Fall 2013

Development Of Isotags For Nmr Based Metabolite Profiling And Applications

Fariba Tayyari
Purdue University

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Fariba Tayyari

Entitled DEVELOPMENT OF ISOTAGS FOR NMR BASED METABOLITE PROFILING AND APPLICATIONS

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

David H. Thompson co-chair
Chair

M. Daniel Raftery co-chair

Chittaranjan Das

Kavita Shah

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): David H. Thompson

M. Daniel Raftery

Approved by: R. E. Wild 9-9-2013
Head of the Graduate Program Date

DEVELOPMENT OF ISOTAGS FOR NMR BASED METABOLITE PROFILING AND APPLICATIONS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Fariba Tayyari

In Partial Fulfillment of the

Requirements of the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

I dedicate this work to my parents and my husband, to honor their love, patience, and support

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my adviser, Dr. Daniel Raftery, who has supported me throughout my thesis with his inspiring and delicate guidance, advice, and expertise. I am indebted to him more than he knows.

I am thankful to Dr. Nagana Gowda for his valuable guidance and contributions from the first day I joined the Raftery group. Dr. Haiwei Gu, whose help is greatly appreciated. I also would like to thank Dr. Sulma Mohammed for reading the last chapter of my thesis and her comments and guidance. Many thanks go in particular to Dr. Thompson, Dr. Das and Dr. Shah. I am much indebted to them for their valuable advice and using their precious times to read this thesis.

I would also like to express thanks to the member of the Purdue Interdepartmental NMR Facility Dr. John Harwood, Jerry Hirschinger and Donna Bertram for their support and assistance over the years. My gratitude also goes to Debbie Packer, Betty Hatfield, Lynn Rider and Elizabeth Hewitt for their assistance and in ensuring my enrollment. The great help of Tim Selby and Konrad Kliewer of the John Amy Center for Instrumentation (JAFCI) for their help making my stay a pleasant one are greatly appreciated.

I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

I convey special acknowledgement to my family. My wonderful parents deserve special mention for their support, enthusiastic encouragement and unconditional love. My Father, Dr. Faramarz Tayyari, is the person who showed me the joy of rational pursuit ever since I was a child. My Mother, Ebtehaj Haghighi, is the one who sincerely raised me with her caring and love to set high goals for myself. My grandmothers, Zari Mehrabpour and Keshvar Maleki, for their unconditional love. Reza, thanks for being supportive and caring brother. You always know how to make me laugh. My sister, Faryan, who have always been present for me in all my paths in life. My sister, Farnoosh, for her kindness, love and the encouragement that she always offered. My father and mother in-law for their kindness and interest in my progress and accomplishments.

At last, not the least, I owe my loving thanks to my dear husband, Mohammad Chegeni, words fail me to convey my gratitude to his dedication, love and persistent confidence in me. I owe him for being generously let his intelligence, passions, and ambitions collide with mine.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

Tables Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

LIST OF SCHEMES

ABSTRACT

Tayyari, Fariba. Ph.D., Purdue University, December 2013. Development of Isotags for NMR Based Metabolite Profiling and Applications. Major Professor: Daniel Raftery.

NMR spectroscopy is a powerful analytical tool for both qualitative and quantitative metabolite profiling analysis. However, accurate quantitative analysis of biological systems especially using one-dimensional NMR has been challenging due to signal overlap. In contrast, the enhanced resolution and sensitivity offered by chemoselective isotope tags have enabled new and enhanced methods for detecting hundreds of quantifiable metabolites in biofluids using NMR spectroscopy or mass spectrometry. In this thesis we show improved sensitivity and resolution of NMR experiments imparted by ^{15}N and ^{13}C isotope tagging which enables the accurate analysis of plasma metabolites.

To date, isotope tagging has been used in conjunction with a single analytical platform. The inability to detect the same metabolites using the complementary analytical techniques of NMR and mass spectrometry has hindered the correlation of data derived from the two powerful platforms for applications such as biomarker discovery or the identification of unknown metabolites. To address this problem, we describe a smart isotope tag, ^{15}N-cholamine, which possesses two important properties: an NMR sensitive isotope, and a permanent charge for MS sensitivity. Finally, we present a study on

metabolite profiling using intact breast cancer tissue samples in which we exploit the combined strength of NMR and multivariate statistical methods for metabolite profiling.

CHAPTER 1. INTRODUCTION TO METABOLITE PROFILING

1.1 Introduction

Metabolomics or metabolite profiling describes the advanced study of metabolism in biological systems. In one sense, metabolite profiling has been used for decades for human disease diagnosis,[1] and the initial studies were actually introduced in early 1970s.[2] However, the instrumentation available at that time made definitive studies very challenging. The term "Metabonomics" was introduced 28 years later by Nicholson *et al.* as "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification",[3] and represented a re-birth of the field in some sense. Metabolomics has also been described as "a comprehensive and quantitative analysis of all metabolites" in biological systems and the "systematic study of the unique chemical fingerprints that specific cellular processes leave behind".[4,5] Scientists are often used these terms interchangeably. The Metabolome was first introduced by Oliver et al. in 1998 as the complete set of metabolites synthesized by an organism.[6]

Metabolites represent the downstream products of genes and proteins and as such they integrate the biological information and changes to that system that occur over time. Metabolites are not only building blocks of many other biological components such as proteins, they also have many other important roles in biological systems such as

signaling, enzymes activity, and providing energy (ATP). The metabolome is dynamic, and the level of metabolites in a biological system respond sensitively to a variety of changes, including influences by drugs, disease or any other stress caused by a cell's environment at a certain time.

Given the sensitivity to cellular perturbations, metabolite profiling can be used to study a variety of biological questions, including the differences between groups such as disease and healthy at a molecular level and convey extensive information to the field of "molecular epidemiology." Metabolomics has some advantages over other 'omics such as genomics and proteomics. First metabolite profiling provides detailed information related on phenotype, while genomics and to a lesser extent the proteome provide more information about genotype.[7] Second, the number of metabolites represented in the human metabolome is significantly less than the large number of genes and proteins in the human genome or proteome, which makes metabolomics more discriminating.[8,9] Also, since a given metabolite is same in different organisms, more generic techniques can be used for metabolomics and extended across species.[8] Lastly, the technology involved in measuring metabolomics is faster and less expensive than for genomics or proteomics.[8]

In this chapter, the entire workflow of metabolite profiling, as illustrated in Figure 1.1, will be reviewed with a focus on human samples.

1.2 Biological Samples and Sample Preparation

Metabolomics approaches can be applied to a variety of different biological matrices, including biofluids, tissues, and cells. It is critical to choose an appropriate

preparation method, including all the steps from the time the sample is collected until the data acquisition is started. The method has to be reliable and reproducible, and based on the type of the sample and technique to be used for analysis.

Since the metabolome is dynamic, quenching of the metabolism is necessary, such as using cold methanol, trifluoroacetic acid, or even freezing with liquid nitrogen, followed by appropriate storage at cold temperatures. Other preparation steps are dependent on the analytical technique to be used and the type of the sample or organism of interest. These steps include extraction, separation, isolation, purification, derivatization, buffer addition, and etc.

1.2.1 Samples

Metabolite profiling studies have been reported on many different biological systems such as cells, tissues and a number of biofluids including plasma, serum, urine, bile, cerebrospinal fluid, seminal fluid, amniotic fluid, synovial fluid, and saliva.[10-18] Among all the samples serum, plasma and urine have been most often targeted for metabolomics studies because the sample is non-invasively collected and thus easily obtained.

1.2.1.1 Blood Samples

Blood is the most readily accessible biofluid and has been subjected to many metabolomics studies since it is rich in metabolites and potentially provides vital information on almost every type of cell. In a living organism blood serum or plasma provides a direct global view of the metabolic status. Blood has two parts, a cellular

component including red and white blood cells and plasma, a liquid carrier in which these cells and other molecular species are suspended. [19] Differences between plasma and serum occur from the way they are prepared from blood. Plasma is obtained by adding anti-coagulants to the blood sample; then by centrifuging the mixture, the cellular portion is precipitated and the plasma supernatant is obtained. However, to obtain serum the blood is allowed to clot without adding any anti-coagulant. Plasma is more viscose compare to serum because the clotting proteins remain in the supernatant.[20] While serum and plasma have very similar metabolite profiles; however, differences in the distributions of specific metabolites and even the number of metabolites in each fluid has been reported.[21] Soluble proteins still remain in both plasma and serum, and depending on the analytical technique used, additional preparation steps are required for further purification or deproteination.

1.2.1.2 Urine Samples

Ease and non-invasive sampling of urine add to the other advantages of urine including a relatively high concentration of metabolites and low concentration of proteins, which make urine a favorable target for metabolomics studies. However, urine suffers from high salt content that triggers problems in some analytical methods.

A number of confounding factors can cause differences in the concentrations of urine metabolites between individuals, including the volume of urine, amount of the water and physiological conditions such as age, gender, weight, personal diet and environmental effects.[22] Urinary metabolites also have different concentrations over time for a single individual since they reflect the conditions that have been introduced to the

body such as food intake, medication, and exercise. Nevertheless, the urinary metabolite profile can provide an immense amount of information on biological status.

Urine sample preparation usually is minimal and can be performed after the sample is collected. However, if the urine samples need to be stored, they can be at or below −25 °C without preservatives. If sterility is difficult to maintain or unachievable, and for the case of temporary storage at 4 °C, the addition of preservatives is necessary. In these situations 0.1% sodium azide, is preferred over preservation with sodium fluoride. [22]

1.2.1.3 Tissue Samples

Tissue samples have been used for metabolite profiling for a number of years.[23-24] However, metabolic profiling of intact tissue has become of increasing interest,[25-27] as the study of intact tissues can help researchers better understand the molecular basis of diseases.[25,28] Tissue extraction methods, such as using methanol/water/chloroform destruct the cell structure, and pull out the metabolites into either the aqueous or organic layer, depending on their solubility. Typically these two solutions are analyzed separately. In contrast, the analysis of intact tissue samples is non-destructive and preparation is fast and minimal (addition of deuterated water, buffer, etc.). In addition to the other advantages of using intact tissues, these samples can be used later for clinical studies.

1.2.1.4 Cell ad Cell Medium Samples

Metabolite profiling of cells have been used to discriminate between different type of tumors and cell lines.[29,30] Cell samples also provide valuable information for studying biochemical pathways and mechanisms. Numerous studies has been conducted on different types of cancer cells and they show diverse biochemical pathways.[31,32] To quench cell metabolism, liquid nitrogen or cold methanol are the most common methods.[33,34] Cell sample preparation should be prudently done to save the integrity of the sample and to limit leakage of intracellular metabolites, especially for the types of cells that are highly sensitive to osmotic changes of medium.[34,35] Metabolite profiling of cell culture medium are not as common as cells. It can be used to evaluate the impact of the medium on the cell culture performance.[36]

1.2.2 Sample Preparation for NMR Experiments

One of the main advantages of NMR over mass spectrometry is the ease of sample preparation. Except for experiments that require samples to be derivatized, sample preparation for NMR experiments is minimal. Blood samples including plasma and serum contain proteins and lipids which cause broad lines in the NMR spectrum. To detect small molecules several spectral editing methods are available and widely used. Only in some cases are samples deproteinized prior adding them into NMR tubes. Urine samples benefit from their low concentrations of proteins and high concentrations of low-molecular-weight compounds; therefore sample preparation is minimized. Tissue preparation methods depend on the type of experiment performed, which group of metabolites are of interest, and the instrumentation used: liquid state NMR, HR-MAS,

solid state NMR, etc.[25,37,38] Overall, and especially beyond NMR analysis, extraction has been the most common method for tissue sample preparation. Among the extraction methods solid-liquid extraction is more common. In this method a solvent, typically deuterated water, is added to the solid material, followed by shaking, vortexing or stirring to enhance the contact between them. For HR-MAS, intact tissues are used and only deuterated water or buffer, with or without reference, is added. Different protocols are available for quenching cell metabolism and perform cell lysis.

1.3 NMR Spectroscopy for Metabolomics Applications

Among the techniques used for metabolomics studies, NMR and Mass spectrometry (MS) are the two most common.[39] Both techniques have advantages and disadvantages, and i general they are complementary to each other. NMR is highly selective, non-destructive, requires minimal sample preparation, is robust, and the data generated by NMR are very quantitative and reproducible. However, NMR is costly, suffers from poor sensitivity, and the complex spectra typically show a high degree of spectral peak overlap. Later in this chapter and thesis, some the methods to enhance sensitivity and resolution will be discussed.

1.3.1 Water Signal Suppression

Water comprise abundantly in most of the biological samples. In NMR-based metabolomics studies overwhelming signal of the water resonance is the obstacle. Without water suppression, water gives rise to a large residual signal and attenuate or

eliminate metabolites signals. Therefore, it is very common and desirable in NMR metabolomics studies to suppress the water resonance signal.

The water resonances can be suppressed by the use of appropriate standard NMR solvent suppression methods. Several pulse sequences exist for this purpose Including, Presaturation (PRESAT), water suppression enhanced through T1effects (WET) sequence, WATER suppression by GrAdient Tailored Excitation (WATERGATE), presaturation utilizing relaxation gradients and echoes (PURGE), and excitation sculpting, etc.[40-44]

Among these pulse sequences presaturation technique is more commonly used in metabolomics studies. The pulse sequence consists of two pulses, first a selective pulse to saturate water frequency (~ 4.8 ppm), it is a continuous low-power radio frequency pulse over the few second time period of an acquisition delay, usually 1-2 sec. Second one is a non-selective high power 90º pulse to excite other protons except saturated water protons.[41] In a spectrum generated by this pulse sequence water signal is still exist, because of the water portion that is not perfectly irradiated, although it is highly suppressed.

1.3.2 1D NMR Spectroscopy Pulse Sequences Used for Metabolomics

In metabolomics studies it is important to select a proper NMR acquisitions are from available NMR experiments, including 1D NOESY with water suppression, CPMG 1D ^{1}H with water suppression, 1D selective TOCSY, DOSY, etc. These methods have advantages and disadvantages, therefore selection of a method depends on the origin of the sample and metabolite molecular groups of interest.

1.3.2.1 1D NOESY

1D NOESY (nuclear Overhauser enhancement spectroscopy) is one of the most commonly used pulse sequences in metabolomics. 1D NOESY is usually coupled with PRESAT to attenuate the water peak and called NOESYPRESAT.[45] In this sequence the water resonance is irradiated two times, first during relaxation delay, RD, (1-3 sec.), and again during the mixing time, $t_{m,}$ (100-150 msec.).[45] With this method, the water signal in biofluid samples is attenuated by a factor of 10^5 or more. The spectra acquired by NOESY show broad signals from macromolecules such as proteins and lipids in addition to the narrow spectral features from small molecules. Therefore, the 1D NOESY sequence is more suitable for urine than serum or plasma samples, since urine contains a low concentration of proteins.

1.3.2.2 1D CMPG

Blood samples including plasma and serum, cells and tissue all contain macromolecules, which display broad lines in the NMR spectrum due to their slow tumbling rates that limit spectral averaging. Therefore, a more suitable pulse sequence for the analysis of these samples should be able to suppress the broad signals resulting from macromolecules. The most common sequence used to suppress macromolecular signals is the CPMG (Carr-Purcell-Meiboom-Gill) spin echo pulse.[46] In this sequence there is a 90° phase shift between the 90° initial pulse along the x-axis and the following 180° pulses along the y-axis. This phase shift creates echoes that all form along y-axis and eliminate most errors of the 180° pulses. Fourier transformation of the CPMG free induction decay produces high resolution NMR spectra, which are highly reproducible and quantitative.

Much like the NOSEY sequence, most of the time CPMG is coupled with PRESAT, and is known as PRESAT-CPMG.

1.3.2.3 1D Selective TOCSY

1D NMR spectra of biological samples are crowded and highly congested; therefore some of the smaller signals are overlapped or buried under other, larger signals. Selective total correlation spectroscopy (Selective-TOCSY) is a NMR experiment that can be used to identify distinct spin-systems, i.e., a group of spins from the sample molecule which are connected by J-couplings.[47] During the TOCSY evolution period, the NMR magnetization spreads from one 1H spin to the rest of members of the spin system via J-coupling, eventually throughout the molecule. Selective TOCSY, in which a single peak is excited and the coupled spins are revealed in the spectrum, has been used in biological samples and showed promising results.[48,49] This technique can help to detect and quantify metabolites with concentrations 10–100 times below those of the major components.[48]

1.3.3 2D NMR Spectroscopy Pulse Sequences Used for Metabolomics

1D ^{1}H NMR spectra are often complicated because of the overlapping signals due to the large number of metabolites in biological samples.[50] 2D NMR experiments are sometimes able to improve the resolution by dispersion signals in two different dimensions. Several 2D NMR experiments are used in metabolomics studies, including heteronuclear single quantum coherence spectroscopy (HSQC) and heteronuclear multiple bond correlation spectroscopy (HMBC), 2D J-resolved spectroscopy, correlation

spectroscopy (COSY), total correlation spectroscopy (TOCSY), etc. However, the disadvantages of 2D NMR experiments, including longer acquisition times, larger data size and difficult data analysis compared to 1D NMR decrease their use and interest for metabolomics studies.[51] 2D-HSQC is the only 2D NMR experiment used for the work in this thesis, and therefore only the 2D-HSQC experiment is discussed in this chapter.

1.3.3.1 2D HSQC

HSQC displays correlations between heteronuclear spins across a single chemical bond, and thus allows the identification of the directly connected nuclei.[52] Usually, the detected correlation is between ^{1}H and either ^{15}N or ^{13}C nuclei. For instance, 2D ^{1}H–^{15}N HSQC NMR correlates ^{1}H and ^{15}N NMR resonances from the same molecule, and as a result only those nitrogen nuclei that are coupled to protons are visible in the spectrum.[53,54] Several studies have been done for identification and quantification of metabolites by tagging them with ^{15}N or ^{13}C and using HSQC NMR experiments.[55-57]

1.3.4 Enhancement in Sensitivity and Resolution

Over the past 10 years a number of improvements in NMR instrumentation have provided significant gains in performance, both in resolution and sensitivity. Higher resolution and sensitivity are greatly improved by the use of higher magnetic field strengths,[41] and higher magnetic fields such as 18.8 Tesla (800 MHz for ^{1}H NMR) or above are attractive for metabolomics studies.[58,59] In addition, cryogenic (cryo) probes can significantly increase the signal-to-noise ratio (SNR) by reducing the level of thermal noise in the detection coil circuit. By cooling the probe and its electronics from room

temperature to 20 K, the SNR improves almost four fold. One of the disadvantages of NMR compared to MS in metabolomics studies is the amount of sample required. Micro-coil probes can be used to reduce the amount of sample required for NMR studies and thus limit this problem. Various micro-coil probes have been designed and constructed for metabolomics studies.[60-62] Coupling the NMR instrument with liquid chromatography is another way to increase the resolution, while a combined LC-MS-NMR approach can assist the identification of unknown metabolites. In this approach LC fractions are split between MS and NMR instruments for parallel analysis.[63]

Quantifying and identifying metabolites are some of the most important and challenging aspects of metabolite profiling, and therefore methods that improve NMR resolution and sensitivity are of great interest. Since biological samples are rich in metabolites, conventional 1D NMR methods suffer from the often overwhelming overlap of metabolite signals. In many cases, low concentrated metabolites are not even detected since they are buried under the other, larger signals. The relatively small dispersion of the ^{1}H NMR chemical shifts and the large number of spin-spin couplings between metabolites are the main causes of signal overlap. One of the strategies that has been demonstrated to improve NMR chemical shift dispersion is to detect heteronuclei such as ^{15}N, ^{13}C. ^{13}C and ^{15}N spectra are simpler, as they have fewer couplings; however, these nuclei have low natural abundances, and therefore conventional methods to detect these nuclei suffer from low sensitivity.

In contrast, isotope labeling can enhance the sensitivity. Isotope labeling has been performed *in vivo* and *in vitro* in biological systems. In vivo isotope labeling has been accomplished by feeding animals with ^{13}C-enriched diets,[64] as well as numerous cell

studies that use ^{13}C labeled glucose,[65] for example, to provide detailed mechanistic information. In vitro isotope labeling studies are often easier to perform and focus on improving the measurement technology to differentiate biological samples. One attractive approach involves the labeling specific classes of metabolites in biofluids with ^{13}C and ^{15}N, which has been performed with promising results.[66,55-57] For example, an isotope-labeled acetylation reaction using ^{13}C-acetic anhydride can label the amines present in samples, and later the tagged metabolites are detected using ^{1}H-^{13}C 2D HSQC.[66] Amino acids and carboxylic acids can labeled with ^{15}N-ethanolamine, and the amide can be detected by ^{1}H-^{15}N 2D HSQC.[55] This approach showed an improvement in the limit of detection such that over 100 metabolites in human urine and serum could be be detected.

1.4 Data Processing and Statistics

1.4.1. Pre- Data Processing

NMR spectral data processing is a critical step in metabolomics profiling and should be done with care. However, before data processing to be subjected to multivariate and/or univariate analyses, it has to be preprocessed. Accurate results cannot be achieved without appropriate preprocessing.

Preprocess steps after NMR data acquisition depend on the study and NMR experiment, but typically include phase and baselines correction, peaks alignment, bucketing (binning) or peak integration, normalization, and scaling.[67-68] These steps help eliminate or decrease the undesired spectral variations over samples.

1.4.1.1 Phasing and Baseline Correction

Phase and baseline corrections are the first steps in preprocessing following data acquisition. Baseline distortions can be corrected by automated baseline correction or manually, using a variety of processing software packages such as Topspin (Bruker Biospin, Billerica, MA), VnmrJ (Varian, Palo Alto, CA), MestReNova (Mestrelab Research, Santiago deCompostela, Spain), and KnowItAll (Bio-Rad Laboratories, Hercules, CA).

1.4.1.2 Peak Alignment and Bucketing

Small chemical shift changes can be caused by different factors, including instrumental effects (see below), changes in sample pH or ion concentrations, or the presence of other metabolites or macromolecules. It should be noted that the chemical shift changes caused by these factors do not affect all chemical shifts for all metabolites at the same level. Chemical shift offsets caused by the NMR instrument can be fixed and aligned by using a reference signal such as DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid), TSP (trimethylsilyl propionate) at 0.00 ppm or a peak from one of the metabolites. For example some metabolomics studies use the methyl doublet peaks of alanine around 1.48 ppm as a reference.[69-71] Peak alignment can be performed manually or by using special algorithms such as RSPA (recursive segment-wise peak alignment).[72] Usually, the next step following the peak alignment is either bucketing or peak integration. For bucketing, the NMR spectrum is divided into a set of desired spectral width segments (a typical width is 0.04 ppm), which are known as a bucket or bin. Bucketing integrates the peak areas over that spectral segment, and ensures a consistent measurement of the same

resonance for all samples by reducing the effect of pH variations or ion concentrations on chemical shifts.[73] A data matrix is produced by bucket segments, can be saved into a text file or exported to a worksheet directly to be used further data analysis. Data sets that use bucketed spectra are easier to work with, as they typically have only about 250 – 400 buckets per spectrum. An alternative to bucketing is the direct integration of spectral peaks. Although this is more time consuming, it can be more accurate and used to derive absolute concentrations of specific metabolites. Typically, though, the number of integrated metabolites is less, ranging from 20-40 for serum or tissue samples and up to about 80 metabolites for urine.

1.4.1.3 Normalization and Scaling

Metabolomics data typically have to be normalized, otherwise the samples may not be comparable to each other. For the normalization process different approaches can be used.[68,74] A common method for normalization is integral normalization known also as constant sum normalization. This approach normalizes each spectrum to a constant total integrated intensity across the whole region. Another method uses a standard with a known concentration that is to each sample; normalization of the spectrum is based on the standard.[73] For tissue samples, the data can be normalized with respect to the weight of the tissue.[68]

Metabolites in biological samples have a wide range of concentrations. The absolute variation in metabolite levels is related to their concentrations.[68] Scaling is therefore used to avoid the selection of only high concentration metabolites as biomarker candidates because they tend to exhibit high variations. Scaling therefore helps to

emphasize lower concentration species that might otherwise be overlooked. A number of scaling methods are available to use such as Pareto scaling, variance scaling, Log scaling, range scaling, vast scaling, and level scaling.[67,68,75,76]

1.4.2 Statistical Methods

Two general approaches are used to perform statistical analysis on the rich metabolomics NMR data, exploratory analysis and confirmatory analysis. Exploratory data analysis is used to find patterns in the data set using methods such as principal component analysis (PCA) and hierarchical clustering analysis (HCA). However, confirmatory data analysis makes an implicit use of the group labels. Methods for confirmatory analysis include the univariate Student's t-test, and multivariate methods that include partial least square-discriminant analysis (PLS-DA), orthogonal signal correction-PLS-DA (O-PLS-DA), logistic regression and many others.[77]

1.4.2.1 Unsupervised Methods

The exploratory methods commonly used in metabolomics are known as "unsupervised" since patterns are discovered without assigning the spectra to classes such as "disease" or "healthy." This approach draws distinctions between groups of samples with respect to their chemical compositions, i.e., metabolite concentration level changes among the samples. The most commonly used multivariate method in metabolomics is PCA.[78-83] PCA transforms the multidimensional data and builds linear multivariate models. The PCA model is built on the basis of orthogonal vectors (Eigenvectors). The principal components (PCs), are the eigenvectors of X, PCs are initially calculated from

the original data set X by a diagonalization of the covariance data matrix. The PCs are then ranked by their eigenvalues in decreasing order. The first PC (PC1) thus describes the direction of the largest variations generated across the set of spectra, the second PC (PC2) then describes the direction for the largest portion of the remaining variation, and similarly for the rest of the PCs. Except for the first several PCs, the other remaining PCs mostly contain noise. Therefore, the first few PCs are typically used to display the variation in a "score plot." Separation of samples along a particular PC is explained using the loading plot; it shows the contribution of each variable in the original data to a principle component. Loading plot assists with the identification of potential biomarkers.

1.4.2.2 Supervised Methods

Popular confirmatory methods include partial least square discriminant analysis (PLS-DA), orthogonal signal correction partial least square discriminant analysis (O-PLS-DA), t-test, logistic regression, k-nearest neighbor (KNN), and soft independent modeling by class analogy (SIMCA). The class information for each sample is assigned prior to the analysis, and therefore these methods are known as supervised analysis.

PLS is one of the most common used supervised methods. PLS fits the data matrix of predictors X and class matrix or vector of responses Y, and find a linear regression model to the new coordinate system. Equation 1.1 shows the general equation for the linear regression model. Y is the predicted outcome value for the linear regression model with the b, regression coefficients, 1 to z; b_0 is the Y intercept while the values for the X, predictor variables, are 1to z.

$$Y = b_o + b_1X_1 + b_2X_2 + \ldots + b_zX_z \qquad \text{(eq. 1.1)}$$

Each orthogonal axis in PLS is named a latent variable (LV). PLS-DA is the combination of PLS and discriminant analysis (DA), where DA is a statistical method for determination of a linear combination of features to predict to which class a case (or sample) belongs.[84,85] PLS-DA modeling can be used to find the difference between pre-assigned sample groups and recognize variables responsible for the class separation. Qualitative values including class or category of samples are included in Y matrix. Typically, Y is set up as a "dummy matrix" where, for example, Disease = 1 and Healthy = 0. The PLS-DA method in particular, and most supervised methods in general, suffer from data over-fitting, therefore it is crucial to perform the result validation otherwise.[86] Cross validation is the most common method for validation, in which results are applied to a new set of observations that was not used to build the model.[87] Leave-one-out cross-validation (LOOCV) is one of the most commonly used validation methods for PLS-DA models in metabolomics studies.[88-90] In this technique one single observation from the original sample set is used to evaluate the predictive accuracy of the model, while all the other samples are used as the training data to build the model. Each sample is used once as the validation data in the cross validation process.

1.4.2.3 Univariate Analysis

Univariate methods are often employed to identify significant metabolites that are altered between different groups. To identify significant difference, P-values, calculated from the Student's *t*-test, Welch's t-test, Mann-Whitney U test, or other variants are used to explains the probability of two means, are used.[91] Values less than 0.05 are generally considered statistically significant. However, the analysis of multiple metabolites results

in a process of making multiple comparisons between the sample classes, which often leads to an overly optimistic set of significant metabolites, of which many may be false discoveries. A number of methods are used to try to correct for this effect and generally involve the calculation of False Discovery Rate.[92-95]

Box-and-whisker plots (box plots) are a descriptive statistic for illustrating groups of numerical data. Box plots show differences between populations with no assumptions regarding the statistical distribution. Figure 2 illustrates a single box plot in some detail. The receiver operating characteristic (ROC) curve describes the function of a variable in binary classification. In binary classification tests, sensitivity and specificity are the key statistical measurments.[85,96,97] Several additional terms are useful for defining sensitivity and specificity, including the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) rates. If a disease is proven in a patient, and diagnostic test also indicates the disease, the result from test is considered a true positive (TP). Similarly, if the diagnostic test shows the patient does not have a disease and also proven that the person does not have it, the result of the test is a true negative (TN). If the diagnostic test indicates the disease in a person who does not actually have the disease, the result is a false positive (FP). And the result is a false negative (FN) if a disease is proven in a patient but the result of the diagnostic test indicates no disease. Equations 1.2, 1.3, and 1.4 describe sensitivity, specificity, and accuracy, respectively base on the terms of TP, TN, FP, and FN.[94]

$$\text{Sensitivity} = TP/(TP + FN) \quad \text{(eq. 1.2)}$$

$$\text{Specificity} = TN/(TN + FP) \quad \text{(eq. 1.3)}$$

$$\text{Accuracy} = (TN + TP)/(TN+TP+FN+FP) \quad \text{(eq. 1.4)}$$

The area under the ROC curve (AUC), measurement describes the overall accuracy of a test. An AUC of 0.5 describes equal distribution between two groups, i.e., a random result equivalent to flipping a coin, and the test based on it is valueless for discrimination. An AUC of more than 0.9 is considered to be an excellent test, and more than 0.8 considered good.[98]

1.5 Thesis Summery

The rest of this thesis focuses on the development of new methods and applications in NMR-based metabolomics. In Chapter 2, derivatization approaches were used for quantitative analysis of human blood plasma. Improved sensitivity and resolution of NMR experiments imparted by ^{15}N and ^{13}C isotopes enhanced the metabolite detection pool and accuracy of plasma metabolite analysis. Furthermore the approach can be extended to many additional metabolites in almost any biological mixture for high throughput analysis. A new "smart isotope tag," ^{15}N-cholamine, which possesses two important properties: an NMR sensitive isotope, and a permanent charge for MS sensitivity, is discussed in Chapter 3. This unique approach enables effective detection of the carboxyl-containing metabolome by both analytical methods. In Chapter 4, HR-MAS was used in a metabolomics study to identify altered concentrations of small-molecule metabolites in triple negative breast cancer (TNBC). In this chapter the metabolite

profiles of African Americans and Caucasians were studied and compared. It is shown that NMR-based metabolomics has good potential for identifying altered metabolism in the aggressive TNBC that is observed especially in African American women.

1.6 References

1. Gates, S. C.; Sweeley, C. C. *Clin. Chem.* **1978**, *24*, 1663-1673.

2. Horning, E. C.; Horning M. G. M *Clin Chem.* 1**971**, *17*, 802–809.

3. Nicholson, J. K.; Lindon J. C.; Holmes, E. *Xenobiotica.* **1999**, *29*, 1181–1189.

4. Fiehn, O. *Comp. Funct. Genom.***2001**, *2*, 155–168.

5. Daviss B.*The Scientist* **2005**, *19 (8)*, 25–28.

6. Oliver, S.G.; Winson, M.K.; Kell, D.B.; Baganz, R. 1998. Trends Biotechnol. **1998**, *16*, 373–378.

7. Gieger, C.; Geistlinger, L.; Altmaier, E.; Hrabé, M.; Kronenberg, F.; Meitinger, T.;Mewes H. W.; Wichmann, H.E.; Weinberger, K. M.; Adamski, J.; Illig, T.; Karsten, S. *PLoS Genet.* **2008**, *4(11),* e1000282.

8. Álvarez-Sánchez, B.; Priego-Capote, F.; Luque de Castro, M. D. *Trends Anal. Chem.* **2010**, *29*, 111–119.

9. Bain, J. R.; Stevens, R. D.; Wenner, B. R.; Ilkayeva, O.; Muoio, D. M.; Newgard, C.B. *Diabetes* **2009,** *58,* 2429-2443.

10. Jain, M.; Nilsson, R.; Sharma, S.; Madhusudhan, N.; Kitami, T.; Souza, A. L.; Mootha, V. K. *Science* **2012**, *336,*1040–1044.

