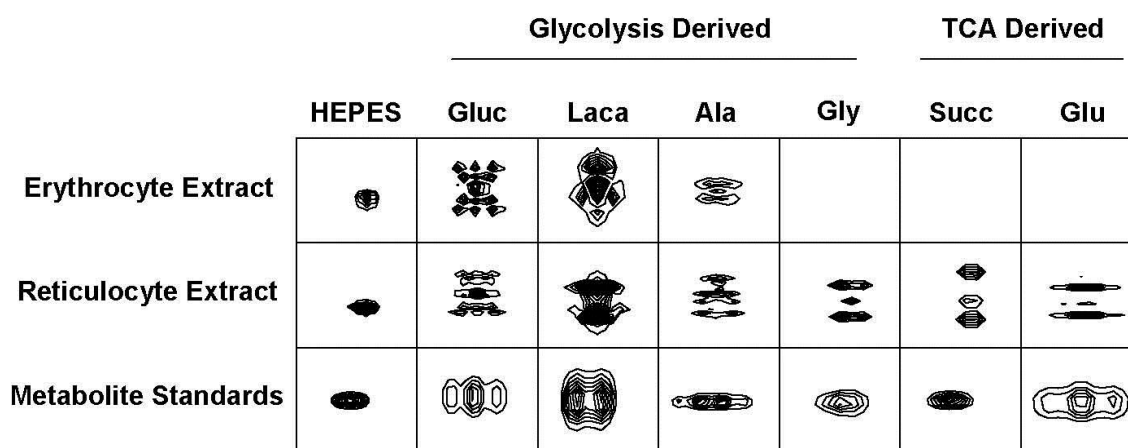

Figure 9.3. Metabolites observed in high resolution (2000 increments) ^{1}H-^{13}C NMR spectra of erythrocyte and reticulocyte (band 3 +/+) extracts. Extracts were prepared from 25% hematocrit cell suspensions incubated for 12 h in [U-^{13}C] glucose. Metabolite standards and HEPES signals have a peak shape indicative of natural abundance ^{13}C labeling whereas metabolites in extracts show ^{13}C-^{13}C coupling indicative of isotopic enrichment. The intensities of the HEPES, glucose, and lactate signals were reduced to show their peak shape. Abbreviations: TCA, tricarboxylic acid cycle; HEPES, (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; Gluc, glucose; Laca, lactic acid; Ala, alanine; Gly, glycine; Succ, succinate; Glu, glutamate.
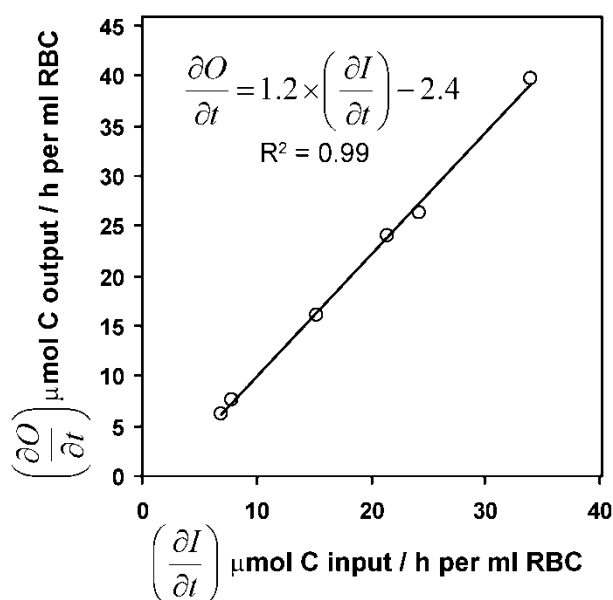
9.5. Discussion

This study demonstrated oxygen-dependent alterations in glycolytic, 2,3-DPG, and PPP fluxes of reticulocytes and showed that these metabolic phenomena are unaltered in cells missing the band 3 protein. Because the GE complex cannot form in band 3 -/- reticulocytes (Campanella, Chu et al. 2008), our data suggest a robust band 3-indepenent regulatory mechanism in reticulocytes. This finding contradicts our recent study showing strong evidence for GE-band 3 regulation in erythrocytes (Lewis, Campanella et al. 2009). These conflicting lines of evidence indicate significant metabolic differences between the human erythrocytes used in our previous study and the mouse reticulocytes used in this study.

One well established distinction between erythrocyte and reticulocyte metabolism is aerobic glucose metabolism via the TCA cycle activity and oxidative phosphorylation. As was shown here (Fig. 9.3), and elsewhere (Ghosh and Sloviter 1973; Siems, Muller et al. 1982; Kim, Koury et al. 1991), reticulocytes are capable of aerobic glucose metabolism whereas erythrocytes are restricted to the anaerobic pathway. If metabolic fluxes in reticulocytes are controlled primarily by demand for ATP, which has been previously argued (Hamasaki, Asakura et al. 1970; Ghosh and Sloviter 1973; Rapoprot, Berger et al. 1976; Mulquiney, Bubb et al. 1999; Mulquiney and Kuchel 1999; Mulquiney and Kuchel 1999), then the increased glycolytic fluxes of deoxygenated cells can be attributed to a switch from aerobic to anaerobic glucose metabolism. Although this regulatory mechanism is intuitive in reticulocytes from species (e.g. rabbits) that rely heavily on aerobic metabolism (Siems, Muller et al. 1982), it is unclear why an aerobic/anaerobic metabolic shift would produce the large metabolic alterations observed in murine reticulocytes, which commit less than 6% of incoming glucose to the oxidative pathway (Ghosh and Sloviter 1973; Kim, Koury et al. 1991). Although we observed nearly stoichiometric conversion of glucose to lactate (Fig. 9.4) and TCA activity at a level comparable to alanine and glycine biosynthesis (Fig. 9.3), deoxygenated stimulated a 60% increase in glycolytic flux.

The regulatory mechanism controlling oxygen-dependent metabolic alterations is unclear to us, and others who have studied murine reticulocytes (Ghosh and Sloviter 1973; Kim, Koury et al. 1991). With respect to GE-band 3 regulation, however, our data are unequivocal; reticulocytes lacking the GE protein have glycolytic and PPP fluxes equivalent to wild type cells and show the normal oxygen-dependent metabolic alterations. In summary, reticulocytes have a robust band 3-independ mechanism for controlling oxygen-dependent changes in glycolytic and PPP fluxes.

Figure 9.4. Rates of carbon uptake versus rates of glycolytic and PPP carbon produced by reticulocytes and erythrocytes under oxygenated and deoxygenated conditions. Carbon uptake rates are based on stoichiometrically adjusted rates of glucose consumption (6 x glucose uptake), carbon output rates represent stoichiometrically adjusted rates of glycolytically derived metabolite production (3 × [lactate + alanine +

pyruvate production]) and CO_2 lost through the oxidative branch of the PPP (PC × glucose uptake). Data points reflect rates observed in (upper two points) band 3 -/- reticulocytes, (middle two points) band 3 +/+ reticulocytes, and (lower two points) band 3 +/+ erythrocytes under oxygenated and deoxygenated conditions. The slope of the regression line (1.2) indicates the observed moles of carbon produced per mole of input carbon. The intercept of the regression line indicates carbon present in the system prior to the start of the experiment.

The graph shows:

$$\frac{\partial O}{\partial t} = 1.2 \times \left(\frac{\partial I}{\partial t}\right) - 2.4$$

$R^2 = 0.99$

Y-axis: $\left(\frac{\partial O}{\partial t}\right)$ μmol C output / h per ml RBC

X-axis: $\left(\frac{\partial I}{\partial t}\right)$ μmol C input / h per ml RBC

CHAPTER 10

Glutathione synthesis via aminotransferases in human erythrocytes

Adapted from:

James J. Ellinger, Ian A. Lewis, and John L. Markley. Glutathione synthesis via aminotransferases in human erythrocytes. (to be submitted)

This study was a follow up to my original investigation of oxygen-dependent metabolic regulation in erythrocytes (Chapter 8). I had observed significant *de novo* production of alanine in the original study, which is non-intuitive because erythrocytes do not have the capacity for new protein synthesis. However, red cells do require glutamate for glutathione synthesis and, strangely, lack a transporter for this amino acid. I speculated that erythrocytes were using alanine aminotransferase to supply the cell's demand for glutamate. James J. Ellinger, a fellow graduate student in the Markley lab, and I conducted this study to determine the extent to which aminotransferases contribute to glutathione synthesis. This study also illustrates a powerful application of bioanalytical metabolomics. Cell lysates were incubated with various substrates and the products were determined by multidimensional NMR. Our metabolomics approach allowed us to trace a complex coupled reaction that resulted from incubating lysates with oxaloacetate. This strategy could provide a robust mechanism for tracing novel metabolic pathways in more exotic systems. With the exception of the MS work and data analysis, which was done exclusively by Mr. Ellinger, I shared most of the duties with Mr. Ellinger.

10.1. Abstract

Glutathione plays a central role in repairing oxidative damage to red blood cells (RBCs). Erythrocytes require a continual supply of glutamate to support glutathione synthesis, but are unable to transport this amino acid across their cell membrane. Although the source of the RBC glutamate pool has been studied extensively, all of the proposed mechanisms for glutamate assimilation have been disputed. In this study, we screen for aminotransferase activity in RBCs using ^{1}H-^{13}C nuclear magnetic resonance (NMR) and multiple reaction mode mass spectrometry (MRM-MS). We show that both alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are present in erythrocytes. Furthermore, we find that ALT contributes significantly to the total glutamate pool of intact cells (89% after 21 h) and that ALT-derived glutamate is subsequently used for glutathione synthesis. These findings provide a direct mechanism for the sole unresolved step in erythrocyte glutathione synthesis.

10.2. Introduction

Glutathione (GSH) plays a central role in repairing oxidative damage to red blood cells (RBCs). Alterations in glutathione levels have been linked to a variety of human disorders, including sickle cell disease, hemolytic anemia, and pulmonary hypertension (Wu, Fang et al. 2004). Despite fifty years of research, the RBC glutathione synthesis pathway has not been fully characterized. Although all of the enzymes required for synthesizing glutathione from its component amino acids (glutamate, cysteine, glycine) are present in erythrocytes (Majerus, Brauner et al. 1971; Minnich, Smith et al. 1971), the RBC membrane is essentially impermeable to glutamate (Winter and Christensen 1964; Sass 1968; Young, Wolowyk et al. 1987). Various studies have cited either peptide transport (King and Kuchel 1985) or glutamine deamination (Griffith 1981; Ellory, Preston et al. 1983; Niihara, Zerez et al. 1997; Morris, Suh et al. 2008) as potential mechanisms for glutamate assimilation. However, the evidence supporting these mechanisms has been questioned (King and Kuchel 1985; Kuchel, King et al. 1987).

Recently, we observed considerable *de novo* alanine production in RBCs incubated with ^{13}C-enriched glucose (Lewis, Campanella et al. 2009). This result, which has been previously reported (Manuel y Keenoy, Conget et al. 1991), suggests that washed RBCs have significant aminotransferase activity. Since both α-ketoglutarate and alanine cross the RBC membrane (Sass 1968; Griffith 1981; Young, Wolowyk et al. 1983), transamination could provide a direct mechanism for *de novo* glutamate synthesis (Fig. 10.1). Although RBCs are known to contain both aspartate aminotransferase (Raica and Sauberlich 1964) (AST; EC 2.6.1.1) and alanine aminotransferase (Raica and Sauberlich 1964; Ubbink, Bissbort et al. 1989) (ALT; EC 2.6.1.2), the contribution of these enzymes to glutathione synthesis has not been documented.

Given active alanine production and α-ketoglutarate transport in RBCs, we hypothesized that glutathione synthesis relies on *de novo* glutamate synthesis produced by intracellular aminotransferases. In this study, we verify *de novo* alanine production, demonstrate both ALT and AST activity in erythrocyte lysates, and show that ^{13}C-enriched α-ketoglutarate is converted to glutamate and incorporated into glutathione by intact RBCs. These data indicate that aminotransferases play a significant role in glutathione biosynthesis.
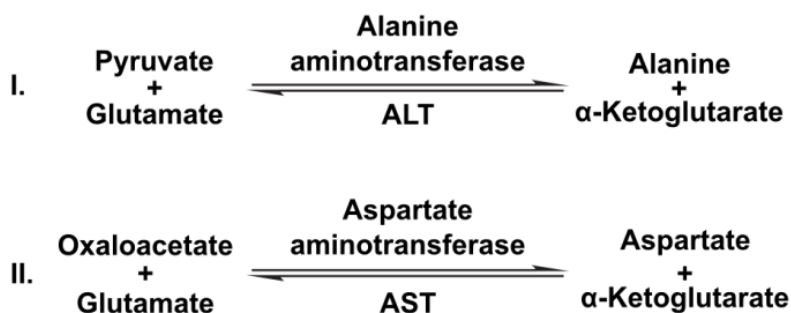

Figure 10.1. Reaction mechanisms for I) alanine aminotransferase (ALT) and II) aspartate aminotransferase (AST). Unlike glutamate, both alanine and α-ketoglutarate cross the RBC membrane; ALT and AST provide a mechanism for *de novo* glutamate production that could support glutathione biosynthesis.

10.3. Materials and Methods

Preparation of RBCs and lysates. Fresh blood was collected by venipuncture from healthy human volunteers ($n = 3$) into heparinized vacutainers. RBCs were isolated by centrifugation (10 min at 3,000 × g), and the buffy coat was discarded. Samples were washed three times in isotonic HEPES buffer (25 mM HEPES, 1 mM NaH_2PO_4, 106 mM NaCl, 19 mM KCl, 1 mM $CaCl_2$, pH 7.4), combined into a single pool, and resuspended at 20% hematocrit in HEPES buffer containing 20 IU/mL Penicillin/Streptomycin (Gibco). Lysates were prepared from washed RBCs suspended at 36% hematocrit. RBCs were hemolyzed by sonication for 1 min and the resulting lysates were diluted to the equivalent of 25% hematocrit with isotonic HEPES buffer containing metabolite standards (see aminotransferase activity assays in RBC lysates). All sample preparations were conducted at 4 °C to minimize metabolic activity.

Metabolic activity assays in intact RBCs. Washed RBCs were prepared in isotonic HEPES buffer containing either (i) 5 mM [U-^{13}C]-glucose (Isotec) or (ii) 5 mM glucose, 5 mM alanine, and 5 mM [^{13}C$_{1,2,3,4}$]-α-ketoglutarate (Cambdridge Isotope Laboratories). Each RBC suspension (i and ii) were split into three replicate samples, which were incubated over periods of 12 (i) and 21 (ii) h at 37 °C. Aliquots (1 mL) were collected from samples after 0, 1.5, 3, 6, 12, and 21 h of incubation and flash frozen in liquid nitrogen. Metabolites were then extracted from each sample (as described below) and analyzed by nuclear magnetic resonance (NMR) and mass spectrometry (MS) for the presence of isotopically enriched molecules. Enriched metabolites were identified and quantified by NMR, and the isotopic enrichments were determined by MS. Samples containing [U-^{13}C]-glucose (i) provided information about glycolytically-related metabolism, whereas samples containing [^{13}C$_{1,2,3,4}$]-α-ketoglutarate (ii) provided information on aminotransferase reactions.

Aminotransferase activity assays in RBC lysates. RBC lysates were incubated at 37 °C for 6−12 h in isotonic HEPES buffer containing 200 µM pyridoxal-5′-phosphate (PLP; the

cofactor required by aminotransferases) and one of the following pairs of substrates (5 mM each): pyruvate + glutamate for alanine aminotransferase (ALT) assays, alanine + α-ketoglutarate for reverse ALT assays, oxaloacetate + glutamate for aspartate aminotransferase (AST) assays, aspartate + α-ketoglutarate for reverse AST assays, and pyruvate + one of each of the amino acids listed in Fig. 10.3 for aminotransferase screening assays. Aliquots (900 µL) of each sample were collected after 0, 1, 3, 6 and 12 h of incubation and flash frozen in liquid nitrogen. Metabolites were then extracted (as described below) and analyzed by NMR.

Metabolite extraction. Samples were placed in a boiling water bath for 7.5 min to lyse cells and halt enzymatic activity. Boiled lysates were spun at 16,000 × g to pellet cellular debris. For samples to be analyzed by LC-MS/MRM, a 200 µL aliquot of each supernatant was transferred to a fresh tube and stored at -80 °C until analysis. For samples to be analyzed by NMR, an 800 µL aliquot of each supernatant was dried in a SpeedVac Concentrator (Thermo Scientific), and the resulting residue was dissolved in 800 µL D_2O containing 500 µM NaN_3 and 500 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

NMR spectroscopy. NMR spectroscopy was carried out at the National Magnetic Resonance Facility at Madison. Two-dimensional sensitivity enhanced ^1H-^{13}C HSQC spectra were collected on a Varian 600 MHz spectrometer equipped with a cryogenic probe. Spectra were collected, following 16 transients to achieve steady state, as four averaged transients with 128 increments in the second dimension (50 increments for the *in vivo* labeling studies). The acquisition time was 300 ms (3,000 data points), following an initial delay of 1 s; the carbon spectral width was 70 ppm. Time-domain data were Fourier transformed with a shifted exponential sine bell window function, phased, and chemical shift referenced to DSS using custom NMRPIPE (Delaglio, Grzesiek et al. 1995) scripts written in-house.