11. Jung, K., Reszka, R., Kamlage, B., Bethan, B., Stephan, C., Lein, M. and Kristiansen, G. *Int. J. Cancer. 2013*. In Press.

12. Liu, L.; Aa, J.; Wang, G.; Yan, B.; Zhang, Y.; Wang, X.; Zhao, C.; Cao, B.; Shi, J.; Li, M.; Zheng, T.; Zheng, Y.; Hao, G.; Zhou, F.; Sun, J.; Wu, Z. *Anal. Biochem.* **2010**, *406 (2)* 105– 112.

13. Xie GX, Chen TL, Qiu YP, Shi P, Zheng XJ, Su MM, et al. Urine metabolite profiling offers potential early diagnosis of oral cancer. *Metabolomics* **2012**, *8*, 220–231.

14. Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2008**, *871,*164-173.

15. Lynch MJ, Masters J, Pryor JP, Lindon JC, Spraul M, Foxall PJD, Nicholson JK. Ultra high field NMR spectroscopic studies on human seminal fluid, seminal vesicle and prostatic secretions. *J. Pharm. Biomed. Anal.* **1994,** *12,* 5–19.

16. Graça, G.; Duarte, I. F.; Goodfellow, B. J.; Carreira, I. M.; Couceiro, A. B.; Domingues, M. R.; Spraul, M.; Tseng, L. H.; Gil, A. M. *Anal. Chem.* **2008**, *80 (15)*, 6085– 6092

17. Damyanovich, A. Z.; Staples, J. R.; Marshall K. W. *Osteoarthr. Cartilage* **1999,** *7,*165–172.

18. Wei, J., Xie, G., Zhou, Z., Shi, P., Qiu, Y., Zheng, X., Chen, T., Su, M., Zhao, A. and Jia, W. *Int. J. Cancer* **2011,** *129*, 2207–2217.

19. Fox S. I. Human physiology. Boston, Mass.: WCB/McGraw-Hill. 1999, 364–367.

20. West, J. B.; editor. Best and Taylor's Physiological Basis of Medical Practice. 11th ed. Baltimore MD, USA: Waverly Press, Inc. **1985**, 334–336.

21. Wedge, D. C.; Allwood, J. W.; Dunn, W.; Vaughan, A. A.; Simpson, K.; Brown, M.; Priest, L.; Blackhall, F. H.; Whetton, A. D.; Dive, C.; Goodacre, R. *Anal. Chem.* **2011**, 83, 6689– 669.

22. Lauridsen, M.; Hansen, S. H.; Jaroszewski, J. W.; Cornett, C. *Anal. Chem.* **2007**, *79 (3)*, 1181-1186.

23. Gribbestad, I. S.; Fjösne, H. E.; Haugen, O. A.; Nilsen, G.; Krane, J.; Petersen, S. B.; Kvinnsland, S. . *Anti. cancer Res.* **1993**, *13,* 1973–1980.

24. Beckonert, O.; Monnerjahn, J.; Bonk, U.; Leibfritz, D. *NMR Biomed.* **2003**, 16, 1–11.

25. Sitter; B.; Bathen, T.; Hagen, B.; Arentz, C.; Skjeldestad, F. E.; Gribbestad, I. S. *Magn. Res. Mater. Phy.* **2004,** *16*, 174-181.

26. Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2010**, *5*, 1019–1032.

27. Li, M.; Song, Y.; Cho, N.; Chang, J.; Koo, H. R.; Yi, A.; Kim, H.; Park, S.; Moon W. K. *PLoS ONE* **2011**, *6,* e25563.

28. Schenetti, L.; Mucci, A.; Parenti, F.; Cagnoli, R.; Righi, V.; Tosi, M. R.; Tugnoli, V. *Concept. Magnetic Res.* **2006**, *28A*, 430-443.

29. Tate, A. R.; Crabb, S.; Griffiths, J. R.; Howells, S. L.*;* Mazucco, R. A.; Rodrigues, L. M.; Watson, D. *Anticancer Res.* **1996**, *16*, 1575–1579.

30. Tate A. R.; Griffiths, J. R.; Martinez-Perez, I.; Moreno, A.; Barba, I.; Cabanas, M. E.; Watson, D.; Alonso, J.; Bartumeus, F.; Isamat, F.; Ferrer, I.; Vila, F.; Ferrer, E.; Capdevilla, A.; Arus, C. *NMR Biomed.* ***1998****,11*, 177–191.

31. Florian, C. L.; Preece, N. E.; Bhakoo, K. K.; Williams, S. R.; Noble M. D. *NMR Biomed.* **1995**, *8*, 253–264.

32. Florian, C. L.; Preece, N. E.; Bhakoo, K. K.; Williams, S. R.; Noble M. D. *Cancer Res.* **1995**, *55*, 420–427.

33. Winder, C. L.; Dunn, W. B.; Schuler, S.; Broadhurst, D.; Jarvis, R.; Stephens, G. M.; Goodacre, R. *Anal. Chem.* 2008, *80*, 2939-2948.

34. Bolten, C. J.; Wittmann, C. *Biotechnol. Lett.* **2008**, *30*, 1993-2000.

35. Faijes, M.; Mars, A. E.; Smid, E. J. *Microb. Cell Fact.* **2007,** *6*, 27.

36. Zang, L.; Frenkel, R.; Simeone, J.; Lanan, M.; Byers, M.; Lyubarskaya, Y. *Anal. Chem.***2011,** *83(13),* 5422–5430.

37. Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G.; *Clin. Cancer Res*. **2009,** *15(2)*, 431-440.

38. Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R. *Anal. Biochem.* **2008**, *372(2)*, 204-212.

39. Lindon, J. C.; Nicholson, J. K. Annu. Rev. *Anal. Chem.* **2008**, *1*, 45-69.

40. Liu, M. L.; Mao, X. A.; Ye, C. H.; Huang, H.; Nicholson, J. K.; Lindon, J. C. *J. Mag. Resonan.* **1998**, *132*, 125-129.

41. Hoult, D. I. J. *Magn. Reson.**1976, 21,*** 337- 347.

42. Ogg, R. J.; Kingsley, P. B.; Taylor, J. S. *J. Magn. Reson.***1994**,*104(1),* 1-10.

43. Simpson, A. J;. Brow, S. A. J. Magn. Reson., **2005,** *175,* 340–346

44. Hwang, T. L.; Shaka, A. J. J. Mag. Resonan. *A* **1995**,*112*, 275-279.

45. Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant R. D.; Lindon, J. C. *Anal. Chem.* **1995**, *67*, 793-811.

46. Meiboom, S., Gill, D. *Rev.Sci. Instrum.* **1958**, *29,* 688–691.

47. Kessler, H.; Oschkinat, H.; Griesinger, C. J. Magn. Reson. 1986, 70, 106−133.

48. Sandusky, P.; Raftery, D. *Anal. Chem.* **2005**, *77*, 7717-7723.

49. Sandusky, P.; Raftery, D. *Anal. Chem.* **2005**, *77*, 2455-2463.

50. Nicholson, J. K.; Wilson, I. D. *Prog. Nucl. Mag. Res. Sp*. **1989**, *21*, 449-501.

51. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2*, 2692-2703.

52. Bodenhausen, G.; Ruben, D.J. *Chem. Phys. Lett.* **1980**, *69,* 185–189.

53. Koskela, H.; Heikkila¨, O.; Kilpela¨inen, I.; Heikkinen, S. *J. Magn. Reson*. **2010**, *202*, 24–33.

54. McKenzie,J. S.; Charlton, A. J.; Donarski, J. A.; MacNicoll, A. D.; Wilson, J. C. *Metabolomics,* 2010, *6*, 574–582.

55. Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009**, *81*, 4882-4888.

56. Ye, T.; Zhang, S.; Mo, H.; Tayyari, F.; Gowda, G. A.; Raftery, D. *Anal. Chem.* **2010**, *82*, 2303-2309.

57. Gowda, G. A. N.; Tayyari, F.; Ye, T.; Suryani, Y.; Wei, S. W.; Shanaiah, N.; Raftery, D. *Anal. Chem.* **2010**, *82*, 8983-8990.

58. Bertram, H. C.; Malmendal, A.; Petersen, B. O.; Madsen, J. C.; Pedersen, H.; Nielsen, N. C.; Hoppe, C.; Molgaard, C.; Michaelsen, K. F.; Duus, J. O. *Anal. Chem.* **2007**, *79*, 7110-7115.

59. Bernini, P.; Bertini, I.; Luchinat, C.; Nepi, S.; Saccenti, E.; Schäfer, H.; Schütz, B.; Spraul, M.; Tenori, L. *J. Proteome Res.* **2009**, *8*, 4264-4271.

60. Bergeron, S. J.; Henry, I. D.;Santini, R. E.; Aghdasi, A.; Raftery, D. Magn. *Reson. Chem.* **2008**, *46*, 925-929.

61. Henry, I. D.; Park, G. H. J.; R. Kc; Tobias, B.; Raftery, D. Concepts in Magnetic Resonance Part B-Magnetic Resonance Engineering 2008, 33B, 1-8.

62. Kc, R.; Henry, I. D.; Park, G. H. J.; Raftery, D. *J. Magn. Reson.* **2009**, 197, 186-192.

63. Lindon, J. C.; Nicholson, J. K.; Wilson, I. D. *J. Chromatogr. B* **2000**, *748*, 233-258.

64. Chikayama, E.; Suto, M.; Nishihara, T.; Shinozaki, K.; Kikuchi, J. *PLoS One* **2008**, *3*, e3805.

65. Cuperlovic-Culf, M.; Barnett, D. A.; Culf, A. S.; Chute, I. *Drug Discov. Today* **2010***, 15(15),* 610-621.

66. Shanaiah, N.; Desilva; M. A.; Gowda, G. A. N.; Raftery, M. A.; Hainline, B. E.; Raftery, D. P. *Natl. Acad. Sci. USA* **2007**, *104*, 11540-11544.

67. Zhang, S. C.; Gowda, G. A. N.; Ye, T.; Raftery, D. *Analyst* **2010**, *135*, 1490-1498.

68. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C., *Anal. Chem.*, **2006**, *78*, 2262–2267.

69. Monleón, D.; Morales, J. M.; Gonzalez-Segura, A.; Gonzalez-Darder, J. M.; Gil-Benso, R.; Cerdá-Nicolás, M.; López-Ginés, C. *Cancer Res.* **2010**, *70(21)*, 8426 – 8434.

70. Zhang, J.; Liu, L.; Wei, S.; Gowda, G. A. N.; Hammoud, Z.; Kesler, K. A.; Raftery, D. *J. Thorac. Cardiovasc. Surg.* **2011**, *141(2)*, 469– 475.

71. Zhang, J.; Bowers, J.; Liu, L.; Wei, S.; Gowda, G. A.; Hammoud, Z.; Raftery, D. *PLoS One* **2012**, *12*, e30181.

72. Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M. D.; Crockford, D.;Volynkin, V.; Holmes, E.; Davies, D. B.; Nicholson J. K. *Anal. Chem.* **2009**, *81*, 56-66.

73. Lämmerhofer, M.; Weckwerth, W. Front Matter, in Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany **2013**.

74. Spraul, M.; Neidig, P.; Klauck, U.; Kessler, P.; Holmes, E.; Nicholson, J. K.; Sweatman, B. C.; Salman, S. R.; Farrant, R. D.; Rahr, E.; Beddell, C. R.; Lindon, J. C. *J. Pharmaceut. Biomed.* **1994**, *12(10)*, 1215-1225.

75. Ebbels, T. M.; Lindon, J. C.; Coen, M. *Methods Mol. Biol.* **2011**, *708*, 365-388.

76. Van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. *BMC Genomics* **2006**, *7, 142.*

77. Zhang, S.; Zheng, C.; Lanza, I. R.; Nair, S.; Raftery, D.; Vitek, O. *Anal. Chem.* **2009**, *81,*6080–6088.

78. Krzanowski, W. J. Principles of Multivariate Analysis: A User's Perspective. Revised ed.; Oxford University Press, **1988**.

79. Duda, R. O.; Hart, P. E.; Stork, D. G. Pattern Classification. 2nd ed.; Wiley **2001**.

80. Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. Chemometrics. Wiley- Interscience **1986**.

81. Holmes, E.; Foxall, P. J.; Nicholson, J. K.; Neild, G. H.; Brown, S. M.; Beddell, C. R.; Sweatman, B. C.; Rahr, E.; Lindon, J. C.; Spraul, M.; Neild, P. *Anal. Biochem.* **1994**, *220(2)*, 284-296.

82. Beckwith-Hall, B. M.; Nicholson, J. K.; Nicholls, A. W.; Foxall, P. J.; Lindon, J. C.; Connor, S. C.; Abdi, M.; Connelly, J.; Holmes, E. *Chem. Res. Toxicol.* **1998**, *11(4)*, 260-272.

83. el-Deredy, W. *NMR Biomed.* **1997**, *10(3)*, 99-124.

84. Wold, S.; Sjostrom, M.; Eriksson, L. *Chemometr. Intell. Lab. Syst.* **2001**, *58*, 109-130.

85. Barker, M.; Rayens,W. *J. Chemometr.* **2003**, *17*, 166-173.

86. Westerhuis, Johan, A.; Hoefsloot, H. C. J.; Smith, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J.; van Duijnhoven, J. P. M.;van Dorsten, F. A. *Metabolomics* **2008**, *4*,81–89.

87. Anderssen, E., Dyrstad, K., Westad, F., & Martens, H. Reducing over-optimism in variable selectio n by cross-model validation. *Chemometr. Intell. Lab.* **2006**, *84(1–2)*, 69–74.

88. Stretch, C.; Eastman, T.; Mandal, R.; Eisner, R.; Wishart, D. S.; Mourtzakis, M.; Prado, C. M.; Damaraju, S.; Ball, R. O.; Greiner, R.; Baracos, V. *J. Nutr.* **2012**, *142*,14-21.

89. Gu, H.; Pan, Z.; Xi, B.; Asiago, V.; Musselman, B.; Raftery, D. *Anal. Chim. Acta.* **2011**,*686*, 57–63.

90. Guan, W.; Zhou, M.; Hampton, C. Y.; Benigno, B. B.; Walker, L. D.; Gray, A.; McDonald, J. F.; Fernández, F. M. *BMC Bioinform.* **2009,** *10*, 259.

91. Goodpaster, A. M.; Romick-Rosendale, L. E.; Kennedy, M. A. *Anal. Biochem.* **2010**, *401*, 134-143.

92. Bonferroni, C. E. *In Studi in Onore del Professore Salvatore Ortu Carboni* **1935**, 13-60.

93. Bonferroni, C. E. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **1936**, *8*, 3-62.

94. Benjamini, Y.; Hochberg, Y. *J. Roy. Stat. Soc. B.* **1995**, *57(1)*, 289–300.

95. Storey J. D. *J. Roy. Stat. Soc. Ser. B Met.* **2002**, *64*, 479-498

96. Metz, C. E. *Semin. Nucl. Med.* **1978**, *8*, 283-298.

97. Raubertas, R. F.; Rodewald, L. E.; Humiston, S. G.; Szilagyi, P. G. *Med. Decis. Making* **1994**, *14*, 169-174.

98. Zhu, W.; Zeng, N.; Wang, N. **2010,** http://www.nesug.org/Proceedings/nesug10/hl/hl07.pdf. Access Aug. 1st, 2013.

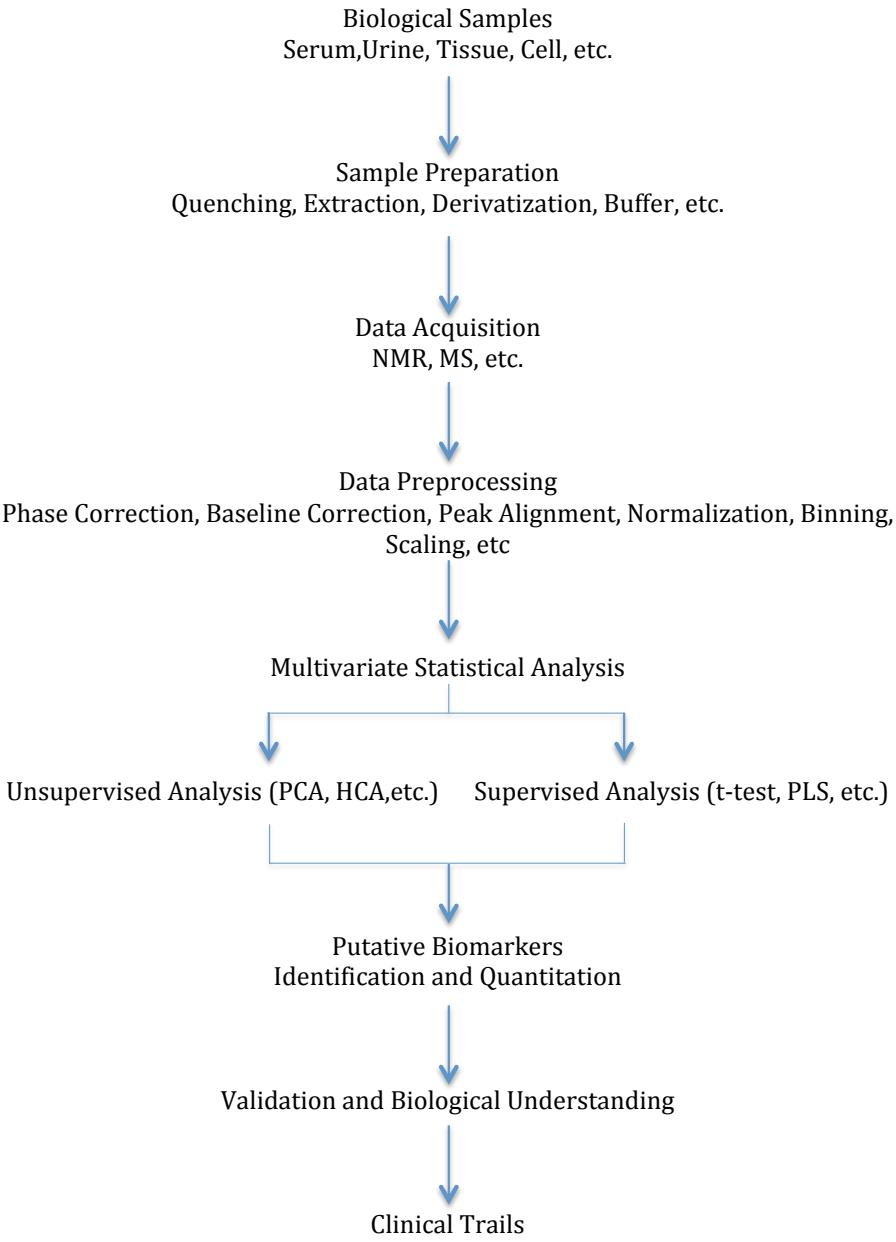


Figure 1.1 The general scheme of metabolic profiling to for disease biomarkers discovery

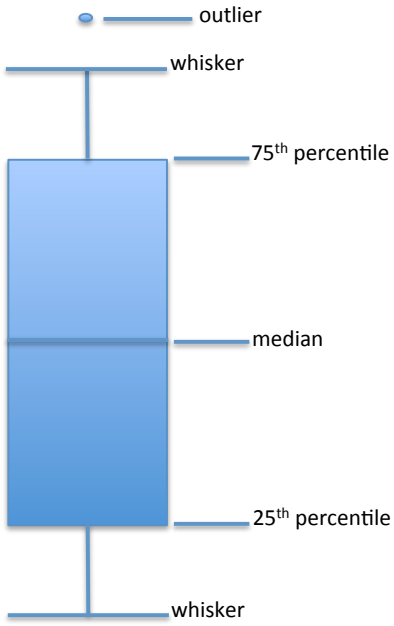


Figure 1.2 Box-and-whisker plot. Horizontal line in the middle portion of the box indicates the median value; bottom and top boundaries of the box indicate the lower and upper quartiles, respectively; whiskers at the bottom and top are 5th and 95th percentiles, respectively, and circle indicates an the outlier.

CHAPTER 2. QUANTITATIVE ANALYSIS OF BLOOD PLASMA METABOLITES USING ISOTOPE ENHANCED NMR METHODS

2.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is increasingly used in metabolomics for the analyses of multiple metabolites in biofluids and tissue. Metabolomics promises a number of important applications in biomedicine including a more detailed understanding of biological processes, the discovery of the biomarkers associated with numerous diseases, pharmaceutical development and toxicology.[1-6] It is increasingly recognized that NMR is a very attractive methodology for quantitative metabolomics because of its high reproducibility and quantitative nature. In particular, the improved resolution of two-dimensional (2D) NMR methods is considered very useful for metabolomics applications[7-11]. However, a major drawback of 2D NMR is that the cross-peak volume in the spectrum is influenced by numerous experimental or intrinsic parameters including the non-uniform excitation profile of the radio frequency pulses, number and duration of the pulses, inter- pulse delays, relaxation times and the magnitude of indirect spin-spin couplings. The high sensitivity of peak intensities (or volumes) to these parameters has limited the use of 2D (particularly ^{1}H homonuclear 2D) experiments for quantitative analysis in metabolomics.

To overcome the limitations of quantitation and to improve the analysis, new higher resolution 2D approaches utilizing ^{1}H-^{13}C heteronuclear 2D experiments (HSQC)

have been proposed.[12,13] One approach is to utilize the information from 2D ^{1}H-^{13}C NMR spectra of standard compounds obtained under identical conditions and relate the peak heights in the samples to standard mixtures[12]; another is to utilize calibration curves obtained using ^{1}H-^{13}C HSQC spectra for individual metabolites to determine the metabolite concentrations.[13] Most recently, a more general approach that does not require measurements of standard compounds was proposed.[14] This method utilizes correction factors derived theoretically from the solution of the Bloch equations and the analysis of product operator formalism incorporating longitudinal (T_1) and transverse (T_2) relaxation parameters, ^{1}H-^{13}C heteronuclear J-coupling and various delays used in the pulse sequence.

A major drawback of using the ^{1}H-^{13}C HSQC experiment for quantitative analysis, is the lack of sensitivity arising from low metabolite concentration and natural abundance of ^{13}C (1.1 % natural abundance). To compensate this limitation, unusually long acquisition times (nearly 10 hrs or more) are typically required since the NMR sensitivity scales with the square-root of the number of scans. Moreover, although ^{1}H-^{13}C HSQC greatly enhances resolution when compared with 1D NMR, given the complexity of the biological samples, the resolution obtainable from a single 2D experiment is not always adequate for analyzing a large number of metabolites.

In the present study, with the goal of circumventing the current drawbacks of limited resolution and sensitivity, we utilize a combination of isotope tagging approaches and 2D NMR methods to accurately analyze human plasma metabolites. A number of the most common metabolites in blood plasma were quantified using this approach after validating the experimental protocols using a mixture of synthetic compounds.

Metabolites containing carboxyl and amino groups were tagged with ^{15}N or ^{13}C, respectively, before detection by 2D NMR. We have recently shown the proof-of-principle approaches to introduce isotope tags using simple chemical derivatization methods and that the NMR spectra of the tagged metabolites improve both resolution and sensitivity.[15-18] The combination of advanced isotope tagging methods with conventional 1D and 2D NMR methods as described in the present study enables the quantitative analysis of a large number of metabolites in human blood on a routine basis.

2.2 Materials and Methods

2.2.1 Chemicals and Samples

Twenty-eight metabolite standards (Table 2.1), 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), maleic acid, ethanolamine (all from Sigma–Aldrich, St. Louis, MO), 4-(4,6- dimethoxy [1,3,5] triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM) (Acros, Geel, Belgium), ^{15}N-ethanolamine, ^{13}C-formic acid (Cambridge Isotope Laboratories, Andover, MA), N, N-dicyclohexylcarbodiimide and N-hydroxysuccinimide (Sigma-Aldrich) were used without further purification. An ultra-pure primary quantitative standard, tris(hydroxymethyl)aminomethane, (99.9%) was obtained from Mallinckrodt Baker Inc. Phillipsburg, NJ. Human blood plasma (10 × 1 mL) was procured from the National Institute of Standards and Technology (NIST, Gaithersburg, MD). Frozen plasma samples were transported to Purdue under dry ice and stored at −80 °C until used for the analysis.

2.2.2 Calibration of the Standard Solutions

Twenty-eight metabolites that commonly occur in human blood plasma were selected based on entries in the human metabolite database (HMDB),[19] the analysis of isotope□labeled ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC spectra, as well as 1D and 2D ^{1}H-^{1}H TOCSY spectra of a human plasma sample. Stock solutions (20 mM) for the synthetic analogues of all these 28 metabolites (Table 2.2) and internal standards, DSS (5 mM), maleic acid (20 mM) and ethanolamine (20 mM), were prepared. The concentration of the DSS solution was calibrated using ^{1}H NMR against a primary stoichiometric standard, tris(hydroxymethyl)aminomethane (22.4 mM), prepared in the lab. The calibrated DSS solution was then used to calibrate all other standard solutions. Briefly, solutions of the synthetic analogues of the 28 metabolites and internal standards were divided into 5 groups as shown in Table 2.3. The grouping of samples in Table 2.3 was such that the ^{1}H NMR peaks for at least one proton from each compound in the group were isolated for the measurement of the peak integral. For each group, the one-dimensional (1D) ^{1}H NMR spectrum was recorded, and based on the integrated area of the isolated peaks with reference to DSS, exact concentrations of the standard solutions were determined (Table 2.2).

2.2.3 Mixture Analyzing

A mixture of the 28 synthetic analogues of the metabolites was prepared using the stock solutions such that the final concentration of each standard compound was matched approximately to its expected concentration in human blood plasma (Table 2.1).[19] From

this mixture, three identical sets of samples (Set 1, Set 2 and Set 3) were prepared, Figure 2.1 shows a flow diagram illustrating the steps followed in the analysis of the standard metabolite mixture using isotope enhanced NMR methods; each set consisted of four solutions, 2× 500 µL and 2× 1000 µL of the mixture. All solutions were then dried under vacuum. To the samples from Set 1, maleic acid solution (92 nmol) was added as an internal reference and the solutions were diluted to 500 µL using doubly distilled water. The carboxylic acid class of metabolites was then tagged with ^{15}N-labeled ethanolamine following the established procedure.[17] To each sample from Set 2, ethanolamine solution (200 nmol) was added as an internal reference and the solution diluted to 500 µL using doubly distilled water. Amines and amino acids were subjected to ^{13}C isotope tagging using a ^{13}C-formic acid reaction following the established procedure.[18] Finally, to the samples from Set 3, DSS (9.44 nmol) was added as an internal reference and reconstituted in 560 µL of phosphate buffer (pH=7.4) in D_2O and transferred to 5 mm NMR tubes for 1D and ^{1}H-^{1}H 2D TOCSY NMR experiments.

2.2.4 General Procedure for Tagging Metabolites

2.2.4.1 ^{15}N-Ethanolamine Tagging Procedure

3 µL ^{15}N-ethanolamine (50 µmol) was added to the sample, the pH adjusted to 7.0 with 1 M HCl and DMT-MM (21 mg) was added to initiate the reaction.[20,21] The mixture

was continuously stirred at room temperature for 4 hrs to complete the reaction. The pH was then adjusted to 5.0 by adding 1 M HCl or NaOH and the solutions were diluted to 600 μL by adding water or D_2O prior to detection by ^{1}H-^{15}N 2D NMR.[22,17]

2.2.4.2 ^{13}C –formic Acid Tagging Procedure

2 μL of ^{13}C-formic acid (0.05 mmol) and 5 mg of N-hydroxysuccinimide (0.04 mmol) were dissolved in 100 μL tetrahydrofuran. 9 mg of N, N-dicyclohexylcarbodiimide (0.04 mmol) in 50 μL tetrahydrofuran was added to the mixture and stirred at room temperature.[23] After 15 min, the reaction mixture was centrifuged to remove insoluble urea; the supernatant containing ^{13}C-N-formyloxysuccinimide was then added to the mixture of synthetic analogues of the metabolites along with 50 μL of 2 M $NaHCO_3$ (0.1 mmol) aqueous solution. The reaction was stirred at room temperature for 4 h and dried under vacuum. The residue was redispersed in 500 μL D_2O, the pH was adjusted to 7.0 by adding 1M HCl and then transferred to a standard 5mm NMR tube for analysis using ^{1}H-^{13}C 2D NMR.[18]

2.2.5 Sample Prepration for NMR Experiments

2.2.5.1 Plasma Deproteination

Cold methanol (4° C; 9.6 mL) was added to 4.8 mL of the NIST plasma, vortexed, and then kept for 30 min at -20 °C. The precipitated protein pellet was removed after centrifuging at 13,200 g for 10 min. The supernatant was divided into 12 equal parts and divided into three groups, each group consisting of four samples. Flow chart show in

Figure 2.2 depicting the steps followed in the analysis of metabolites in the NIST plasma sample using isotope enhanced NMR methods.

2.2.5.2 Isotope Tagging of Plasma Metabolites and Sample Preparation for NMR

In each group, two samples served as controls and the remaining two were spiked with 400 μL of the stock solution mixture of 28 synthetic samples. All three groups of samples were then dried in vacuum. One group was used to label metabolites containing carboxylic acid groups with ^{15}N-ethanolamine, and the second group was used to label metabolites containing amine groups with ^{13}C-formic acid, after the addition of internal standards, either maleic acid or ethanolamine, appropriately. Identical procedures were used for isotope tagging the plasma metabolites. To the third group of samples, DSS (9.44 nmol) was added as an internal reference and reconstituted in 560 μL of phosphate buffer (pH=7.4) in D_2O. Samples from all the three preparations were transferred to 5 mm NMR tubes after adjusting the pH and solution conditions as described earlier for the mixture of standards.

2.2.6 NMR Experiments

NMR experiments were performed at 298 K on a Bruker Avance-III-800 equipped with a room temperature ^{1}H inverse detection Z-gradient probe or a Bruker DRX-500 spectrometer equipped with a ^{1}H inverse detection Z-gradient cryo-probe. 1D NMR experiments for the five groups of standard samples (Table 2.3), the mixture of 28 synthetic analogues, and the plasma extracts were performed using a one pulse sequence with residual water signal suppression by pre-saturation during relaxation delay. Thirty-two scans with 64k time domain data points were collected with a sufficiently long

recycle delay (20 s) to ensure complete recovery of the magnetization between scans. For the ^{15}N isotope tagged samples, ^{1}H -^{15}N 2D HSQC experiments were performed employing an INEPT transfer delay of 5.5 ms corresponding to a $^{1}J_{NH}$ of 90 Hz. Spectral widths of approximately 10 kHz in ^{1}H and 5 kHz in ^{15}N dimensions were used for the 800 MHz experiments. For ^{13}C isotope tagged samples, sensitivity-enhanced ^{1}H-^{13}C 2D HSQC experiments were performed employing an INEPT transfer delay of 2.5 ms corresponding to a $^{1}J_{C\text{-}H}$ of 200 Hz. Spectral widths of approximately 10 kHz for the ^{1}H dimension and 600 Hz for ^{13}C were used at 800 MHz. For both ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC 2D experiments, 128 free induction decays were collected along the indirect (t_1) dimensions using 4 transients per increment and 2s or 3s recycle delay, resulting in a total acquisition time of 18 min for the ^{1}H-^{15}N HSQC and 28 min for the ^{1}H-^{13}C HSQC. Phase-sensitive data were obtained using echo-anti-echo mode with nitrogen (for ^{1}H-^{15}N HSQC) or carbon (for ^{1}H-^{13}C HSQC) decoupling during acquisition (t_2 dimension) using the GARP (Globally Optimized Alternating-phase Rectangular Pulses) sequence. ^{1}H-^{1}H 2D TOCSY experiments were performed for the neat (non-derivatized) samples with a spectral width of 6 kHz (500 MHz) or 12 kHz (800 MHz) in both the dimensions. The residual water signal was suppressed by presaturation. 400 free induction decays were collected with t_1 increments using 8 transients per increment and 2s recycle delay, resulting in a total acquisition time of 116 min (500 MHz) or 111 min (800 MHz).

All 1D data were Fourier transformed with a 0.3 Hz line broadening function. The 2D data were zero-filled to 1,024 points in the t_1 dimension after forward linear prediction to 512 points and Fourier-transformed after multiplying by a squared sine-bell

window function shifted typically by π/4 or π/2 along both the dimensions. All NMR data were processed with Bruker Topspin 2.0 on a Redhat Linux platform and Bruker XWINNMR 3.5 on a SGI / IRIX platform. An automatic baseline correction using a polynomial of degree 5 was used to correct the baseline in both 1D and 2D spectra.