Our methods for identifying and quantifying metabolites by ^{1}H-^{13}C NMR have been described elsewhere (Lewis, Schommer et al. 2007; Lewis, Campanella et al. 2009). Briefly, metabolites were identified using the Madison Metabolomics Consortium Database (MMCD) (Cui, Lewis et al. 2008); assignments were verified by overlaying NMR spectra of standards from the BioMagResBank (BMRB) (Markley, Anderson et al. 2007); metabolite concentrations were calculated from peak intensities on the basis of calibration curves from metabolite standards prepared at 2, 5, and 10 mM. The rNMR software package (Lewis, Schommer et al. 2009) was used in performing all NMR data analyses.

Mass spectrometry. Mass spectra were collected at the University of Wisconsin Biotechnology Center Mass Spectrometry Facility. An aliquot (5 µL) of each metabolite extract was analyzed by liquid chromatography (LC) MS on an Applied Biosystems 3200 Q TRAP LC-MS/MS system equipped with an Agilent 1100 series capillary LC pump and an electrospray ionization (ESI) source. Online LC used a 4.6 mm × 150 mm Phenomenex Luna hydrophobic interaction chromatography (HILIC) column (200-Å pore size, 5-µm particle size) with a constant flow rate of 200 µL/min. Samples were eluted over a 48-min ammonium formate (50 mM, pH 5.4; buffer A) to acetonitrile gradient. The elution gradient was constructed as follows: 90% A at time zero, 70% A at 20 min, holding at 70% A for 2 min, 10% A at 22 min, holding at 10% A for 1 min, back to 90% A at 23 min, and isocratic at 90% A until 48 min. ESI-MS was performed in positive ion multiple reaction mode (MRM). Peak picking and integration were accomplished with Analyst software (Applied Biosystems). Metabolite identification, retention times, and ion fragment patterns, were verified by reference to standard compounds. The Kombyonyx isotope calculator (www.kombyonyx.com) was used to calculate isotopomer distributions.

Regression analyses and statistics. Rates of metabolite consumption and production were calculated by linear regression of metabolite concentration as a function of time. Regression analyses were based on all time points, except for samples incubated with

aspartate + α-ketoglutarate and oxaloacetate + glutamate. Kinetics observed in the latter samples were nonlinear because of spontaneous decarboxylation of oxaloacetate to pyruvate (Hatch and Heldt 1985). For these samples, initial rates were determined from the time points between 0 and 1 h. Samples incubated with [$^{13}C_{1,2,3,4}$] α-ketoglutarate contained contaminating [$^{13}C_{1,2,3,4}$] glutamate, which was determined from the 0 h time points and used as a baseline for subsequent measurements. Glutamate kinetics from cells incubated with [$^{13}C_{1,2,3,4}$] α-ketoglutarate were derived from linear regression of the first three time points. All of the *p* values presented here were derived from a two-tailed equal variance *t*-test.

10.4. Results

***De novo* alanine synthesis by RBCs.** Previous reports (Manuel y Keenoy, Conget et al. 1991; Lewis, Campanella et al. 2009) of alanine biosynthesis were verified by incubating RBCs for 12 h in 5 mM [U-^{13}C] glucose and monitoring the appearance of ^{13}C-enriched alanine. As expected, ^{1}H-^{13}C NMR analyses of extracts showed time-dependant accumulation of ^{13}C alanine (Fig. 10.2). The observed rate of ^{13}C alanine production (0.022 ± 0.001 μmol m^{-1} RBC h^{-1}) was consistent with our previous findings and accounted for 1.1% of the total ^{13}C output (Table 10.1). LC-ESI-MS analyses indicated that 35% of the total alanine pool was uniformly ^{13}C labeled after 12 h of incubation with [U-^{13}C] glucose, whereas samples harvested at 0 h showed natural abundance ^{13}C levels (Table 10.2).
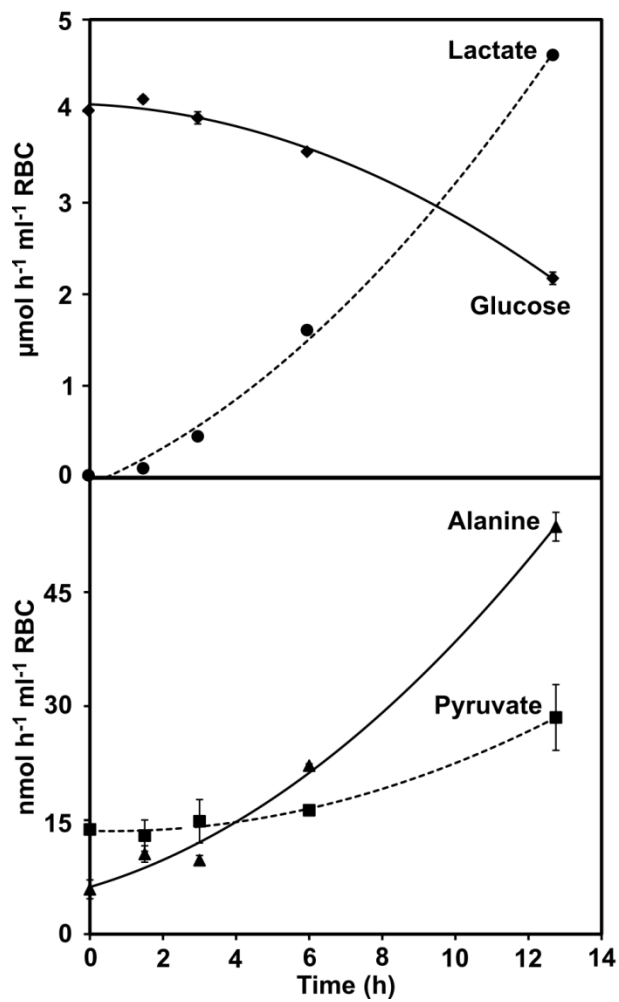
Figure 10.2. Glucose, lactate, alanine, and pyruvate metabolism observed in human RBCs. Samples were incubated with 5 mM [U-^{13}C] glucose and metabolites were quantified by ^{1}H-^{13}C NMR. Error bars represent S.E.M. ($n = 3$).

Table 10.1. Rates of glucose, lactate, alanine, pyruvate and glutamate metabolism in RBCs

	Biosynthetic rates (µmol ml^{-1} RBC hr^{-1})[a]					
Metabolites added	Glucose	Lactate	Alanine	Pyruvate	Glutamate	α-KG
[U-^{13}C]-glucose	-0.78 ± 0.03	1.89 ± 0.01	0.022 ± 0.001	0.01 ± 0.002	n.o.	n.o.
[^{13}C$_{1-4}$]-α-KG, glucose, alanine	-0.80 ± 0.01[†]	1.35 ± 0.02[†]	-0.51 ± 0.07[†]	0.57 ± 0.04[†]	0.174 ± 0.01[†]	-0.160 ± 0.064[†]

Table 10.2. Isotopomer distribution of alanine in RBCs incubated with [U-^{13}C] glucose

	Percent mass isotopomer distribution in alanine[a]				
Time (h)	Monoisotopic	M+1	M+2	M+3	M+4
0	95.64 ± 0.39	2.43 ± 0.63	n.d.	1.94 ± 0.51	0
12	60.58 ± 1.05**	4.04 ± 0.04*	n.d.	35.38 ± 1.10**	0
Expected distribution due to natural abundance	95.92	3.62	0.44	0.01	0

[a] Mean ± S.E.M., n = 3; [†] 21 h time point was excluded in the rate calculation; n.o. = not observed;

n.d. = not determined; * $p < 0.05$; ** $p < 0.001$

Aminotransferase activity in RBC lysates. To screen for aminotransferase activity, RBC lysates were incubated for 6 h in HEPES buffered saline containing 5 mM pyruvate, 200 μM pyridoxal-5′-phosphate (PLP), and 5 mM target amino acid. ^{1}H-^{13}C NMR analysis of the lysates showed significant alanine and α-ketoglutarate production only in the sample incubated with glutamate as the target amino acid (Fig. 10.3).
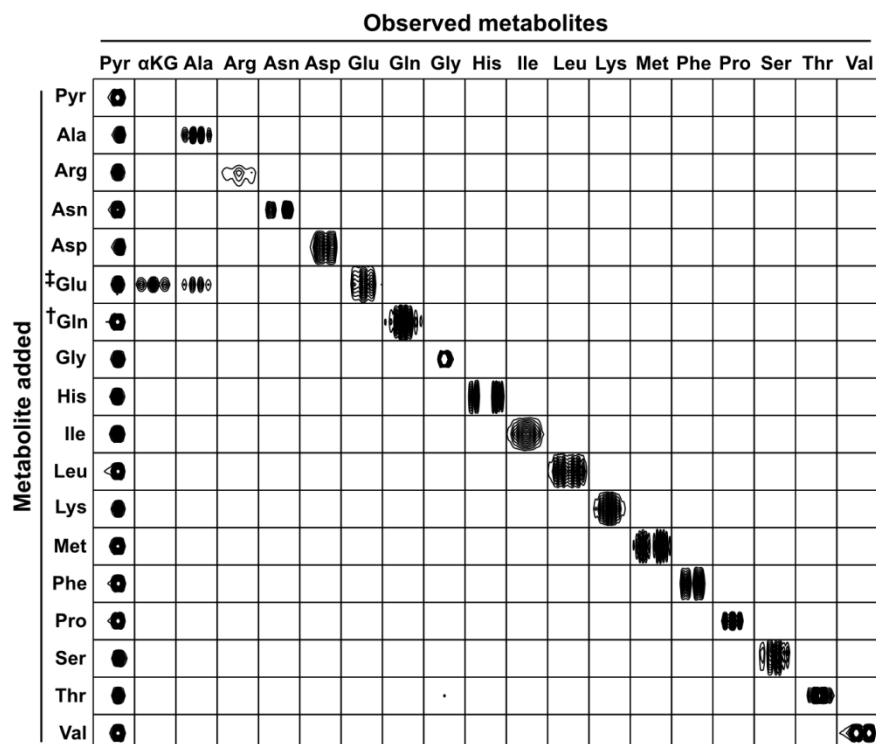

Figure 10.3. Aminotransferase screening assay. RBC lysates were incubated with 200 μM pyridoxal-5′-phosphate, 5 mM pyruvate, and 5 mM of a target amino acid. Each square shows the unique ^{1}H-^{13}C NMR sub-spectrum used to identify and quantify metabolite in the lysates. Row labels denote the compounds added to samples; column labels denote observed metabolite signals. Abbreviations: Pyr, pyruvate; αKG, α-ketoglutarate; amino acids are represented by standard three letter code. ‡ Samples prepared with glutamate showed signals from glutamate and the two products. † Samples incubated with glutamine showed no detectable levels of GA activity by NMR; GA activity was detectable by MS.

Conversion of glutamate + pyruvate to α-ketoglutarate + alanine in the presence of PLP is consistent with the established alanine aminotransferase (ALT) reaction mechanism (Fig. 10.1) and indicates aminotransferase activity in the lysates. However, ALT is present in serum, and contaminating enzyme from the wash medium could influence our results. To control for contaminating ALT, both serum and the final RBC wash buffer were assayed for ALT activity using the method we applied to the RBC lysates. Although ALT activity was observed in serum samples (2.82 ± 0.32 µmol alanine produced over 12 h), the final RBC wash buffer showed no detectable ALT activity (Fig. 10.4). These findings demonstrate that our assay is sufficiently sensitive to detect serum levels of ALT and that the ALT activity observed in RBC lysates did not result from extracellular ALT.

To determine the activity of AST and ALT in erythrocytes, RBC lysates were incubated for 12 h with 200 µM PLP and the respective substrates for the two aminotransferases (Fig. 10.1). Both the forward and reverse reaction rates were determined for each aminotransferase. ^{1}H-^{13}C NMR analysis of the lysates showed efficient conversion of α-ketoglutarate to glutamate in the presence of either aspartate or alanine, indicating AST and ALT activity, respectively. Similarly, glutamate was efficiently converted to α-ketoglutarate in the presence of either pyruvate or oxaloacetate (Fig. 10.5). As expected for the ALT and AST enzymes (Yagi, Kagamiyama et al. 1982; Manuel y Keenoy, Conget et al. 1991), rate constants observed for the forward and reverse reactions were comparable (Table 10.3). However, the rate constant for glutamate production via the AST pathway was 12.7 times higher than that via the ALT pathway.
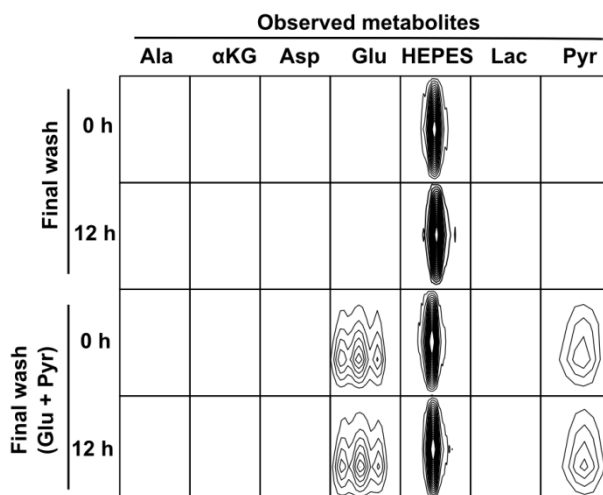
Figure 10.4. Metabolic activity of final wash buffer used for RBC preparation. ^{1}H-^{13}C NMR signals indicating no detectable aminotransferase activity after 12 h of incubation. Abbreviations: αKG, α-ketoglutarate; Lac, lactate; Pyr, pyruvate; amino acids are represented by their three letter codes.
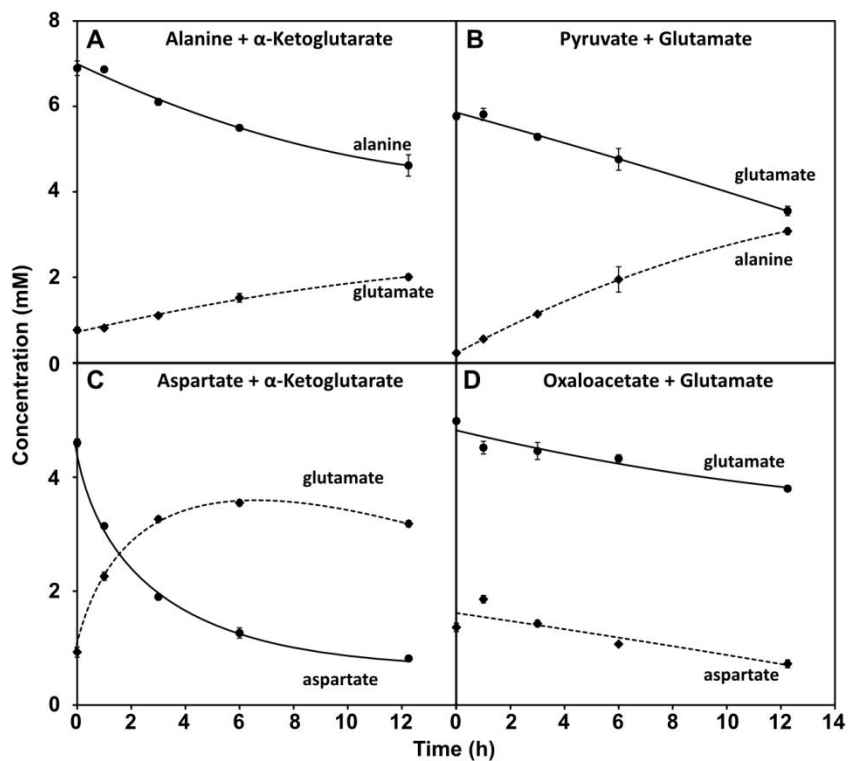
Figure 10.5. Aminotransferase reactions catalyzed by (A,B) alanine aminotransferase and (C,D) aspartate aminotransferase observed in human RBC lysates. Samples were incubated with 5 mM of each substrate and analyzed by ^{1}H-^{13}C NMR. The title of each plot indicates which substrates were added. Error bars represent S.E.M. ($n = 3$).

Table 10.3. Rates of ALT and AST aminotransferase reactions observed in RBC lysates.

Metabolites added	Observed rates (µmol mL^{-1} lysate hour^{-1})[a]				
	Alanine	α-KG	Aspartate	Glutamate	Pyruvate
alanine + α-ketoglutarate[e]	-0.78 ± 0.11	-0.77 ± 0.08	n.o.	0.42 ± 0.03	0.46 ± 0.03
glutamate + pyruvate[e]	0.94 ± 0.03	0.57 ± 0.01	n.o.	-0.76 ± 0.04	-0.46 ± 0.06
aspartate + α-ketoglutarate[f]	0.49 ± 0.17	-3.23 ± 0.75[c]	-5.85 ± 0.30[c]	5.32 ± 0.45[c]	0.73 ± 0.05
glutamate + oxaloacetate[f]	-0.73 ± 0.05	2.13 ± 0.54[c]	1.98 ± 0.53[c]	-1.87 ± 0.28[c]	-0.18 ± 0.04

[a] Mean ± S.E.M., n = 3; n.o. = not observed; [c] rate was calculated using the first two time points; [e] ALT pathway; [f] AST pathway

Interpretation of the AST reaction is complicated by the rapid spontaneous degradation of oxaloacetate to pyruvate (Hatch and Heldt 1985). RBC lysates incubated with aspartate + α-ketoglutarate showed rates of alanine production comparable to samples incubated with pyruvate + glutamate (Table 10.3). This non-intuitive finding is the product of a three-part coupled reaction involving conversion of aspartate to oxaloacetate via AST, spontaneous degradation of oxaloacetate to pyruvate, and the subsequent conversion of pyruvate to alanine via ALT.