Peaks in the 1D and 2D NMR spectra were assigned to various metabolites based on literature reports.[17-19] Integrals for well resolved peaks in 1D and 2D spectra were obtained with respect to the peak for the internal standard DSS, maleic acid or ethanolamine. Integral limits for each peak in the 2D spectra were selected such that the selected region encompassed the whole peak and that no other peak interfered with the selection. Once chosen for each type of 2D spectrum, the same sets of integral limits were used for all other samples. Concentrations of the plasma metabolites were determined by comparing the peak integrals from the spectra obtained with and without spiking with the synthetic analogues, and also by directly comparing the peak integrals of the plasma metabolites with those from the standards. The accuracy, reproducibility and errors were estimated from two to eight measurements, depending on the detection of the resolved peak for a particular metabolite in one or more types of spectra, for both synthetic mixtures and plasma samples. The ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC and ^{1}H-^{1}H TOCSY experiments and the data analyses were performed by independent persons.

2.3 Results

The standard solutions of the synthetic analogues of the plasma metabolites and internal standards (maleic acid and ethanolamine), prepared based on their weights, were calibrated using 1D ^{1}H NMR. The actual concentration of the standard solutions prepared

based on the weight varies depending on the purity and hygroscopic nature of the compounds and hence, it is important to calibrate the standard solutions especially for accurate quantitative analysis. The DSS solution, which was first calibrated using a primary stoichiometric standard, tris(hydroxymethyl)aminomethane, was used for calibrating all the standard solutions (Table 2.3). The difference between the concentrations determined based on sample weight and calibration using 1D NMR varied as much as 10% for all but three metabolites, which varied up to nearly 20% (Table 2.2) due to hygroscopic nature of the metabolites or sample impurities.

2.3.1 Analysis of Synthetic Metabolite Mixture

A mixture of 28 metabolites was analyzed using both ^{15}N and ^{13}C isotope tagging approaches. Fig. 2.3 shows 2D spectra of the mixture of 28 compounds with ^{15}N and ^{13}C isotope tagging, as well as without tagging. The integrated 2D peak volumes were obtained and then used to calculate the metabolite concentrations. Fig. 2.4 shows the concentration of the compounds thus determined. As can be seen in the figure, an excellent match between the metabolite concentrations derived from NMR methods and the actual values was obtained. Further, as shown in Figure 2.5, a correlation of the NMR derived values with the expected values showed a very good agreement for all low and high concentration metabolites ($R^2 > 0.99$).

2.3.2 Quantitation of Plasma Metabolites

The 1D ^{1}H NMR spectrum of the plasma sample obtained without isotope labeling is highly complex, with only a relatively small number of metabolite signals being isolated from other signals as shown in figure 2.6. 2D HSQC spectra of plasma samples tagged with ^{15}N and ^{13}C isotopes provide resolved peaks for a much larger number of carboxylic acid and amine containing metabolites. Fig. 2.7 shows 2D spectra of the plasma obtained with and without ^{15}N or ^{13}C tagging. The 2D TOCSY spectrum of the same plasma sample also provided a number of well resolved peaks (Fig. 2.7c). However, unlike the HSQC spectra, the TOCSY spectrum showed a number of redundant peaks for the same metabolite, which increases the complexity of the spectrum.

Quantitation of the plasma metabolites followed an identical procedure used for the determination of the concentrations of metabolites in the synthetic mixture. The integrated peak areas/volumes in the 1D/2D spectra of the neat and the spiked plasma samples were obtained and the metabolite concentrations determined. Twenty-seven metabolites that were identified in human plasma were analyzed in duplicate measurements with and without ^{15}N and ^{13}C isotope tagging. Fig. 2.8 depicts the concentration of the metabolites thus determined. The ^{1}H, ^{13}C and ^{15}N chemical shifts for the blood plasma metabolites analyzed in this study are shown in Table 2.4 and the derived concentrations using a combination of four different NMR methods is shown in Table 2.5. Further, the concentration of the carboxylic acid and amino metabolites were also determined by directly comparing the 2D peak integrals with those for the corresponding standard compound. Comparison of the metabolites concentration

determined with and without spiking is shown in Fig. 2.8. Notably, the values determined from both approaches agree well.

2.5 Discussion

^{1}H NMR spectroscopy is an attractive tool for the quantitative analysis of multiple metabolites from intact biological samples. Considering its ease of use, reproducibility, and high- throughput capabilities, 1D ^{1}H NMR spectroscopy is often used for metabolomics-based studies. However, it is challenging to analyze the 1D NMR spectrum of plasma for absolute quantitation as it contains a large number of overlapping signals due to hundreds of metabolites present at variable concentrations. The multiplicity of the signals due to *J*-coupling makes 1D ^{1}H NMR spectra of plasma particularly challenging. The interference from macromolecules such as proteins and lipids adds to the complexity and causes baseline distortions in the spectra. Such overlap and baseline issues substantially affect the accuracy of the quantitative analysis using 1D NMR. To offset such limitations, a majority of the studies that use 1D NMR resort to comparisons of the relative intensities of the 1D NMR signals between disease and healthy samples. While the use of relaxation edited techniques such as the Carr-Purcell-Meiboom-Gill (CPMG) experiment, serum/ plasma deproteinization, and line fitting approaches significantly improves the analysis of metabolites,[24] such methods are not ideal. Diffusion-sensitized 1D NMR spectroscopy, which uses data from two separate 1D experiments, one obtained using low diffusion gradients and the other using high gradients to suppress macromolecular background signals effectively, was shown to be

useful for the quantitative analysis of blood plasma metabolites.[25] However, spectral overlap still significantly limits the number of metabolites that can be analyzed. 2D NMR promises quantitative analysis of a large number of metabolites on a routine basis. An important requirement is that the cross-peaks in 2D spectra should be devoid of overlaps for reliable quantitative results; however, this criterion is not often met for a large number of metabolites by a single 2D experiment due to the extremely high complexity of plasma. The advantage of the new 2D NMR approaches used here is that the use chemoselective isotope tags greatly reduces the complexity of the spectra, since only a single peak is observed for the metabolites with a single functional group (see Figures 2.3 and 2.7). The reduced complexity of the spectra due to the absence of less interesting chemical signals is particularly important for the analysis of low- concentration metabolites (Figures 2.4 and 2.8). This method, however, does not work for the analysis of lipoproteins, which represents a major class of metabolites in blood plasma and which have been effectively analyzed using a multivariate deconvolution approach.[26,27]

An important criterion for the quantitative analysis method to be robust is that it does not require the use of spiking standards for each sample. To test this, we also determined the concentration of ^{15}N and ^{13}C isotope tagged metabolites in plasma by comparing the 2D peak integrals with those from the synthetic analogues. It may be interesting to note that, as shown in Figure 2.9, the values determined using both ^{15}N and ^{13}C isotope tagging agree well with those determined on the basis of spiking with synthetic analogues. Therefore, it is sufficient to obtain the integral for each synthetic analogue only once, which can be used for the analysis of any number of samples. Utilization of 2D HSQC experiments involving the isotopes has the additional advantage

since both the magnitude of the coupling and the relaxation properties of the nuclear pairs (^{15}N/ ^{13}C and ^{1}H) do not appreciably vary across the metabolites of interest and, hence, provide the relative cross peak intensities that are less sensitive to instrumental settings.

In this study, we quantified 27 metabolites with an average CV of 2.4% for 17 metabolites and 5.6% when all the metabolites were considered. When the results from all the four NMR methods were combined for the same metabolites, the average CV's were 4.8% and 8.7%, respectively. We note that, as the metabolite library expands, we can quantify additional metabolites from the same and already acquired 2D data by comparison of the peak integrals with those from the standards. Mass spectrometry (MS), another very useful method for quantitative analysis, is highly sensitive and provides quantitative information on a larger number of metabolites. However, MS invariably involves the combination of a separation method such as gas chromatography or liquid chromatography for accurate analysis and often renders the obtained results to be sensitive to the specific column and separation parameters and especially the ionization conditions. In addition, a standard compound is needed for each quantified metabolite.

In conclusion, this investigation presents quantitative analysis of over 25 plasma metabolites using ^{15}N and ^{13}C isotope tagging methods. Carboxylic acids and amines represent a majority of the metabolites in body fluids, and their analysis by isotope tagging significantly enhances the detectable metabolic pool for biomarker discovery applications. The combination of improved sensitivity and resolution and the reduced time required when compared to natural abundance heteronuclear NMR methods are attractive for the routine and accurate analysis of metabolites in complex biological samples. Although, the isotope tagging methods use 2D NMR experiments, each 2D

experiment requires only 30 min or less (~10 min with a cryoprobe), and hence, the approach can be useful for high throughput analysis of human plasma as well as other biological fluids. Further, combination of the isotope tagging approach with the latest advancements in NMR technology, such as detection using microcoil probes, for example, can significantly minimize the volume of biofluid samples required for routine analysis.

2.5 References

1. Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29,* 1181– 1189.

2. Clayton, T. A.; Lindon, J. C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Provost, J. P.; Le Net, J. L.; Baker, D.; Walley, R. J.; Everett, J. R.; Nicholson, J. K. *Nature* **2006**, *440*, 1073–1077.

3. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2* (11), 2692–2703.

4. Pan, Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, *3*87, 525–527.

5. Nagana Gowda, G. A.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Mol. Diagn.* **2008**, *8*, 617–633.

6. Holmes, E.; Wilson, I. D.; Nicholson, J. K. *Cell* **2008**, *13*4 (5), 714–717.

7. Dumas, M-.E.; Canlet, C.; Andre, F.; Vercauteren, J.; Paris, A. *Anal. Chem.* **2002**, *74*, 2261–2273.

8. Tang, H.; Wang, Y.; Nicholson, J. K.; Lindon, J. C. *Anal. Biochem.* **2004**, *325*, 260–272.

9. Dumas, M-.E.; Canlet, C. *J. Proteome Res.* **2005**, *4*, 1493–1502.

10. Xi, Y.; de Ropp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P. *Metabolomics* **2006**, *2*, 221–233.

11. Fonville,J.M.;Maher,A.D.;Coen,M.;Holmes,E.;Lindon,J.C.;Nicholson, J. K. *Anal. Chem.* **2010**, *82*, 1811–1821.

12. Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79* (24), 9385– 9390.

13. Gronwald, W.; Klein, M. S.; Kaspar, H.; Fagerer, S. R.; Nu¨rnberger, N.; Dettmer, K.; Bertsch, T.; Oefner, P. J. *Anal. Chem.* **2008**, *80* (23), 9288– 9297.

14. Rai, R. K.; Tripathi, P.;Sinha, N.*Anal.Chem.***2009**, *81*(24),10232–10238.

15. Shanaiah, N.; Desilva, A.; Nagana Gowda, G. A.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (28), 11540–11544.

16. DeSilva, M. A.; Shanaiah, N.; Nagana Gowda, G. A.; Rosa-Pe´ rez, K.; Hanson., B. A.; Raftery, D. *Magn. Reson. Chem.* **2009**, *47*, S74–S80.

17. Ye, T.; Mo, H.; Shanaiah, N.; Nagana Gowda, G. A.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009**, *81* (12), 4882–4888.

18. Ye, T.; Zhang, S.; Mo, H.; Tayyari, F.; Nagana Gowda, G. A.; Raftery, D. *Anal. Chem.* **2010**, *82* (6), 2303–2309.

19. Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res.* **2007**, *35* (Database issue), D521–D526.

20. Kunishima, M.; Kawachi, C.; Morita, J.; Terao, K.; Iwasaki, F.; Tani, S. *Tetrahedron* **1999**, *55*, 13159-13170.

21. Kunishima, M.; Kawachi, C.; Hioki, K.; Terao, R.; Tani, S. *Tetrahedron* **2001**, *57*, 1551- 1558.

22. Zhang, Y. Z.; Paterson, Y.; Roder, H. *Protein Sci.* **1995**, *4*, 804-814.

23. Hecht, S. M.; Werner, D. *J. Chem. Soc. Perkin Trans. 1*, **1973**, 1903-1906.

24. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. *Anal. Chem.* **2006**, *78*, 4430–4442.

25. De Graaf, R. A.; Behar, K. L. *Anal. Chem.* **2003,** 75, 2100-2104.

26. Otvos, J. D.; Jeyarajah, E. J.; Bennett, D. W. *Clin Chem.* **1991**, *37* (3), 377–386.

27. Otvos, J. D. *Clin Lab.* **2002**, *48*, 171–180.

Table. 2.1 Synthetic analogues of metabolites used for the quantitative analyses of human plasma metabolites

Serial number	Standard compound	Estimated approximate blood plasma concentrations[a] (μM)	Actual NMR calibrated concentrations in the mixture (μM) used for spiking[b]
1	3-hydroxybutyrate	60	59.2 ± 0.9
2	acetate	80	73.7 ± 1.1
3	l-alanine	300	287.1 ± 4.5
4	l-arginine	80	90.1 ± 1.4
5	citrate	30	28.1 ± 0.4
6	creatinine	40	31.6 ± 0.6
7	formate	40	32.5 ± 0.5
8	l-glutamic acid	50	53.8 ± 0.3
9	l-glutamine	300	292.1 ± 2.8
10	l-glycine	200	172.6 ± 2.7
11	l-histidine	80	83.7 ± 0.6
12	l-isoleucine	50	50.7 ± 0.2
13	lactate	1000	959.5 ± 15.1
14	l-leucine	80	80.7 ± 1.2
15	l-lysine	100	75.3 ± 2.3
16	l-methionine	20	19.2 ± 0.03
17	l-phenylalanine	70	71.1 ± 1.1
18	l-proline	100	93.3 ± 1.5
19	l-threonine	100	90.0 ± 1.9
20	L-tryptophan	30	29.1 ± 0.4
21	l-tyrosine	80	80.9 ± 0.3
22	l-valine	200	183.2 ± 1.5
23	succinate	10	9.9 ± 0.1
24	betaine	50	44.6 ± 0.7
25	4-hydroxy proline	50	53.4 ± 0.8
26	l-serine	100	93.3 ± 0.7
27	l-asparagine	40	43.9 ± 0.7
28	taurine	30	27.7 ± 0.1

a Obtained from the combination of database search and comparison of the relative peak integrals in the NMR spectra.
b The errors are standard deviations from two measurements.

Table. 2.2 Concentration of synthetic analogues of human plasma metabolites before and after calibration using ^{1}H NMR

Serial Number	Metabolites	Concentration from the weight (mM)*	Concentration after calibration by 1H NMR (mM)*
1	3-Hydroxybutyrate	20.02	20.69 ± 0.32
2	Acetate	19.62	19.33 ± 0.30
3	L-Alanine	19.67	20.07 ± 0.31
4	L-Arginine	20	23.62 ± 0.37
5	Citrate	20.09	19.67 ± 0.31
6	Creatinine	20.15	16.18 ± 0.30
7	Formate	20.38	17.02 ± 0.26
8	L-Glutamic acid	20.18	22.39 ± 0.12
9	L-Glutamine	20.15	20.33 ± 0.19
10	L-Glycine	20.58	18.09 ± 0.28
11	L-Histidine	20.1	21.83 ± 0.17
12	L-Isoleucine	20.24	21.07 ± 0.09
13	Lactate	20.21	20.12 ± 0.31
14	L-Leucine	20.37	21.16 ± 0.33
15	L-Lysine	20.11	15.98 ± 0.49
16	L-Methionine	19.95	19.96 ± 0.03
17	L-Phenylalanine	20	21.29 ± 0.33
18	L-Proline	20.16	19.57 ± 0.31
19	L-Threonine	19.98	18.95 ± 0.41
20	L-Tryptophan	20	20.36 ± 0.32
21	L-Tyrosine	19.87	21.04 ± 0.08
22	L-Valine	20.07	19.10 ± 0.15
23	Succinate	20.03	20.77 ± 0.33
24	Betaine	20.15	18.69 ± 0.30
25	4-hydroxy-proline	20	22.39 ± 0.35
26	L-Serine	19.98	19.45 ± 0.14
27	L-Asparagine	20.01	23.01 ± 0.36
28	Taurine	19.94	19.17 ± 0.07
30	Ethanolamine	20	20.34 ± 0.95
31	Maleic acid	20.07	20.85 ± 0.31
32	DSS	5	4.70 ± 0.02

Table. 2.3 Synthetic analogues of human plasma metabolites grouped for concentration calibration using ^{1}H NMR

Group 1	Group 2	Group 3	Group 4	Group 5
3-Hydroxybutyrate	L-Arginine	L-Glutamine	Creatinine	Succinate
Acetate	L-Histidine	L-Isoleucine	L-Glutamic acid	L-Serine
L-Alanine	Lactate	L-Threonine	L-Lysine	Taurine
Citrate	L-Phenylalanine	L-Tryptophan	L-Methionine	
Formate	L-Tyrosine	Ethanolamine	L-Proline	
L-Glycine	L-Valine			
L-Leucine	4-Hydroxy-proline			
Betaine	Maleic acid			
L-Asparagine				

Table. 2.4 ^{1}H, ^{13}C, and ^{15}N chemical shifts of the peaks used in the analysis of NIST plasma metabolites

$^{1}H-^{15}N$ HSQC				$^{1}H-^{13}C$ HSQC			
label	metabolite	^{1}H (ppm)	^{15}N (ppm)	label	metabolite	^{1}H (ppm)	^{13}C (ppm)
2	acetate	8.01	120.93	3	l-alanine	7.96	163.4
3	l-alanine	8.23	113.7	10	l-glycine	8.03	164.15
5	citrate	8.04	122.62	11	l-histidine	7.97	163.56
8	l-glutamic acid	8.21	115.58	12	l-isoleucine	8.03	163.81
10	l-glycine	8.12	114.78	18	l-proline	8.1	163.1
11	l-histidine	8.25	116.42	19	l-threonine	8.12	164.12
17	l-phenylalanine	8.13	117.19	20	l-tryptophan	7.88	163.51
19	l-threonine	8.28	117.52	22	l-valine	8.05	163.88
20	l-tryptophan	8.03	116.9	25	4-hydroxy proline	8.04	165.21
21	l-tyrosine	8.18	117.25	26	l-serine	8.07	163.92
22	l-valine	8.3	118.43	27	l-asparagine	8.01	163.57
24	betaine	8.57	124.28	28	taurine	7.99	164.23

^{1}H 1D NMR			$^{1}H-^{1}H$ TOCSY			
label	metabolite	^{1}H (ppm)	label	metabolite	^{1}H (ppm) F2 dimension	^{1}H (ppm) F1 dimension
1	3-hydroxybutyrate	1.19	4	l-arginine	1.68	3.23
2	acetate	1.91	5	citrate	2.65	2.51
3	l-alanine	1.47	11	l-histidine	7.07	7.06
7	formate	8.45	12	l-isoleucine	3.66	0.98
9	l-glutamine	2.13	13	lactate	1.32	1.32
11	l-histidine	7.06	14	l-leucine	0.95	1.7
12	l-isoleucine	1.01	15	l-lysine	3.02	1.48
16	l-methionine	2.13	19	l-threonine	3.57	1.33
17	l-phenylalanine	7.42	20	l-tryptophan	7.53	7.72
20	l-tryptophan	7.74	21	l-tyrosine	6.9	7.18
21	l-tyrosine	6.89	22	l-valine	3.6	2.27
22	l-valine	1.03				

Table. 2.5 NIST plasma metabolite concentrations obtained using a combination of NMR experiments with or without isotope tagging

label	metabolite	NIST plasma concentration (μM)[a]	label	metabolite	NIST plasma concentration (μM)a
1	3-hydroxybutyrate[e]	99.3 ± 13.1	16	l-methionine[e]	16.6 ± 2.8
2	acetate[be]	142.0 ± 3.0	17	l-phenylalanine[be]	50.6 ± 3.8
3	l-alanine[bce]	279.4 ± 18.9	18	l-proline[c]	127.6 ± 13.1
4	l-arginine[d]	155.9 ± 16.0	19	l-threonine[bcd]	107.3 ± 16.7
5	citrate[bd]	40.1 ± 2.3	20	l-tryptophan[bcde]	45.9 ± 8.1
7	formate[e]	51.2 ± 2.1	21	l-tyrosine[bde]	58.9 ± 7.2
8	l-glutamic acid[b]	69.3 ± 5.3	22	l-valine[bcde]	159.7 ± 11.6
9	l-glutamine[e]	368.5 ± 2.3	24	betaine[b]	27.3 ± 2.8
10	l-glycine[bc]	204.5 ± 31.2	25	4-hydroxy proline[bc]	11.5 ± 1.3
11	l-histidine[bcde]	63.1 ± 5.7	26	l-serine[c]	95.8 ± 15.0
12	l-isoleucine[cde]	48.2 ± 2.4	27	l-asparagine[c]	33.4 ± 2.8
13	lactate[d]	2403.6 ± 127.6	28	taurine[c]	32.4 ± 0.8
14	l-leucine[d]	100.1 ± 0.1	29	glucose[e]	8778.5 ± 62.8
15	l-lysine[d]	190.8 ± 21.9			

a The errors are standard deviations.
b Obtained from $^{1}H-^{15}N$ HSQC.
c Obtained from $^{1}H-^{13}C$ HSQC.
d Obtained from $^{1}H-^{1}H$ TOCSY.
e Obtained from ^{1}H 1D NMR.
Two samples were used for each type of experiment resulting in two, four, six, or eight independent measurements for each metabolite.

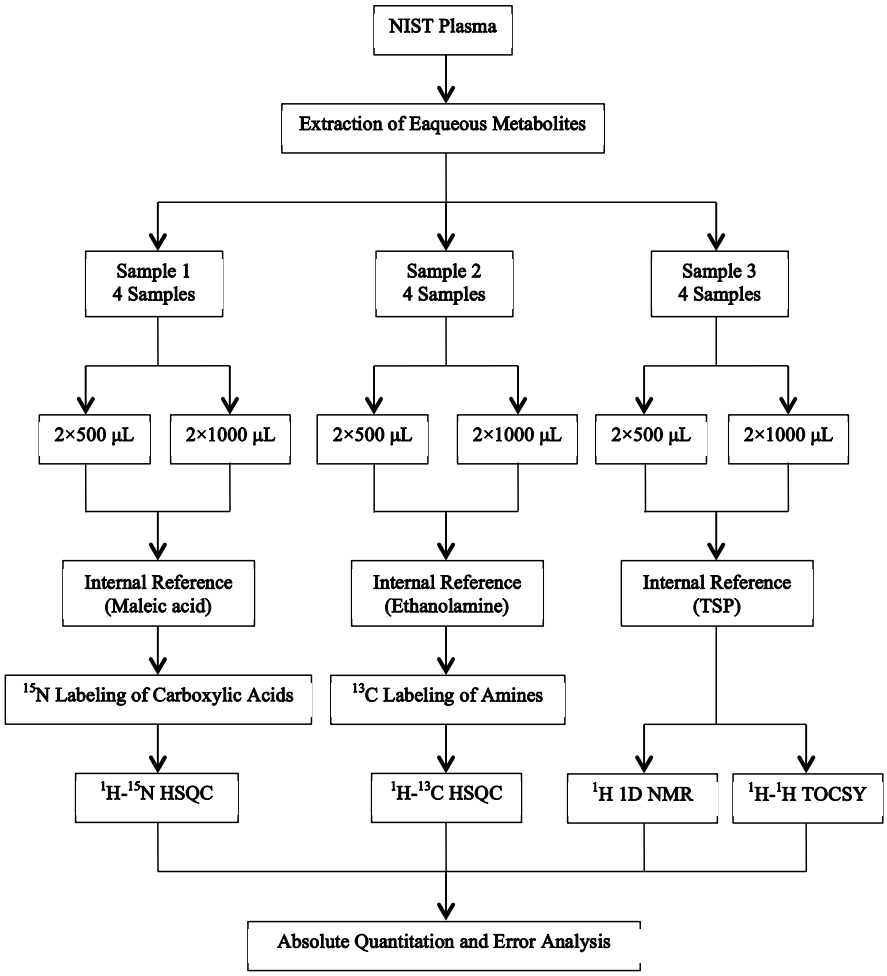


Figure 2.1 Flow chart depicting the steps followed in the analysis of the standard metabolite mixture using isotope enhanced NMR methods.

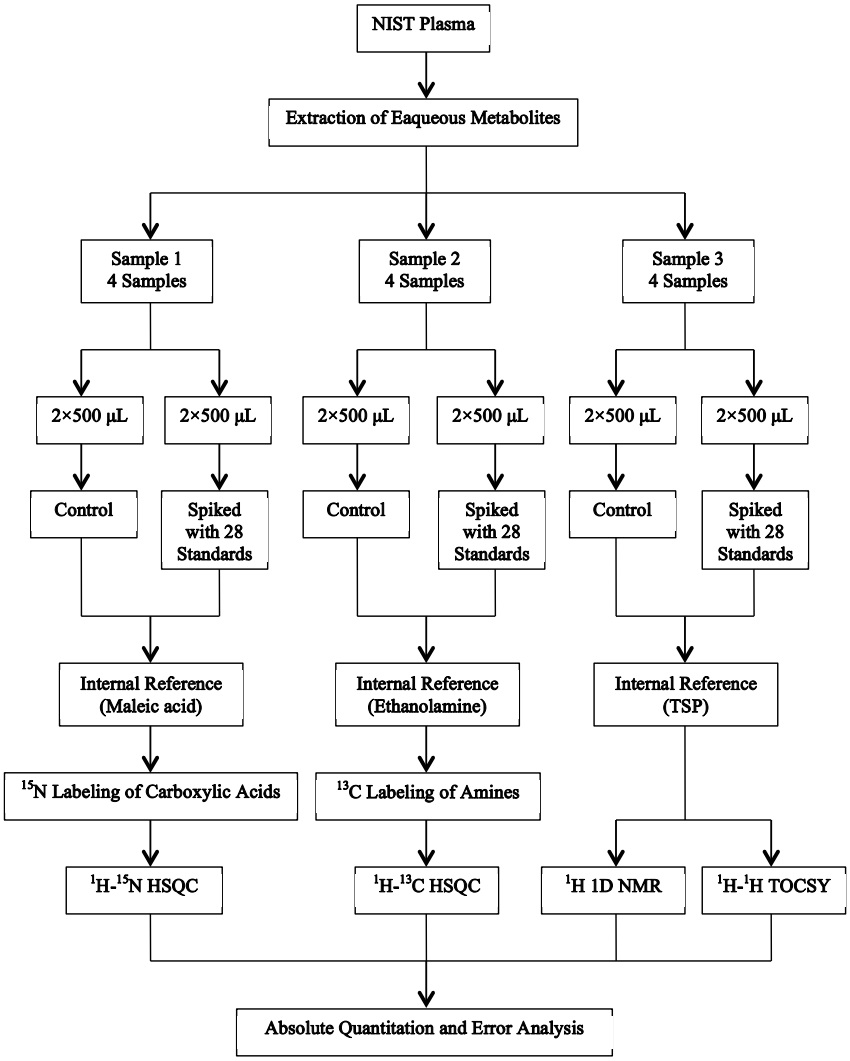


Figure 2.2 Flow chart depicting the steps followed in the analysis of metabolites in the NIST plasma sample using isotope enhanced NMR methods.

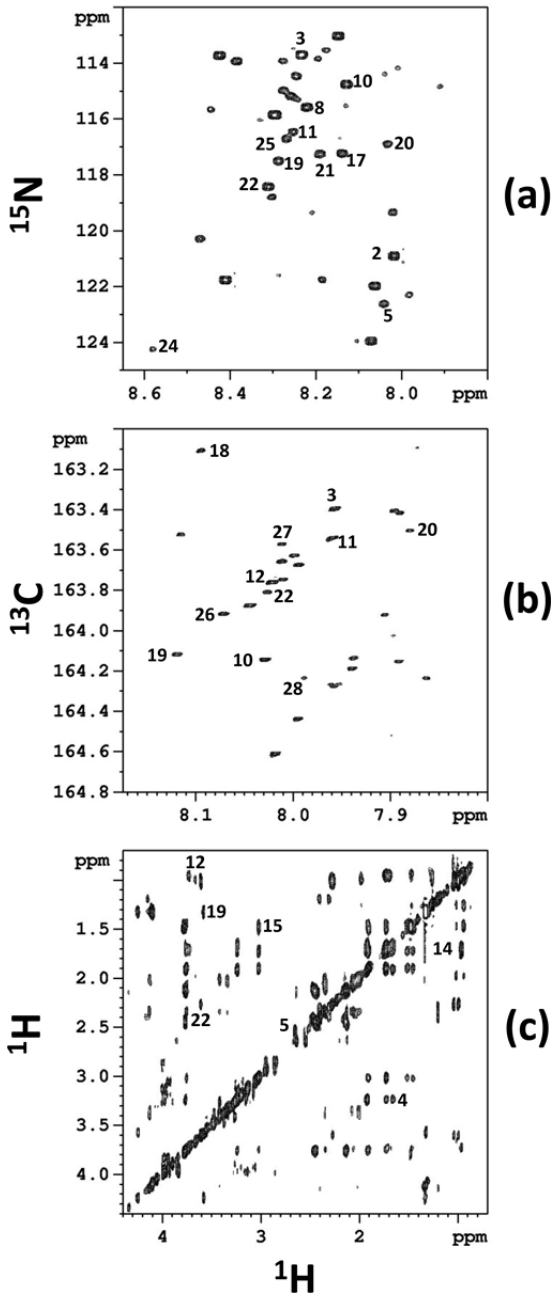


Figure 2.3 2D spectra of mixtures of 28 synthetic compounds obtained with or without isotope tagging: (a) $^{1}H-^{15}N$ HSQC spectrum with ^{15}N tagging of carboxylic acids, (b) $^{1}H-^{13}C$ HSQC spectrum with ^{13}C tagging of amines and amino acids, and (c) $^{1}H-^{1}H$ TOCSY spectrum of the neat mixture. All the spectra were obtained on an 800 MHz spectrometer. The labeled peaks correspond to the numbered metabolites in Table 2.1.

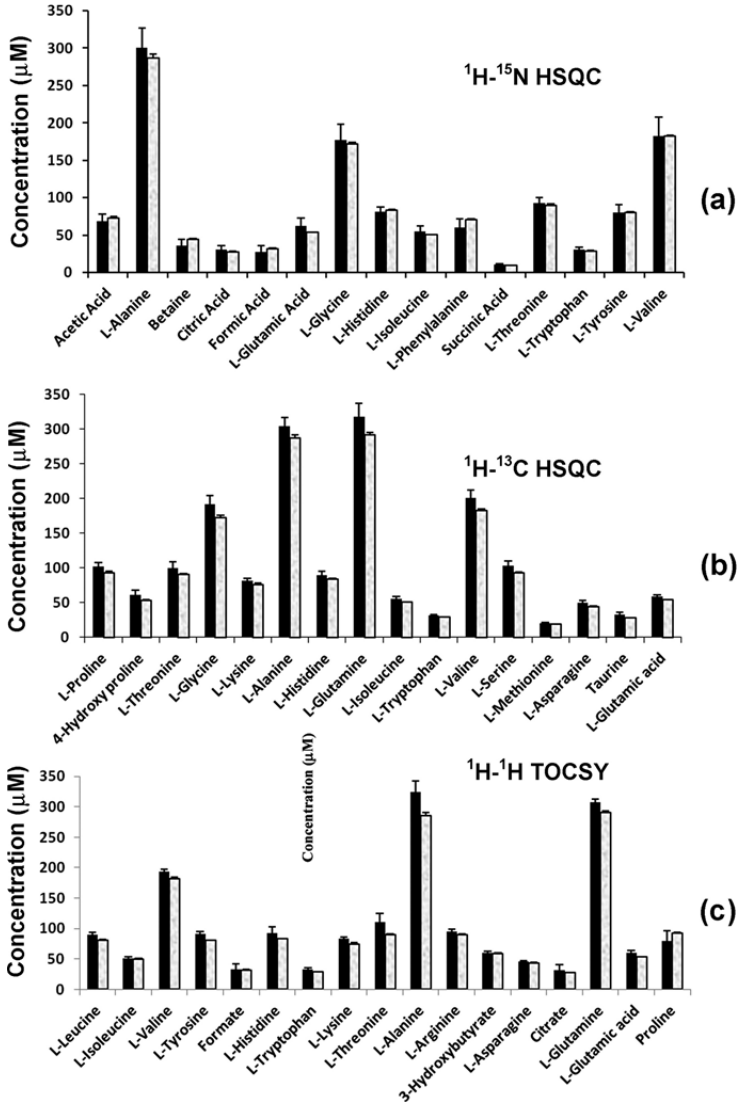


Figure 2.4 Concentration of 28 standard metabolites obtained by combining 2D NMR experiments with and without ^{15}N or ^{13}C tagging: (a) obtained from ^{1}H–^{15}N HSQC NMR after ^{15}N tagging; (b) obtained from ^{1}H–^{13}C HSQC NMR after ^{13}C tagging, and (c) obtained from ^{1}H–^{1}H TOCSY NMR of the neat mixture. The shaded bar on the right in each pair represents the actual concentration of the metabolite.

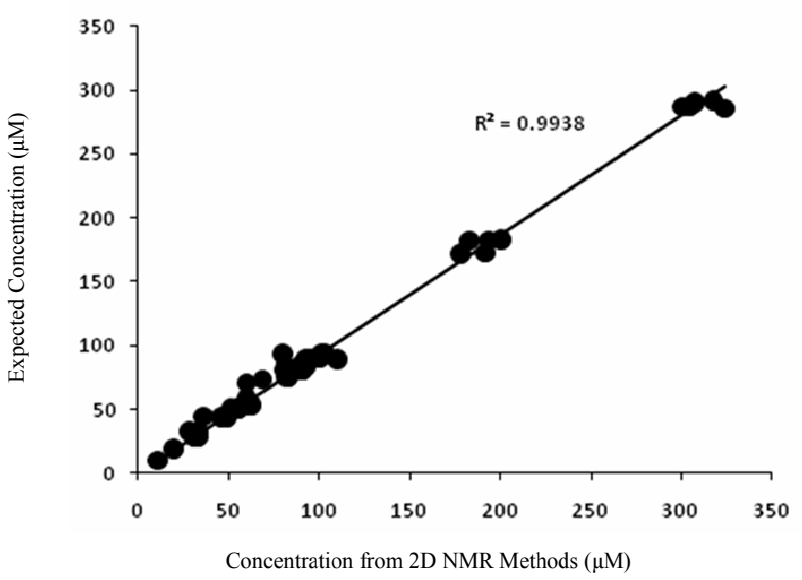


Figure 2.5 Correlation of the concentration of the metabolites determined by a combination of 2D experiments with or without ^{15}N or ^{13}C tagging with the expected values.

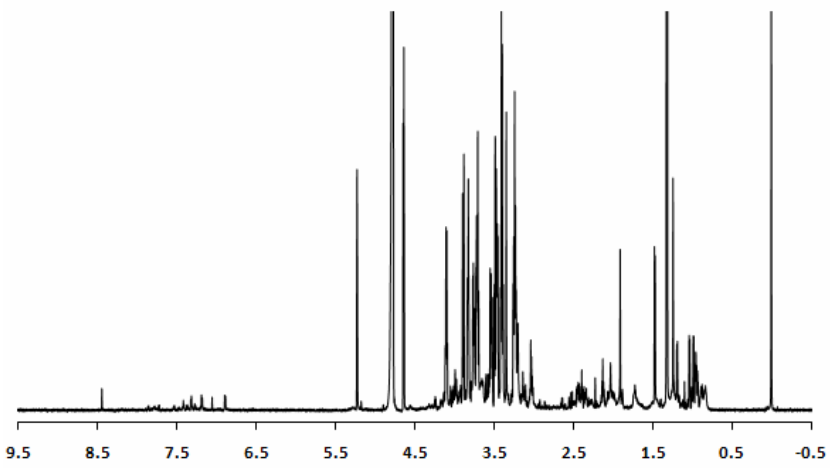


Figure 2.6 1D ^{1}H NMR spectrum of aqueous metabolites of NIST plasma obtained on a Bruker 500 MHz NMR spectrometer.

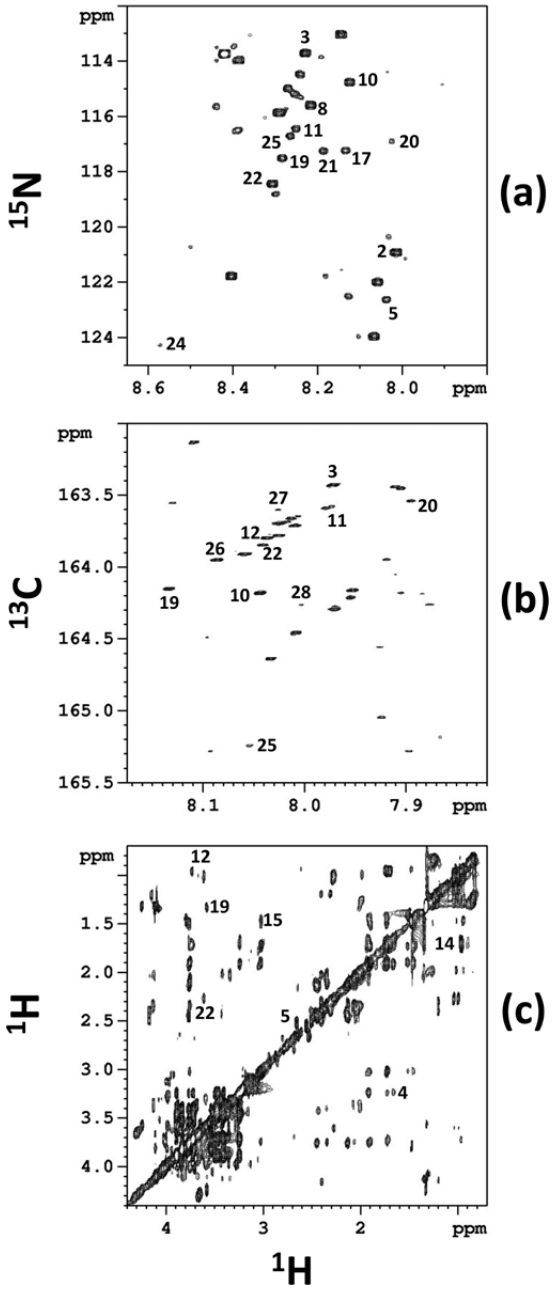


Figure 2.7 2D spectra of NIST plasma obtained with and without isotope tagging: (a) $^{1}H-^{15}N$ HSQC spectrum obtained after ^{15}N tagging of carboxylic acids, (b) $^{1}H-^{13}C$ HSQC spectrum obtained after ^{13}C tagging of amines and amino acids, and (c) $^{1}H-^{1}H$ TOCSY spectrum of the neat mixture. All the spectra were obtained on an 800 MHz spectrometer. The labeled peaks correspond to the numbered metabolites in Table 2.1.

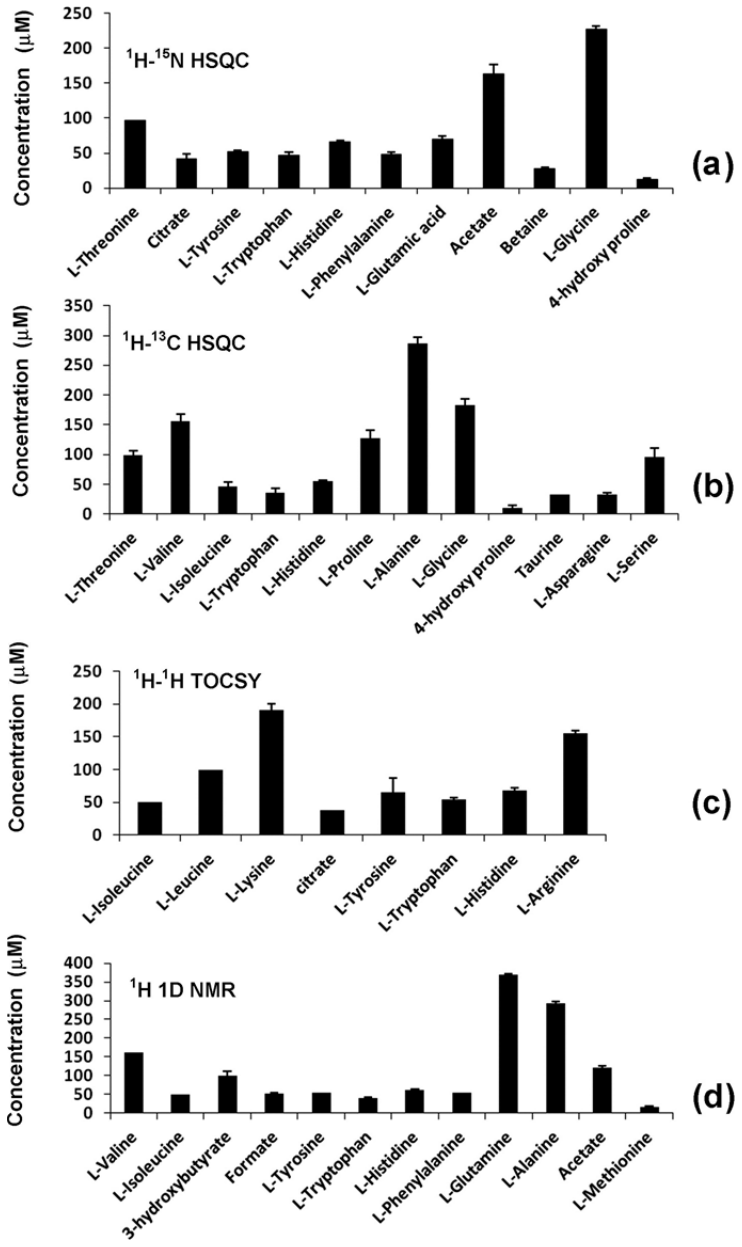


Figure 2.8 Concentrations of metabolites in the NIST plasma obtained using 1D/2D NMR experiments with and without isotope tagging: (a) obtained from $^{1}H-^{15}N$ HSQC NMR after ^{15}N tagging, (b) obtained from $^{1}H-^{13}C$ HSQC NMR after 13C tagging, (c) obtained from $^{1}H-^{1}H$ TOCSY NMR of neat plasma, and (d) obtained from 1D NMR of the neat plasma sample.

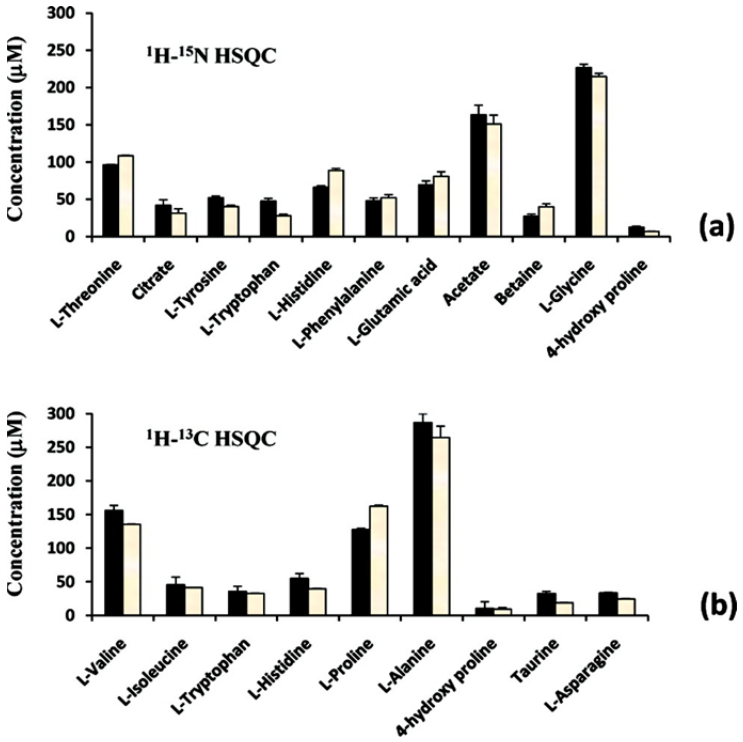


Figure 2.9 Comparison of the concentrations of (a) carboxylic acid and (b) amino group containing metabolites in the NIST plasma obtained with spiking (left bars) and without spiking (right bars) with the standard compounds.

CHAPTER 3: METABOLITE PROFILING OF THE CARBOXYL-CONTAINING METABOLITES WITH SMART ISOTOPE TAGGING

3.1 Introduction

The metabolomics field has witnessed an exponential growth since the last decade due to its potential applications in numerous disciplines including biomedicine, toxicology, food and nutrition, drug development and environmental science.[1-5] Commonly used analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy and/or mass spectrometry (MS) have evolved in response to the high demand for resolving the complexity of biological mixtures and identifying the large pool of quantifiable metabolites. However, despite numerous advances, the biological complexity still often outweighs the capabilities of these advanced analytical methods; no single technique currently is capable of detecting all metabolites in a single experiment. Each analytical method is sensitive to certain classes of metabolites, and depending on the nature of the metabolites of interest, generally one or sometimes a combination of NMR or MS techniques are used to profile as many metabolites as possible and thereby derive the biological meaning. A major hurdle of such an approach is that the metabolite data obtained from NMR and LC-MS or GC-MS methods for the same or similar samples are often not directly comparable. The inability to compare and correlate data from different analytical techniques for the same or similar samples is a significant challenge that prevents drawing meaningful conclusions from the vast amount of metabolite data

existing in the literature and exploiting the combined strength of NMR and MS for unknown metabolite identification. The main contributing factors for this bottleneck are the limited NMR sensitivity, complex spectral signatures and variable MS ionization efficiency or suppression.

The use of chemo-selective tags provides an avenue to improve the sensitivity of metabolite detection by both NMR and MS methods. For example, the sensitivity of MS detection is shown to be enhanced by three orders of magnitude or more by tagging metabolites with a chemo-selective tags containing a permanent charge.[6-10] Because of the permanent charge, the tagged metabolites are effectively detected with high sensitivity and better quantitative accuracy, irrespective of the pH or nature of the solvents used to separate metabolites before detection by MS. Separately, based on differential dansylation using $^{12}C/^{13}C$-dansyl chloride, absolute or relative quantitation of amine and phenol containing metabolites has been achieved with a sensitivity enhancement of three orders of magnitude.[11,12] Similarly, NMR-sensitive isotope based chemo-selective tags have been shown to detect many quantifiable metabolites with high sensitivity and resolution by NMR.[13-17] Using ^{15}N-ethanolamine as the tag, for example, over a hundred carboxyl-containing metabolites have been detected by ^{1}H-^{15}N two-dimensional NMR with high resolution and sensitivity.[13] However, while metabolites can be detected with high sensitivity by both MS and NMR separately using chemoselective tags, the inability to compare and correlate the data from the two methods is a major bottleneck in the metabolomics field.

The ability to more easily detect the same metabolites by both NMR and MS methods would offer new avenues to compare data between MS and NMR platforms and

to exploit the combined strength of the complimentary methods. Towards this goal, we introduce a new "smart isotope tag" approach, using ^{15}N-cholamine in this case, which possesses the characteristics of high NMR sensitivity and resolution through its isotope enrichment and high MS sensitivity through its permanent positive charge (see schematic Figure 3.1). The tag combines the strengths of individual chemo-selective tags, demonstrated previously and separately for NMR and MS detection,[6, 13] and offers news avenues to exploit the combined strength of these powerful and complementary techniques for areas such as metabolite profiling and unknown metabolite identification.

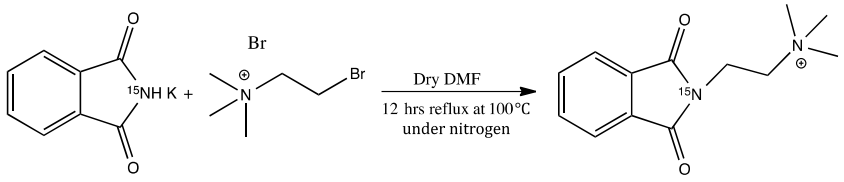
3.2 Materials and Methods

3.2.1 Chemicals and Biofluids

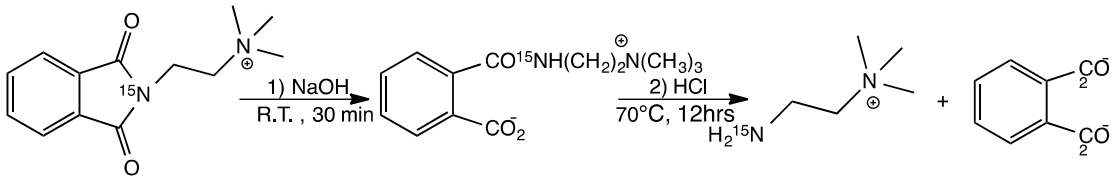
A total of 48 carboxyl-containing metabolite standards (Table 1), (2-bromoethyl)trimethylammonium bromide, dimethylformamide (DMF), methanol, acetonitrile, isopropanol, acetone, hydrochloric acid (HCl), sodium hydroxide (NaOH) (all from Sigma-Aldrich, St. Louis, MO), 4-(4,6-dimethoxy[1.3.5]triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) (Acros Organic, Pittsburgh, PA), ^{15}N-phthalimide potassium and deuterium oxide (Cambridge Isotope Laboratories, Andover, MA) were used without further purification. Human serum samples were obtained from Innovative Research, Inc. (Novi, MI) and urine from healthy volunteers, in accordance with the Internal Review Board at Purdue University. Deionized (DI) water was from in-house Synergy Ultrapure Water System from Millipore (Billerica, MA).

3.2.2 Design and Synthesis of the Smart Isotope Tag ^{15}N-cholamine

Synthesis of ^{15}N-cholamine involved a two-step reaction and followed the Gabriel synthesis procedure with modifications as described below to yield the pure product.[18,19] The first step involved reacting potassium ^{15}N-phthalimide with (2-bromoethyl)trimethylammonium bromide in DMF to obtain the ^{15}N substituted phthalimide intermediate (Scheme 1). The second step involved alkaline and acid hydrolyses of the ^{15}N substituted phthalimide to yield the smart isotope tag, ^{15}N-cholaimne (Scheme 2).



Scheme 3.1 Synthesis of ^{15}N substituted phthalimide



Scheme 3.2 Alkaline and acid hydrolyses of the ^{15}N substituted phthalimide to yield ^{15}N-cholamine

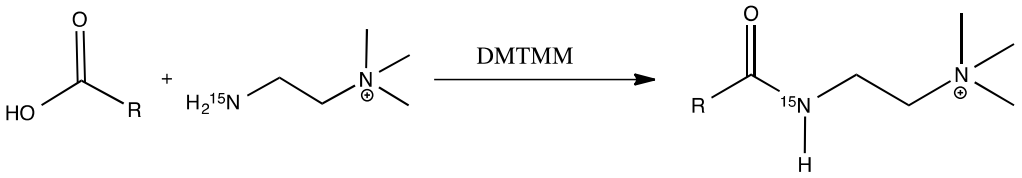
Briefly, for the synthesis of ^{15}N substituted phthalimide (Scheme 1), (2-bromoethyl)trimethylammonium bromide (9.5 mmol, 2.35 g) was mixed with

^{15}N-phthalimide potassium (10 mmol, 1.86 g) in a 250 mL round bottom flask and dry DMF (100 mL) was added under nitrogen gas. The mixture was then refluxed at 100 °C with stirring for 12 h. The supernatant from the reaction mixture was separated and the solvent was removed using a rotary evaporator.[18] The resulting crude residue was washed thrice using acetonitrile (5 mL each time), twice with acetone (2 mL each time) followed by washing again once with acetonitrile (3 mL) to obtain the pure ^{15}N-substituted phthalimide. ^{1}H NMR spectra in D_2O at each step were monitored to assess the purity of the intermediate product. For the synthesis of ^{15}N-cholamine, in the second step, the ^{15}N-substituted phthalimide (538 mg) (Scheme 1) was dissolved in DI water (24 mL); 1 N NaOH (2.69 mL) was added to the solution and the mixture was left at room temperature with stirring for 30 min to complete the alkaline hydrolysis (Scheme 2).[19] Subsequently, 12 N HCl (1.8 mL) was added to the solution, the temperature was raised to 70°C and left for 12 h with stirring to complete the acid hydrolysis (Scheme 2).[19] The solvent was then removed using a rotary evaporator. The resulting crude residue was washed thrice with acetonitrile (4 mL each time) followed by washing thrice with 25:75 water/acetone mixture (2 mL each time) to yield the pure product, ^{15}N-cholamine. ^{1}H NMR spectra in D_2O at each step were monitored to assess the purity of the final product.

3.2.3 Tagging Metabolites Using the Smart Isotope Tag ^{15}N-cholamine

^{15}N-cholamine (5mg, 50 μmol) was added to 500 μL sample in an eppendorf tube, the pH of the mixture was adjusted to 7.0 with 1 M HCl or NaOH. 21 mg DMTMM was added to initiate the reaction.[13,20,21] The mixture was stirred at room temperature for 4 h to complete the reaction. The general reaction for tagging metabolites with the smart

isotope tag is shown in Scheme 3. To maintain ^{15}N amide protonation, the pH was adjusted to 5.0 by adding 1 M HCl or 1 M NaOH, and the solution volume was adjusted to 580 µL by adding DI water and 30 µL of D_2O prior to NMR detection. Serum was deproteinized using methanol prior to metabolite tagging and urine was used with no pretreatment.[13]



Scheme 3.3 General reaction for tagging carboxyl-containing metabolites with the smart isotope tag- ^{15}N-cholamine

3.2.4 NMR Spectroscopy

For each sample, 550 µl was mixed with 30 µl D_2O and placed in a 5 mm NMR tube. NMR experiments were performed on a Bruker DRX 500 MHz or Avance III 800 spectrometer equipped with a room temperature or cryoprobe probe, respectively, suitable for ^1H inverse detection with Z-gradients at 298 K. A one pulse sequence with or without solvent signal suppression using presaturation was used for ^1H 1D NMR experiments. The sensitivity-enhanced ^1H-^{15}N 2D HSQC experiments employed an INEPT transfer delay of 6 ms corresponding to the J_{NH} of 90 Hz. Spectral widths for the ^1H and ^{15}N dimensions were approximately 8 kHz and 3 kHz, respectively. 128 free induction decays of 1,024 data points each were collected in the indirect (t_1) dimension with 1 or 4 transients per increment. Nitrogen decoupling during the direct acquisition (t_2

dimension) was achieved with the GARP (Globally Optimized Alternating-Phase Rectangular Pulses) sequence. The resulting 2D data were zero-filled to 1,024 points in the t_1 dimension after forward linear prediction to 256 or 512 points. A 45° shifted sine-bell window function was applied to both dimensions before Fourier transformation. Chemical shifts were referenced to the ^{1}H signal of TSP for the 1D spectra or the derivatized formic acid signal (^{1}H: 8.05 ppm; ^{15}N: 123.93 ppm) in the HSQC spectra. Bruker Topspin versions 3.0 or 3.1 software packages were used for NMR data acquisition or processing.

3.2.5 Mass spectrometry

LC-MS and LC-MS/MS experiments were performed using an Agilent 1200 SL-LC system coupled online with an Agilent 6520 Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA). The sample (8 μL) was injected onto an Agilent Poroshell 120 EC-C18 column (30x50 mm, 2.7-micron), which was heated to 50 °C. The flow rate was 0.5 mL/min. Mobile phase A was 5 mM ammonium acetate in water, and mobile phase B was 0.1% water in ACN. The mobile phase composition was initially kept isocratic at 3% B for 1 min, then increased to 90% B over 4 min; after another 4 min at 90% B, the mobile phase composition returned to 3% B. Electrospray ionization (ESI) was used in positive mode, and the voltage was 3.5 kV. The mass analyzer was scanned over a range of 50–1000 m/z. The collision energy for auto LC-MS/MS experiments was fixed at 10 V, targeting pre-selected compounds.

3.3 Results and Discussion

The smart isotope tag, ^{15}N-cholamine, designed, developed and used in this study retains all the characteristics of the ^{15}N-ethanolamine tag including the solubility of the tagged metabolites in aqueous media, large one-bond J-coupling between ^{1}H and ^{15}N of ~90 Hz for efficient polarization transfer between ^{1}H and ^{15}N nuclei, and wide chemical shift dispersion for different metabolites in the resulting 2D NMR spectra.[13] In addition, and importantly, ^{15}N-cholamine possesses a permanent positive charge, which enables efficient positive mode detection of the same carboxyl-containing metabolites by MS, irrespective of the pH or solvent conditions of the eluting media, commonly used for chromatographic separation before detection by MS.[6]

Synthesis of ^{15}N-cholamine involved a two-step reaction and followed the Gabriel synthesis procedure with suitable modifications to yield the pure product.[18, 19] As seen in the ^{1}H NMR spectrum (Figure 3.2), the pure intermediate compound, ^{15}N substituted phthalimide, was obtained after the first step of the synthesis. Hydrolysis of this compound yielded the ^{15}N-cholamine in pure form as can be ascertained from its ^{1}H NMR spectrum (Figure 3.3; peaks at 3.16; 3.48; 3.64 ppm). The accurate MS and MS/MS spectra for ^{15}N-cholamine, shown in Supplementary Figure 3.4, help further verify the identity and purity of the synthesized smart isotope tag (m/z= 104.120).

The compound was then used to tag 48 metabolites that were selected for their prominence as carboxyl-containing metabolites in biofluids that represent a variety of metabolic pathways. The general reaction for tagging metabolites with the smart isotope tag is shown in Scheme 3. To maintain ^{15}N amide protonation, the pH was adjusted to 5.0 by adding 1 M HCl or 1 M NaOH, and the solution volume was adjusted to 580 µL by

adding DI water and 30 μL of D_2O prior to NMR detection. Serum was deproteinized using methanol prior to metabolite tagging and urine was used with no pretreatment.[13]

The ^{1}H and ^{15}N chemical shift data derived from the 2D NMR experiments, after tagging with ^{15}N cholamine, are shown in Table 3.1. Because the ^{15}N-cholamine and the previously used ^{15}N-ethanolamine differ only in their terminal group, the tagging efficiency, reproducibility and chemical shift values for metabolites with ^{15}N-cholamine tag were similar to those obtained using the ^{15}N-ethanolamine tag.[13]

Importantly, as anticipated based on the ^{15}N-ethanolimane tagging approach shown earlier in our laboratory,[13] the ^{15}N-cholamine tagging of metabolites in human serum provided a rich NMR spectrum due to the large number of carboxyl-containing metabolites normally present in blood (Figure 3.5). The low natural abundance of ^{15}N (0.37%) ensures that none of the nitrogen containing metabolites interferes with the detection of carboxyl-metabolites through ^{15}N-cholamine tag. Each peak in the spectrum corresponds to different metabolite from the carboxylic acid class. However, metabolites, which contain more than one carboxyl group, provide additional peaks depending on the number of carboxyl groups and molecular symmetry. In addition, metabolites such as lactate, which possess α-hydroxyl groups, show more than one peak for the same metabolite as we described earlier using the ^{15}N-ethanolmaine tag.[13] Some of the peaks assigned based on the chemical shift values for the standard compounds are marked with the corresponding number shown in Table 3.1 and Figure 3.6. Similarly, tagging of metabolites in human urine with ^{15}N-cholamine also enabled the detection of peaks corresponding to well over a hundred carboxylic acid group containing metabolites

(Figure 3.7). Peaks tentatively assigned based on the values for the standard compounds are marked by their numbers shown in Table 3.1 and Figure 3.6.

The ^{15}N-cholamine tagging of metabolites in aqueous media enabled a sensitivity enhancement of up to three orders of magnitude or more in the MS detection of carboxyl metabolites. The derivatized metabolites could be detected easily in positive ion mode as compared to the same metabolites detected in negative ion mode without the tag. For example, the sensitivity for pyruvic acid detected in positive ion mode after ^{15}N-cholamine tagging was enhanced by a factor of about 1500 when compared to that detected for the same metabolite without the ^{15}N-cholamine tag, in negative ion mode. Figure 3.8 shows typical mass spectra for formic acid and pyruvic acid after tagging with ^{15}N-cholamine. The enhancement in sensitivity is primarily due to the high ionization efficiency imparted by the permanent positive charge of the ^{15}N-cholamine and is in agreement with results by Smith and co-workers for fatty acid analysis using the heavy and light forms of cholamine.[6] In that study, reactions of metabolites with cholamine were made in organic solution in contrast to the aqueous media used here. The ^{15}N-cholamine derivatized serum samples were then analyzed by LC-MS. As anticipated, due to the presence of the permanent positive charge, tagged metabolites could be readily detected in positive ion mode with high sensitivity. Sensitivity enhancement by a factor of up to nearly 3000 could be achieved for tagged acids. The extracted ion chromatograms for a few typical carboxylic acids detected in serum with ^{15}N-cholamine tag are shown in the Supplementary Figure 3.9.

One potential issue is the effect on chromatographic retention time caused by the addition of the cholamine tag. However, separation of the tagged metabolites using

HILIC columns offers opportunity to effectively separate before detection using MS. For example, the results of separation of a mixture of standard carboxylic and amino acids performed using a HILIC column, without attempting to optimize chromatography conditions, indicate that ^{15}N-cholamine tagged metabolites can be separated effectively (Figure 3.10). More broadly, we can contemplate the use of dual purpose smart tags for other NMR-MS combinations. For GC-MS, the addition of a charged species will likely cause problems related to reduced volatility; however, a different tag, such as ^{13}C or even ^{29}Si labeled silyl-type tags can be contemplated.[23] Another alternative is the use of smart tags for capillary electrophoresis (CE) coupled to MS, which is increasingly of interest in metabolomics.[24] In fact, positively charged derivatization agents (based on pyridinum containing compounds) have been demonstrated for the use of metabolite profiling of carboxylic acids by CE-MS.[25] Thus, the potential for the use of smart tags such as cholamine for CE-MS and NMR is quite promising.

In conclusion, we have developed a smart isotope tag, ^{15}N-cholamine, which possesses dual characteristics for metabolite profiling in complex biological mixtures using the powerful analytical techniques of NMR and MS. By combining the individual strengths of the ^{15}N label and permanent charge, the smart isotope tag facilitates detection of carboxyl-containing metabolome by both NMR and LC-MS techniques with high sensitivity. Detection of the same metabolites by both NMR and MS (Figure 3.11), effectively opens unique opportunities for identification of unknown metabolites and direct comparison of metabolite data from the two powerful analytical platforms.

3.6 References

1. Gowda, G.A.N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Diagn.* **2008**, *8(5)*, 617-633.

2. Shintu, L.; Banudin, R.; Navratil, V.; Prot, J. M.; Pontoizeau, C.; Defernez, M.; Blaise B. J.; Domange, C.; PÉRY, A.R.; Toulhoat, P. *Analytical chemistry.* **2012**, *84*, 1840-1848.

3. Wishart, D. S. *Trends in Food Science & Technology.* **2008**, *19*, 482-93.

4. Lindon, J.C.; Holmes, E.; Nicholson, J. K. *Pharmaceutical research.* **2006**, *23*, 1075-1088.

5. Veldhoen, N.; Ikonomou, M. G.; Helbing, C. C. *Ecotoxicology and environmental safety.* **2012**, *76*, 23-38.

6. Lamos, S. M.; Shortreed M. R.; Frey B. L.; Belshaw P. J.; Smith L.M. *Anal. Chem.* **2007,** *79*, 5143-5149.

7. Yang W. C.; Adamec, J.; Regnier, F. E. Enhancement of the LC/MS analysis of fatty acids through derivatization and stable isotope coding. *Anal. Chem.* **2007,** *79,* 5150-5157.

8. Yang, W. C.; Sedlak, M.; Regnier, F. E.; Mosier, N.; Ho, N.; Adamec, J. *Anal. Chem.* **2008.** *80*, 9508-9516.

9. Yang, W. C.; Regnier, F. E.; Silva, D.; Adamec, J. *J. Chromatogr. B.* **2008,** *870*, 233-240.

10. Yang, W. C.; Regnier, F. E.; Jiang Q.; Adamec, J. *J. Chromatogr. A.* **2010,** *1217*, 667-675.

11. Guo, K.; LI, L. *Anal. Chem.* **2009,** *81*, 3919-3932.

12. Wu, Y.; Li, L. *Anal. Chem.* **2013**, *85(12)*, 5755-5763.

13. Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G.A.N.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009,** *81(12)*, 4882–4888.

14. Ye T.; Zhang S.; Mo H.; Tayyari F.; Nagana Gowda G. A.; Raftery D. *Anal Chem.* **2010,** *82(6)*, 2303–2309.

15. Shanaiah N.; Desilva M. A.; Nagana Gowda G. A.; Raftery M. A.; Hainline B. E.; Raftery D. *Proc Natl Acad Sci U S A.* **2007,** *104*, 11540–11544.

16. DeSilva M. A.; Shanaiah N.; Gowda G. A. N.; Rosa-Perez K.; Hanson BA.; Raftery D. *Magn. Reson. Chem.* **2009**, *47*, S74-S80.