Glutamate and glutathione synthesis in intact RBCs. Intracellular aminotransferases provide a direct route for *de novo* synthesis of glutamate and glutathione. To test the extent to which ALT influences intracellular glutamate and glutathione pools of intact cells, RBCs were incubated for 21 h with glucose, alanine and [$^{13}C_{1,2,3,4}$] α-ketoglutarate (5 mM each). Metabolic rates were measured by ^{1}H-^{13}C NMR, and isotopic enrichments were determined by LC-MS/MRM. Intact RBCs produced ^{13}C-enriched glutamate at an average rate of 0.17 μmol mL^{-1} RBC h^{-1} and consumed alanine at a rate of 0.51 μmol mL^{-1} RBC h^{-1} (Table 10.1). MS analysis showed that 89% ($p < 0.001$ relative to natural abundance) of the RBC glutamate pool was [$^{13}C_{1,2,3,4}$] labeled after 21 h (Table 10.4), indicating active *de novo* glutamate synthesis by intact cells. Although glutathione biosynthetic rates were below the NMR detection limit, MS analysis showed significant ($p < 0.05$ relative to natural abundance) $^{13}C_{1,2,3,4}$ isotopic enrichment in the glutamate moiety of glutathione at the 21 h time point and time-dependent enrichment of glutathione over the course of the experiment (Table 10.4).

Table 10.4. Isotopomer distribution of glutamate and glutathione in RBCs incubated with alanine, glucose, and $[^{13}C_{1,2,3,4}]$ α-ketoglutarate

	Percent Mass isotopomer distribution of glutamate[a]				
Time (h)	**Monoisotopic**	**M+1**	**M+2**	**M+3**	**M+4**
0	95.34 ± 1.66	4.66 ± 1.66	0	0	0
6	15.87 ± 0.66**	0.72 ± 0.08**	0.10 ± 0.10	2.52 ± 0.37*	80.79 ± 0.31**
12	10.39 ± 0.69**	0.41 ± 0.01**	0	3.08 ± 0.55*	86.12 ± 0.90**
21.25	7.47 ± 0.24**	0.24 ± 0.01**	0	2.82 ± 0.14*	89.44 ± 0.32**
Expected distribution due to natural abundance	93.40	5.64	0.91	0.05	0.00

	Percent Mass isotopomer distribution of glutathione[a]				
Time (h)	Monoisotopic	M+1	M+2	M+3	M+4
0	85.56 ± 2.48	11.37 ± 2.88	1.47 ± 0.86*	1.59 ± 0.81	0
6	86.59 ± 1.07	7.99 ± 0.19*	1.67 ± 0.45*	0.33 ± 0.33	3.42 ± 0.33**
12	80.93 ± 0.76	8.08 ± 0.60	2.11 ± 0.34*	0.80 ± 0.42	8.09 ± 0.15**
21.25	78.35 ± 0.93*	7.96 ± 0.67	1.03 ± 0.52*	1.38 ± 0.18*	11.28 ± 1.19*
Expected distribution due to natural abundance	82.93	10.90	5.43	0.62	0.10

[a] Mean ± S.E.M., $n = 3$; * $p < 0.05$; ** $p < 0.001$

Glutamine aminohydrolase (GA; EC 3.5.1.2) has been purified from human RBCs,[25] and several studies have attributed intracellular glutamate pools to the deamination of glutamine by GA. To determine the relative contributions of GA and ALT to glutamate synthesis, washed RBCs were incubated for 12 h in HEPES buffered saline containing either (i) glucose, alanine and $[^{13}C_{1,2,3,4}]$ α-ketoglutarate or (ii) glucose and $[U^{13}C-^{15}N]$ glutamine (5 mM each). ALT activity was determined by $^1H-^{13}C$ NMR, whereas GA activity was measured by MS. The GA-derived glutamate production rate (0.0039 ± 0.00065 µmol mL^{-1} RBC h^{-1}) was 30 times lower than that observed for ALT-derived glutamate (Table 10.1). This finding supports previous studies (King and Kuchel 1985) that reported minimal glutamine aminohydrolase activity in washed RBCs.

10.5. Discussion

Human erythrocytes require a continual supply of glutamate to support glutathione synthesis, but are unable to transport this amino acid across their cell membrane (Winter and Christensen 1964; Sass 1968; Young, Wolowyk et al. 1987). Although the source of the RBC glutamate pool has been studied extensively (Srivastava and Beutler 1967; Kuchel, King et al. 1987), the proposed mechanisms for glutamate assimilation fail to account for intracellular glutamate levels (King and Kuchel 1985; Young, Wolowyk et al. 1987). In this study, we show that 89% of the intracellular glutamate pool is ALT-derived after 21 h. As expected, ALT-derived glutamate is ultimately incorporated into glutathione. These findings provide direct evidence supporting a role for ALT in maintaining RBC glutamate and glutathione levels *in vivo*.

We also observed aspartate aminotransferase activity at rates exceeding ALT by 12.7 fold. In contrast to our ALT findings, the biological role of AST is unclear. The RBC membrane is thought to be impermeable to aspartate (Maede, Inaba et al. 1983; Fincham, Mason et al. 1987), and mature erythrocytes lack the enzymes required for oxaloacetate synthesis. However, all of the aspartate transport studies we are aware of were conducted with non-human RBCs. If

aspartate were transported in human RBCs, then AST would provide an efficient mechanism for *de novo* glutamate biosynthesis. Moreover, the high activity of AST could be explained by the lower concentrations of aspartate relative to alanine in human serum. Alternatively, AST activity may be a developmental holdover from reticulocytes, which synthesize oxaloacetate as a part of the tricarboxylic acid cycle (Gasko and Danon 1972) and presumably synthesize aspartate *de novo*.

Glutamine aminohydrolase is often cited as the main contributor to intracellular levels of glutamate in RBCs (Griffith 1981; Ellory, Preston et al. 1983; Niihara, Zerez et al. 1997; Morris, Suh et al. 2008). While we observed GA activity consistent with the literature (Griffith 1981), GA cannot account for steady-state glutamate concentration. Rates of GSH biosynthesis, even by conservative estimates (Griffith 1981) exceed GA activity by 7.6 fold (Fig. 10.6). This is clear evidence that GA is a minor contributor to the glutamate pool. Similarly, the AST contribution is also modest because of the limitation in aspartate transport (Maede, Inaba et al. 1983; Fincham, Mason et al. 1987). In contrast, ALT activity was found to exceed GA capacity by 30 fold, a value significantly greater than the demand for GSH. These findings indicate that GSH synthesis predominantly relies on ALT activity, which is further supported by the observation that cells incubated without α-ketoglutarate transport alanine at a rate equivalent to demand for GSH (Fig. 10.6).

In summary, we have evaluated three possible pathways for glutamate production in RBCs. Of these pathways, only ALT provides a plausible mechanism for supporting GSH synthesis. Given that human RBCs are readily permeable to cysteine and glycine (Harvey and Clive Ellory 1989) and that all of the glutathione biosynthetic steps are well established (Majerus, Brauner et al. 1971; Minnich, Smith et al. 1971), these data provide a simple mechanism for the sole unresolved step in glutathione biosynthesis.
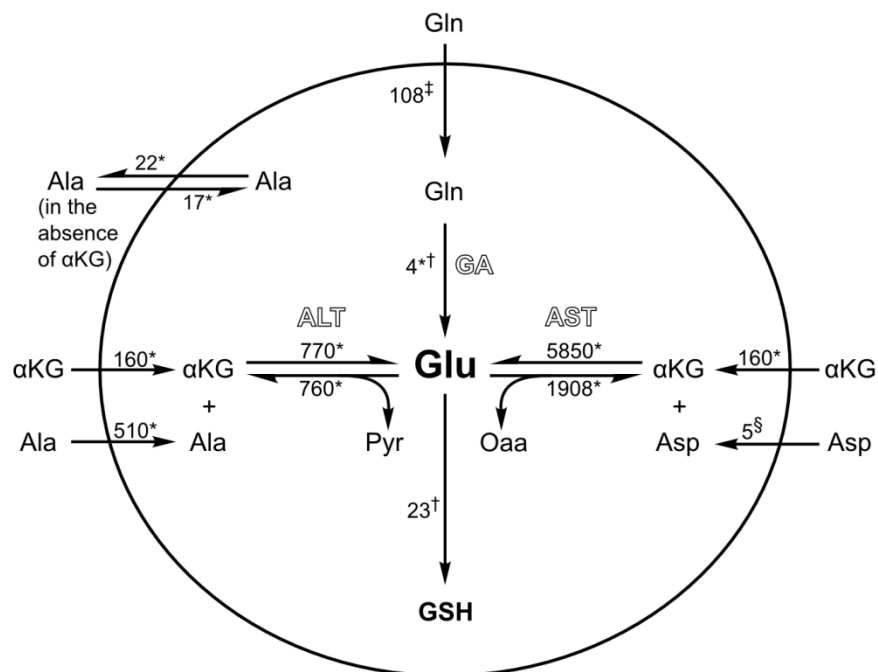
Figure 10.6. Three pathways for *de novo* glutamate synthesis in RBCs. Rates (μM h^{-1}) indicate NMR, MS and literature values for transport kinetics, alanine aminotransferase, aspartate aminotransferase, glutamine aminohydrolase, and glutathione biosynthesis. Abbreviations: Pyr, pyruvate; αKG, α-ketoglutarate; Oaa, oxaloacetate; amino acids are represented by their three letter codes. [*]Data from the current study; [†]data found in (Griffith 1981); [‡]data found in (Niihara, Zerez et al. 1997); [§]data found in (Maede, Inaba et al. 1983; Fincham, Mason et al. 1987).

CHAPTER 11

Concluding remarks

Over the course of my Ph.D. research, I have developed a range of tools and techniques that make multidimensional NMR a practical option for routine metabolomics studies. The MMCD and BMRB databases (Chapter 3) allow metabolites to be efficiently and unambiguously identified, FMQ by NMR and DETOCS allow metabolites to be accurately quantified (Chapters 4 and 6), and rNMR greatly simplifies data analysis (Chapter 5). Collectively, these tools allow 90% of the NMR-observable metabolites to be rapidly quantified with less than 3% error. I have integrated the new technology into a general research strategy, bioanalytical metabolomics, which is thoroughly documented (Chapter 2) and has been successfully applied to hypothesis-driven research (Chapters 8-10). I have also developed a new device for preparing samples (Appendix II) and explored calibration-independent methods for quantifying metabolites (Chapter 7).

Despite these advances, bioanalytical metabolomics has considerable room for improvement. A significant proportion of signals in every study go unassigned, bioinformatics tools suffer from unacceptable false discovery rates, accurate quantification requires *ad hoc* correction of quantitatively unreliable data, resonance assignments rely on visual inspection, and specialized expertise is required to collect, analyze, and interpret multidimensional data. Although my tools have simplified comprehensive analyses, bioanalytical metabolomics is still more difficult than traditional statistics-based analyses.

The ability to detect, identify, and quantify whatever organic compounds happened to be present in a complex mixture is a unique attribute of NMR. Although MS is more sensitive, NMR provides and unbiased window into biological samples. My ultimate vision for bioanalytical metabolomics is a one-button analysis tool that will accurately quantify the 50 most abundant compounds present in any biological extract. There are no fundamental barriers standing between the present technology and this ultimate goal, just a series of practical problems waiting to be solved.

APPENDIX I

Metabolic analysis of *Arabidopsis thaliana* histidine kinase mutants

This project was conducted in collaboration with Dr. Michael R. Sussman's laboratory (UW-Madison). The Sussman lab is interested in an *Arabidopsis thaliana* histidine kinase (*ATHK1),* which they believe to be involved in osmotic regulation. To study the role of *ATHK1* in osmotic regulation, the Sussman lab prepared several mutants, including *ATHK1* knockout (KO) and *ATHK1* over expressing (OE) plants (Wohlbach, Quirino et al. 2008). Studies conducted by the Sussman lab indicate that KO plants are more sensitive to osmotic challenge than wild type (WS) plants, whereas OE plants are less sensitive to osmotic challenge than WS (Wohlbach, Quirino et al. 2008). Microarray analyses of the mutants identified several differentially regulated genes related to embryogenesis and metabolism. Two of the upregulated genes in OE plants, *P5CS1* and *SUS1,* are involved in the biosynthesis of proline and sucrose, respectively. Proline and sucrose have previously been reported as osmoprotectants in plants (Zhu 2002). These findings suggest that *ATHK1* over expressing (OE) plants are more tolerant of osmotic challenge because they accumulate organic osmolytes. To find direct evidence for this hypothesis, the Sussman lab collaborated with the NMRFAM to measure the concentrations of metabolites present in WS, OE, and KO plants. All of the plant tissues analyzed in this study were provided by Dr. Dana J. Wohlbach, an alumnus of the Sussman laboratory. I conducted all of the sample preparation with the assistance of Eric Swanson, an undergraduate student. Dr. Jessie Q. Bond (UW-Madison) collected the magic angle spinning (MAS) NMR spectra whereas I collected and analyzed all of the solution spectra. Ultimately, this study did not find any consistent metabolic phenotypes across multiple seed batches that could account for the osmotic tolerance of *ATHK1* over expressing plants.

A.I.1. Analysis of oil content in seeds

One of my original projects with the Sussman laboratory was to measure the oil content of *A. thaliana* seeds. Because many of the differentially regulated genes identified in *ATHK1* mutants relate to embryogenesis, we speculated that WS, KO, OE plants would show differential lipid profiles as seeds. To measure lipid content, Dr. Bond and I analyzed whole seeds using magic angle spinning (MAS) NMR. MAS NMR is typically used for NMR analysis of solids. Although oils found in seeds are in a semi-liquid state, MAS is useful because it averages magnetic inhomogeneity in seeds. The technical aspects of this study worked well; MAS NMR spectra of seeds showed sharp signals from oleic acid, linoleic acid, and glycerol (Fig. A.I.1). However, there were no detectable differences in the types or quantities of oils present in the seeds (Fig. A.I 2.).
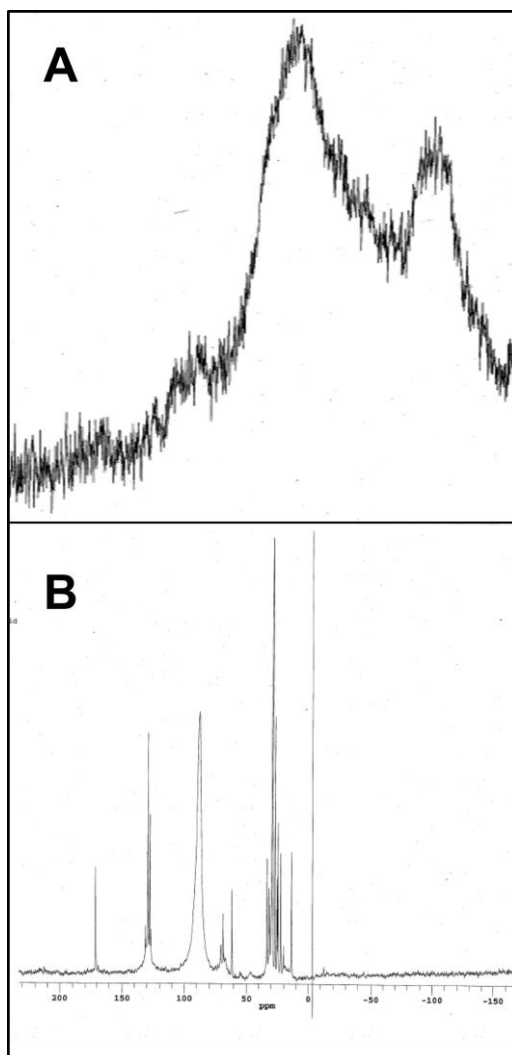
Figure A.I.1. One dimensional ^{1}H NMR spectra of wild type (WS) *A. thaliana* seeds A) without and B) with magic angle spinning (MAS). Chemical shifts of the observed signals matched those of oleic acid, linoleic acid, and glycerol. Proton spectra were collected at 299.5 Mhz in 512 complex data points with a 4kHz spectral width, 128 transits, and a recycle delay of 0.5 s.
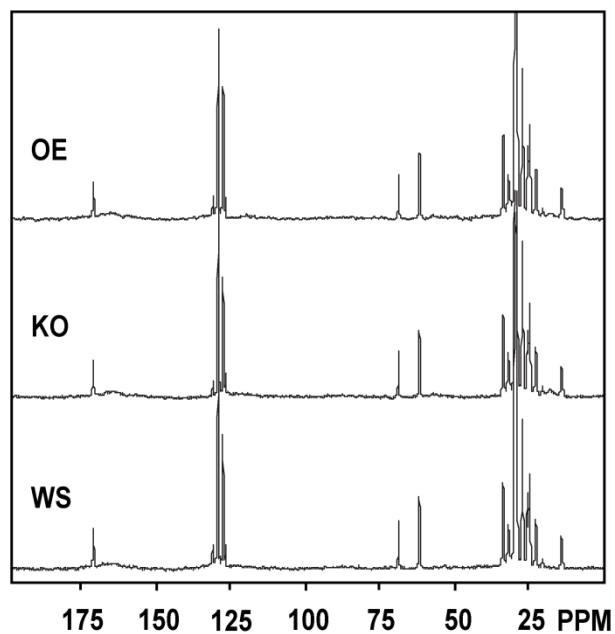
Figure A.I.2. One dimensional ^{13}C MAS NMR spectra of ATHK1 over expressing (OE), ATHK1 knockout (KO), and wild type (WS) seeds. No quantitative differences were observed between the oils (oleic acid, linoleic acid, and glycerol) observed in the seeds. Carbon spectra were collected at 75.3 Mhz in 1600 complex data points with 20 kHz spectral width, 16384 transits, and a recycle delay of 0.5 s.