17. Gowda, G. A. N.; Tayyari, F.; Ye, T.; Suryani, Y.; Wei, S.; Shanaiah, N. Raftery, D. Quantitative analysis of blood plasma metabolites using isotope enhanced NMR methods. *Anal. Chem.* **2010,** *82*, 8983-8990.

18. Iida K.; Ohtaka K.; Kajiwara M. *J. Label. Compd. Radiopharm.* **2007,** *50(4)*, 251-253.

19. Niyaz Khan M. *J. Org. Chem.,* **1996,** *61(23)*, 8063–8068.

20. Kunishima, M.; Kawachi, C.; Monta, J.; Terao, K.; Iwasaki, F.; Tani, S. *Tetrahedron* **1999,** *55*, 13159-13170.

21. Kunishima, M., Kawachi, C., Hioki, K., Terao, R. & Tani, S. *Tetrahedron* **2001,** *57*, 1551-1558.

22. Appiah-Amponsah E.; Owusu-Sarfo K.; Nagana Gowda G. A.; Ye T.; Raftery D. *Metabolites* **2013**, *3(3)*, 575-591.

23. Schraml J. *Prog. NMR Spect.* **1990**, *22*, 289-348.

24. Ramautar R.; Somsen G.W.; de Jong G.J. *Electrophoresis* **2013**, *34(1)*, 86-98.

25. Yang, W.-C.; Regnier, F. E.; Adamec, J. *Electrophoresis* **2008**, *29(22)*, 4549-4560.

Table. 3.1 ^{1}H and ^{15}N NMR chemical shifts for ^{15}N-cholamine tagged carboxyl-containing metabolites measured with reference to the formic acid peak

Label	Name	^{1}H (ppm)	^{15}N (ppm)	Label	Name	^{1}H (ppm)	^{15}N (ppm)
1	Cis-Aconitic acid	8.5	118.24	**23**	2-hydroxyisobutyric acid	7.95	117.51
		8.14	121.47	**24**	DL-Isocitric acid	8.40	117.15
		8.06	119.49			8.11	120.77
		8.07	120.21			8.28	122.78
		8.23	116.00			8.04	117.88
		8.14	120.81	**25**	Isoleucine	8.37	118.19
2	Adipic acid	8.05	120.57	**26**	Isovaleric acid	8.07	121.92
3	DL-Alanine	8.30	114.39	**27**	α-Ketoglutaric acid	8.69	116.34
4	4-Aminobenzoic acid	8.25	111.35			8.63	111.84
5	Arginine	8.34	115.96	**28**	Lactic acid	8.23	114.18
6	Asparagine	8.31	116.03			8.49	114.45
7	Aspartic acid	8.15	120.01	**29**	Leucine	8.34	115.24
		8.38	115.27	**30**	Lysine	8.33	115.88
		8.31	115.6	**31**	Maleic acid	8.39	120.39
		8.16	121.35	**32**	Malic acid	8.28	122.83
8	Betaine	8.55	122.69			8.29	115.14
9	Citric acid	8.20	121.46	**33**	Malonic acid	8.19	121.44
		8.07	123.95	**34**	Methionine	8.36	116.08
		7.87	121.88	**35**	Oxalic acid	8.47	117.13
10	Cysteine	8.35	115.93	**36**	Oxaloacetic acid	8.35	112.67
11	Cystine	8.5	115.22			8.63	111.40
12	Formic acid	8.05	123.93	**37**	L-phenylalanine	8.21	118.85
13	Fumaric acid	8.42	122.68	**38**	L-proline	8.35	115.58
		8.56	124.24	**39**	Propionic acid	7.95	118.85
14	Glucuronic acid	8.38	119.54	**40**	Pyroglutamic acid	8.29	115.88
15	Glutamic acid	8.28	115.99	**41**	Pyruvic acid	8.63	111.39
		8.05	120.42			8.35	112.72
16	Glutamine	8.35	115.90	**42**	Serine	8.17	117.63
17	Glycine	8.2	115.45	**43**	Succinic acid	7.97	119.56
18	Glycolic acid	8.22	114.97	**44**	Succinyl-COA	7.78	123.45
		8.37	115.19	**45**	L-threonine	8.34	117.79
19	Hippuric acid	8.2	115.62	**46**	L-tryptophan	7.98	119.37
20	Histidine	8.36	116.60	**47**	Tyrosine	8.27	118.05
21	3-Hydroxybutyric acid	8.07	122.20	**48**	Valine	8.38	118.20
22	4-Hydroxy-L-proline	8.5	115.89				
		8.36	117.62				

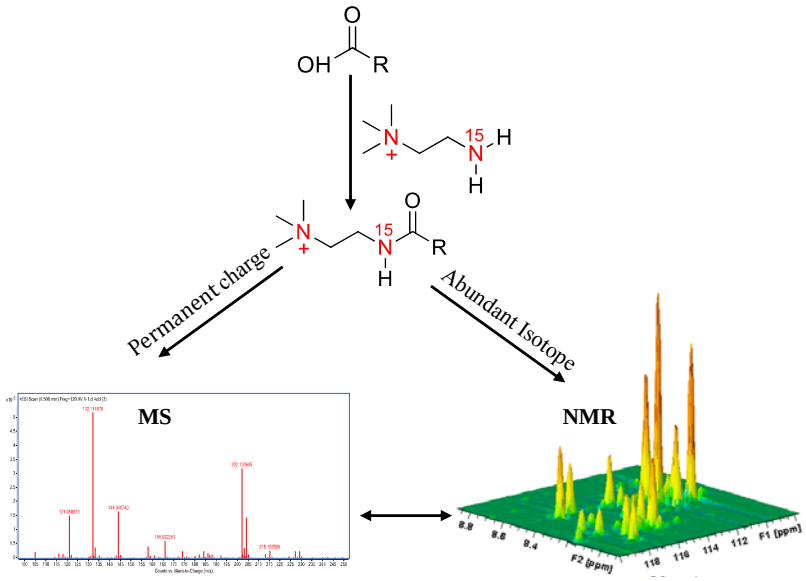


Figure 3.1 Schematic figure illustrating the "smart isotope tag" approach used to detect the same metabolites using NMR and MS with high sensitivity. Tagging carboxyl-containing metabolites with ^{15}N-cholamine enables their detection by both NMR and MS with high sensitivity.

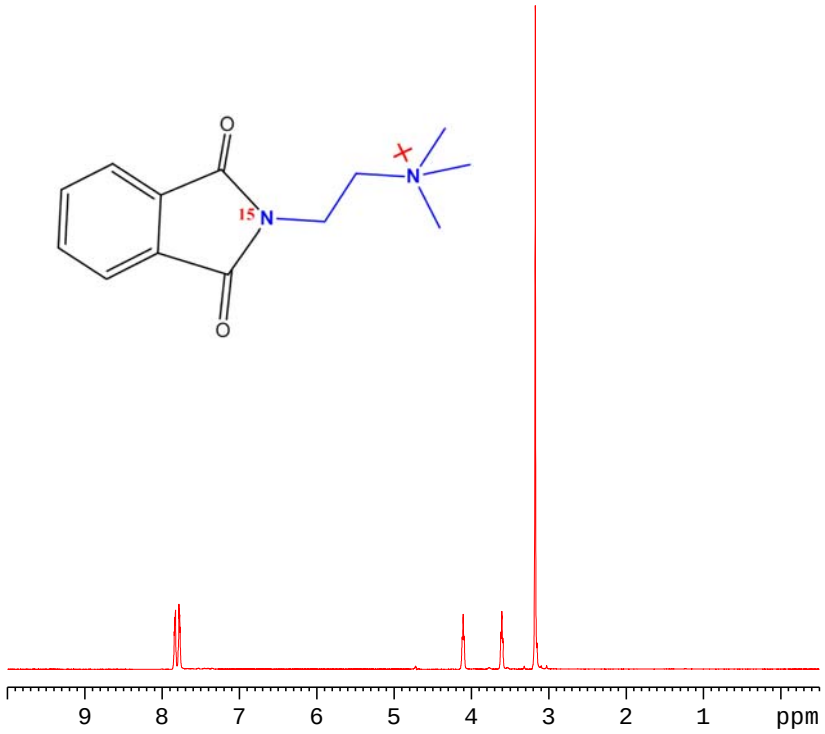


Figure 3.2 ^{1}H NMR spectrum of ^{15}N-substituted phthalimide intermediate compound, obtained for the synthesis of ^{15}N-cholamine, recorded on a Bruker DRX 499 MHz NMR spectrometer.

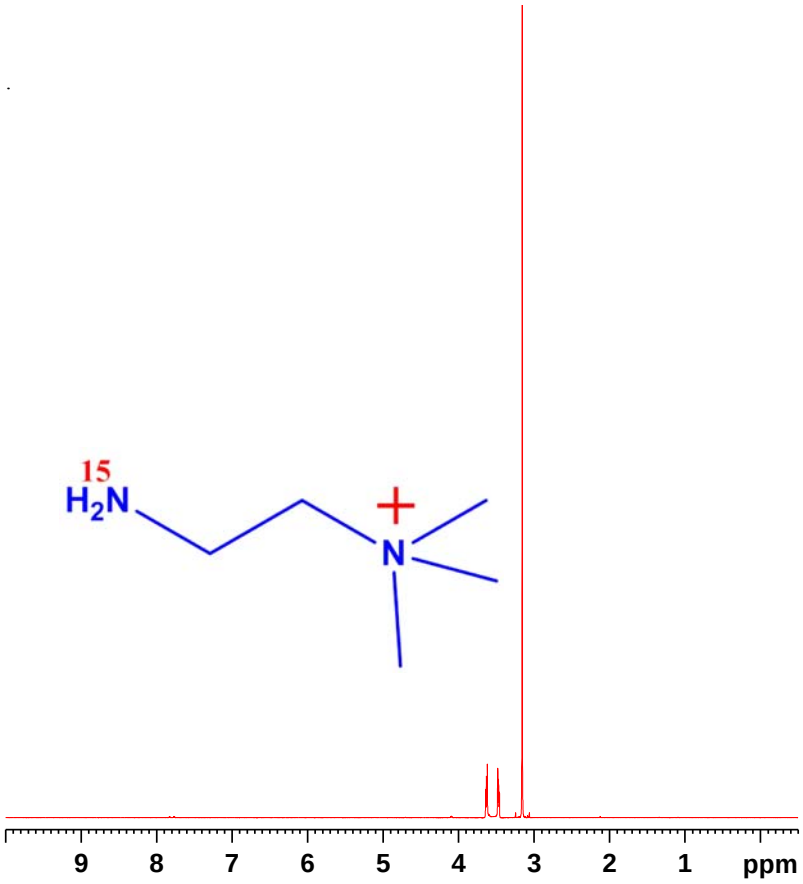


Figure 3.3 ^{1}H NMR spectrum of the synthesized ^{15}N-cholamine obtained on a Bruker Avance III 800 MHz NMR spectrometer.

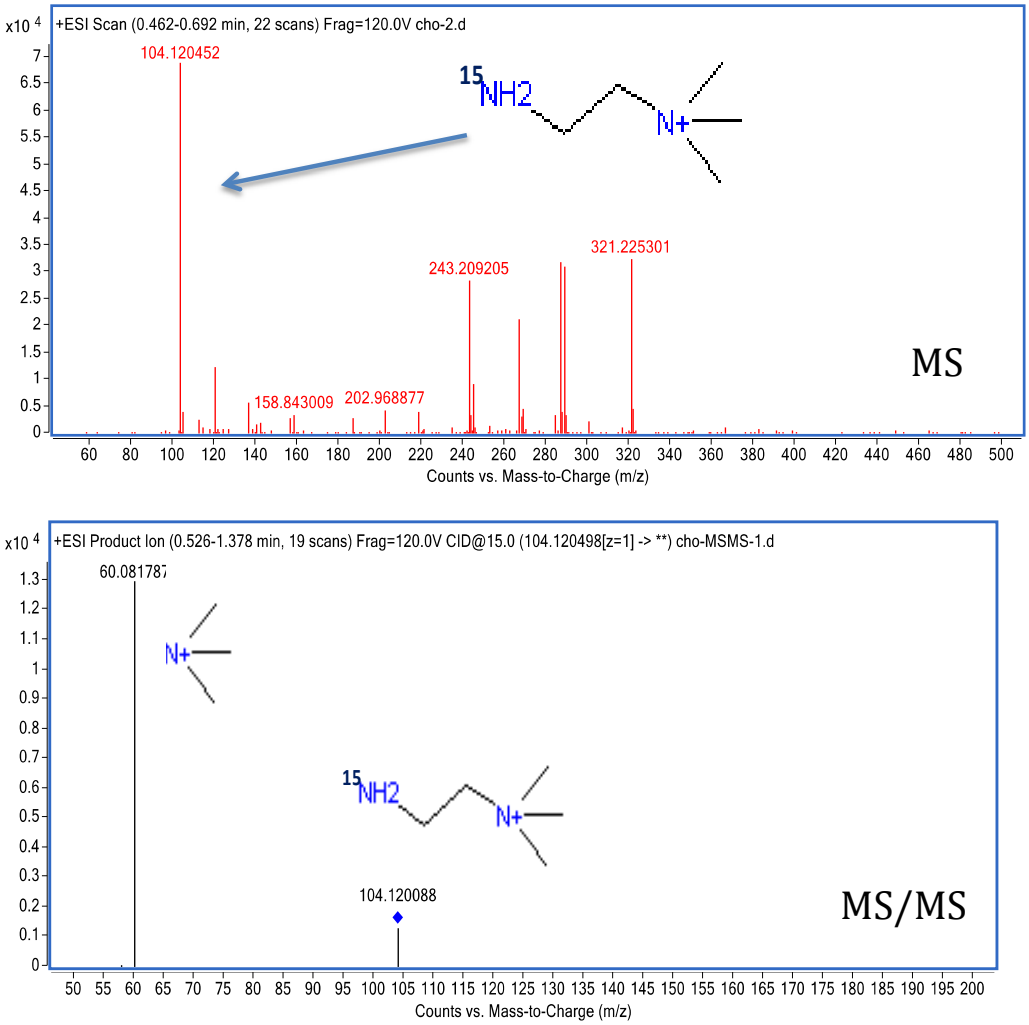


Figure 3.4 MS and MS/MS spectra of the synthesized ^{15}N-cholamine.

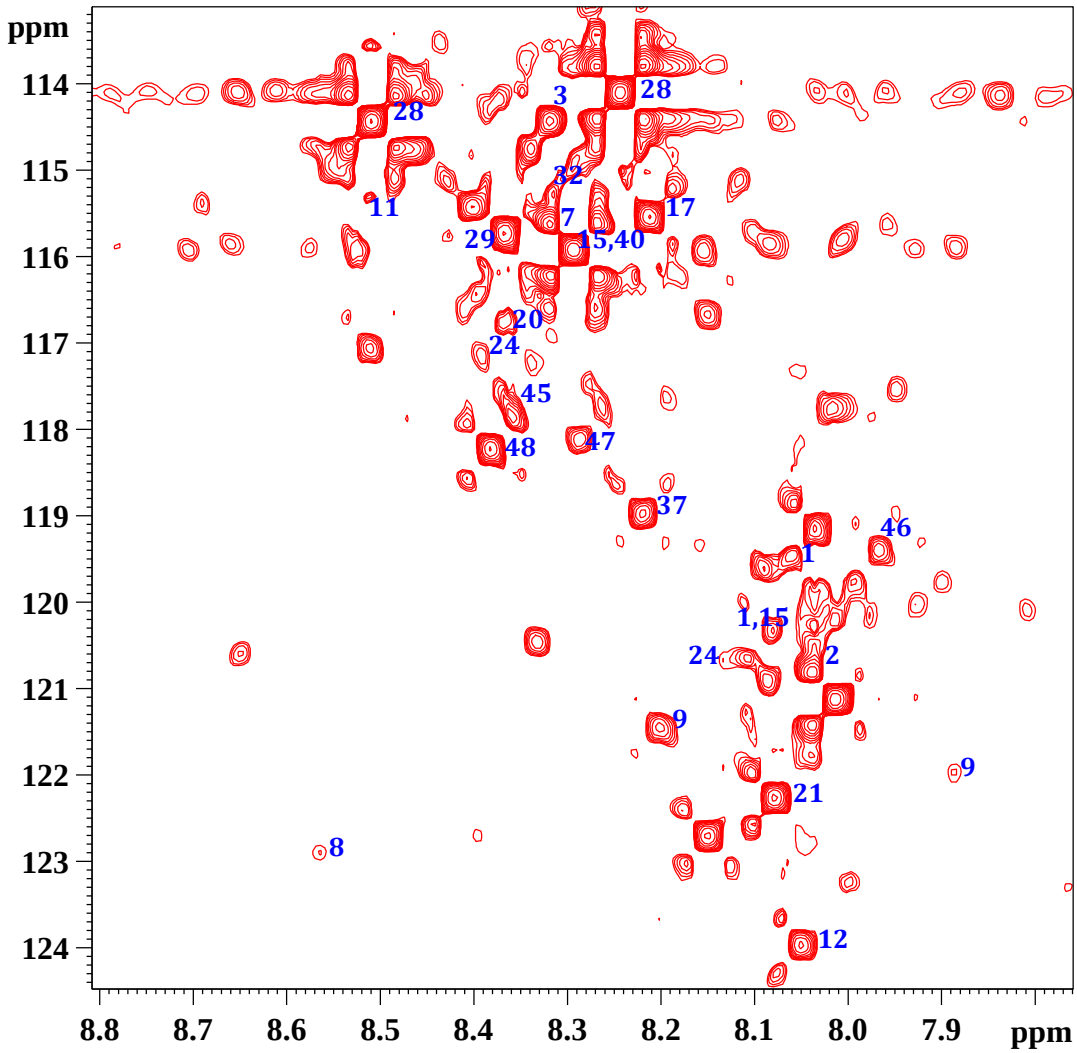


Figure 3.5 A portion of the ^{1}H-^{15}N HSQC spectrum of human serum tagged with ^{15}N-cholamine. 1: aconitic acid; 2: adipic acid; 3: alanine; 7: aspartic acid; 8: betaine; 9: citric acid; 11: cystine; 12: formic acid; 15: glutamic acid; 17: glycine; 20; histidine; 21: 3-hydroxybutyric acid; 24: isocitric acid; 28: lactic acid; 29: leucine; 32: malic acid; 37: phenylalanine; 40: pyroglutamic acid; 45: threonine; 46: tryptophan; 47: tyrosine; 48: valine.

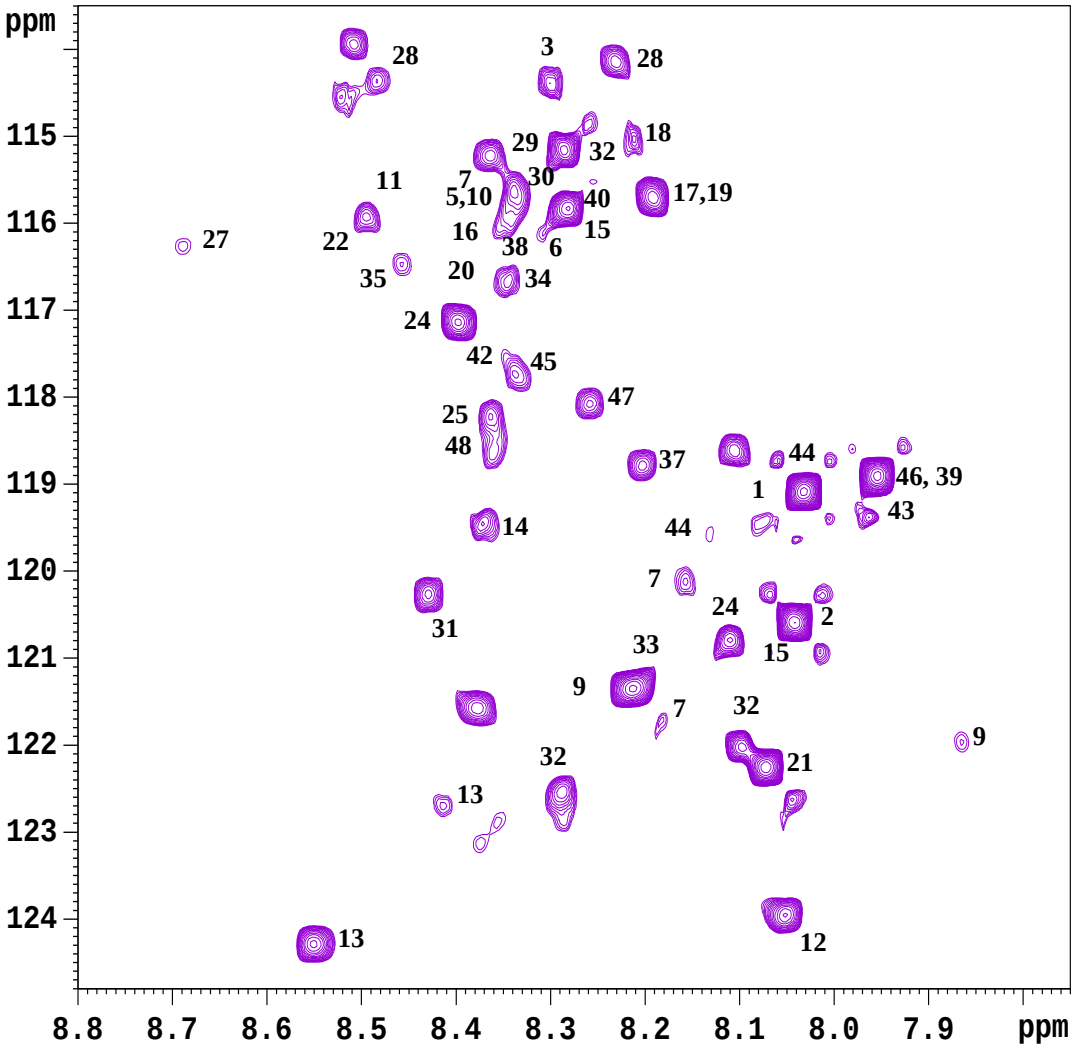


Figure 3.6 A portion of the ^{1}H-^{15}N HSQC spectrum of a mixture of standard compounds at various concentrations obtained after tagging with ^{15}N-cholamine. The peak numbers correspond to the compounds shown in Table 3.1.

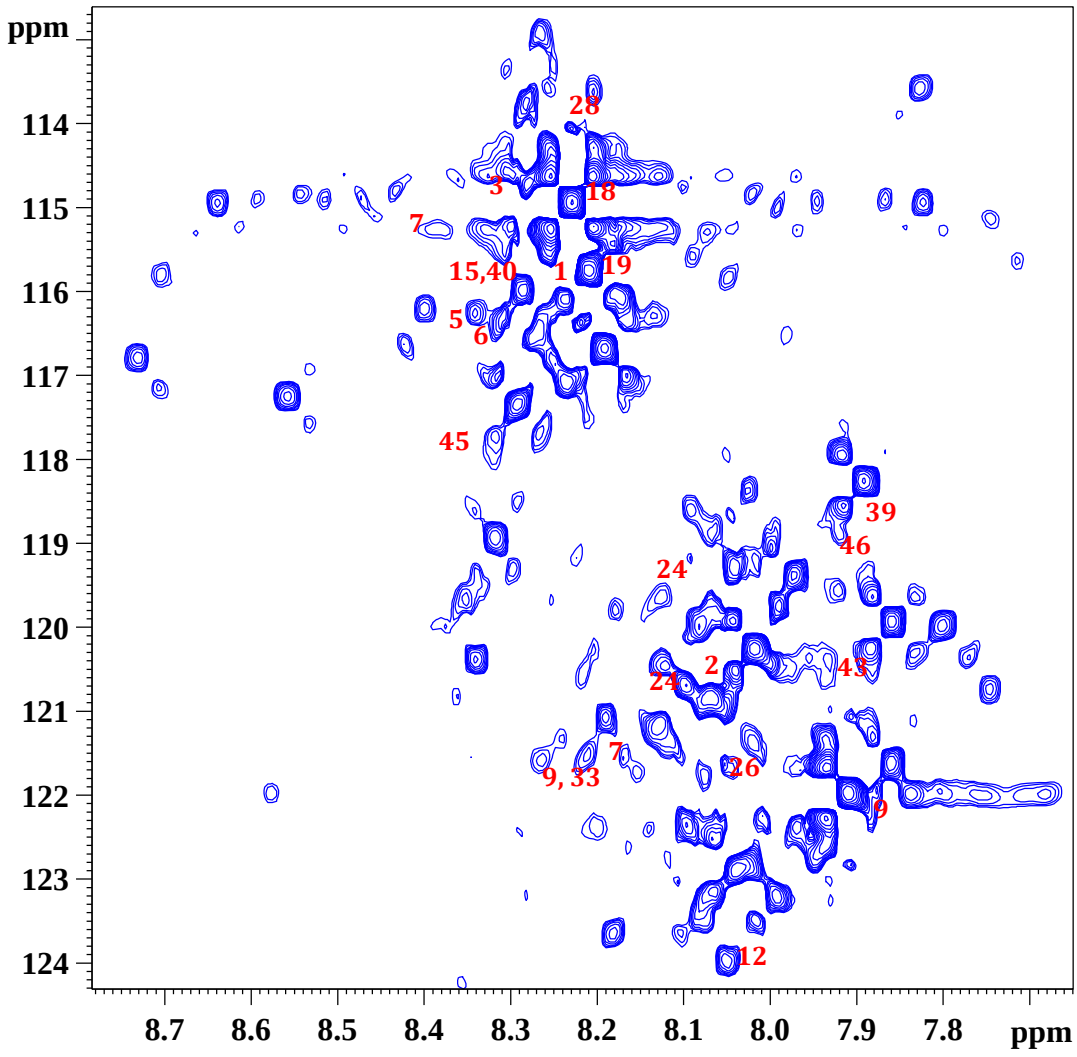


Figure 3.7 A portion of the ^{1}H-^{15}N HSQC spectrum of human urine tagged with ^{15}N-cholamine. 1: aconitic acid; 2: adipic acid; 3: alanine; 5: arginine; 6: asparagine; 7: aspartic acid; 9: citric acid; 12: formic acid; 15: glutamic acid; 18: glycolic acid; 19: hippuric acid; 24: isocitric acid; 28: lactic acid; 33: malonic acid; 39: propionic acid; 40: pyroglutamic acid; 43: succinic acid; 45: threonine; 46: tryptophan.

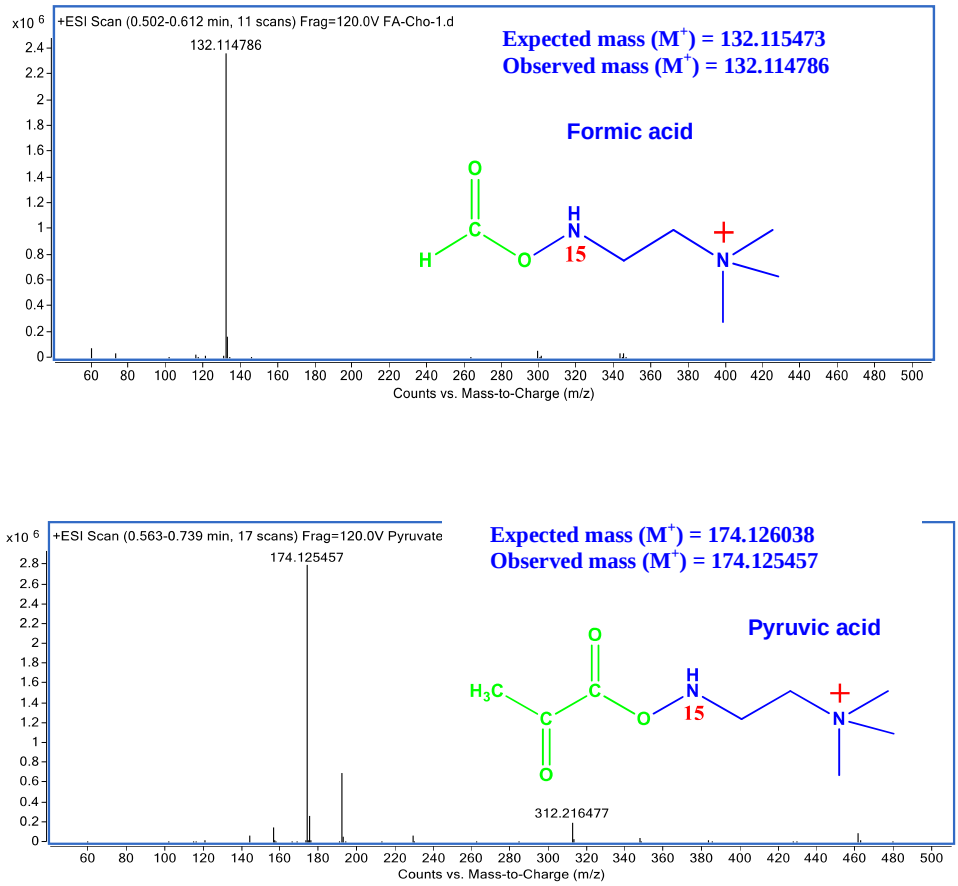


Figure 3.8 Typical LC-QTOF-MS spectra for formic acid and pyruvic acid obtained after tagging with the smart isotope tag, ^{15}N-cholamine. The permanent charge on the tagged metabolites enables their sensitive detection; the observed peak is from the intact tagged metabolite.

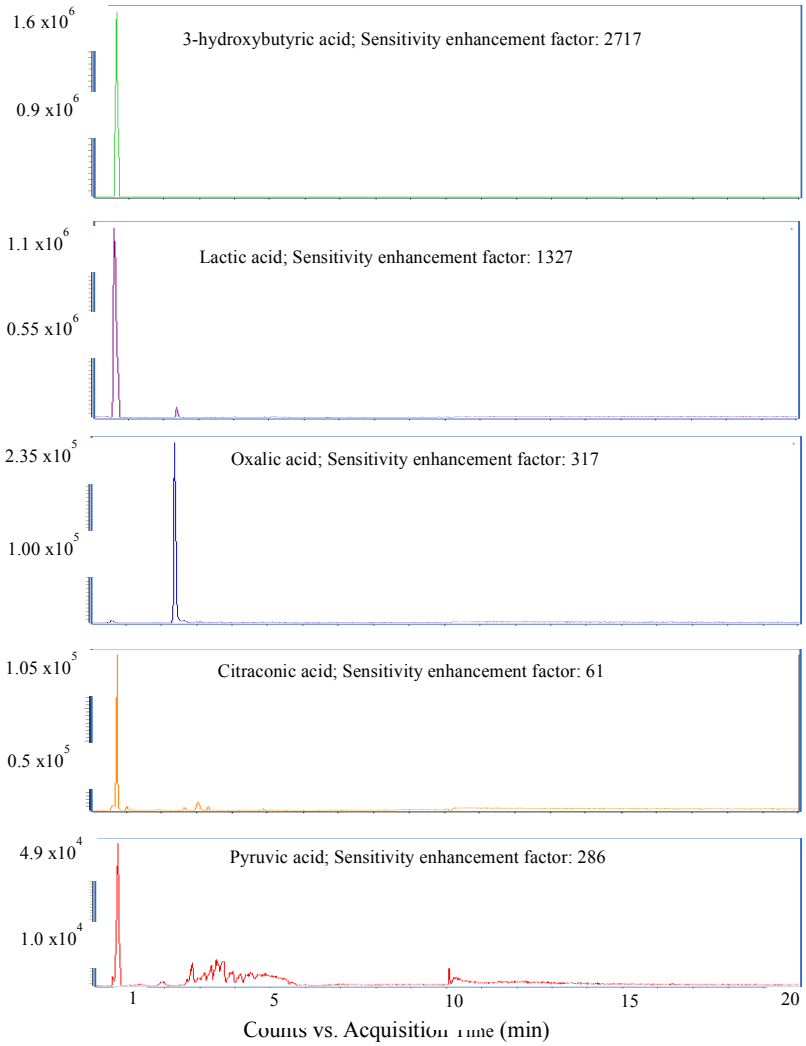


Figure 3.9 Accurate mass extracted ion chromatograms for a few carboxylic acids detected in serum in positive ion mode after tagging with ^{15}N-cholamine. The sensitivity enhancement factor indicates the ratio of peak area obtained with ^{15}N-cholamine tag to the peak area for the same acid detected without tagging (in negative ion mode), in the same serum sample.

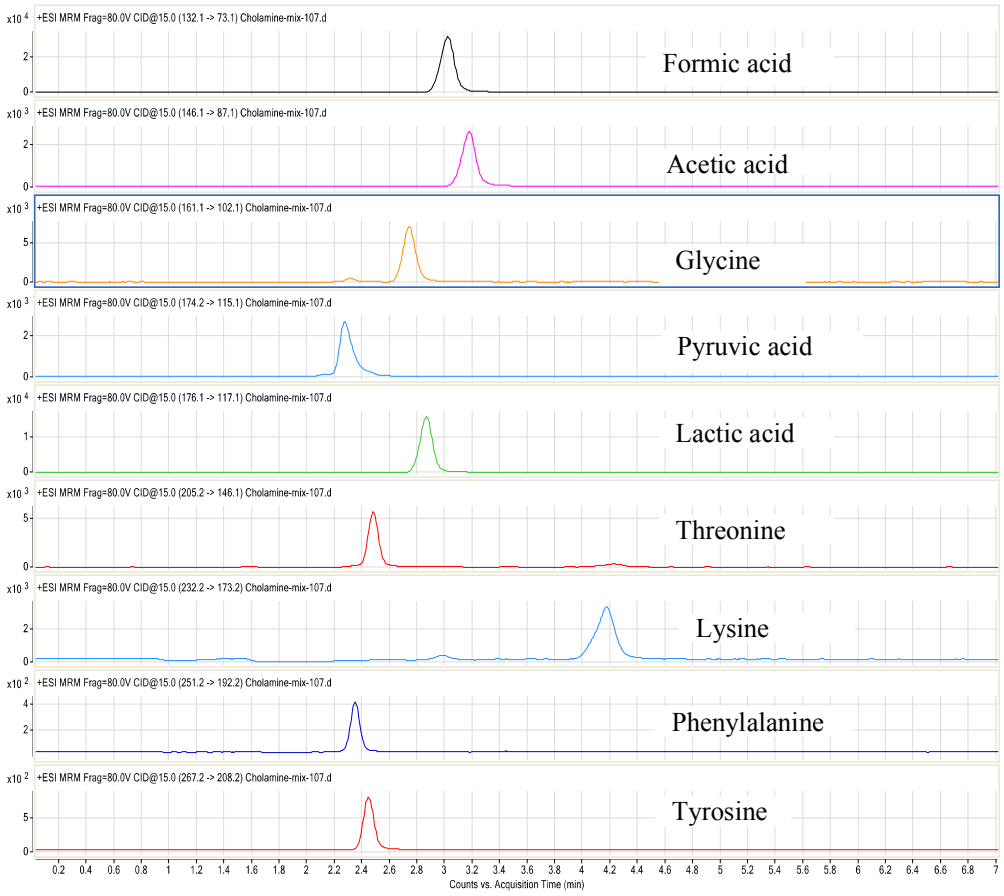


Figure 3.10 MRM chromatograms for a mixture of cholamine tagged carboxylic and amino acids detected after separation using an HILIC column, without attempting to optimize chromatography conditions. Considering that all metabolites have the same permanently charged cholamine tag, the separation achieved in a quick experiment which is still not well optimized may be remarkable.

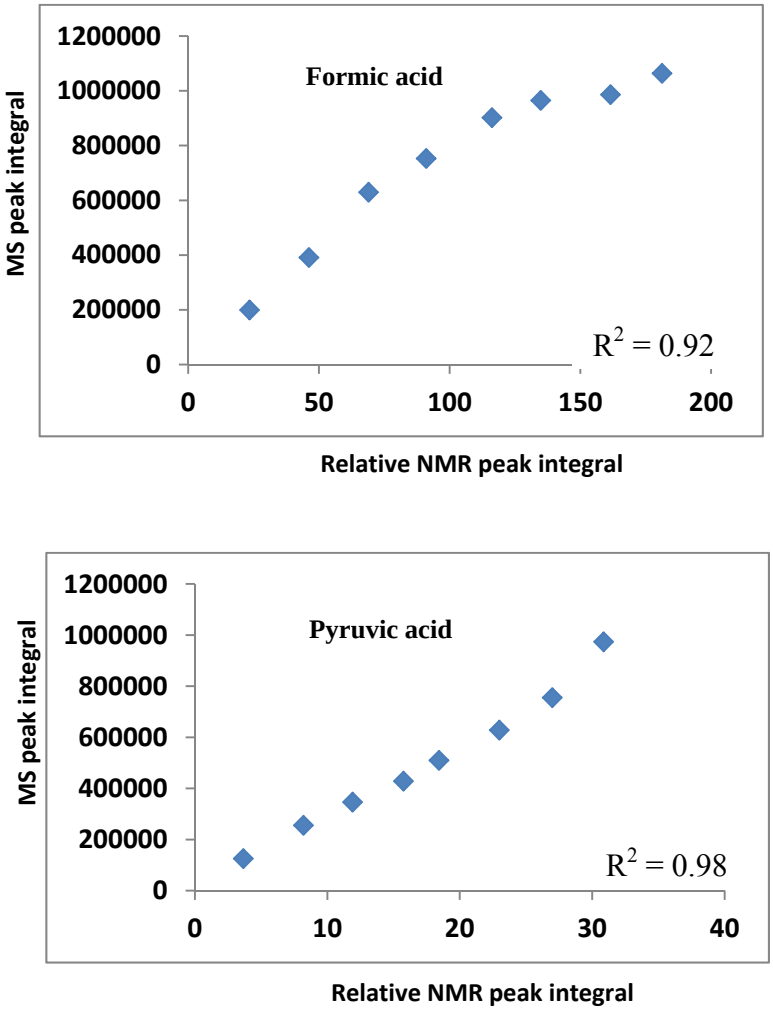


Figure 3.11 Two examples comparing the MS and NMR peak integral intensities for formic and pyruvic acids at different concentrations. Eight mixtures with random concentrations of various synthetic compounds were tagged with ^{15}N-cholamine and analyzed using NMR and MS methods. Good correlation between the NMR and MS measurements, as seen in the two figures, suggest the potential of using the new tagging approach for direct comparisons of the data from the two analytical platforms.

CHAPTER 4. APPLICATION OF HIGH-RESOLUTION MAGIC ANGLE SPINNING NUCLEAR MAGNETIC RESONANCE (HR-MAS NMR) SPECTROSCOPY FOR BREAST CANCER METABOLITE PROFILING OF AFRICAN AMERICAN COMPARE TO CAUCASIAN WOMEN

4.1 Introduction

Breast cancer is a heterogeneous group of diseases that are immunohistochemically subtyped by cancer cell expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER-2). These subtypes are recognized as: ER or PR positive and HER-2 negative subtype (ER+, PR+, HER-2-); ER or PR positive and HER-2 positive subtype (ER+, PR+, HER-2+); and ER, PR, and HER-2 negative (triple-negative breast cancer).[1,2] These subtypes not only differ in hormonal status and HER-2 expression but also clinically in their prognosis and response to therapy[3] as well as incidence rates. The incidence rates of triple negative cancer are higher in African American compared to Caucasian women, while Caucasian women have higher rates of ER-positive subtypes.[4]

Triple negative breast cancer (TNBC) constitutes about 10-20% of diagnosed breast cancer and is also more prevalent in younger women. Despite the small percentage of TNBC, it causes a disproportionate number of breast cancer deaths. This high mortality is due to the aggressive nature of the subtype that includes earlier relapses and a distinct pattern of metastasis.[5]

Because of the lack of hormonal receptors (ER and PR) and HER-2 receptor, TNBC is not responsive to treatment with hormone or anti-HER-2 monoclonal antibody therapy. Currently, TNBC treatment is limited to systemic cytotoxic chemotherapy. Interestingly, patients with TNBC treated with neoadjuvant chemotherapy and who show a pathologic complete response (pCR) show significant improvements in both disease-free survival and overall survival compared with patients with residual invasive disease. The prognosis for patients who experience pCR is excellent and equivalent to those with other breast cancer types who experience pCR. However, patients with TNBC who did not experience pCR with the same chemotherapy have a poorer prognosis.[6] In the pursuit of identifying specific targeted therapies for TNBC, preclinical studies have recently identified a few potential molecular targets such as epidermal growth factor receptor (EGFR), SRC, MET and poly ADP ribose polymerase 1/2 (PARP1/2); however, drug candidates developed for these targets have underperformed at some point during their clinical testing.[5]

Therefore, it is important to find effective treatments for TNBC. A better understanding of the biology of TNBC will aid in identifying new molecules for specifically targeting this disease. Molecular targets such as protein and genes have been explored for early detection and treatment of TNBC without much success; however, recently metabolites are being examined as an alternative and promising approach. Cancer metabolic profiling enables identification of small-molecule metabolites in biofluids and tissues that are sensitive to altered pathology of stimuli including disease processes. Small molecule metabolites in biological samples such as blood, urine and tissues have been examined using the powerful analytical techniques of nuclear magnetic

resonance (NMR) and mass spectrometry (MS) and their combination with multivariate statistical methods.[7] Since these metabolites are sensitive to subtle differences in pathological status, metabolites profiling may identify altered pathways and key enzymes that could be targeted therapeutically.

In this study, we used a metabolomics approach to identify altered small-molecule metabolites in TNBC. Intact cancer tissue samples (n=47) and normal adjacent tissue (n=35) from 47 patients (30 African Americans, 17 Caucasian; normal adjacent 18 African Americans, 17 Caucasian) were obtained before neoadjuvant chemotherapy, and were studied using high resolution magic angle spinning (HR-MAS) ^{1}H-NMR and multivariate statistics methods.

4.2 Materials and Methods

4.2.1 Chemicals and Patients Samples

Deuterium oxide (D_2O, 99.9% D) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA).

A total of 82 human breast tissue samples (tumor and normal adjacent) were collected from patients operated on at the Indiana University School of Medicine Teaching Hospitals, Indianapolis IN, the University of Chicago, Chicago, IL and Arnett Clinic in Lafayette, IN. Samples were frozen in liquid nitrogen prior to storage and shipping to Purdue University for analysis.

4.2.2 Sample Preparation

Frozen tissue samples were removed from their Eppendorf storage tubes using tweezers, placed in a petri dish and kept on dry ice during the preparation. Tissues samples were cut into an appropriate size, with weights between 11.4 and 22.4 mg, such that they could fit into an HR-MAS NMR sample tube for analysis. To provide a field-frequency lock and for air removal 50µl D_2O was also transferred into each sample tube, and then into the HR-MAS rotor and analysis.

4.3 NMR Experiments and Data Processing

4.3.1 NMR Experiments

1D ^{1}H NMR experiments on the tissue samples were performed on a Bruker Avance-III-800 spectrometer equipped with an HR-MAS probe. NMR data were acquired using the 1D CPMG (Carr-Purcell-Meiboom-Gill) pulse sequence with water presaturation and the following parameters: number of scan=128; number of dummy scans=16; number of time domain points =32K; spectral width=15.24 ppm; relaxation delay=2 sec; acquisition time=1.34 sec; number of CPMG 180° pulses=400.

4.3.2 Data Processing and Statistics

After acquisition, the NMR data were Fourier transformed after apodization, and the spectra were then phased, baseline corrected and referenced to the lipid peak δ=0.909 ppm using Bruker Topspin 3.0 software. The data were then binned to 4K buckets of

equal width (0.0034 ppm) to minimize errors due to any fluctuations of chemical shifts arising from pH or ion concentration variations using MestReNova 7.0 software. The resulting data generated from MestReNova were transferred into Excel (Microsoft Office Excel 2011). Spectral regions within the range of 0 to 9.0 ppm were used for the analysis after removing the residual water peak in the range of 4.5 to 5.0 ppm. To identify significantly different spectral bins for each metabolite between tumors and normal adjacent, the unpaired Student's *t*-test was used. P-values ≤ 0.05 were considered to be statistically significant. The binned NMR data were imported into Matlab (R2008a, Mathworks, Natick, MA) installed with a PLS-DA toolbox (version 4.1, Eigenvector Research, Inc., Wenatchee, WA) to classify tumor and normal adjacent groups. The R statistical package (version 3.0.0) was used to generate box-and-whisker plots and receiver operating characteristics (ROC) curves.

4.4 Results

4.4.1 Biomarker Discovery and Evaluation

Clinicopathological characteristics of patients and samples are summarized and listed in Table 4.1. Representative HR-MAS NMR spectra of the 82 breast tissue samples (47 tumors and 35 normal adjacent) are shown in Figure 4.1. The 29 metabolites considered for multivariate classification models are listed in the Table 4.2 and indicated in Figure 4.1.

4.4.1.1 Effect of Age and Race

Supervised multivariate statistics using partial least squares discriminant analysis (PLS-DA) models were developed to evaluate the combination of potential biomarker candidates. The 29 metabolites in Table 4.2 were selected as the variables to build the PLS-DA model. Leave-one-out cross-validation was performed to obtain the best model and avoid over-fitting. Metabolites with $p<0.05$ are listed in Table 4.3 for tumor vs. adjacent normal tissue for all ages, age>50, and age<50.. The PLS-DA score plots derived from tumor versus normal adjacent tissue for all ages, age>50, and age<50 are shown in Figure 4.2A, 4.2B, and 4.2C, respectively. Box-and-whisker plots for metabolites with $p<0.05$ for the older and younger age groups are shown in Figure 4.3 and 4.4, respectively.

4.4.1.2 Effect of Estrogen Receptor Status

To study the effect of estrogen status on the breast cancer metabolite profile, PLS-DA was again used to build a multivariate model to evaluate the combination of potential biomarker candidates. To build the PLS-DA model the 29 metabolites in Table 4.2 were again selected as the variables. Leave-one-out cross-validation was done to obtain the best model and avoid over-fitting. Metabolites with $p<0.05$ are listed in Tables 4.5 and 4.6 for ER-negative and ER-positive tissue samples, respectively. The PLS-DA score plots of tumor versus normal adjacent ER-negative and ER-positive tissue samples are shown as in Figure 4.5A and 4.5B, respectively. Box-and-whisker plots are shown for the metabolites with $p<0.05$ for both groups in Figure 4.6 and 4.7, respectively.

For a better understanding of the effect of estrogen receptor status on the metabolic profiles of African Americans and Caucasians, PLS-DA models consisting of all the metabolites in Table 4.2 was constructed to compare tumor versus normal adjacent tissues. Figure 4.8A and 4.8B are the PLS-DA metabolite profile score plots for African Americans with ER-negative and ER-positive, respectively. Figure 4.8C and 4.8D show ROC curves generated from the PLS-DA models of Figure 4.8A and 4.8B, respectively. Box and whisker plots of the relative levels of metabolites with $p<0.05$ for ER-negative and ER-positive African American women are shown in Figures 4.9 and 4.10, respectively.

Uridine was the only metabolite in the group of ER-positive Caucasian women with a $p<0.05$ ($p = 1.06*10^{-2}$); no significant p-value for ER-negative Caucasian women was observed. The box and whisker plot for uridine in tissue samples from ER-positive Caucasian women is shown in Figure 4.111.

4.5 Discussion

The feasibility of HR-MAS NMR spectroscopy of breast tissue was investigated to distinguish the metabolic differences between tumor and normal adjacent tissue classified based on estrogen receptor status, age and race. HR-MAS NMR has been used before in several cancer studies including colon, brain, prostate and breast cancer.[8-11] In this study, we investigated this technique to be used as an adjunct tool for differentiation of tumor from normal adjacent breast tissue. Based on our results, we believe this methodology, when combined with multivariate statistical analysis, has the potential for use as a powerful complementary tool to the current clinical histopathological methods

for better diagnosis and prognosis of breast cancer. This method is fairly fast, and therefore could be used to quickly assess tissue samples around the time of surgery; since NMR is nondestructive technique, the same sample can be used later for histopathology.

Our results showed that more metabolites change significantly when comparing tumor to normal adjacent in women with ages above 50 compared to those below 50 years old. Myo-inositol and phosphocholine were significantly (p<0.05) increased in tumor compare to normal adjacent tissues in both age groups. Choline, lactate, glutamate, taurine, methionine, alanine, threonine, glycine, tyrosine, ATP (adenosine triphosphate), glutathione, unsaturated lipid, lipid, valine, phenyalanine, glutamine, and α-glucose were significantly altered for patients above 50 years old. However, for patients below 50 years old only uridine in addition to myo-inositol and phosphocholine was changed.

Data from Caucasian and African American women were also studied separately. For Caucasians, most metabolites were not changed significantly whereas they were for African Americans. These results could be affected by the sample collection methods. Samples obtained from the University of Chicago mainly belonged to African Americans, and thus this set was most likely consistent.

Estrogen receptor (ER) refers to a group cytoplasmic proteins existing in normal estrogen target tissues such as uterus and breast. The ER status in invasive carcinomas is very important for breast cancer prognosis. Studies show that women with ER-positive tumors generally have a better prognosis.[12,13] Moreover, in addition to the differences in morphology, ER-positive and ER-negative tumor tissues they have different metabolite profiles due differences in their altered metabolic pathways.[14-16] Glutathion, myo-inositol, taurine and lactate were significantly altered in tumor compare to normal adjacent tissues.

Our results show that fewer metabolites change concentration in tumor compared to normal adjacent tissues for ER-negative patients. Signals from lipid (0.909 ppm), unsaturated lipid (5.333ppm) and α-glucose decreased in ER-negative tumors , the rest of the metabolites shown in Figure 4.6 and 4.7 for ER-negative and positive patients, respectively, were increased in tumor compare to normal adjacent tissue samples.

Results from comparing metabolite *p*-values between races showed that these changes mostly occurred between African Americans. Uridine was the only metabolite that changed significantly ($p<0.05$) among ER-positive Caucasians. Lactate, taurine and glutathione were elevated in tumor for both ER-positive and ER-negative Caucasians. However, among African Americans, glutathione was just changed significantly in the ER-positive group ($p <0.05$). Glutathione is an intracellular antioxidant, and plays an important role in cellular defense.[17] Perry *et al.* also reported that glutathione levels were elevated in breast cancer tumor tissue compared to normal tissue.[18] Previous studies have also showed that taurine and lactate are elevated in breast cancer tumors,[19,20] and our study agrees with these results, especially for women above age 50.

Uridine was elevated in tumors among ER-positive Caucasians below 50; however, our results do not support previous studies on human breast cancer tumors that showed an increase in the activity of Uridine phosphorylase (UPase).[21,22] UPase is the enzyme responsible for the reversible phosphorolysis of uridine to uracil. Therefore, based on the previous results uridine concentration should be lower in tumor tissues compared to normal adjacent.

Higher concentrations of choline and phosphocholine in breast cancer tumors have been reported by a number of studies.[23-26] Eliyahu *et al.* showed elevated levels of

phosphocholine in breast cancer cells occurs as more choline is transported into the cells. The activity of intracellular choline kinase increases several fold above that of CTP:PCho cytidylyltransferase, and converts phosphocholine to cytidyldiphosphate-choline (CDP-choline).[26] We did not find significant change in the intensity level of these two metabolites for all patients with ER-negative tumors. However, for African Americans with ER-negative tumors, we did see a statistically significant change for choline. A study by Sitter *et al.* showed that the ratio of phosphocholine to choline is lower in ER-negative compare to ER-positive tumors.[25] A study by Shin *et al*. using Magnetic Resonance (MR) spectroscopy showed significant differences in the total choline compounds between ER-positive and ER-negative (p=0.007), as did a study by Choi *et al.* using HR-MAS. But our results was not in agreement with their results (p=0.34 for choline and p=0.74 for phosphocholine).[27,28] Likewise, several studies also indicated no significant change between for choline compounds in ER-positive versus ER-negative tumors (p=0.23).[14,29]

Findings from this study showed the intensity of numerous metabolites were changed (p<0.05) in tumors from African American women compared to Caucasians, but not significant change between them. Nevertheless, using PLS-DA results, and considering all the metabolites listed in Table 4.2, not just low p-value metabolites, showed that the HR-MAS analysis combined with multivariate statistical analysis could completely distinguish ER-negative and ER-positive tumors from normal adjacent tissues (see Figure 4.8). This result indicates that there are many smaller changes that when combined help define the altered metabolism observed in breast tumors.

The approach described in this study has some limitations. Since samples were collected in 3 locations by different individuals, differences in the collection methodology may have affected our results. It is possible that with additional number of samples and a tighter control of sample collection, better results could be achieved, especially for age<50 for both African Americans and Caucasians.

In conclusion, we showed that HR-MAS ^{1}H NMR combined with multivariate statistical analysis can be used as a powerful technique for identifying metabolic differences between different tumor types. Because of the effect called field cancerization, characterizing the altered metabolite profiles of normal adjacent tissue is as important as those for the tumor itself.[30] This effect has been previously confirmed by several studies.[31,32] Results from this study show a wide variation in the metabolite levels for tumors whereas the distribution of metabolites in normal adjacent tissue samples are much narrower. Further prospective studies with more number of samples especially for ages below 50 are needed to identify out the differences between ages in different hormone receptor status and races.

4.6 References

1. Bernstein, L.; Lacey, J. V. *J. Natl. Cancer Ins.* **2011**, *103*, 451-453.

2. Carey, L. A.; Perou, C. M.; Livasy, C. A.; Dressler, L. G.; Cowan, D.; Conway, K.; Karaca, G.;Troester M. A.; Tse, C. K.; Edminston, S.; Deming, S. L.; Geradys, J.; Cheang, M. C.; Nielson, T. O.; Moorman, P. G.; Earp, H. S.; Millikan, R. C. *JAMA*, **2006**, *295(21)*, 2492-2502.

3. Onitilo, A. A.; Engel, J. M.; Greenlee, R. T.; Mukesh, B. N. *Clin. Med. Res.* **2009**, *7(1-2)*, 4-13.

4. Clarke, C. A., Keegan, T. H., Yang, J., Press, D. J., Kurian, A. W., Patel, A. H., Lacey, J. V. *J. Natl. Cancer Inst.* **2012**, *104(14)*, 1094-1101.

5. Duffy, M. J.; McGowan, P. M.; Crown, J. *Int. J. Cancer* **2012**, *131(11)*, 2471-2477.

6. Baser,O.; Wei, W.; Xie, L.; Henk, H. J.; Teitelbaum, A. *Community Oncol*. **2012**, *9(1),* 8–14.

7. Gu, H.; Gowda, G. N.; Raftery, D. *Future Oncology* 2012, *8(10)*, 1207-1210.

8. Chan, E. C.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C. *J. Proteome Res*. **2009**, *8*, 352−361.

9. Martinez-Bisbal, M. C.; Marti-Bonmati, L.; Piquer, J.; Revert, A.; Ferrer, P.; Llacer, J. L.; Piotto, M.; Assemat, O.; Celda, B. *NMR Biomed.* **2004**, *17*, 191−205.

10. Tomlins, A. M.; Foxall, P. J. D.; Lindon, J. C.; Nicholson, J. K.; Lynch, M. J.; Everett, M. S. *Anal. Commun.* **1998**, *35*, 113-115.

11. Cao, M. D.; Sitter, B.; Bathen, T. F.; Bofin, A.; Lønning, P. E.; Lundgren, S. and Gribbestad, I. S. *NMR Biomed.* **2012**, *25*, 369–378.

12. Maynard, P. V.; Davis, C. J.; Blamey, R. W.; Elston, C. W.; Johnson, J.; Griffiths, K. *Br J Cancer* **1978**, *38(6)*, 745–748.

13. Fisher, B.; Costantino, J.; Redmond, C.; Poisson, R.; Bowman, D.; Couture, J.; et al. *N. Engl. J. Med.* **1989**, *320*, 479-484.

14. Chen, J. H.; Baek, H. M.; Nalcioglu, O.; and Su, M. Y. *J. Magn. Reson. Imaging* **2008**, *27*, 825–833.

15. Ferguson, A. T.; Davidson, N. E. *Crit. Rev. Oncog.* **1997**, *8*, 29-46.

16. Mandal, S.; Protiti Khan, Li, and Davie, J. R.; Gunduz, M. (Ed.), **2011**, ISBN: 978-953-307-714-7, InTech, DOI: 10.5772/22987. Available from: http://www.intechopen.com/books/breast-cancer-carcinogenesis-cell-growth-and-signalling-pathways/metabolomics-and-transcriptional-responses-in-estrogen-receptor-positive-breast-cancer-cells.

17. Ortega, A. L.; Mena, S.; Estrela, J.M. *Cancers* **2011**, *3*, 1285-1310.

18. Perry R. R.; Mazetta, J.; Levin, M.; Barranco S.C. *Cancer* **1993**, *72*, 783-787.

19. Bathen, T. F.; Geurts, B.; Sitter, B.; Fjosne, H. E.; Lundgren, S.; Buydens, L. M.; Gribbestad, I. S.; Postmag, G.; Giskeodegard, G. F. *PloS one* **2013**,*8*, e61578.

20. Giskeødegård G. F.; Lundgren, S.; Sitter, B.; Fjosne, H. E.; Postma, G.; Buydens, L.; Gribbestad, I. S.; Bathen, T. F. *NMR Biomed.* **2012**, *25*, 1271-1279.

21. Liu, M.; Cao, D.; Russell, R.; Handschumacher, R. E.; Pizzorno G. *Cancer Res.* **1998**, *58*, 5418-5424.

22. Kanzaki, A.; Takebayashi, Y.; Bando, H.; Eliason, J. F.; Watanabe, S. S.; Miyashita, H.; Fukumoto, M.; Toi, M.; Uchida, T. *Int. J. Cancer* **2002**, *97*, 631-635.

23. Glunde, K.; Bhujwalla, Z. M.; Ronen, S. M. *Nat. Rev. Cancer* **2011**, *11*, 835-848.

24. Ackerstaff, E.; Glunde, K.; Bhujwalla, Z. M. *J. Cell. Biochem.* **2003**, *90*, 525–533.

25. Sitter, B.; Bathen, T. F.; Singstad, T. E.; Fjosne, H. E.; Lundgren, S.; Halgunset, J.; Gribbestad, I. S. *NMR Biomed.* **2010**, *23*, 424−31.

26. Eliyahu, G.; Kreizman, T.; Degani, H. *Int. J. Cancer* **2007**, *120*, 1721–1730.

27. Shin, H. J.; Baek, H. M.; Cha, J. H.; Kim, H. H. *Am. J. Roentgenol.* **2012**, *198*, W488-W497.

28. Choi, J. S.; Baek, H. M.; Kim, S.; Kim, M. J.; Youk, J. H.; Moon, H. J.; Kim, E. K.; Han, K. H.; Kim, D. H.; Kim, S. I.; Koo, J. S. *PLoS ONE* **2012**,*7*, e51712.

29. Baek, H. M.; Chen, J. H.; Nalcioglu, O.; SU, M. Y. *Ann. Oncolo.* **2010**,*21*, 663-665.

30. Slaughter, D. P.; Southwick, H. W.; Smejkal, W. *Cancer* **1953**, *6*, 963−8.

31. Yakoub, D.; Keun, H. C.; Goldin, R.; Hanna, G. B. *Cancer research* **2010**,70, 9129-9136.

32. Dakubo, G. D.; Jakupciak, J. P.; Birch-Machin, M. A.; Parr, R. L. *Cancer Cell Int.* **2007**, *7:*2.

Table 4.1 Clinicopathological characteristics of women with invasive breast cancer.

Patient Characteristics	Number
Total Patients	47
African American women	30 (T=30; N=18)
Caucasian women	17 (T=17; N=17)
Pathology	
Invasive carcinoma grade I and DCIS	3
Invasive carcinoma grade II	10
Invasive carcinoma grade III	25
Unknown	9
Patient Age	
<50	10
>50	37
African American women	
<50	7
>50	23
Caucasian women	
<50	3
>50	14
ER Status	
ER-	18
ER+	29
African American women	
ER-	13
ER+	17
Caucasian women	
ER-	5
ER+	12

T: tumor.
N: Normal adjacent.
ER: Estrogen receptor

Table. 4.2 Quantified metabolites used for multivariate classification models.

Number	Metabolite	Chemical Shifts (ppm)	Number	Metabolite	Chemical Shifts (ppm)
1	Acetate	1.927	15	Lactate	4.124
2	Alanine	1.479	16	Lipid 1	0.909493
3	Arginine/Lysine/ Leucine	1.786-1.649	17	Lipid 2	2.777
4	ATP	6.125	18	Methionine	2.648
5	α-glucose	5.238	19	Myo-inositol	3.536
6	β-glucose	4.652	20	Phenylalanine	7.448-7.298
7	Choline	3.21	21	Phosphocholine	3.229
8	Creatine	3.032	22	Taurine	3.431
9	Formate	8.411	23	Threonine	4.261
10	Glutamate	2.355	24	Tyrosine(1^{st} peak)	6.891
11	Glutamine	2.455	25	Tyrosine(2^{nd} peak)	7.189
12	Glutathione	2.561	26	Unsaturated lipid	5.333
13	Glycine	3.567	27	Uridine	5.907
14	Isobutyrate	1.146	28	Valine	1.048

Table. 4.3 Metabolites significantly ($p<0.05$) expressed between tumor and adjacent normal tissues.

Metabbolites	T vs. N (all ages)	Metabbolites	T vs. N (above 50)	Metabbolites	T vs. N (below 50)
myo-inositol	1.78E-04	choline	2.40E-04	uridine	2.17E-02
taurine	5.50E-04	lactate	2.69E-04	myo-inositol	3.99E-02
lactate	6.49E-04	glutamate	7.84E-04	phosphocholine	4.87E-02
phosphocholine	8.82E-04	myo-inositol	1.43E-03		
glutathione	2.57E-03	taurine	3.43E-03		
glutamine	4.67E-03	methionine	4.00E-03		
ATP	6.22E-03	alanine	4.19E-03		
glutamate	8.76E-03	phosphocholine	4.89E-03		
choline	1.15E-02	threonine	6.50E-03		
glycine	3.21E-02	glycine	6.90E-03		
alanine	3.40E-02	tyrosine 2	1.15E-02		
uridine	3.42E-02	tyrosine 1	1.16E-02		
tyrosine 1	4.06E-02	ATP	1.17E-02		
creatine	4.26E-02	glutathione	1.21E-02		
tyrosine 2	4.60E-02	unsaturated lipid	1.23E-02		
		lipid	1.45E-02		
		valine	1.53E-02		
		phenyalanine	1.64E-02		
		glutamine	2.31E-02		
		α-glucose	4.43E-02		

T: tumor
N: normal adjacent

Table 4.4 Metabolites significantly expressed ($P<0.05$) in African American women: tumors vs. adjacent normal.

Metabolites	Tumor African Americans vs. African American normal adjacent
lactate	6.90E-04
myo-inositol	1.15E-03
choline	1.16E-03
glutamate	1.21E-03
l-methionine	2.49E-03
taurine	3.82E-03
Alanine	4.52E-03
tyrosine 1	4.86E-03
glycine	5.27E-03
phosphcholine	5.53E-03
ATP	8.01E-03
tyrosine 2	8.13E-03
threonine	8.20E-03
lysine, leucine, arginine	8.74E-03
valine	1.08E-02
phenylalanine	1.15E-02
unsaturated lipid	1.29E-02
lipid	1.93E-02
glutamine	2.00E-02
glutathione	2.59E-02

Table. 4.5 Metabolites differentially expressed (P<0.05) in ER-negative samples (tumor vs. adjacent normal) regardless of race.

Metabolites	ER- tumor vs. ER- normal adjacent
unsaturated lipid	7.60E-03
lipid	1.14E-02
myo-inositol	1.34E-02
glutathione	1.83E-02
taurine	1.96E-02
α-glucose	2.02E-02
ATP	2.62E-02
lactate	3.40E-02
creatine	4.59E-02

Table. 4.6 Metabolites significantly expressed (P<0.05) in ER-positive samples (Tumors vs. adjacent normal) regardless of race

Metabolites	ER^+ tumor vs. ER^+ normal adjacent
lactate	2.75E-03
choline	4.61E-03
phosphcholine	6.94E-03
myo-inositol	6.96E-03
glutamate	8.01E-03
methionine	8.21E-03
threonine	1.20E-02
taurine	1.49E-02
uridine	1.64E-02
Alanine	2.17E-02
glutamine	2.70E-02
glutathione	3.07E-02
glycine	3.29E-02
valine	4.51E-02
tyrosine 2	4.57E-02

Table. 4.7 Metabolites significantly altered (P<0.05) in ER negative African American women: tumor vs. adjacent normal tissues.

Metabolites	ER- tumor vs. ER- normal adjacent
unsaturated lipid	6.79E-03
a-glucose	7.97E-03
lipid	9.95E-03
lactate	2.04E-02
glycinne	2.17E-02
threonine	2.20E-02
Alanine	2.32E-02
glutamate	2.38E-02
glutathione	2.62E-02
taurine	2.84E-02
choline	2.90E-02
tyrosine 2	2.97E-02
tyrosine 1	3.27E-02
ATP	3.55E-02
valine	3.61E-02

Table. 4.8 Metabolites differentially altered (P<0.05) in ER-positive African Americans: tumor vs. adjacent normal.