A.I.2. Time-dependent alterations of metabolites in vegetative tissues

The object of the study was to establish the optimal time point for harvesting tissues in subsequent analyses. Approximately 25 mg of sterilized seeds were sewn in shaker flasks containing 75 mL of liquid MS media (half-strength Murashige and Skoog salts (MS) (Sigma), 2.5 mM MES, 1% sucrose). The seeds were grown with shaking under continuous light. After 7 days growth, 150 mM NaCl (final concentration) or water (control) was added to the flasks, and seeds were grown for another 0, 4, 8, 16, 32 hours. Flasks were harvested individually, dried by blotting on paper-towels, and pooled into samples containing approximately 5 g tissue each. Four replicates for each condition were collected. Samples were flash frozen in liquid nitrogen and stored at -80°C until sample extraction.[4] Metabolite extraction, NMR sample preparation, and metabolite identification were conducted using the methods presented in Chapter 2.

At the time of this study, I had not developed a method for calibrating 2D NMR signals (Chapter 4). Consequently, I was unable to measure molar concentrations of metabolites directly. The procedure I used for quantifying metabolites here is similar to the method presented in Chapter 6. Sample-specific quantitative defects resulting from osmolarity and line shape were normalized by expressing data relative to MES, which is accumulated in plants grown in this buffer (Fan, Colmer et al. 1993). Resonance-specific quantitative problems (e.g. off-resonance effects and differential T_1 relaxation) were normalized by expressing each resonance relative to the mean MES normalized intensity for the corresponding resonances observed in untreated wild type extracts. This quantitative method produced relatively little technical error and showed time-dependent alterations in metabolite levels of osmotically stressed plants (Fig. A.I.3). As expected, proline showed significant time-dependent accumulation (Fig. A.I.4).

---

[4] All plant preparations were conducted by Dr. Wohlbach. The method presented here was adapted from a personal communication from Dr. Wohlbach.
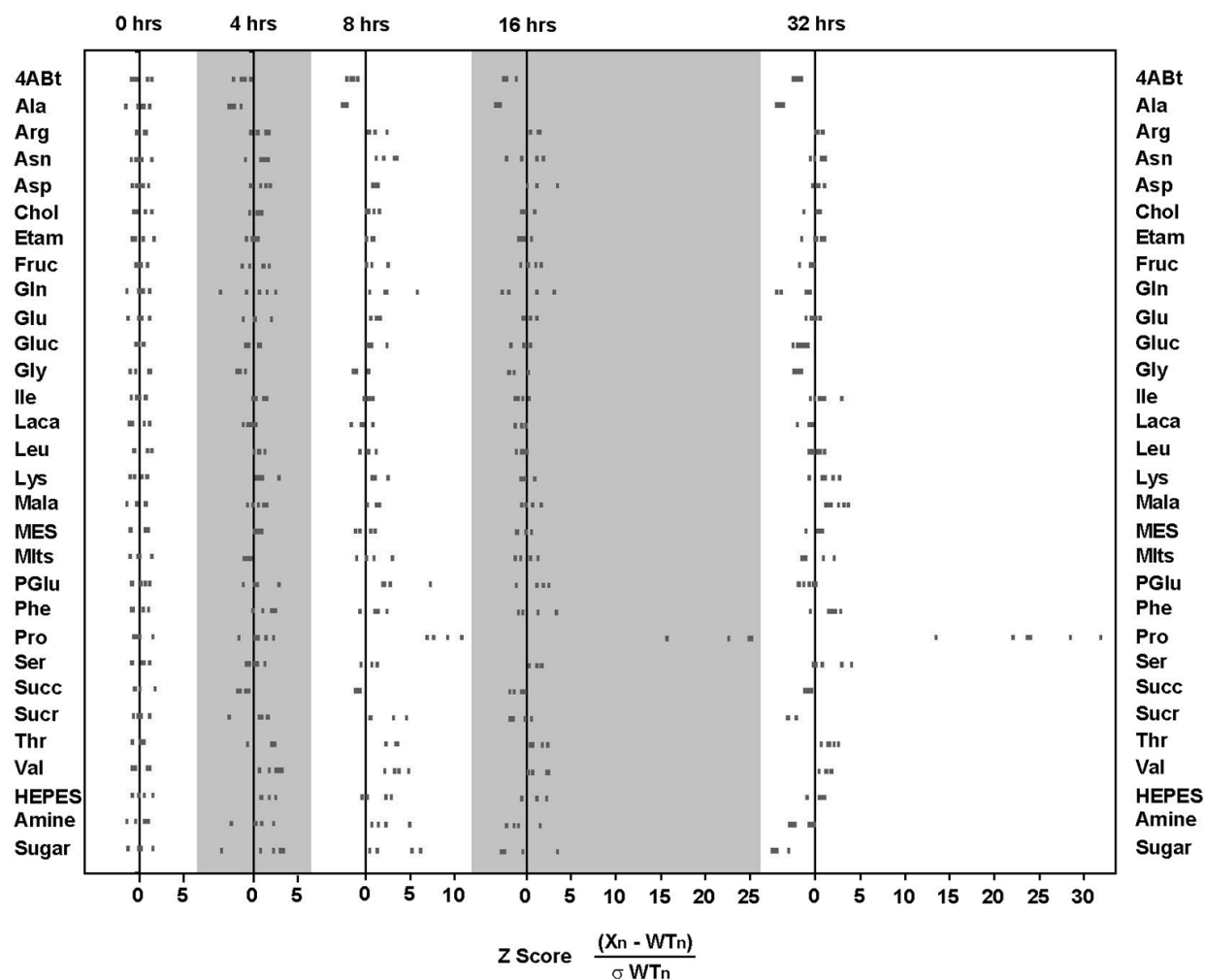
Figure A.I.3. Time-dependent changes of metabolites observed in wild type *A. thaliana* tissue extracts after 0, 4, 8, 16, and 32 h of incubation in 150 mM NaCl. Metabolites were identified and quantified by 2D ^{1}H-^{13}C NMR. All metabolites levels are expressed as *Z* scores relative to the 0 h time point. Abbreviations: 4ABt, 4-aminobutyrate; Ala, alanine; Asn, asparagine; Asp, aspartate; Chol, choline; Etam, ethanolamine; Fruc, fructose; Gln, glutamine; Glu, glutamate; Gluc, glucose; Gly, glycine; Ile, isoleucine; Laca, lactate; Leu, leucine; Lys, lysine; Mala, malate; MES, 2-(N-morpholino)ethanesulfonic acid; Mlts, maltose; PGlu, pyroglutamate; Phe, phenylalanine; Pro, proline; Ser, serine; Succ, succinate; Sucr; sucrose; Thr, threonine; Val, valine; HEPES, (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid); Amine, overlapped signals from multiple amines; Sugar, overlapped signals from multiple sugars.
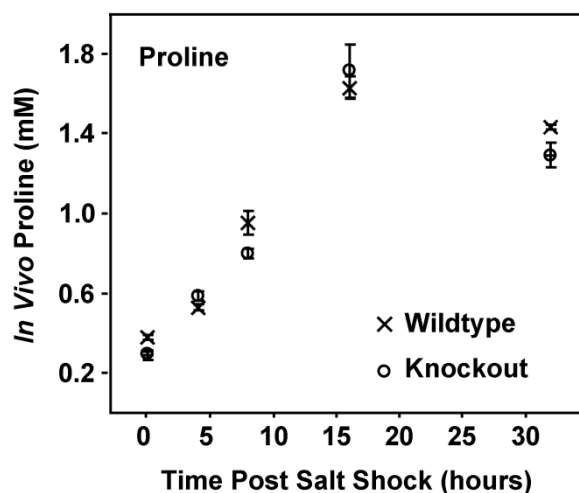
Figure A.I.4. Time-dependent levels of proline observed in wild type and *ATHK1* knockout *A. thaliana* tissue extracts after 0, 4, 8, 16, and 32 h of incubation in 150 mM NaCl. Error bars indicate standard error. In this study, proline was the only metabolite with a large time-dependent metabolic response. No significant differences were detected between wild type and knockout plants. *In vivo* concentrations were estimated using standards collected well after the original data acquisition and were calibrated to MES levels observed in unprocessed *Arabidopsis* homogenates. My current quantitative methods recommend that standards be collected at the same time as the test samples (Chapter 2). Consequently, molar concentrations reported here may be incorrect. Relative levels shown here, however, are accurate.

A.I.3. Metabolic profiles in vegetative tissue of *ATHK1* mutants

Including the data presented in the preceding section (A.I.2), I conducted four independent analyses of metabolites present in the vegetative tissues of salt-challenged *Arabidopsis* mutants. Three of these studies were conducted with liquid-grown plants and one was conducted with plants grown on agar plates. All of the analyses were based on independent seed batches. Ultimately, the data I derived from these studies are inconclusive. Although I observed low technical and biological error within each study, metabolic phenotypes were inconsistent across seed batches. In this section, I present a summary of the 3264 individual metabolite measurements made in the four studies. All of the methods used for these studies are the same as those presented in section A.I.2, with exception that all the data are derived from plants harvested after 16 h of salt stress.
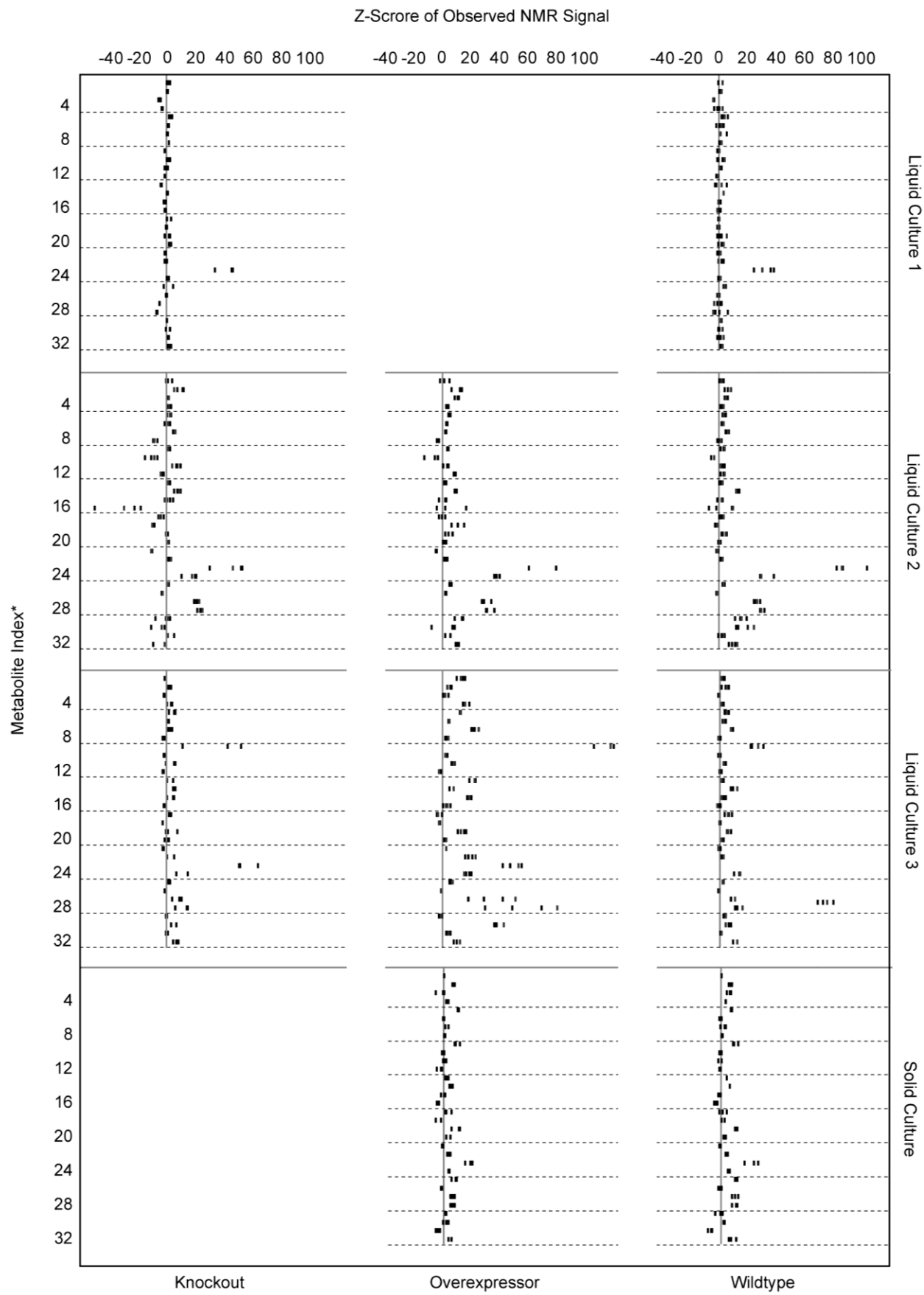
Z-Scrore of Observed NMR Signal

Figure A.I.5. Metabolite levels observed in vegetative tissue of *A. thaliana* after 16 h of incubation in a high salt (150 mM NaCl) solution. The metabolic responses of wildtype (WS), *ATHK1* knockout, and *ATHK1* over expressing plants are shown. These data were compiled from four independent studies conducted with separate batches of seeds. Three studies were conducted with plants grown in liquid culture, and one study was conducted with plants grown on agar plates. Only two of the studies (Liquid culture 1 and 2) evaluated all of genotypes. Metabolite levels are expressed as Z-scores relative to the mean metabolite signals observed untreated (low salt) wildtype cultures. *Metabolites quantified in these studies include: 1) adenosine, 2) agmatine, 3) alanine, 4) overlapped signals from multiple amines, 5) arginine, 6) asparagine, 7) aspartate, 8) choline, 9) citrate, 10) ethanolamine, 11) fructose, 12) 4-aminobutyrate, 13) glutamine, 14) glutamate, 15) glucose, 16) glycine, 17) isoleucine, 18) lactate, 19) maltose, 20) MES, 21) unidentified trimethylamine, 22) pyroglutamate, 23) proline, 24) serine, 25) overlapped methylene signals, 26) succinate, 27) sucrose, 28) overlapped signals from multiple sugars, 29) threonine, 30) unidentified amine, 31) uridine, and 32) valine. Additional metabolites were observed, but the concentrations of these molecules fell below the limit for accurate quantification in some studies.
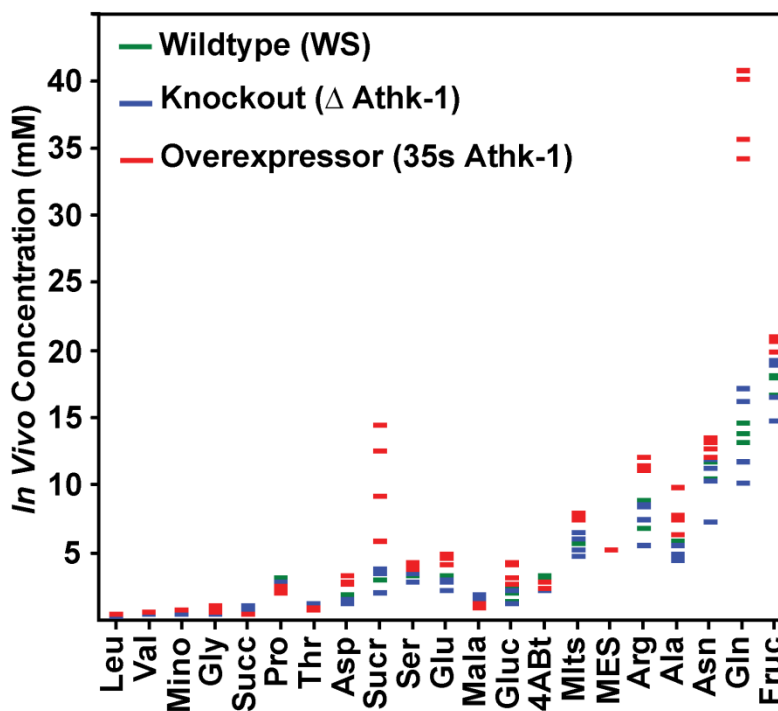
Figure A.I.6. Metabolite concentrations observed in vegetative tissues of wildtype and *ATHK1* mutants after 16 h exposure to 150 mM NaCl. The data presented here are from a single seed batch (Liquid culture 2 from Fig. A.I.5). *In vivo* concentrations were estimated using the methods described in Fig. A.I.4. The clear glutamine phenotype shown in this study was not observed in other seed batches. Abbreviations: Leu, leucine; Val, valine; Mino, myo-inositol; Gly, glycine, Succ, succinate; Pro, proline; Thr, threonine; Asp, aspartate; Sucr, sucrose; Ser, serine; Glu, glutamate; Mala, malate; Gluc, glucose; 4ABt, 4-aminobutyrate; Mlts, maltose, Arg, arginine; Ala, alanine; Asn, asparagine; Gln, glutamine; and Fruc, fructose.
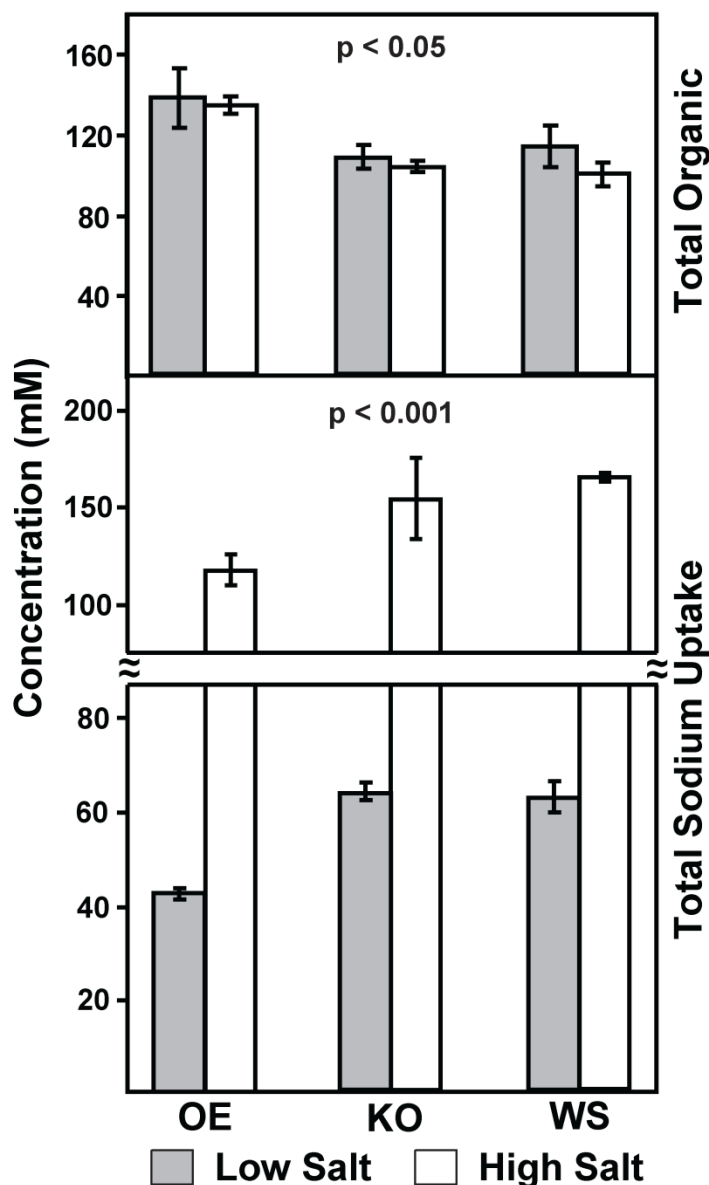
Figure A.I.7. Elemental analysis of vegetative tissues exposed to 16 h of incubation in high salt (150 mM NaCl) medium. Sodium levels were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES) whereas total carbon was measured by NMR (the sum of concentrations presented in Figure A.I.6). No significant differences were observed between *ATHK1* knockout (KO) and a wildtype (WS). However, both sodium and total organic levels differed significantly between *ATHK1* over expressing plants (OE) and wild type.