Metabolites	ER^+ tumor vs. ER^+ normal adjacent
myoinositol	6.79E-03
phosphcholine	9.09E-03
choline	1.91E-02
lactate	1.96E-02
methionine	2.28E-02
glutamate	2.69E-02
taurine	3.97E-02
L-lysine, L-leucine, arginine	4.08E-02
tyrosine 1	4.32E-02

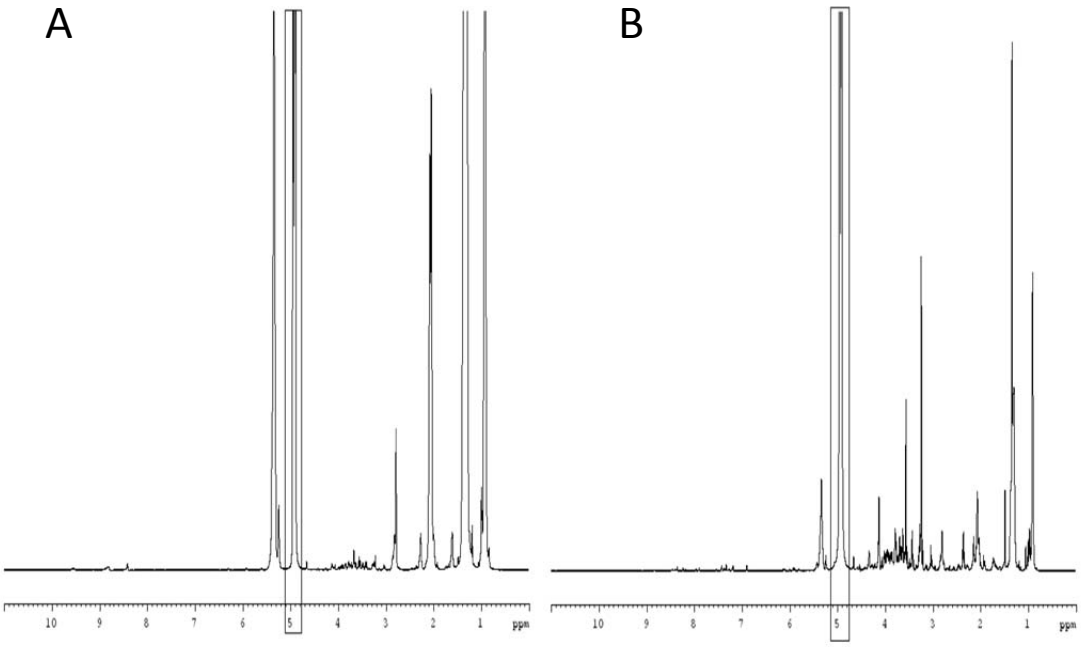


Figure 4.1 Typical HR-MAS tissue spectra from (A) normal adjacent and (B, C, and D) tumor breast cancer tissues.

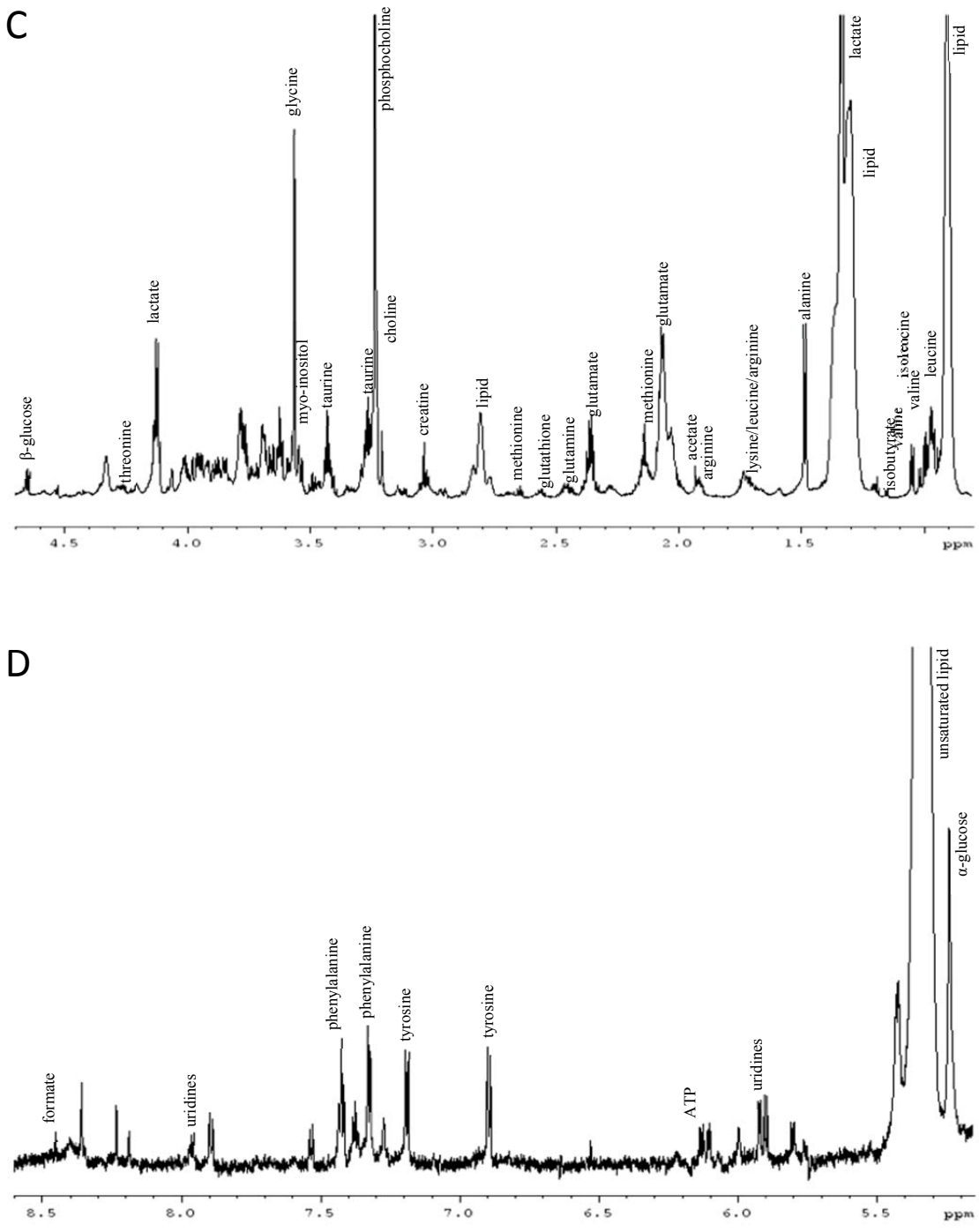


Figure 4.1, continued

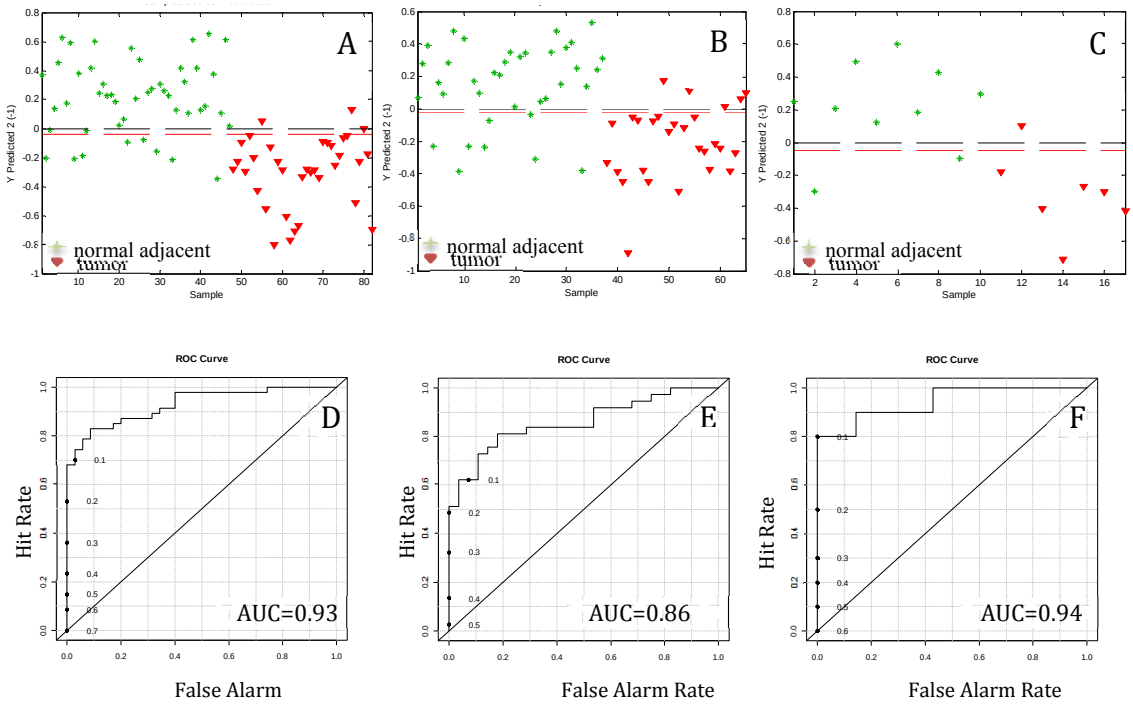


Figure 4.2 Results of PLS-DA models using the 29 metabolites from Table 2: (A) all samples (B) samples with age above 50 years old and (C) samples with age below 50 years old. ROC curves using the cross-validated predicted class values for (D) all samples (E) samples with age above 50 years old and (F) samples with age below 50 years old.

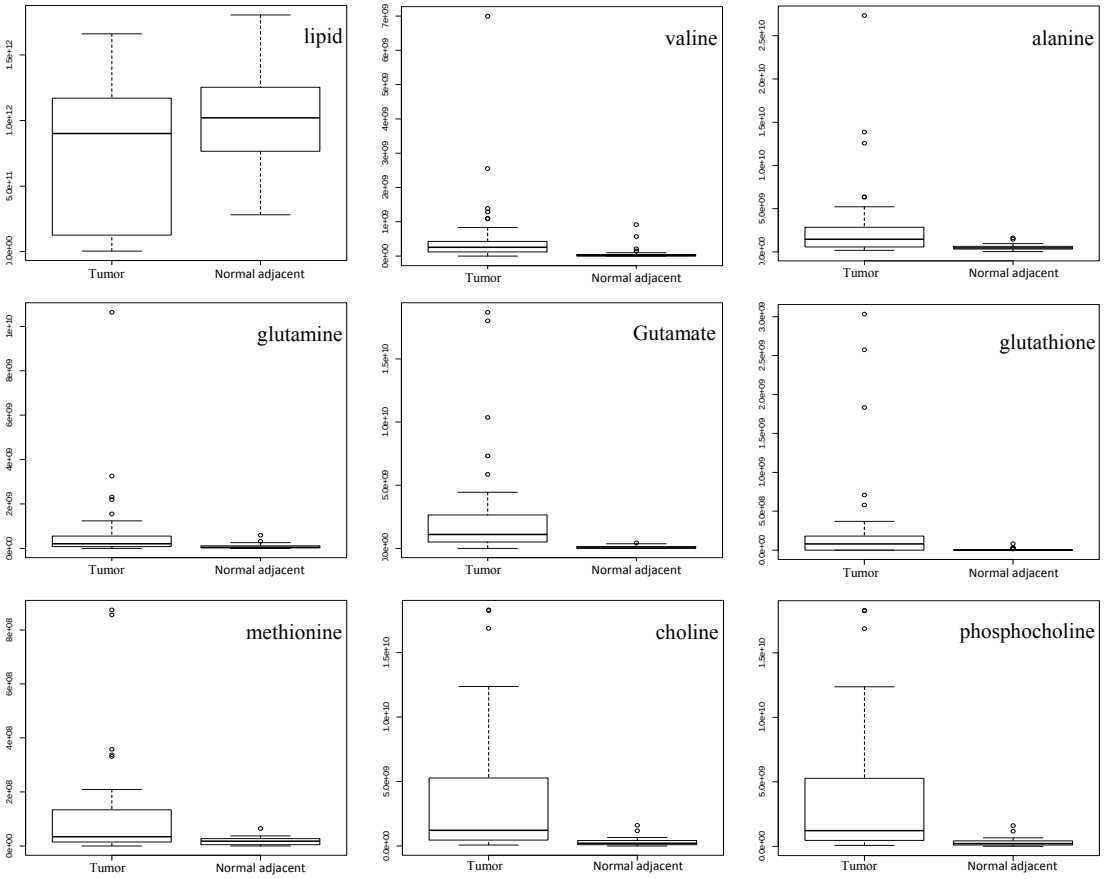


Figure 4.3 Box-and-whisker plots of metabolites with $p<0.05$ illustrating discrimination between tumor and normal adjacent above 50 years old. Horizontal line in the middle portion of the box indicates the median Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively. Open circles show outliers.

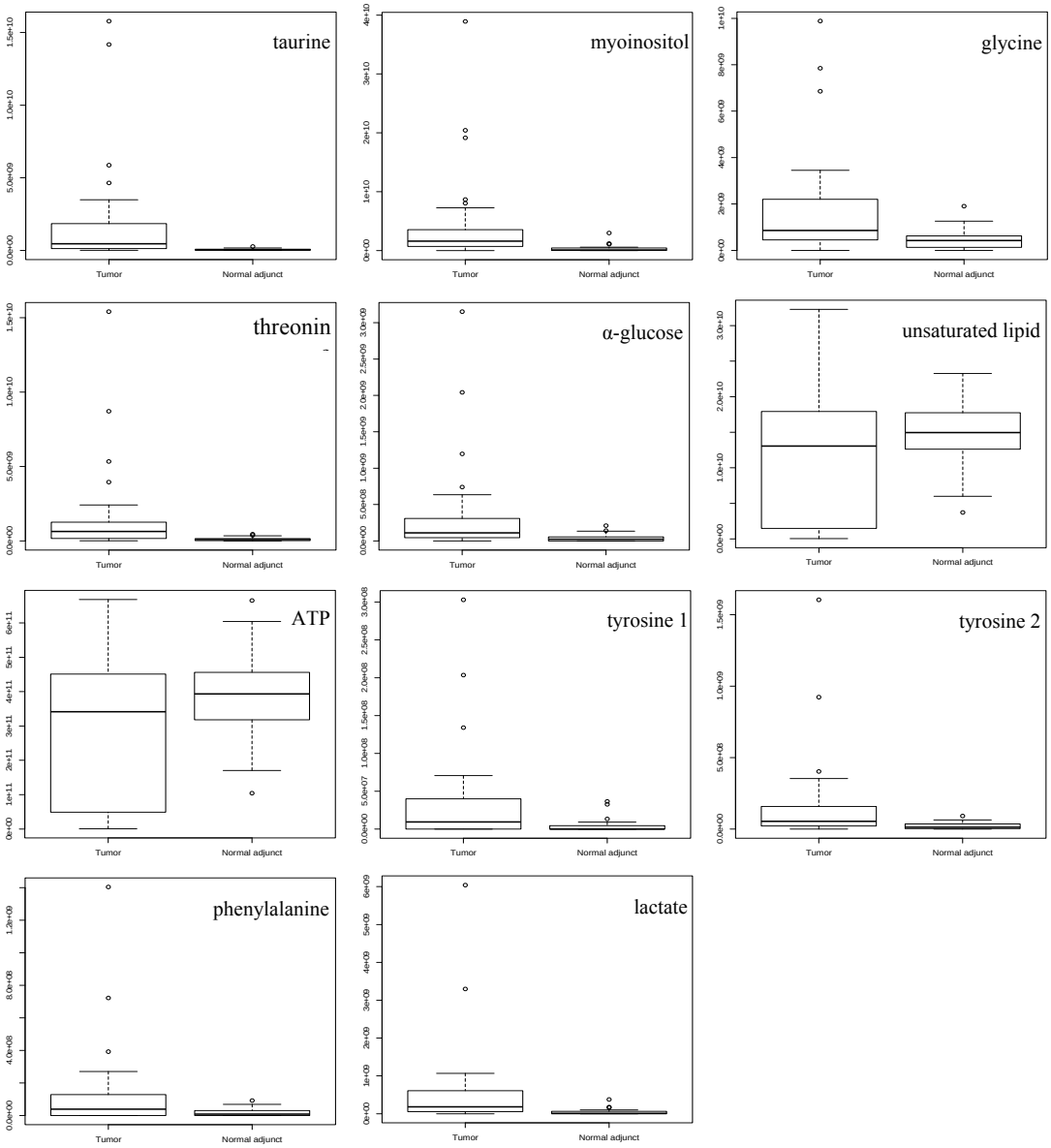


Figure 4.3, Continued.

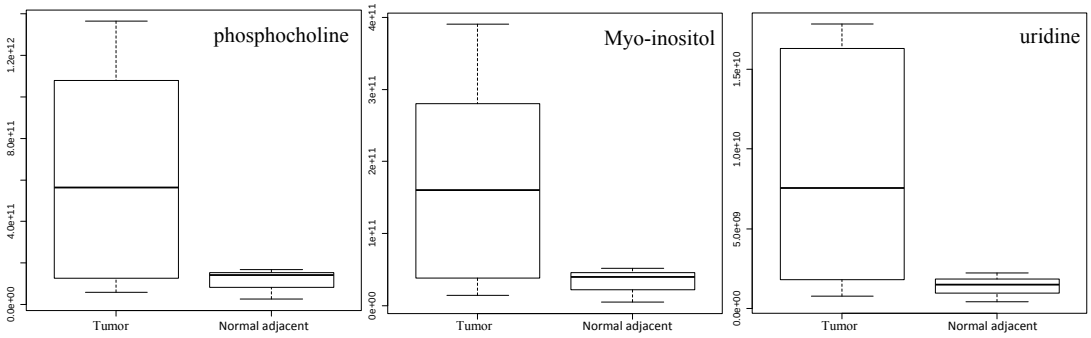


Figure 4.4 Box-and-whisker plots of metabolites with $p<0.05$ illustrating discrimination between tumor and normal adjacent tissues for samples from women below 50 years old. Horizontal line in the middle portion of the box indicates the median. Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively, and the open circles show outliers.

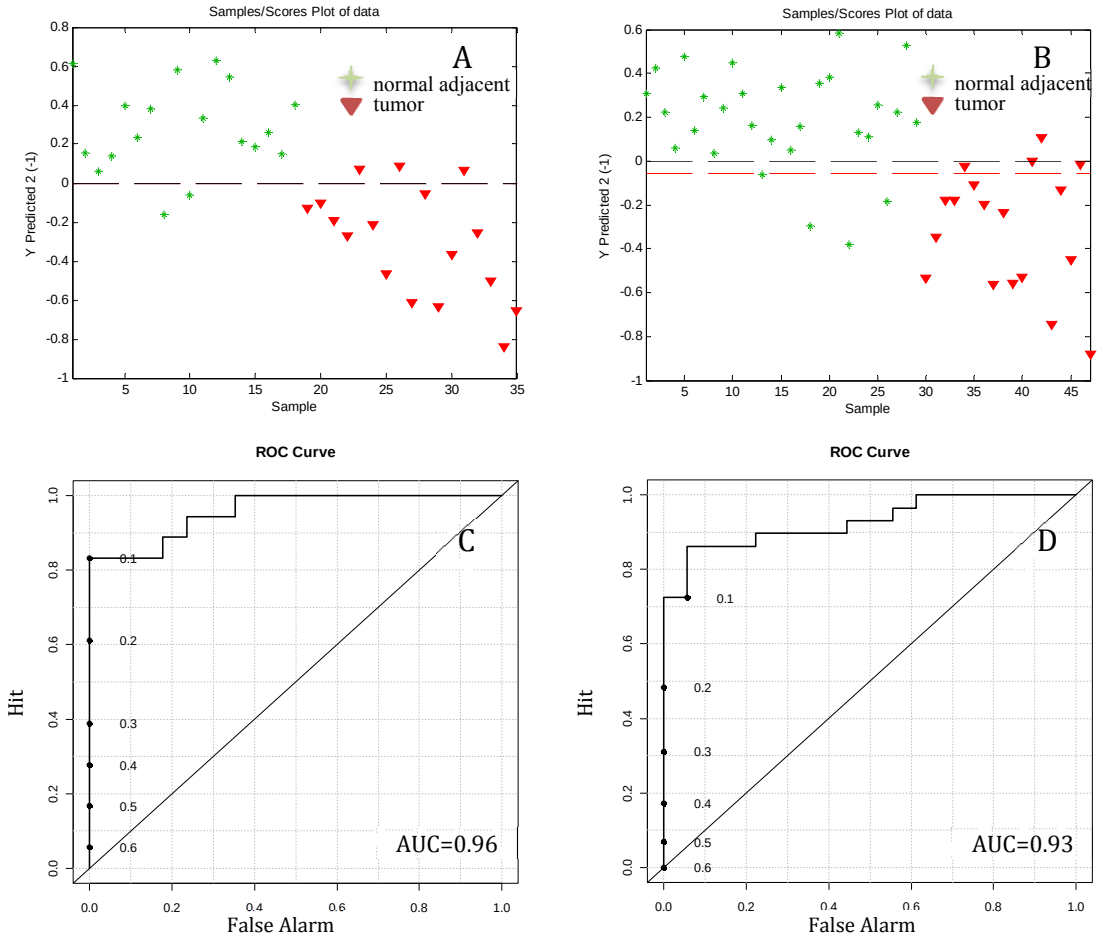


Figure 4.5 Results of the PLS-DA model from the 29 metabolites (A) ER-negative samples, and (B) ER-positive samples. ROC curves using the cross-validated predicted class values for (D) ER-negative samples and (E) ER-positive samples.

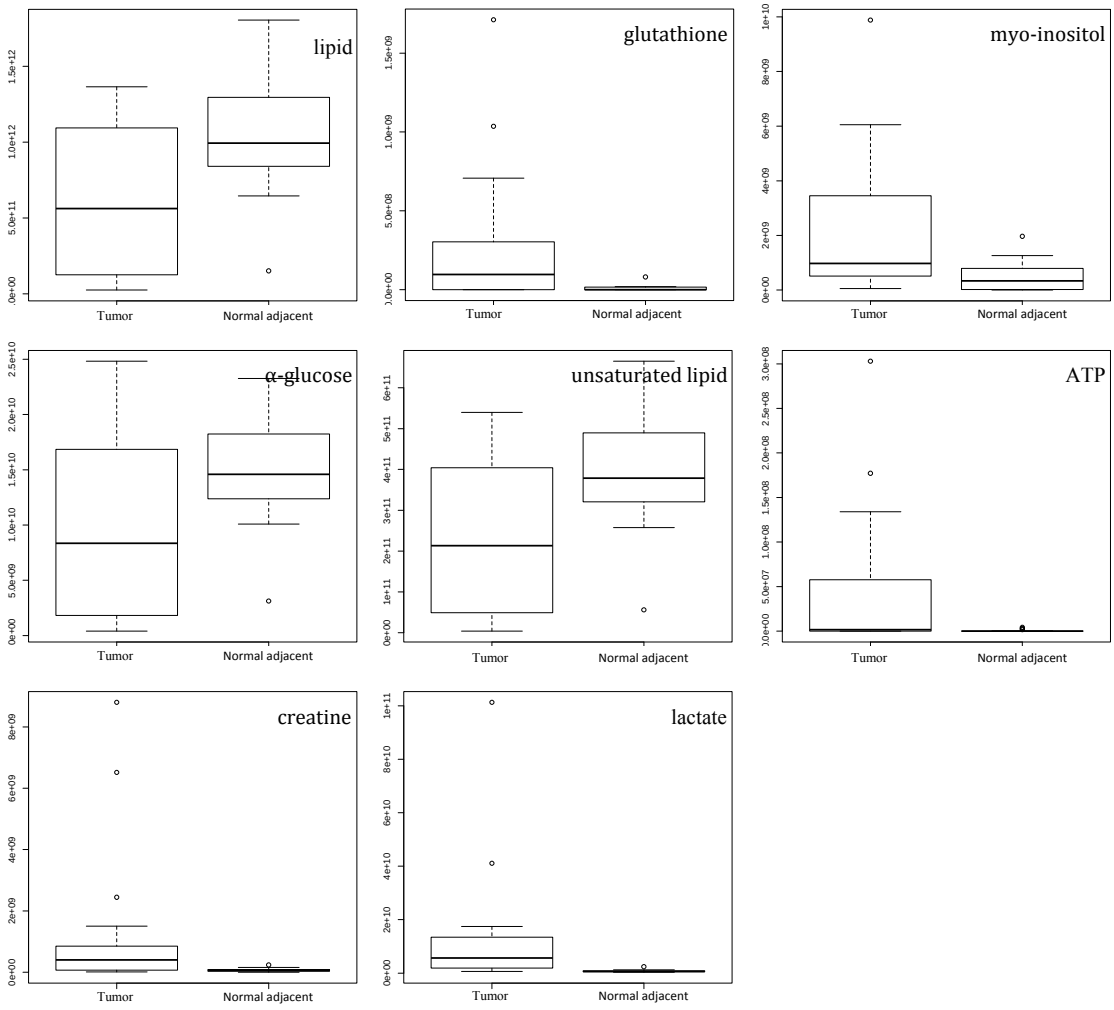


Figure 4.6 Box-and-whisker plots of metabolites with $p<0.05$ illustrating discrimination between tumor and normal adjacent tissues for ER-negative samples. Horizontal line in the middle portion of the box indicates the median. Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively. Open circles show outliers.

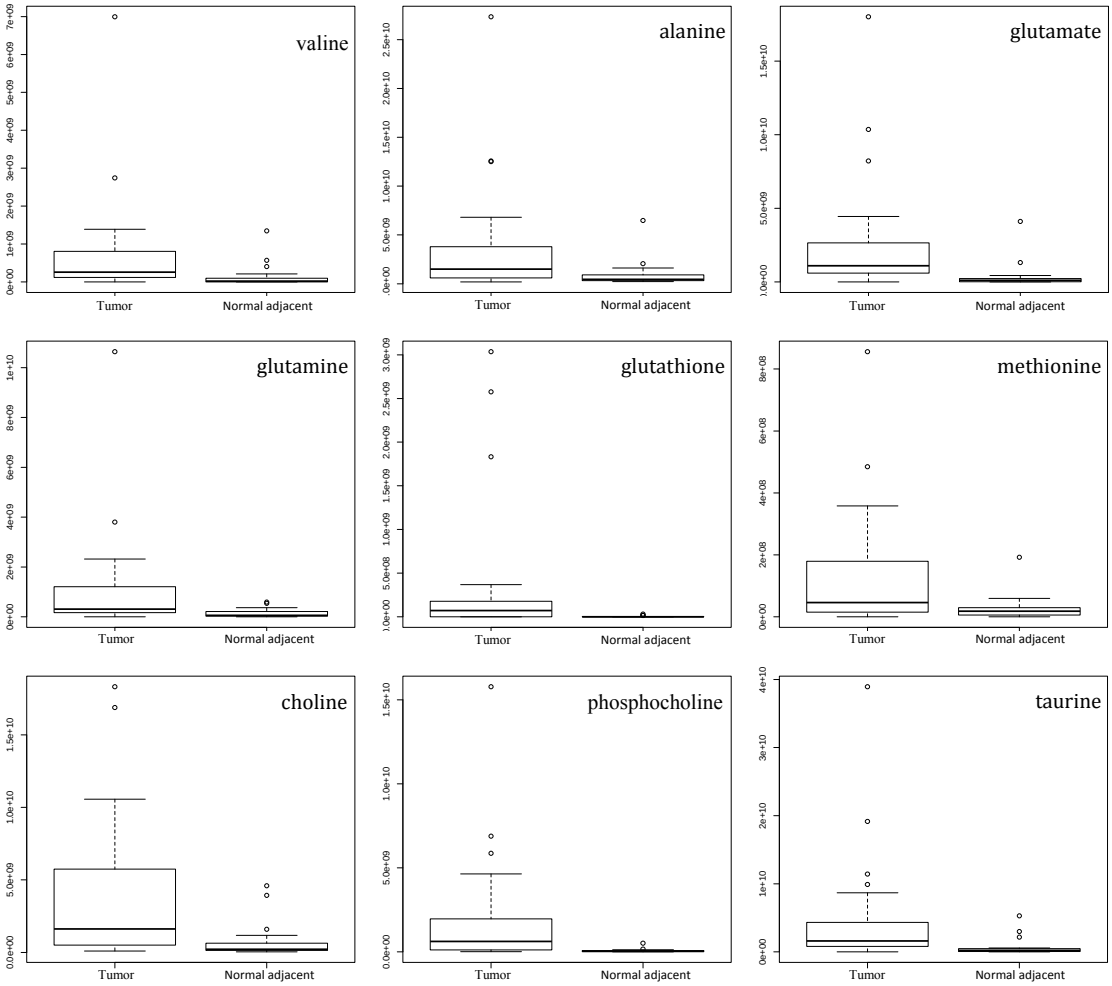


Figure 4.7 Box-and-whisker plots of metabolites with $p<0.05$ illustrating discrimination between tumor and normal adjacent tissues for ER-positive samples. Horizontal line in the middle portion of the box indicates the median. Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively. Open circles show outliers.

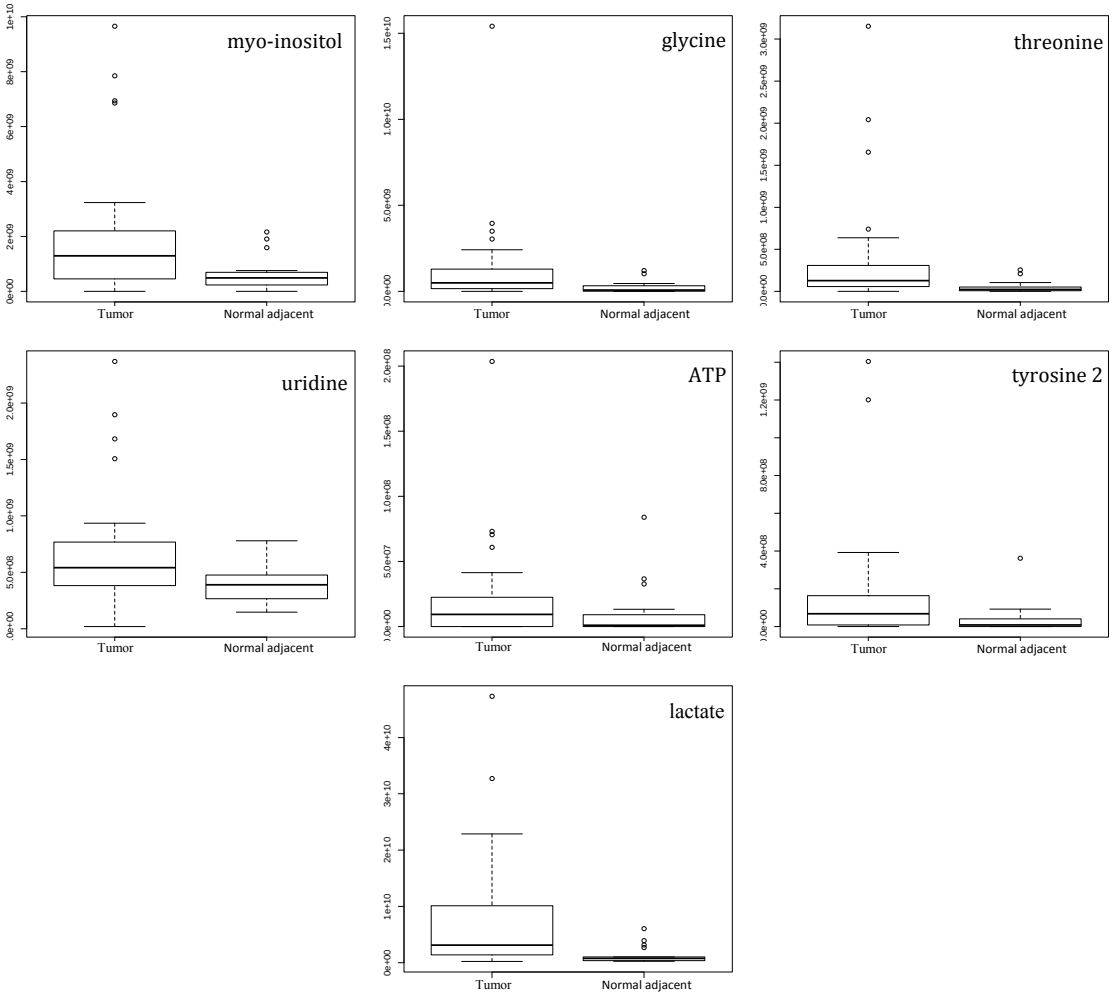


Figure 4.7, Continued.