APPENDIX II

Extraction Contraption: device for isolating metabolites from biological tissues

Adapted from:

Ian A Lewis, Daniel C. Miller, and John L. Markley. Extraction Contraption: Tool for Extracting Metabolites from Biological Tissues. Invention Disclosure Report to the Wisconsin Alumni Research Foundation. (06/06/2009)

Sample preparation is a labor-intensive process that has become the main bottleneck in bioanalytical metabolomics. The extraction contraption is the latest, and most elaborate, of several devices I have constructed to simplify sample preparation. The extraction contraption performs all of the steps required for isolating metabolite from biological tissues. I sketched the original design several years ago, but was unable to build the device because of the complicated machining involved in its fabricating. We hired an engineering student, Daniel C. Miller, in the summer of 2009 to fabricate a prototype. The design has subsequently undergone several updates to improve its function and reduce its cost. A prototype of the updated version has recently been completed, but time constraints have prevented me from adequately testing the finished product. Consequently, I have restricted my discussion here to the details of the device's design and construction.

A.II.1. Abstract

Sample preparation is a critical component of natural products chemistry and metabolomics. Although there are many effective protocols for extracting metabolites from biological tissues, most of the existing tools for preparing samples are poorly suited to studies involving large numbers of samples. We have developed a new tool, the extraction contraption, that simplifies metabolite isolation from biological tissues. The extraction contraption integrates all sample processing steps into a single device, eliminates sample handling, and allows six samples to be processed simultaneously. The extraction contraption relies on three new inventions: 1) valves that allow solvent to be introduced into, and removed from, a ball milling chamber, 2) a pressure-driven fluid delivery system for transferring homogenate from chamber to chamber, and 3) a cluster-based device assembly that allows six samples to be processed simultaneously. In this report, we describe the details of the extraction contraption's design and fabrication.

A.II.2. Introduction

Metabolomics studies often require metabolites to be reproducibly extracted from hundreds of biological samples (Fiehn, Kopka et al. 2000). Although there are many effective protocols available for isolating metabolites from biological tissues, most of the existing extraction tools (e.g. ball mills) were not designed for studies involving large numbers of samples. Typical sample preparation protocols involve homogenization by mortar and pestle or ball milling, aqueous metabolite extraction, ultrafiltration (UF), and lyophilized (Lewis, Shortreed et al. 2010). Each of these steps requires manual transfer of material from one vessel to another. The labor intensive nature of this process makes it difficult to standardize extraction protocol between laboratories. Furthermore, the low-throughput nature of the homogenization technology limits the number of samples that can be processed. Consequently, sample preparation has become a major bottleneck in metabolomics.

We have recently conceived of a tool, the "extraction contraption", which streamlines the sample preparation processes and allows six samples to be prepared in parallel. Our concept is to integrate existing ball milling technology with a pressure-driven fluid delivery system that allows samples to be homogenized, extracted with a wide variety of solvents, filtered, and lyophilized in an integrated extraction vessel. Parallelizing sample preparation and eliminating sample handling has several benefits. The time required to complete an extraction is currently proportional to the number of samples. Parallelizing extractions significantly reduces the requisite time of any given extraction. In addition, integrating homogenization with all of the subsequent extraction steps and eliminating sample handling removes these sources of technical error. Lastly, automated extractions allow independent research labs to replicate extraction procedures more precisely than can be currently achieved via traditional methods.

A.II.3. Overall design

The extraction contraption is composed of six integrated components: (1) milling chamber, (2) vibrational shaker, (3) solvent reservoir, (4) homogenization platform, (5) filtration chamber, and (6) filtration platform. These components allow metabolites to be isolated from unprocessed biological tissues without any manual sample handling. The entire system is run by compressed air at 75 psi (Fig. A.II.1). Biological tissues are placed in the milling chambers, and are homogenized using the vibrational shaker. A pre-allocated volume of extraction solvent is injected under pressure into the milling chamber from the solvent reservoir. The milling chamber supports both wet and dry homogenization with a variety of solvents. Homogenate suspensions are transferred via the fluid delivery system to a filtration chamber, containing a guard column and an ultrafiltration (UF) membrane. The filtrate is collected in lyophilization vial connected to the filtration platform. The extraction contraption allows six samples to be processed simultaneously and is constructed out of autoclavable materials (Fig. A.II.2).
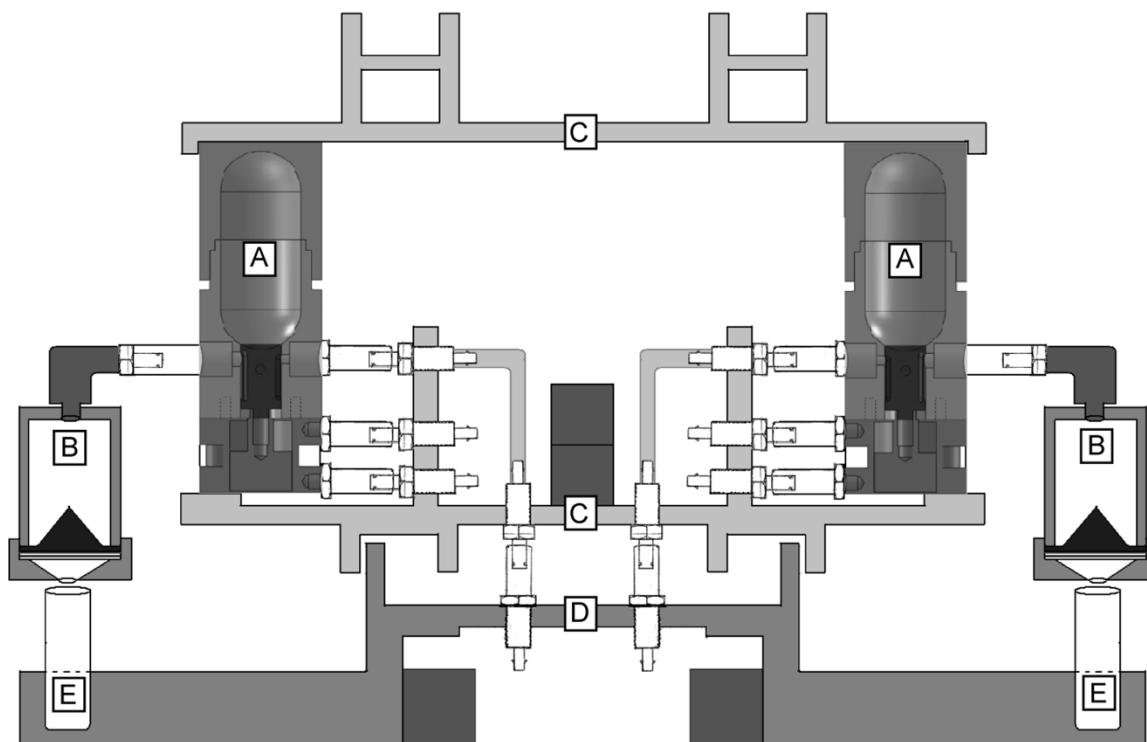
Figure A.II.8.1. Schematic representation of the extraction contraption. The system isolates metabolites from biological tissues. The device is composed of the following six major components: A) milling chambers, B) filtration chambers, C) homogenization platform, D) filtration platform, E) sample collection vilas, and F) a vibrational paint shaker (not shown). The extraction contraption is run from a single compressed air source and allows up to six samples to be homogenized, extracted, and filtered simultaneously.
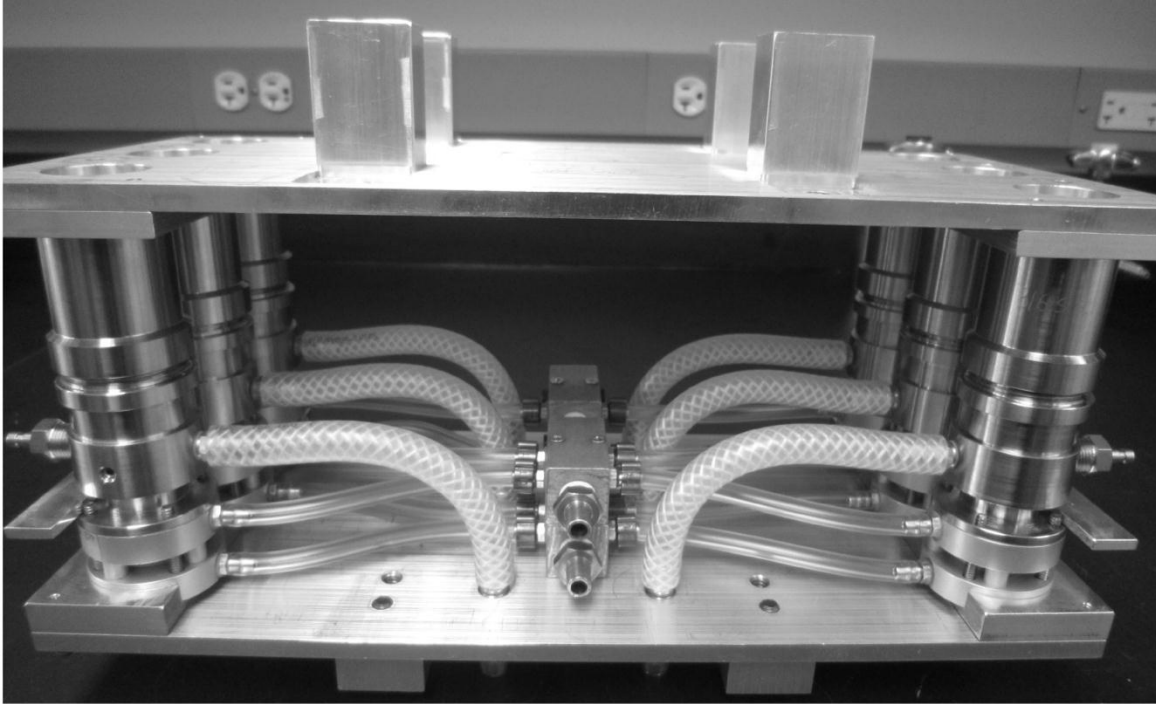
Figure A.II.2. Homogenization platform and milling chambers. The extraction contraption is built from aluminum, stainless steel, and other autoclavable materials. The modular design allows the system to be easily disassembled for cleaning and sterilization and minimizes the number of manual steps required for sample preparation.

A.II.4. Milling chamber

The milling chamber used in the extraction contraption was modeled from a conventional ball mill (Retsch MM400 Mixer Mill). Opposing ends of the milling chamber have a radial contour to accommodate a grinding ball. The milling chamber was designed to be as light as possible to minimize stress on the shaking apparatus. The mass of the milling chamber was reduced by 50% via additional machining to the exterior of the chamber.

We have experimented with a variety of materials for the construction of the milling chamber. These materials include stainless steel types 303, 316, and tempered 440C. Tempered 440C is the hardest of these materials (Atlas 2008), but is impractical in this system

due to cost considerations (~$1500 per chamber). The majority of fabrication costs results from machining time, which is directly proportional to the hardness of the material. For the extraction contraption, we selected a mid grade stainless steel (type 303) for the milling chamber and softer grade (type 316) for the grinding ball. A softer grinding ball minimizes deformation of the milling surface.

Eliminating sample handling requires a method for introducing and removing solvents from the milling chamber. We devised a unique method for introducing solvents (Fig. A.II.3a). A pneumatically operated plunger at the base of the milling chamber controls a valve, which allows fluid to be added or removed from the chamber (Fig. A.II.8.3b). When the plunger is raised, it exposes inlet and outlet ports (Fig. A.II.4a); when lowered, the plunger seals the milling chamber (Fig. A.II.4b). Adequate sealing is ensured by two chemical resistant Viton$^{®}$ O-rings. The orientation of the milling chamber controls fluid delivery into and out of the system. When the milling chamber is inverted, pressurized solvent is injected through an inlet port and air is displaced through an outlet port. When the milling chamber is upright, pressurized air drives homogenate into a filtration chamber. The outlet channel is of sufficient diameter to accommodate viscous homogenate. The plungers in each of the six milling chambers are operated simultaneously by pneumatic pancake cylinders controlled by a single toggle switch.
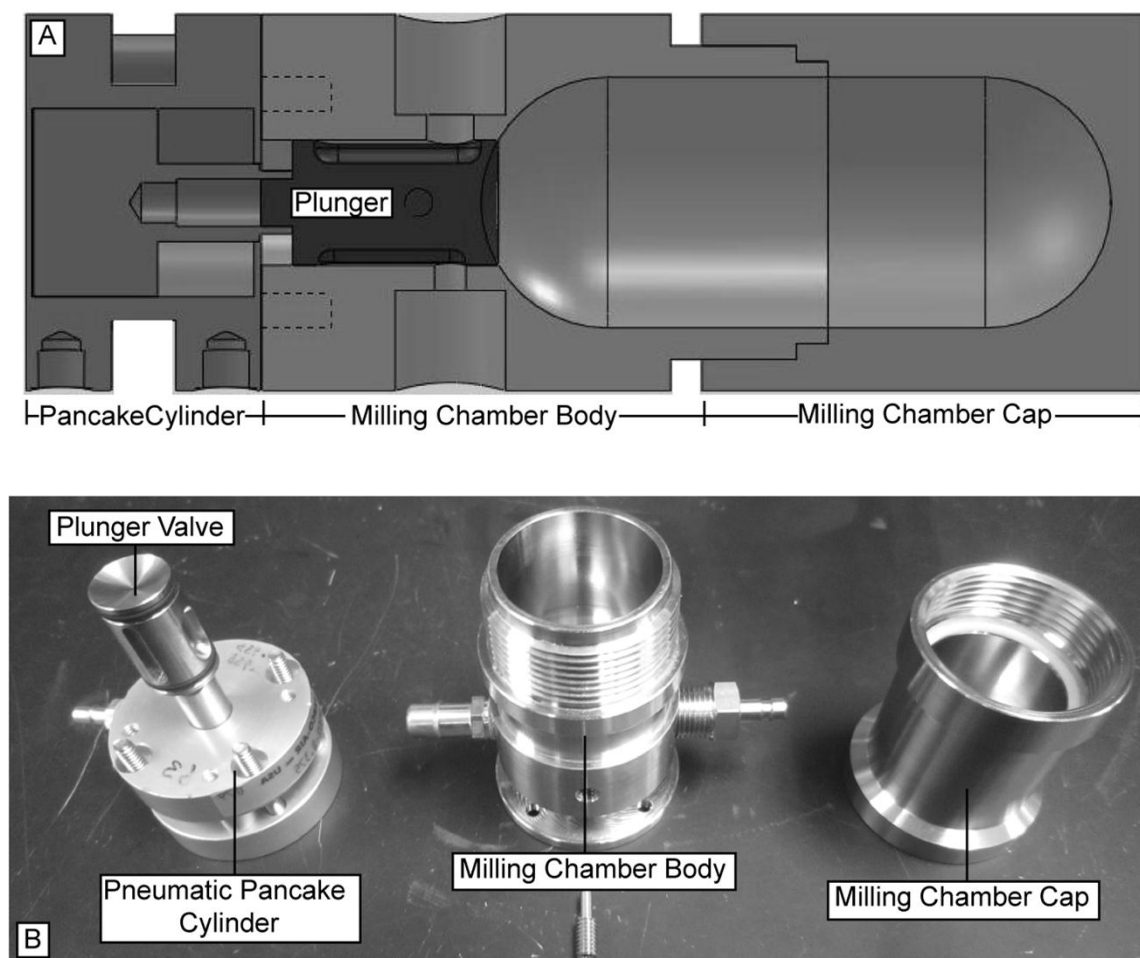
Figure A.II.3. A) Schematic cross section and B) photo of the milling chamber components. The milling chamber supports ball milling of materials and extraction of metabolites with a variety of solvents. The plunger allows pressurized solvent to be introduced and removed from the milling chamber via inlet and outlet ports. The plunger is actuated by a pneumatic pancake cylinder. The milling chamber body and plunger are constructed from stainless steel type 303.
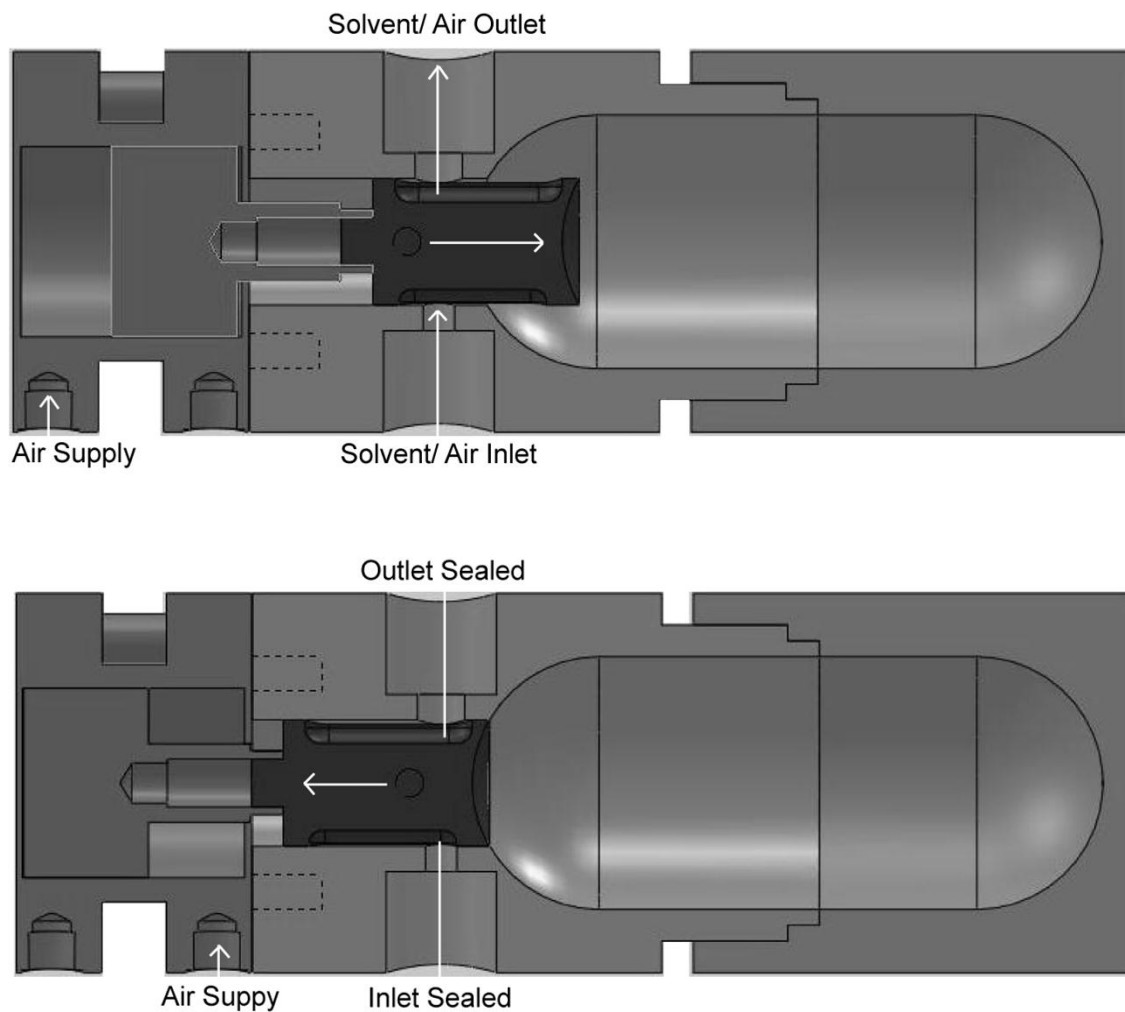
Figure A.II.4. Mechanism for introducing and removing solvents from the milling chamber. A pneumatically operated pancake cylinder allows the plunger to be raised and lowered, which (upper) opens the inlet/outlet ports and (lower) seals the chamber, respectively. The orientation of the milling chamber controls fluid flow into and out of the chamber. When the milling chamber is oriented such that the cap is down (inverted), solvent flows from the inlet into the milling chamber and air is vented through the outlet; when upright, compressed air forces homogenate through the outlet into the filtration chamber.

A.II.5. Vibrational Shaker

Commercial ball milling systems are equipped with vibrational shakers capable of 30 Hz oscillation frequency (e.g. Retsch MM400). Although these systems are highly effective, they are also very expensive. The retail price of the Retsch MM400 vibrational shaker is $7,500. To minimize overall cost, we selected a pneumatic paint shaker ($150; Astro Pneumatic AST4550), which has a 23.3 Hz maximum oscillation frequency. We have partially compensated for the shaker's lower vibrational frequency by increasing the milling chamber's rotational inertia. This is achieved via clamps that extend the milling chambers 5.25" from the center of rotation (Fig. A.II.5). The original paint shaker clamps were modified to accommodate the homogenization platform.

We have tested the homogenization system with a variety of biological tissues, including corn kernels, coffee beans, wildflower seeds, alfalfa sprouts, and switchgrass stem. Qualitative evaluation of the homogenization platform (Fig. A.II.6) indicated that the extraction contraption is comparable to the Retsch MM400 for non-fibrous tissues. Switchgrass stem samples, however, did not homogenize well at the lower oscillation frequency. Future versions of the extraction contraption may implement higher frequency vibrational shakers for use with fibrous samples. However, our current design is favorable because of its lower cost and because the entire system can be run from a single compressed air source.
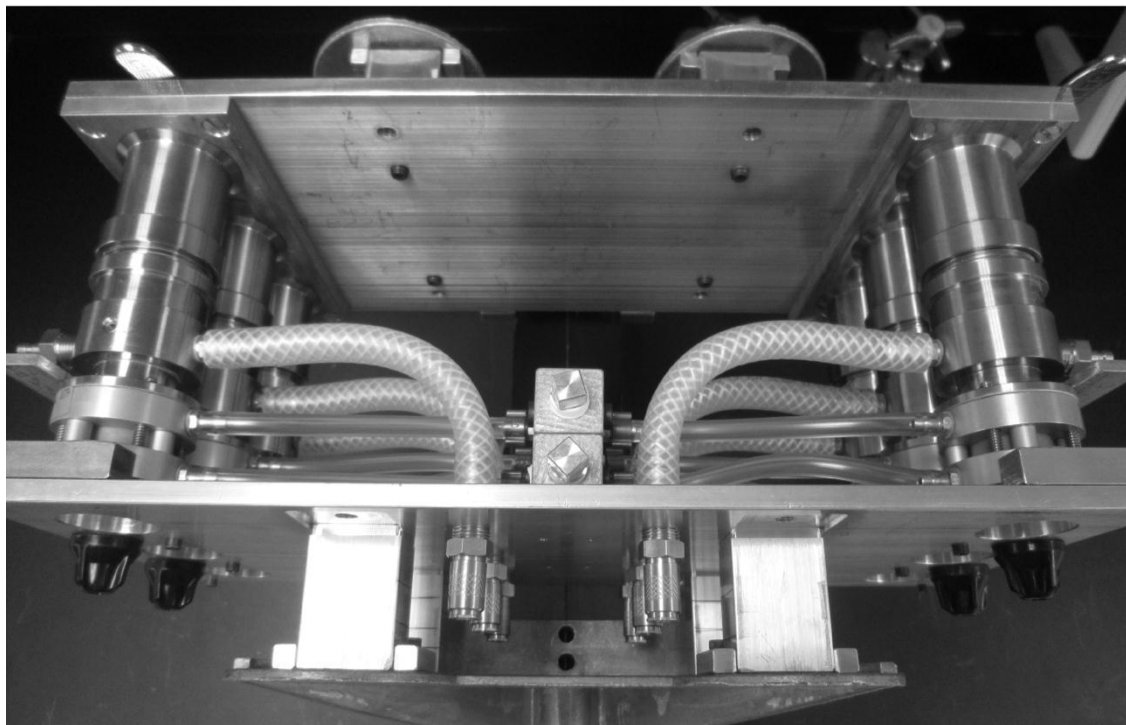
Figure A.II.5. Homogenization platform and vibrational shaker. The homogenization platform extends the milling chambers 5.25" from the center of rotation. The platform is secured to a pneumatic paint shaker via a modified clamp. The paint shaker used for the extraction contraption is capable of a 23.3 Hz oscillatory frequency.

Corn Kernel   Wildflower Seed   Coffee Bean   Switchgrass Stem
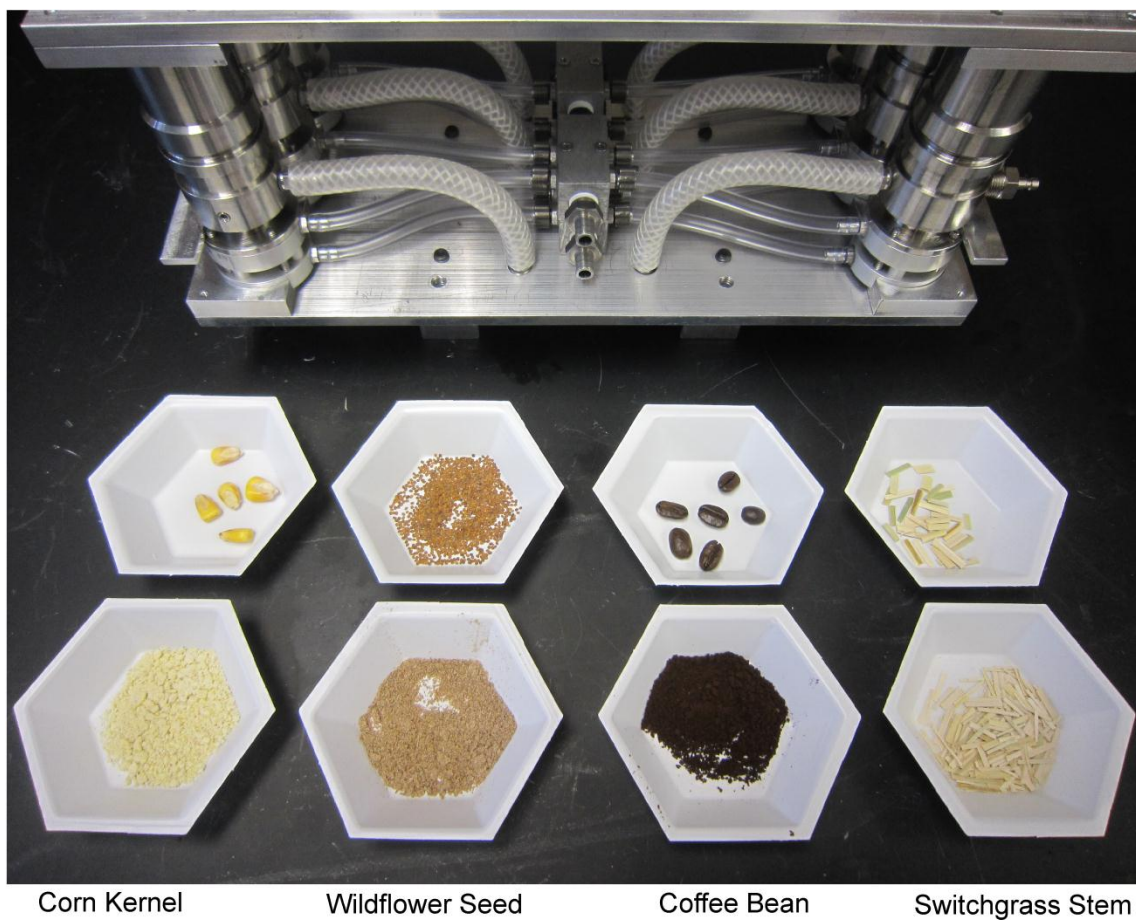
Figure A.II.6. Trial homogenization of various tissues using the extraction contraption. All tissues were ball milled for 30 seconds in a 23.3 Hz vibrational paint shaker. Homogenates were of comparable granularity to those produced from a commercial system (Retch MM400) for all tissues except switchgrass stem.

A.II.6. Solvent reservoir

The solvent reservoir allows a fixed volume of pressurized solvent to be introduced into the milling chamber. It is attached to both the air source and milling chamber via stainless steel quick disconnect couplings. For wet milling, solvent from the reservoir is transferred to the milling chamber prior to homogenization; for dry milling, the solvent is introduced after milling and is used to transfer the homogenate to the filter chamber. The extraction contraption uses six individual reservoirs, each corresponding to a specific milling chamber. Reservoirs were designed to accommodate a volume solvent greater than that required for extraction. This excess capacity can be used to wash the extraction and filtration chambers.

### A.II.7. Homogenization platform

The homogenization platform attaches the milling chambers to the vibrational shaker and serves as fluid delivery system. The platform was constructed from aluminum to minimize weight. Aluminum is also desirable because it is relatively inexpensive and is easy to machine. The body of each milling chamber is fixed to the homogenization platform, where it is supplied with compressed air. The inlet to each milling chamber is routed to stainless steel quick disconnect couplings (Beswick Engineering), which extend through the plates. When upright, these couplings allow the homogenization platform to couple with the filtration platform (Fig. A.II.1). When inverted, these couplings allow the solvent reservoirs to connect to the milling chambers.

A.II.8. Filtration Chamber

The filtration chamber is compatible with solvents ranging from organics to dilute acids or alkaline solutions. The interior of the filtration chamber is lined with a chemical resistant Pyrex® sleeve sealed with Viton® gaskets. The exterior of the filtration chamber was fabricated from

stainless steel type 316. Windows were milled into the exterior to allow visual inspection of the filtration progress (Fig. A.II.7). The entire chamber was designed to safely accommodate the optimal pressure for ultrafiltration (~75 psi). Although the chamber was designed around UF membrane requirements, it will accommodate filters appropriate for organic solvents. One design complication pertaining to UF membranes is that they do not allow air to be passed through them. In order to transfer homogenate into the chamber without a pressure relief vent, the volume of this chamber must be large. We found that a 60 mL capacity is sufficient for transferring 20 mL between the milling and filtration chambers at 70 psi. Because of the large diameter of the UF membrane (44.5 mm), a stainless steel mesh disk was necessary to prevent the filter from rupturing. The UF membrane and mesh disk are sealed with Viton[®] gaskets. We have also designed chambers with smaller volumes (15 mL) for use with homogenates that do not require ultrafiltration (eg. organics). The filtration chamber is attached to the milling chamber via stainless steel quick disconnect couplings equipped with internal Viton[®] seals (Beswick Engineering). The poppet was removed from these couplings in order to minimize fluid resistance between the two chambers.

Figure A.II.7. Filtration chamber for the extraction contraption. Compressed air drives homogenates from the milling chamber to the filter chamber. Extracts are passed through an ultrafiltration (UF) membrane and the filtrates are collected in lyophilization vials (not shown). The filter chambers are lined with a chemical resistant Pyrex® sleeve and windows milled into the filter chamber allow the filtration progress to be observed. The entire system was designed to safely operate at the optimum UF pressure (75 psi).

A.II.9. Filtration platform

The construction of the filtration platform is similar to the homogenization platform. The body is constructed from aluminum and attaches to the homogenization platform via stainless steel quick disconnects. Our design allows the two platforms to be connected in one step. Once coupled, a toggle switch routes compressed air to the homogenization platform via an air manifold (Fig. A.II.1). When activated, air forces homogenate from the milling chamber to the filtration chamber and the pressurized air is then used to drive filtration. Filtrate is collected in lyophilization vials (50 mL Falcon tubes) that are supported by an integrated rack.

BIBLIOGRAPHY

Aharoni, A., C. H. Ric de Vos, et al. (2002). "Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry." OMICS **6**(3): 217-234.

Annesley, T. M. (2003). "Ion suppression in mass spectrometry." Clin Chem **49**(7): 1041-1044.

Atlas Specialty Metals (2009). "Grade Data Sheet." www.atlasmetals.com.au. 2009.

Bax, A. and D. G. Davis (1985). "Homonuclear Hartmann-Hahn magnetization transfer: new one- and two-dimensional NMR methods for structure determination and spectral assignment MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy." Journal of Magnetic Resonance **65**: 355-360.

Bax, A. and D. G. Davis (1985). "Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy." Journal of Magnetic Resonance **65**(2): 355-360.

Benabe, J. E., L. A. Echegoyen, et al. (1987). "Mechanism of Inhibition of Glycolysis by Vanadate." Journal of Biological Chemistry **262**(20): 9555-9560.

Berry, K. A. and R. C. Murphy (2005). "Analysis of cell membrane aminophospholipids as isotope-tagged derivatives." J Lipid Res **46**(5): 1038-1046.

Berthon, H. A., W. A. Bubb, et al. (1993). "13C n.m.r. isotopomer and computer-simulation studies of the non-oxidative pentose phosphate pathway of human erythrocytes." Biochemical Journal **296 (Pt 2)**: 379-387.

Beynon, R. J. and J. M. Pratt (2005). "Metabolic labeling of proteins for proteomics." Mol Cell Proteomics **4**(7): 857-872.

Birkemeyer, C., A. Luedemann, et al. (2005). "Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling." Trends in Biotechnology **23**(1): 28-33.

Braunschweiler, L. and R. R. Ernst (1983). "Coherence Transfer by Isotropic Mixing - Application to Proton Correlation Spectroscopy." Journal of Magnetic Resonance **53**(3): 521-528.

Braunschweiler, L. and R. R. Ernst (1983). "Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy." Journal of Magnetic Resonance **53**: 521-528.

Brown, L. M., A. Pais, et al. (1995). Twentieth Century Physics

Burton, I. W., M. A. Quilliam, et al. (2005). "Quantitative H-1 NMR with external standards: Use in preparation of calibration solutions for algal toxins and other natural products." Analytical Chemistry **77**(10): 3123-3131.

Campanella, M. E., H. Chu, et al. (2005). "Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane." Proc Natl Acad Sci U S A **102**(7): 2402-2407.

Campanella, M. E., H. Chu, et al. (2008). "Characterization of glycolytic enzyme interactions with murine erythrocyte membranes in wild-type and membrane protein knockout mice." Blood **112**(9): 3900-3906.

Carvalho, R. A., F. M. H. Jeffrey, et al. (1998). "C-13 isotopomer analysis of glutamate by heteronuclear multiple quantum coherence total correlation spectroscopy (HMQC-TOCSY)." Febs Letters **440**(3): 382-386.

Cayley, S. and M. T. Record (2003). "Roles of cytoplasmic osmolytes, water, and crowding in the response of Escherichia coli to osmotic stress: Biophysical basis of osmoprotection by glycine betaine." Biochemistry **42**(43): 12596-12609.

Caytan, E., G. S. Remaud, et al. (2007). "Precise and accurate quantitative (13)C NMR with reduced experimental time." Talanta **71**(3): 1016-1021.

Chu, H., A. Breite, et al. (2008). "Characterization of the deoxyhemoglobin binding site on human erythrocyte band 3: implications for O2 regulation of erythrocyte properties." Blood **111**(2): 932-938.

Climent, F., R. Bartrons, et al. (1981). "Effect of Vanadate on Phosphoryl Transfer Enzymes Involved in Glucose-Metabolism." Biochemical and Biophysical Research Communications **101**(2): 570-576.

Cloarec, O., M. E. Dumas, et al. (2005). "Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets." Analytical Chemistry **77**(5): 1282-1289.

Constantopoulos, T. L., G. S. Jackson, et al. (1999). "Effects of salt concentration on analyte response using electrospray ionization mass spectrometry." J Am Soc Mass Spectrom **10**(7): 625-634.

Cui, Q., I. A. Lewis, et al. (2008). "Metabolite identification via the Madison Metabolomics Consortium Database." Nat Biotechnol **26**(2): 162-164.

Dalluge, J. J., S. Smith, et al. (2004). "Potential of fermentation profiling via rapid measurement of amino acid metabolism by liquid chromatography-tandem mass spectrometry." J Chromatogr A **1043**(1): 3-7.

De Rosa, M. C., C. Carelli Alinovi, et al. (2008). "Allosteric properties of hemoglobin and the plasma membrane of the erythrocyte: new insights in gas transport and metabolic modulation." IUBMB Life **60**(2): 87-93.

Delaglio, F., S. Grzesiek, et al. (1995). "NMRPIPE - A Multidimensional Spectral Processing System Based on UNIX Pipes." Journal of Biomolecular NMR **6**(3): 277-293.

Delgado, T. C., M. M. Castro, et al. (2004). "Quantitation of erythrocyte pentose pathway flux with [2-13C]glucose and 1H NMR analysis of the lactate methyl signal." Magn Reson Med **51**(6): 1283-1286.

Doreleijers, J. F., S. Mading, et al. (2003). "BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank." J Biomol NMR **26**(2): 139-146.

Duhm, J. (1975). "Glycolysis in Human Erythrocytes Containing Elevated Concentrations of 2,3-P2-Glycerate." Biochimica et Biophysica Acta **385**(1): 68-80.

Dumas, M. E., E. C. Maibaum, et al. (2006). "Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study." Anal Chem **78**(7): 2199-2208.

Ellory, J. C., R. L. Preston, et al. (1983). "Transport of amino acids for glutathione biosynthesis in human and dog red cells." Biomedica biochimica acta **42**(11-12): S48-52.

Evilia, R. F. (2001). "Quantitative NMR spectroscopy." Analytical Letters **34**(13): 2227-2236.

Fan, T. W. (1996). "Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures " Progress in Nuclear Magnetic Resonance Spectroscopy **28**(2): 161-219.

Fan, T. W., T. D. Colmer, et al. (1993). "Determination of metabolites by 1H NMR and GC: analysis for organic osmolytes in crude tissue extracts." Anal Biochem **214**(1): 260-271.

Fan, T. W., A. N. Lane, et al. (2001). "Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS." Phytochemistry **57**(2): 209-221.

Fan, T. W. M. and A. N. Lane (2008). "Structure-based profiling of metabolites and isotopomers by NMR." Progress in Nuclear Magnetic Resonance Spectroscopy **52**(2-3): 69-117.

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

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

Fincham, D. A., D. K. Mason, et al. (1987). "Heterogeneity of amino acid transport in horse erythrocytes: a detailed kinetic analysis of inherited transport variation." The Journal of physiology **389**(Journal Article): 385-409.

Fischer, R. B. and D. G. Peters (1968). Quantitative Chemical Analysis. Philadelphia, W.B. Saunders Co.

Gasko, O. and D. Danon (1972). "The Metabolism of Maturing Reticulocytes." British journal of haematology **23**(5): 525-533.

Ghosh, A. K. and H. A. Sloviter (1973). "Glycolysis and Pasteur Effect in Rat Reticulocytes." Journal of Biological Chemistry **248**(9): 3035-3040.

Ghosh, A. K. and H. A. Sloviter (1973). "Glycolysis and pasteur effect in rat reticulocytes." J Biol Chem **248**(9): 3035-3040.

Griffith, O. W. (1981). "Glutathione turnover in human erythrocytes. Inhibition by buthionine sulfoximine and incorporation of glycine by exchange." The Journal of biological chemistry **256**(10): 4900-4904.

Guo, K., C. Ji, et al. (2007). "Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples." Anal Chem **79**(22): 8631-8638.

Hall, S. R. (1991). "The STAR File: A New Format for Electronic Data Transfer and Archiving." Journal of Chemical Information and Computing Sciences **31**: 326-333.

Hall, S. R. and A. P. F. Cook (1995). "STAR dictionary definition language: initial specification." Journal of Chemical Information and Computing Sciences **35**: 819-825.

Hall, S. R. and N. Spadaccini (1994). "The STAR File: Detailed Specifications." Journal of Chemical Information and Computing Sciences **34**: 505-508.

Hamasaki, N., T. Asakura, et al. (1970). "Effect of Oxygen Tension on Glycolysis in Human Erythrocytes." Journal of Biochemistry **68**(2): 157-&.

Harrison, M. L., P. Rathinavelu, et al. (1991). "Role of band 3 tyrosine phosphorylation in the regulation of erythrocyte glycolysis." J Biol Chem **266**(7): 4106-4111.

Harvey, C. M. and J. Clive Ellory (1989). Identification of amino acid transporters in the red blood cell. Methods in Enzymology. F. Sidney and F. Becca, Academic Press. **Volume 173:** 122-160.

Hatch, M. D. and H. W. Heldt (1985). "Synthesis, storage, and stability of [4-14C]oxaloacetic acid." Analytical Biochemistry **145**(2): 393-397.

Hegeman, A. D., C. F. Schulte, et al. (2007). "Stable isotope assisted assignment of elemental compositions for metabolomics." Analytical Chemistry **79**(18): 6912-6921.

Heikkinen, S., M. M. Toikka, et al. (2003). "Quantitative 2D HSQC (Q-HSQC) via suppression of J-dependence of polarization transfer in NMR spectroscopy: application to wood lignin." J Am Chem Soc **125**(14): 4362-4367.

Holmes, E., O. Cloarec, et al. (2006). "Probing latent biomarker signatures and in vivo pathway activity in experimental disease states via statistical total correlation spectroscopy (STOCSY) of biofluids: Application to HgCl2 toxicity." Journal of Proteome Research **5**(6): 1313-1320.

Hurlbert, S. H. (1984). "Pseudoreplication and the design of ecological field experiments." Ecological Monographs **54**(2): 187-211.

Hyberts, S. G., G. J. Heffron, et al. (2007). "Ultrahigh-resolution (1)H-(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction." J Am Chem Soc **129**(16): 5108-5116.

Jenkins, J. D., F. J. Kezdy, et al. (1985). "Mode of interaction of phosphofructokinase with the erythrocyte membrane." Journal of Biological Chemistry **260**(19): 10426-10433.

Jenkins, J. D., F. J. Kezdy, et al. (1985). "Mode of interaction of phosphofructokinase with the erythrocyte membrane." J Biol Chem **260**(19): 10426-10433.

Jensen, F. B. (2004). "Red blood cell pH, the Bohr effect, and other oxygenation-linked phenomena in blood O2 and CO2 transport." Acta Physiologica Scandinavica **182**(3): 215-227.

Johnson, R. M. (1994). "Membrane Stress Increases Cation Permeability in Red-Cells." Biophysical Journal **67**(5): 1876-1881.

Katz, J. and H. G. Wood (1960). "The use of glucose-C14 for the evaluation of the pathways of glucose metabolism." Journal of Biological Chemistry **235**: 2165-2177.

Kikuchi, J., K. Shinozaki, et al. (2004). "Stable isotope labeling of Arabidopsis thaliana for an NMR-based metabolomics approach." Plant Cell Physiol **45**(8): 1099-1104.

Kim, H. D., M. J. Koury, et al. (1991). "Metabolic Adaptation during Erythropoietin-Mediated Terminal Differentiation of Mouse Erythroid-Cells." Blood **77**(2): 387-392.

Kind, T. and O. Fiehn (2006). "Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm." BMC Bioinformatics **7**: 234-244.

Kind, T. and O. Fiehn (2007). "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry." BMC Bioinformatics **8**: 105.

King, G. F. and P. W. Kuchel (1985). "Assimilation of alpha-glutamyl-peptides by human erythrocytes. A possible means of glutamate supply for glutathione synthesis." Biochem J **227**(3): 833-842.

Kinoshita, A., K. Tsukada, et al. (2007). "Roles of hemoglobin Allostery in hypoxia-induced metabolic alterations in erythrocytes: simulation and its verification by metabolome analysis." J Biol Chem **282**(14): 10731-10741.

Koek, M. M., B. Muilwijk, et al. (2006). "Microbial metabolomics with gas chromatography/mass spectrometry." Analytical Chemistry **78**(4): 1272-1281.

Kogler, H., O. W. Sorensen, et al. (1983). "Low-Pass J-Filters - Suppression of Neighbor Peaks in Heteronuclear Relayed Correlation Spectra." Journal of Magnetic Resonance **55**(1): 157-163.

Kosaka, H., I. Tyuma, et al. (1983). "Mechanism of autocatalytic oxidation of oxyhemoglobin by nitrite." Biomedica Biochimica Acta **42**(11-12): S144-148.

Koskela, H., I. Kilpelainen, et al. (2005). "Some aspects of quantitative 2D NMR." J Magn Reson **174**(2): 237-244.

Kuchel, P. W., G. F. King, et al. (1987). "No evidence of high capacity alpha-glutamyl-dipeptide transport into human erythrocytes." The Biochemical journal **242**(1): 311-312.

Lafaye, A., J. Labarre, et al. (2005). "Liquid chromatography-mass spectrometry and 15N metabolic labeling for quantitative metabolic profiling." Analytical Chemistry **77**(7): 2026-2033.

Lamos, S. M., M. R. Shortreed, et al. (2007). "Relative quantification of carboxylic acid metabolites by liquid chromatography-mass spectrometry using isotopic variants of cholamine." Anal Chem **79**(14): 5143-5149.

Lane, A. N. and T. W. M. Fan (2007). "Quantification and identification of isotopomer distributions of metabolites in crude cell extracts using H-1 TOCSY." Metabolomics **3**(2): 79-86.

LePetitThevenin, J., O. Nobili, et al. (1997). "Ethanol stimulates lipid biosynthesis in the rat reticulocyte by activating glycerol kinase." Journal of Nutritional Biochemistry **8**(6): 312-315.

Lewis, B. A., S. Cayley, et al. (1990). "Natural Abundance N-14 and C-13 Nmr of Glycine Betaine and Trehalose as Probes of the Cytoplasm of Escherichia-Coli-K12." Journal of Magnetic Resonance **90**(3): 612-617.

Lewis, I. A., M. E. Campanella, et al. (2009). "Role of band 3 in regulating metabolic flux of red blood cells." Proc Natl Acad Sci U S A **106**(44): 18515-18520.

Lewis, I. A., R. H. Karsten, et al. (2010). "NMR Pulse Sequence for Measuring Carbon-13 Isotopic Enrichment of Metabolites in Complex Solutions." Submitted.

Lewis, I. A., R. H. Karsten, et al. (2010). "Method for Controlling Differential T1 Relaxation in NMR Analyses of Metabolites in Complex Solutions " Submitted.

Lewis, I. A., S. C. Schommer, et al. (2007). "Method for determining molar concentrations of metabolites in complex solutions from two-dimensional ^{1}H-^{13}C NMR spectra." Anal Chem **79**(24): 9385-9390.

Lewis, I. A., S. C. Schommer, et al. (2009). "rNMR: open source software for identifying and quantifying metabolites in NMR spectra." Magnetic Resonance in Chemistry **47**: S123-S126.

Lewis, I. A., M. R. Shortreed, et al. (2010). Novel NMR and MS approaches to metabolomics. Handbook of Metabolomics. T. Fan and A. Lane, in press.

Lindon, J. C., E. Holmes, et al. (2001). "Pattern recognition methods and applications in biomedical magnetic resonance." Progress in Nuclear Magnetic Resonance Spectroscopy **39**(1): 1-40.

Lindon, J. C., E. Holmes, et al. (2007). "Metabonomics in pharmaceutical R & D." FEBS Journal **274**(5): 1140-1151.

Lindon, J. C., J. K. Nicholson, et al. (2000). "Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids." Concepts in Magnetic Resonance **12**(5): 289-320.

Low, P. S., D. P. Allen, et al. (1987). "Tyrosine Phosphorylation of Band-3 Inhibits Peripheral Protein-Binding." Journal of Biological Chemistry **262**(10): 4592-4596.

Low, P. S., P. Rathinavelu, et al. (1993). "Regulation of glycolysis via reversible enzyme binding to the membrane protein, band 3." Journal of Biological Chemistry **268**(20): 14627-14631.

Maede, Y., M. Inaba, et al. (1983). "Increase of Na-K-ATPase activity, glutamate, and aspartate uptake in dog erythrocytes associated with hereditary high accumulation of GSH, glutamate, glutamine, and aspartate." Blood **61**(3): 493-499.

Majerus, P. W., M. J. Brauner, et al. (1971). "Glutathione synthesis in human erythrocytes. II. Purification and properties of the enzymes of glutathione biosynthesis." The Journal of clinical investigation **50**(8): 1637-1643.

Malz, F. and H. Jancke (2005). "Validation of quantitative NMR." Journal of Pharmaceutical and Biomedical Analysis **38**(5): 813-823.

Manuel y Keenoy, B., I. Conget, et al. (1991). "Generation of 3HOH from D-[6-3H]glucose by erythrocytes: role of pyruvate alanine interconversion." Biochemical medicine and metabolic biology **46**(1): 59-74.

Maretzki, D., B. Reimann, et al. (1989). "A reappraisal of the binding of cytosolic enzymes to erythrocyte membranes." Trends Biochem Sci **14**(3): 93-96.

Markley, J. L., M. E. Anderson, et al. (2007). New bioinformatics resources for metabolomics. Pacific Symposium on Biocomputing, World Scientific Press**:** 157-168.

Mashego, M. R., L. Wu, et al. (2004). "MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites." Biotechnol Bioeng **85**(6): 620-628.

Massou, S., C. Nicolas, et al. (2007). "Application of 2D-TOCSY NMR to the measurement of specific C-13-enrichments in complex mixtures of C-13-labeled metabolites." Metabolic Engineering **9**(3): 252-257.

Massou, S., C. Nicolas, et al. (2007). "NMR-based fluxomics: Quantitative 2D NMR methods for isotopomers analysis." Phytochemistry **68**(16-18): 2330-2340.

McIntyre, L. M., D. R. Thorburn, et al. (1989). "Comparison of computer simulations of the F-type and L-type non-oxidative hexose monophosphate shunts with 31P-NMR experimental data from human erythrocytes." European Journal of Biochemistry **180**(2): 399-420.

Measures, J. C. (1975). "Role of Amino-Acids in Osmoregulation of Non-Halophilic Bacteria." Nature **257**(5525): 398-400.

Mehta, T., M. Tanik, et al. (2004). "Towards sound epistemological foundations of statistical methods for high-dimensional biology." Nat Genet **36**(9): 943-947.

Mendes, P. (2002). "Emerging bioinformatics for the metabolome." Brief Bioinform **3**(2): 134-145.

Messana, I., M. Orlando, et al. (1996). "Human erythrocyte metabolism is modulated by the O2-linked transition of hemoglobin." FEBS Lett **390**(1): 25-28.

Meyers, E. N., M. Lewandoski, et al. (1998). "An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination." Nat Genet **18**(2): 136-141.

Minnich, V., M. B. Smith, et al. (1971). "Glutathione biosynthesis in human erythrocytes. I. Identification of the enzymes of glutathione synthesis in hemolysates." The Journal of clinical investigation **50**(3): 507-513.

Morris, C. R., J. H. Suh, et al. (2008). "Erythrocyte glutamine depletion, altered redox environment, and pulmonary hypertension in sickle cell disease." Blood **111**(1): 402-410.

Morvan, D., A. Demidem, et al. (2003). "Quantitative HRMAS proton total correlation spectroscopy applied to cultured melanoma cells treated by chloroethyl nitrosourea: Demonstration of phospholipid metabolism alterations." Magnetic Resonance in Medicine **49**(2): 241-248.

Mulquiney, P. J., W. A. Bubb, et al. (1999). "Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations in vivo kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using C-13 and P-31 NMR." Biochemical Journal **342**: 567-580.

Mulquiney, P. J. and P. W. Kuchel (1999). "Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: computer simulation and Metabolic Control analysis." Biochemical Journal **342**: 597-604.

Mulquiney, P. J. and P. W. Kuchel (1999). "Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement." Biochemical Journal **342**: 581-596.

Munro, G. F., J. Morgan, et al. (1972). "Dependence of Putrescine Content of Escherichia-Coli on Osmotic Strength of Medium." Journal of Biological Chemistry **247**(4): 1272-&.

Murashige, T. (1978). "Citation Classic - Revised Medium for Rapid Growth and Bioassays with Tobacco Tissue-Cultures." Current Contents(43): 10-10.

Murphy, J. R. (1960). "Erythrocyte metabolism. II. Glucose metabolism and pathways." J Lab Clin Med **55**: 286-302.

Murthy, S. N., T. Liu, et al. (1981). "The aldolase-binding site of the human erythrocyte membrane is at the NH2 terminus of band 3." J Biol Chem **256**(21): 11203-11208.

Nagy, A., C. Moens, et al. (1998). "Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles." Curr Biol **8**(11): 661-664.

Nicholson, J. K., J. Connelly, et al. (2002). "Metabonomics: a platform for studying drug toxicity and gene function." Nat.Rev.Drug Discov. **1**(2): 153-161.

Nicholson, J. K., J. C. Lindon, et al. (1999). "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data." Xenobiotica **29**(11): 1181-1189.

Niihara, Y., C. R. Zerez, et al. (1997). "Increased red cell glutamine availability in sickle cell anemia: demonstration of increased active transport, affinity, and increased glutamate level in intact red cells." J Lab Clin Med **130**(1): 83-90.

Ninfali, P., A. Accorsi, et al. (1983). "Vanadate Affects Glucose-Metabolism of Human-Erythrocytes." Archives of Biochemistry and Biophysics **226**(2): 441-447.

Northrop, D. B. (1975). "Steady-State Analysis of Kinetic Isotope-Effects in Enzymic Reactions." Biochemistry **14**(12): 2644-2651.

Pan, C. L., G. Kora, et al. (2006). "Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics." Analytical Chemistry **78**(20): 7110-7120.

Pauli, G. F., B. U. Jaki, et al. (2005). "Quantitative 1H NMR: development and potential of a method for natural products analysis." J Nat Prod **68**(1): 133-149.

Pauling, L., A. B. Robinson, et al. (1971). "Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography." Proc Natl Acad Sci U S A **68**(10): 2374-2376.

Peters, L. L., R. A. Shivdasani, et al. (1996). "Anion exchanger 1 (band 3) is required to prevent erythrocyte membrane surface loss but not to form the membrane skeleton." Cell **86**(6): 917-927.

Peters, L. L., R. A. Swearingen, et al. (2004). "Identification of quantitative trait loci that modify the severity of hereditary spherocytosis in wan, a new mouse model of band-3 deficiency." Blood **103**(8): 3233-3240.

Peterson, D. J. and N. M. Loening (2007). "QQ-HSQC: a quick, quantitative heteronuclear correlation experiment for NMR spectroscopy." Magn Reson Chem **45**(11): 937-941.

Pinciroli, V., R. Biancardi, et al. (2004). "The well-characterized synthetic molecule: A role for quantitative H-1 NMR." Organic Process Research & Development **8**(3): 381-384.

Potter, R. F. and A. C. Groom (1983). "Capillary Diameter and Geometry in Cardiac and Skeletal-Muscle Studied by Means of Corrosion Casts." Microvascular Research **25**(1): 68-84.

Radda, G. K. and P. J. Seeley (1979). "Recent studies on cellular metabolism by nuclear magnetic resonance." Annu Rev Physiol **41**: 749-769.

Raica, N., Jr. and H. E. Sauberlich (1964). "Blood Cell Transaminase Activity in Human Vitamin B6 Deficiency." Am J Clin Nutr **15**: 67-72.

Rapoprot, I., H. Berger, et al. (1976). "Response of the glycolysis of human erythrocytes to the transition from the oxygenated to the deoxygenated state at constant intracellular pH." Biochimica et Biophysica Acta **428**(1): 193-204.

Record, M. T., E. S. Courtenay, et al. (1998). "Responses of E-coli to osmotic stress: Large changes in amounts of cytoplasmic solutes and water." Trends in Biochemical Sciences **23**(4): 143-148.

Regnier, F. E. and S. Julka (2006). "Primary amine coding as a path to comparative proteomics." Proteomics **6**(14): 3968-3979.

Rittenberg, D. and G. L. Foster (1940). "A New Procedure for Quantitative Analysis by Isotope Dilution, with Application to the Determination of Amino Acids and Fatty Acids " Journal of Biological Chemistry **133**(3): 737-744.

Rodgers, R. P., E. N. Blumer, et al. (2000). "Stable isotope incorporation triples the upper mass limit for determination of elemental composition by accurate mass measurement." J Am Soc Mass Spectrom **11**(10): 835-840.

Rogalski, A. A., T. L. Steck, et al. (1989). "Association of Glyceraldehyde-3-Phosphate Dehydrogenase with the Plasma-Membrane of the Intact Human Red Blood-Cell." Journal of Biological Chemistry **264**(11): 6438-6446.

Ross, A., M. Salzmann, et al. (1997). "Fast-HMQC Using Ernst Angle Pulses: An Efficient Tool for Screening of Ligand Binding to Target Proteins." Journal of Biomolecular NMR **10**: 389-396.

Roy, S. M., M. Anderle, et al. (2004). "Differential expression profiling of serum proteins and metabolites for biomarker discovery." International Journal of Mass Spectrometry **238**(2): 163-171.

Sadtler (1967). "Anon - Sadtler Standard Spectra . Nmr Chemical Shift Index." Journal of Chemical Documentation **7**(3): 182-&.

Sandusky, P. and D. Raftery (2005). "Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabonomics of amino acids in honey." Anal Chem **77**(8): 2455-2463.

Sass, M. D. (1968). "Glutathione synthesis in cell-free preparations from erythrocytes of different ages." Clinica chimica acta; international journal of clinical chemistry **22**(2): 207-210.

Schrader, M. C., C. J. Eskey, et al. (1993). "A carbon-13 nuclear magnetic resonance investigation of the metabolic fluxes associated with glucose metabolism in human erythrocytes." Biochim Biophys Acta **1182**(2): 162-178.

Seavey, B. R., E. A. Farr, et al. (1991). "A relational database for sequence-specific protein NMR data." J Biomol NMR **1**(3): 217-236.

Shaka, A. J., C. J. Lee, et al. (1988). "Iterative Schemes for Bilinear Operators; Application to Spin Decoupling." Journal of Magnetic Resonance **77**: 274-293.

Shortreed, M. R., S. M. Lamos, et al. (2006). "Ionizable isotopic labeling reagent for relative quantification of amine metabolites by mass spectrometry." Anal Chem **78**(18): 6398-6403.

Shrot, Y. and L. Frydman (2004). "Spatially resolved multidimensional NMR spectroscopy within a single scan." J Magn Reson **167**(1): 42-48.

Shulman, R. G., T. R. Brown, et al. (1979). "Cellular applications of 31P and 13C nuclear magnetic resonance." Science **205**(4402): 160-166.

Shulman, R. G. and D. L. Rothman (2001). "C-13 NMR of intermediary metabolism: Implications for systemic physiology." Annual Review of Physiology **63**: 15-48.

Siems, W., M. Muller, et al. (1982). "Quantification of Pathways of Glucose-Utilization and Balance of Energy-Metabolism of Rabbit Reticulocytes." European Journal of Biochemistry **124**(3): 567-576.

Simons, T. J. B. (1979). "Vanadate - New Tool for Biologists." Nature **281**(5730): 337-338.

Southgate, C. D., A. H. Chishti, et al. (1996). "Targeted disruption of the murine erythroid band 3 gene results in spherocytosis and severe haemolytic anaemia despite a normal membrane skeleton." Nat Genet **14**(2): 227-230.

Srivastava, S. K. and E. Beutler (1967). "Permeability of normal and glucose-6-phosphate dehydrogenase deficient erythrocytes to glutathione." Biochemical and biophysical research communications **28**(5): 659-664.

Stefanovic, M., N. O. Markham, et al. (2007). "An 11-amino acid beta-hairpin loop in the cytoplasmic domain of band 3 is responsible for ankyrin binding in mouse erythrocytes." Proceedings of the National Academy of Sciences of the United States of America **104**(35): 13972-13977.

Steinbeck, C., S. Krause, et al. (2003). "NMRShiftDB - Constructing a free chemical information system with open-source components." Journal of Chemical Information and Computer Sciences **43**(6): 1733-1739.

Sterner, J. L., M. V. Johnston, et al. (2000). "Signal suppression in electrospray ionization Fourier transform mass spectrometry of multi-component samples." J Mass Spectrom **35**(3): 385-391.

Stokvis, E., H. Rosing, et al. (2005). "Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not?" Rapid Commun Mass Spectrom **19**(3): 401-407.

Swanson, M. G., K. R. Keshari, et al. (2008). "Quantification of choline- and ethanolamine-containing metabolites in human prostate tissues using 1H HR-MAS total correlation spectroscopy." Magnetic Resonance in Medicine **60**(1): 33-40.

Thorburn, D. R. and P. W. Kuchel (1985). "Regulation of the Human-Erythrocyte Hexose-Monophosphate Shunt under Conditions of Oxidative Stress - a Study Using Nmr-Spectroscopy, a Kinetic Isotope Effect, a Reconstituted System and Computer-Simulation." European Journal of Biochemistry **150**(2): 371-386.

Tolstikov, V. V. and O. Fiehn (2002). "Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry." Analytical Biochemistry **301**(2): 298-307.

Tsai, I. H., S. N. Murthy, et al. (1982). "Effect of red cell membrane binding on the catalytic activity of glyceraldehyde-3-phosphate dehydrogenase." J Biol Chem **257**(3): 1438-1442.

Tsai, I. H., S. N. P. Murthy, et al. (1982). "Effect of Red-Cell Membrane-Binding on the Catalytic Activity of Glyceraldehyde-3-Phosphate Dehydrogenase." Journal of Biological Chemistry **257**(3): 1438-1442.

Tybulewicz, V. L., C. E. Crawford, et al. (1991). "Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene." Cell **65**(7): 1153-1163.

Ubbink, J. B., S. Bissbort, et al. (1989). "Genetic polymorphism of glutamate-pyruvate transaminase (alanine aminotransaminase): influence on erythrocyte activity as a marker of vitamin B-6 nutritional status." The American Journal of Clinical Nutrition **50**(6): 1420-1428.

Ulrich, E. L., H. Akutsu, et al. (2008). "BioMagResBank." Nucleic Acids Res **36**(Database issue): D402-408.

Ulrich, E. L., J. L. Markley, et al. (1989). "Creation of a Nuclear Magnetic Resonance Data Repository and Literature Database." Protein Sequences and Data Analysis **2**(1): 23-37.

Uthman, E. (1998). Understanding Anemia, University Press of Mississippi.

Vaidyanathan, S., J. J. Rowland, et al. (2001). "Discrimination of aerobic endospore-forming bacteria via electrospray-Ionization mass spectrometry of whole cell suspensions." Analytical Chemistry **73**(17): 4134-4144.

Vallino, J. J. and G. Stephanopoulos (1993). "Metabolic Flux Distributions in Corynebacterium-Glutamicum during Growth and Lysine Overproduction." Biotechnology and Bioengineering **41**(6): 633-646.

Van Dort, H. M., R. Moriyama, et al. (1998). "Effect of band 3 subunit equilibrium on the kinetics and affinity of ankyrin binding to erythrocyte membrane vesicles." Journal of Biological Chemistry **273**(24): 14819-14826.

Viant, M. R. (2003). "Improved methods for the acquisition and interpretation of NMR metabolomic data." Biochem.Biophys.Res.Commun. **310**(3): 943-948.

Vold, R. L., J. S. Waugh, et al. (1968). "Measurement of Spin Relaxation in Complex Systems." Journal of Chemical Physics **48**: 3831-3832.

von Roepenack-Lahaye, E., T. Degenkolb, et al. (2004). "Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry." Plant Physiology **134**(2): 548-559.

Wang, W., H. Zhou, et al. (2003). "Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards." <u>Analytical Chemistry</u> **75**(18): 4818-4826.

Want, E. J., G. O'Maille, et al. (2006). "Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry." <u>Analytical Chemistry</u> **78**(3): 743-752.

Weber, R. E., W. Voelter, et al. (2004). "Modulation of red cell glycolysis: interactions between vertebrate hemoglobins and cytoplasmic domains of band 3 red cell membrane proteins." <u>American Journal of Physiology-Regulatory Integrative and Comparative Physiology</u> **287**(2): R454-R464.

Weininger, D., A. Weininger, et al. (1989). "Smiles .2. Algorithm for Generation of Unique Smiles Notation." <u>Journal of Chemical Information and Computer Sciences</u> **29**(2): 97-101.

Weljie, A. M., J. Newton, et al. (2006). "Targeted profiling: quantitative analysis of 1H NMR metabolomics data." <u>Anal Chem</u> **78**(13): 4430-4442.

Winter, C. G. and H. N. Christensen (1964). "Migration of Amino Acids Across the Membrane of the Human Erythrocyte." <u>The Journal of biological chemistry</u> **239**(Journal Article): 872-878.

Wishart, D. S., D. Tzur, et al. (2007). "HMDB: the Human Metabolome Database." <u>Nucleic Acids Res</u> **35**(Database issue): D521-526.

Wohlbach, D. J., B. F. Quirino, et al. (2008). "Analysis of the Arabidopsis histidine kinase ATHK1 reveals a connection between vegetative osmotic stress sensing and seed maturation." <u>Plant Cell</u> **20**(4): 1101-1117.

Wu, G., Y. Z. Fang, et al. (2004). "Glutathione metabolism and its implications for health." <u>The Journal of nutrition</u> **134**(3): 489-492.

Wu, L., M. R. Mashego, et al. (2005). "Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards." <u>Anal Biochem</u> **336**(2): 164-171.

Xia, J. G., T. C. Bjorndahl, et al. (2008). "MetaboMiner - semi-automated identification of metabolites from 2D NMR spectra of complex biofluids." <u>Bmc Bioinformatics</u> **9**: -.

Yagi, T., H. Kagamiyama, et al. (1982). "Aspartate: 2-oxoglutarate aminotransferase from bakers' yeast: crystallization and characterization." <u>Journal of Biochemistry</u> **92**(1): 35-43.

Yang, W. C., J. Adamec, et al. (2007). "Enhancement of the LC/MS analysis of fatty acids through derivatization and stable isotope coding." <u>Anal Chem</u> **79**(14): 5150-5157.

Yang, W. C., H. Mirzaei, et al. (2006). "Enhancement of amino acid detection and quantification by electrospray ionization mass spectrometry." <u>Analytical Chemistry</u> **78**(13): 4702-4708.

Yannoukakos, D., H. E. Meyer, et al. (1991). "3 Regions of Erythrocyte Band-3 Protein Are Phosphorylated on Tyrosines - Characterization of the Phosphorylation Sites by Solid-

Phase Sequencing Combined with Capillary Electrophoresis." <u>Biochimica et Biophysica Acta</u> **1066**(1): 70-76.

Young, D. J., M. W. Wolowyk, et al. (1987). "Conflicting evidence regarding the transport of alpha-glutamyl-dipeptides by human erythrocytes." <u>The Biochemical journal</u> **242**(1): 309-311.

Young, J. D., M. W. Wolowyk, et al. (1983). "Red-cell amino acid transport. Evidence for the presence of system ASC in mature human red blood cells." <u>The Biochemical journal</u> **216**(2): 349-357.

Zhang, F., L. Bruschweiler-Li, et al. (2008). "Self-consistent metabolic mixture analysis by heteronuclear NMR. Application to a human cancer cell line." <u>Analytical Chemistry</u> **80**(19): 7549-7553.

Zhu, J. K. (2002). "Salt and drought stress signal transduction in plants." <u>Annual Review of Plant Biology</u> **53**: 247-273.