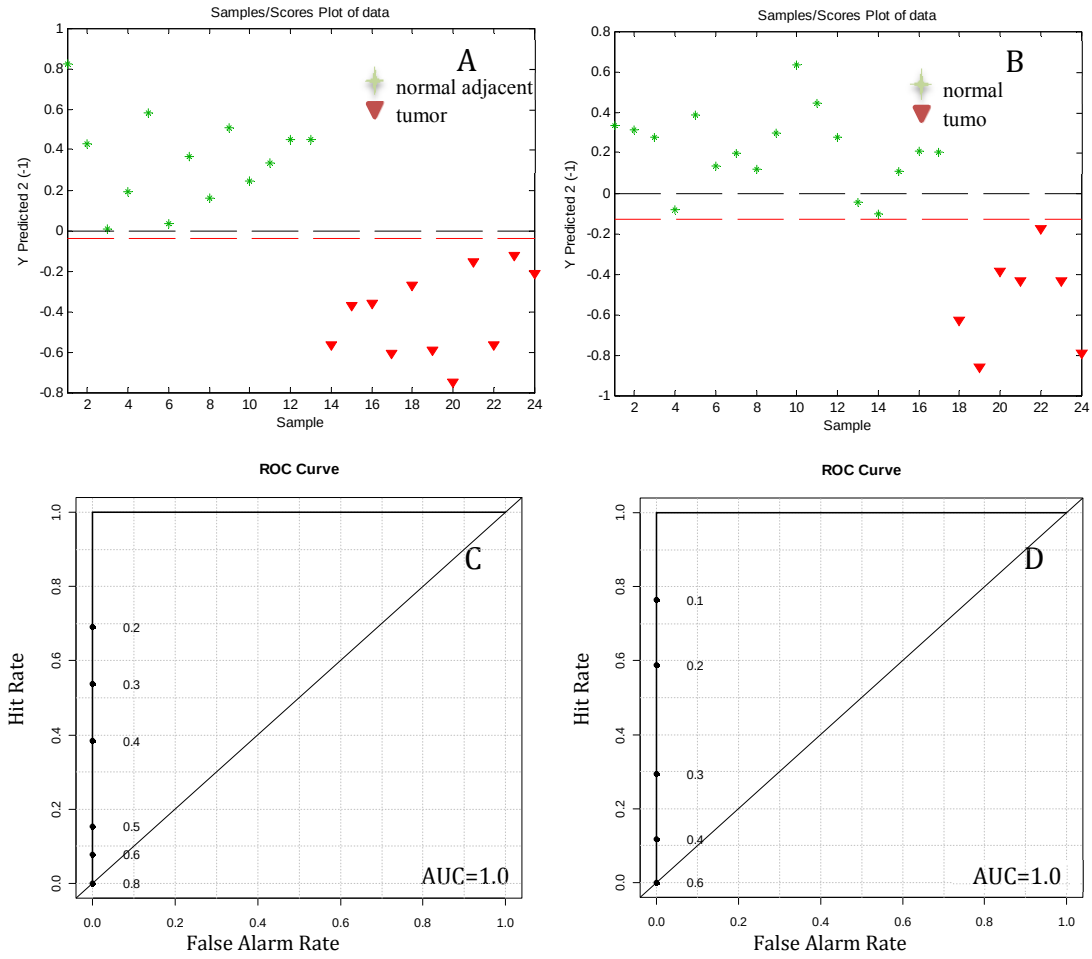


Figure 4.8 Results of the PLS-DA model using the 29 metabolites for African Americans: (A) ER-negative samples and (B) ER-positive samples. ROC curves using the cross-validated predicted class values for African Americans: (C) ER-negative samples and (D) ER-positive samples.

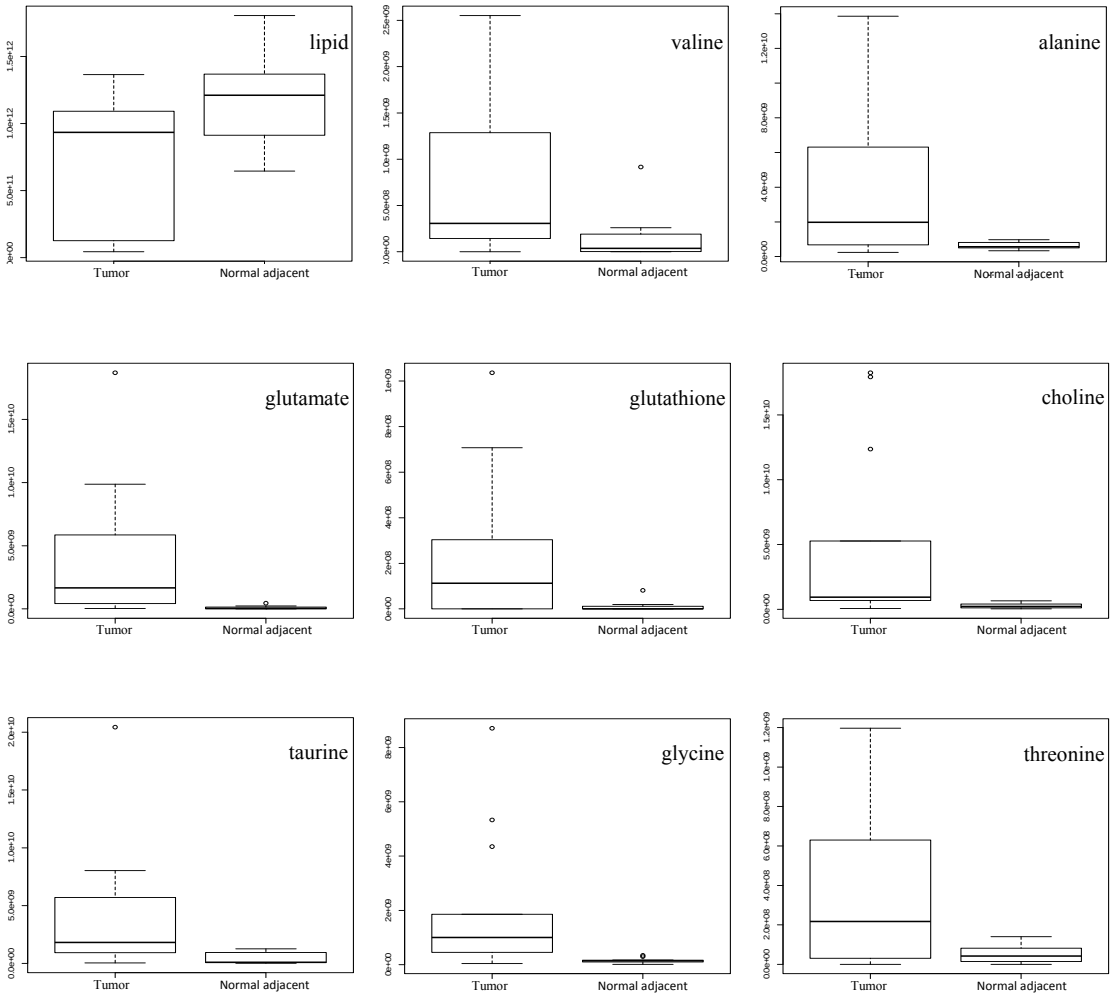


Figure 4.9 Box-and-whisker plots of metabolites with $p<0.05$ illustrating discrimination between tumor and normal adjacent tissues for ER-negative African Americans. Horizontal line in the middle portion of the box indicates the median. Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively. Open circles show outliers.

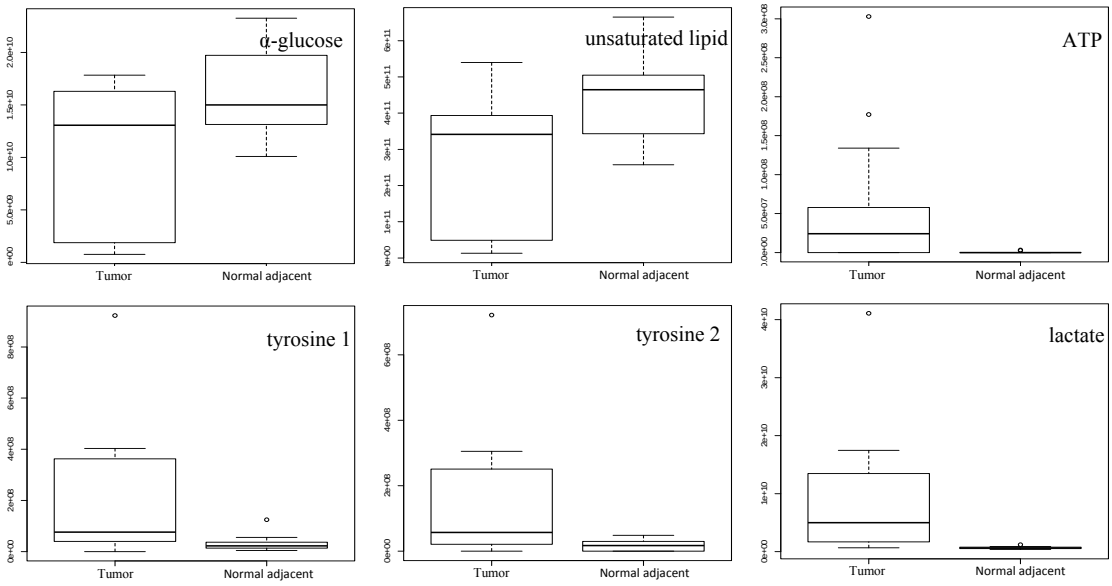


Figure 4.9, Continued.

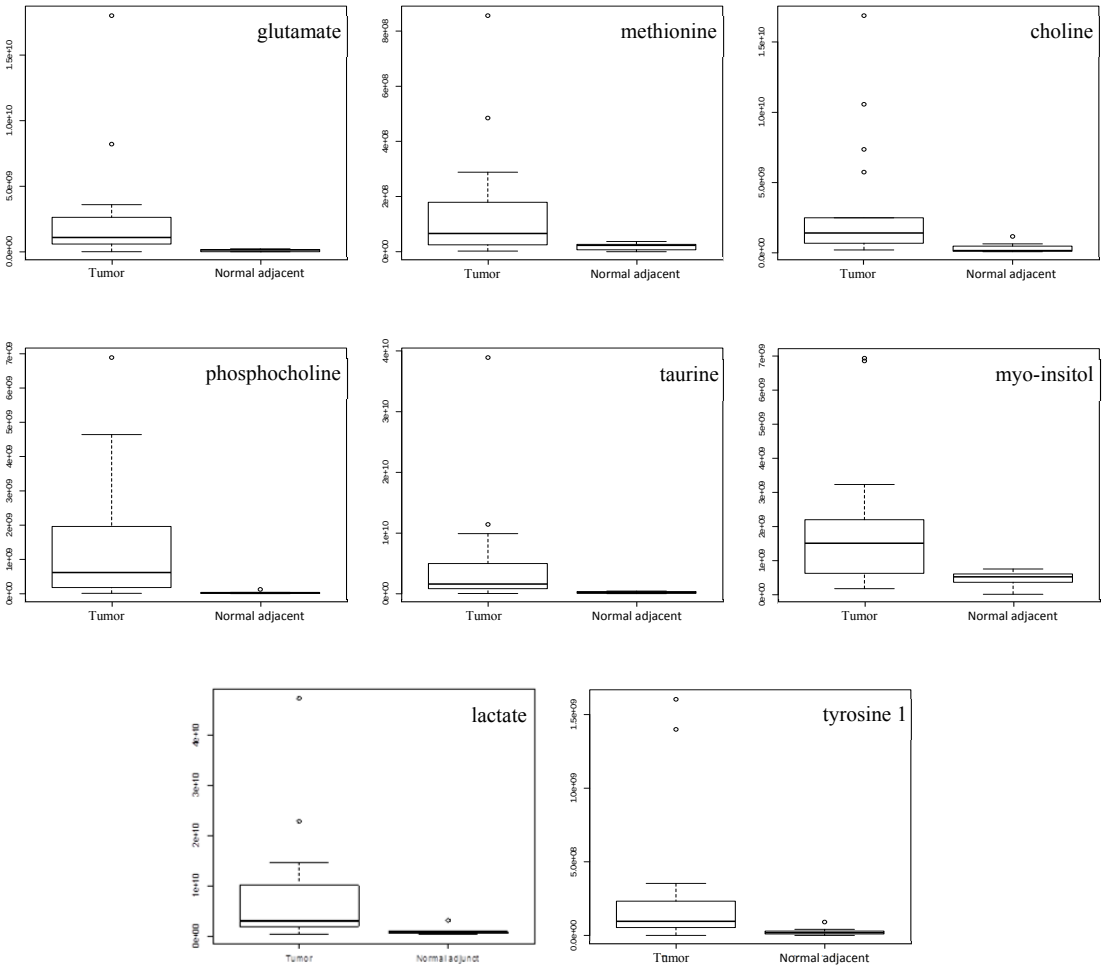


Figure 4.10 Box-and-whisker plots of metabolites with $p<0.05$ illustrating discrimination between tumor and normal adjacent tissues for ER-positive African Americans. Horizontal line in the middle portion of the box indicates the median. Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively. Open circles show outliers.

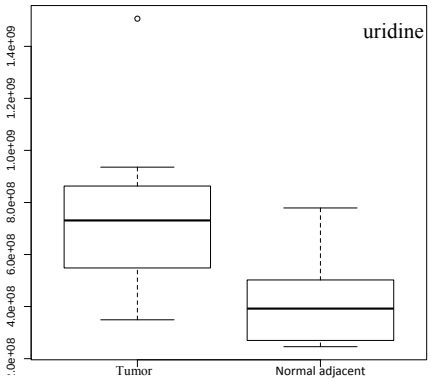


Figure 4.11 Box-and-whisker plots of uridine illustrating discrimination between tumor and normal adjacent tissues for ER-positive Caucasians. Horizontal line in the middle portion of the box indicates the median. Top and bottom boundaries of boxes show the 75th and 25th percentiles, respectively. Upper and lower whiskers show 95th and 5th percentiles, respectively. Open circles show outliers.

VITA

VITA

Fariba Tayyari was born in Iran to Ebtehaj Haghighi and Faramarz Tayyari. She received the degree of Bachelor of Science from Ferdowsi University, Mashhad, Iran, in 2003. She entered the Graduate School in the Department of Chemistry at Ball State University, Muncie, IN, in 2006, under the supervision of Dr. Robert Sammelson, and received a Master's of Science degree in July 2008. Fariba started her Ph.D. studies in the Department of Chemistry at Purdue University, West Lafayette, IN, in 2008. While at Purdue University she joined Professor Daniel Raftery’s research group where she started doing research on NMR-based metabolomics.

PUBLICATIONS

Anal. Chem. **2010,** *82,* 8983–8990

Quantitative Analysis of Blood Plasma Metabolites Using Isotope Enhanced NMR Methods

G. A. Nagana Gowda,[†] Fariba Tayyari,[†] Tao Ye,[†] Yuliana Suryani,[†] Siwei Wei,[†] Narasimhamurthy Shanaiah,[‡] and Daniel Raftery*[,†]

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, and MatrixBio, Inc., 1281 Win Hentschel Blvd., West Lafayette Indiana 47906

NMR spectroscopy is a powerful analytical tool for both qualitative and quantitative analysis. However, accurate quantitative analysis in complex fluids such as human blood plasma is challenging, and analysis using one-dimensional NMR is limited by signal overlap. It is impractical to use heteronuclear experiments involving natural abundance ^{13}C on a routine basis due to low sensitivity, despite their improved resolution. Focusing on circumventing such bottlenecks, this study demonstrates the utility of a combination of isotope enhanced NMR experiments to analyze metabolites in human blood plasma. $^{1}H-^{15}N$ HSQC and $^{1}H-^{13}C$ HSQC experiments on the isotope tagged samples combined with the conventional ^{1}H one-dimensional and $^{1}H-^{1}H$ TOCSY experiments provide quantitative information on a large number of metabolites in plasma. The methods were first tested on a mixture of 28 synthetic analogues of metabolites commonly present in human blood; 27 metabolites in a standard NIST (National Institute of Standards and Technology) human blood plasma were then identified and quantified with an average coefficient of variation of 2.4% for 17 metabolites and 5.6% when all the metabolites were considered. Carboxylic acids and amines represent a majority of the metabolites in body fluids, and their analysis by isotope tagging enables a significant enhancement of the metabolic pool for biomarker discovery applications. Improved sensitivity and resolution of NMR experiments imparted by ^{15}N and ^{13}C isotope tagging are attractive for both the enhancement of the detectable metabolic pool and accurate analysis of plasma metabolites. The approach can be easily extended to many additional metabolites in almost any biological mixture.

Nuclear magnetic resonance (NMR) spectroscopy is increasingly used in metabolomics for the analyses of multiple metabolites in biofluids and tissues. Metabolomics promises a number of important applications in biomedicine including a more detailed understanding of biological processes, the discovery of the biomarkers associated with numerous diseases, pharmaceutical development, and toxicology.[1−6] It is increasingly recognized that because of its high reproducibility and quantitative nature, NMR is a very attractive analytical tool. In particular, the improved resolution of two-dimensional (2D) NMR methods is considered very useful for metabolomics applications.[7−11] However, a major drawback of 2D NMR is that the cross-peak volume in the spectrum is influenced by numerous experimental or intrinsic parameters including the nonuniform excitation profile of the radio frequency pulses, number and duration of the pulses, interpulse delays, relaxation times, and the magnitude of indirect spin−spin couplings. The high sensitivity of peak intensities (or volumes) to these parameters has limited the use of 2D (particularly ^{1}H homonuclear 2D) experiments for quantitative analysis in metabolomics.

To overcome the limitations of quantitation and to improve the analysis, new higher resolution 2D approaches utilizing $^{1}H-^{13}C$ heteronuclear 2D experiments (HSQC) have been proposed.[12,13] One approach is to utilize the information from 2D $^{1}H-^{13}C$ NMR spectra of standard compounds obtained under identical conditions and relate the peak heights in the samples to standard mixtures;[12] another is to utilize calibration curves obtained using $^{1}H-^{13}C$ HSQC spectra for individual metabolites to determine the metabolite concentrations.[13] Most recently, a more general approach that does not require

* To whom correspondence should be addressed. E-mail: raftery@purdue.edu. Tel: (765) 494-6070. Fax: (765) 494-0239.
[†] Purdue University.
[‡] MatrixBio, Inc.

(1) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–1189.
(2) Clayton, T. A.; Lindon, J. C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Provost, J. P.; Le Net, J. L.; Baker, D.; Walley, R. J.; Everett, J. R.; Nicholson, J. K. *Nature* **2006**, *440*, 1073–1077.
(3) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2* (11), 2692–2703.
(4) Pan, Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, *387*, 525–527.
(5) Nagana Gowda, G. A.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Mol. Diagn.* **2008**, *8*, 617–633.
(6) Holmes, E.; Wilson, I. D.; Nicholson, J. K. *Cell* **2008**, *134* (5), 714–717.
(7) Dumas, M-E.; Canlet, C.; Andre, F.; Vercauteren, J.; Paris, A. *Anal. Chem.* **2002**, *74*, 2261–2273.
(8) Tang, H.; Wang, Y.; Nicholson, J. K.; Lindon, J. C. *Anal. Biochem.* **2004**, *325*, 260–272.
(9) Dumas, M-E.; Canlet, C. *J. Proteome Res.* **2005**, *4*, 1493–1502.
(10) Xi, Y.; de Ropp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P. *Metabolomics* **2006**, *2*, 221–233.
(11) Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2010**, *82*, 1811–1821.
(12) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79* (24), 9385–9390.
(13) Gronwald, W.; Klein, M. S.; Kaspar, H.; Fagerer, S. R.; Nürnberger, N.; Dettmer, K.; Bertsch, T.; Oefner, P. J. *Anal. Chem.* **2008**, *80* (23), 9288–9297.

10.1021/ac101938w © 2010 American Chemical Society
Published on Web 09/29/2010

measurements of standard compounds was proposed.[14] This method utilizes correction factors derived theoretically from the solution of the Bloch equations and the analysis of product operator formalism incorporating longitudinal (T_1) and transverse (T_2) relaxation parameters, $^1H-^{13}C$ heteronuclear *J*-coupling, and various delays used in the pulse sequence.

A major drawback of the use of the $^1H-^{13}C$ HSQC experiment for quantitative analysis is the poor signal to noise ratio arising from low metabolite concentration and natural abundance of ^{13}C (1.1% natural abundance). To compensate this limitation, unusually long acquisition times (nearly 10 h or more) are required since the NMR sensitivity scales with the square-root of the number of scans. Moreover, although $^1H-^{13}C$ HSQC greatly enhances resolution when compared with 1D NMR, given the complexity of the biological samples, the resolution obtainable from a single 2D experiment is not always adequate for analyzing a large number of metabolites.

In the present study, with the idea of circumventing the drawbacks of resolution and sensitivity, we utilize a combination of isotope tagging approaches and 2D NMR methods to accurately analyze human plasma metabolites. A number of the most common metabolites in blood plasma were quantified using this approach after validating the experimental protocols using a mixture of synthetic compounds. Metabolites containing carboxyl and amino groups were tagged with ^{15}N and ^{13}C, respectively, before detection by 2D NMR. We have recently shown the proof-of-principle approaches to introduce isotope tags using simple chemical derivatization methods and that the NMR spectra of the tagged metabolites improve both resolution and sensitivity.[15−18] The combination of advanced isotope tagging methods with conventional 1D and 2D NMR methods as described in the present study enables the quantitative analysis of a large number of metabolites in human blood on a routine basis.

EXPERIMENTAL SECTION

Chemicals and Blood Plasma. Twenty-eight metabolite standards (Table 1), 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), maleic acid, ethanolamine (all from Sigma−Aldrich, St. Louis, MO), 4-(4,6-dimethoxy [1,3,5] triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM; Acros, Geel, Belgium), ^{13}C-formic acid (Cambridge Isotope Laboratories, Andover, MA), ^{15}N-ethanolamine, *N*,*N*-dicyclohexylcarbodiimide, and N-hydroxysuccinimide (Sigma-Aldrich) were used without further purification. An ultrapure primary quantitative standard, tris(hydroxymethyl)aminomethane, (99.9%) was obtained from Mallinckrodt Baker Inc. (Phillipsburg, NJ). Human blood plasma (10 × 1 mL) was procured from the National Institute of Standards and Technology (NIST, Gaithersburg, MD). Frozen plasma samples were transported to Purdue University under dry ice and stored at −80 °C until used for the analysis.

Table 1. Synthetic Analogues of Metabolites Used for the Quantitative Analyses of Human Plasma Metabolites

serial number	standard compound	estimated approximate blood plasma concentrations[a] (μM)	actual NMR calibrated concentrations in the mixture (μM) used for spiking[b]
1	3-hydroxybutyrate	60	59.2 ± 0.9
2	acetate	80	73.7 ± 1.1
3	L-alanine	300	287.1 ± 4.5
4	L-arginine	80	90.1 ± 1.4
5	citrate	30	28.1 ± 0.4
6	creatinine	40	31.6 ± 0.6
7	formate	40	32.5 ± 0.5
8	L-glutamic acid	50	53.8 ± 0.3
9	L-glutamine	300	292.1 ± 2.8
10	L-glycine	200	172.6 ± 2.7
11	L-histidine	80	83.7 ± 0.6
12	L-isoleucine	50	50.7 ± 0.2
13	lactate	1000	959.5 ± 15.1
14	L-leucine	80	80.7 ± 1.2
15	L-lysine	100	75.3 ± 2.3
16	L-methionine	20	19.2 ± 0.03
17	L-phenylalanine	70	71.1 ± 1.1
18	L-proline	100	93.3 ± 1.5
19	L-threonine	100	90.0 ± 1.9
20	L-tryptophan	30	29.1 ± 0.4
21	L-tyrosine	80	80.9 ± 0.3
22	L-valine	200	183.2 ± 1.5
23	succinate	10	9.9 ± 0.1
24	betaine	50	44.6 ± 0.7
25	4-hydroxy proline	50	53.4 ± 0.8
26	L-serine	100	93.3 ± 0.7
27	L-asparagine	40	43.9 ± 0.7
28	taurine	30	27.7 ± 0.1

[a] Obtained from the combination of database search and comparison of the relative peak integrals in the NMR spectra. [b] The errors are standard deviations from two measurements.

Calibration of the Standard Solutions. Twenty-eight metabolites that commonly occur in human blood plasma were selected on the basis of entries in the human metabolite database (HMDB)[19] and the analysis of isotope labeled $^1H-^{13}C$ HSQC and $^1H-^{15}N$ HSQC spectra, as well as 1D and 2D $^1H-^1H$ TOCSY spectra of a human plasma sample. Stock solutions (20 mM) for the synthetic analogues of all these 28 metabolites (Supplementary Table S1, Supporting Information) and internal standards, DSS (5 mM), maleic acid (20 mM) and ethanolamine (20 mM), were prepared. The concentration of the DSS solution was calibrated using ^1H NMR against a primary stoichiometric standard, tris(hydroxymethyl)aminomethane (22.4 mM), prepared in the lab. The calibrated DSS solution was then used to calibrate all other standard solutions. Briefly, solutions of the synthetic analogues of the 28 metabolites and internal standards were divided into five groups as shown in Supplementary Table S2, Supporting Information. The grouping of samples in Table S2 was such that the ^1H NMR peaks for at least one proton from each compound in the group were isolated for the measurement of the peak integral. For each group, the one-dimensional (1D) ^1H NMR spectrum was recorded, and

(14) Rai, R. K.; Tripathi, P.; Sinha, N. *Anal. Chem.* **2009**, *81* (24), 10232–10238.
(15) Shanaiah, N.; Desilva, A.; Nagana Gowda, G. A.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (28), 11540–11544.
(16) DeSilva, M. A.; Shanaiah, N.; Nagana Gowda, G. A.; Rosa-Pérez, K.; Hanson., B. A.; Raftery, D. *Magn. Reson. Chem.* **2009**, *47*, S74–S80.
(17) Ye, T.; Mo, H.; Shanaiah, N.; Nagana Gowda, G. A.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009**, *81* (12), 4882–4888.
(18) Ye, T.; Zhang, S.; Mo, H.; Tayyari, F.; Nagana Gowda, G. A.; Raftery, D. *Anal. Chem.* **2010**, *82* (6), 2303–2309.
(19) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res.* **2007**, *35* (Database issue), D521–D526.

on the basis of the integrated area of the isolated peaks with reference to DSS, exact concentrations of the standard solutions were determined (Supplementary Table S1, Supporting Information).

Mixture Analysis. A mixture of the 28 synthetic analogues of the metabolites was prepared using the stock solutions such that the final concentration of each standard compound was matched approximately to its expected concentration in human blood plasma[19] (Table 1). From this mixture, three identical sets of samples (Set 1, Set 2, and Set 3) were prepared (see Supporting Information; Flow Diagram 1); each set consisted of four solutions, 2×500 μL and 2×1000 μL of the mixture. All solutions were then dried under vacuum. To the samples from Set 1, maleic acid solution (92 nmol) was added as an internal reference and the solutions were diluted to 500 μL using doubly distilled water. The carboxylic acid class of metabolites was then tagged with ^{15}N-labeled ethanolamine following the established procedure[17] (see Supporting Information). To each sample from Set 2, ethanolamine solution (200 nmol) was added as an internal reference and the solution was diluted to 500 μL using doubly distilled water. Amines and amino acids were subjected to ^{13}C isotope tagging using a ^{13}C-formic acid reaction following the established procedure[18] (see Supporting Information). Finally, to the samples from Set 3, DSS (9.44 nmol) was added as an internal reference and reconstituted in 560 μL of phosphate buffer (pH = 7.4) in D_2O and transferred to 5 mm NMR tubes for 1D and ^{1}H−^{1}H 2D TOCSY NMR experiments.

Isotope Tagging of Plasma Metabolites. Cold methanol (4 °C; 9.6 mL) was added to 4.8 mL of the NIST plasma, vortexed, and then kept for 30 min at −20 °C. The precipitated protein pellet was removed after centrifuging at 13 200 g for 10 min. The supernatant was divided into 12 equal parts and divided into three groups, each group consisting of four samples (see Supporting Information; Flow Diagram 2). In each group, two samples served as controls and the remaining two were spiked with 400 μL of the stock solution mixture of 28 synthetic samples. All three groups of samples were then dried in vacuum. One group was used to label metabolites containing carboxyl groups with ^{15}N-ethanolamine, and the second group was used to label metabolites containing amino groups with ^{13}C-formic acid, after the addition of internal standards, either maleic acid or ethanolamine, appropriately. Identical procedures were used for isotope tagging the plasma metabolites. To the third group of samples, DSS (9.44 nmol) was added as an internal reference and reconstituted in 560 μL of phosphate buffer (pH = 7.4) in D_2O. Samples from all the three preparations were transferred to 5 mm NMR tubes after adjusting the pH and solution conditions as described earlier for the mixture of standards.

NMR Experiments. NMR experiments were performed at 298 K on a Bruker Avance-III 800 MHz spectrometer equipped with a room temperature ^{1}H inverse detection Z-gradient probe or a Bruker DRX-500 spectrometer equipped with a ^{1}H inverse detection Z-gradient cryo-probe. 1D NMR experiments for the five groups of standard samples (Supplementary Table S2, Supporting Information), the mixture of 28 synthetic analogues, and the plasma extracts were performed using a one pulse sequence with residual water signal suppression by presaturation during relaxation delay. Thirty-two scans with 64 k time domain

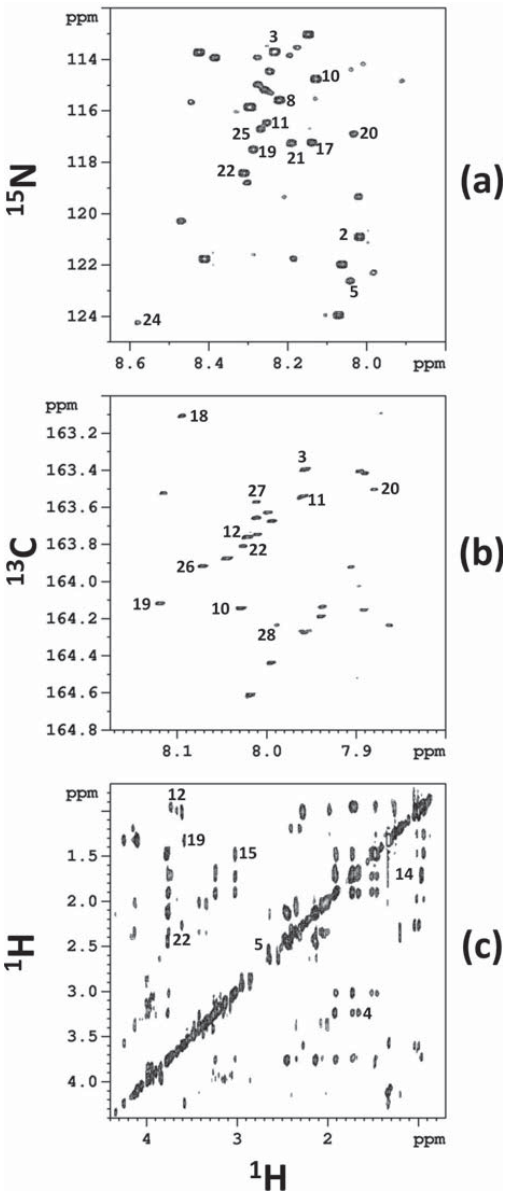


Figure 1. 2D spectra of mixtures of 28 synthetic compounds obtained with or without isotope tagging: (a) ^{1}H−^{15}N HSQC spectrum with ^{15}N tagging of carboxylic acids, (b) ^{1}H−^{13}C HSQC spectrum with ^{13}C tagging of amines and amino acids, and (c) ^{1}H−^{1}H TOCSY spectrum of the neat mixture. All the spectra were obtained on a 800 MHz spectrometer. The labeled peaks correspond to the numbered metabolites in Table 1.

data points were collected with a sufficiently long recycle delay (20 s) to ensure complete recovery of the magnetization between scans. For the ^{15}N isotope tagged samples, ^{1}H−^{15}N 2D HSQC experiments were performed employing an INEPT (insensitive nuclei enhanced by polarization transfer) transfer delay of 5.5 ms corresponding to a $^1J_{NH}$ of 90 Hz. Spectral widths of approximately 10 kHz in the ^{1}H and 5 kHz in the ^{15}N dimensions were used for the 800 MHz experiments. For ^{13}C

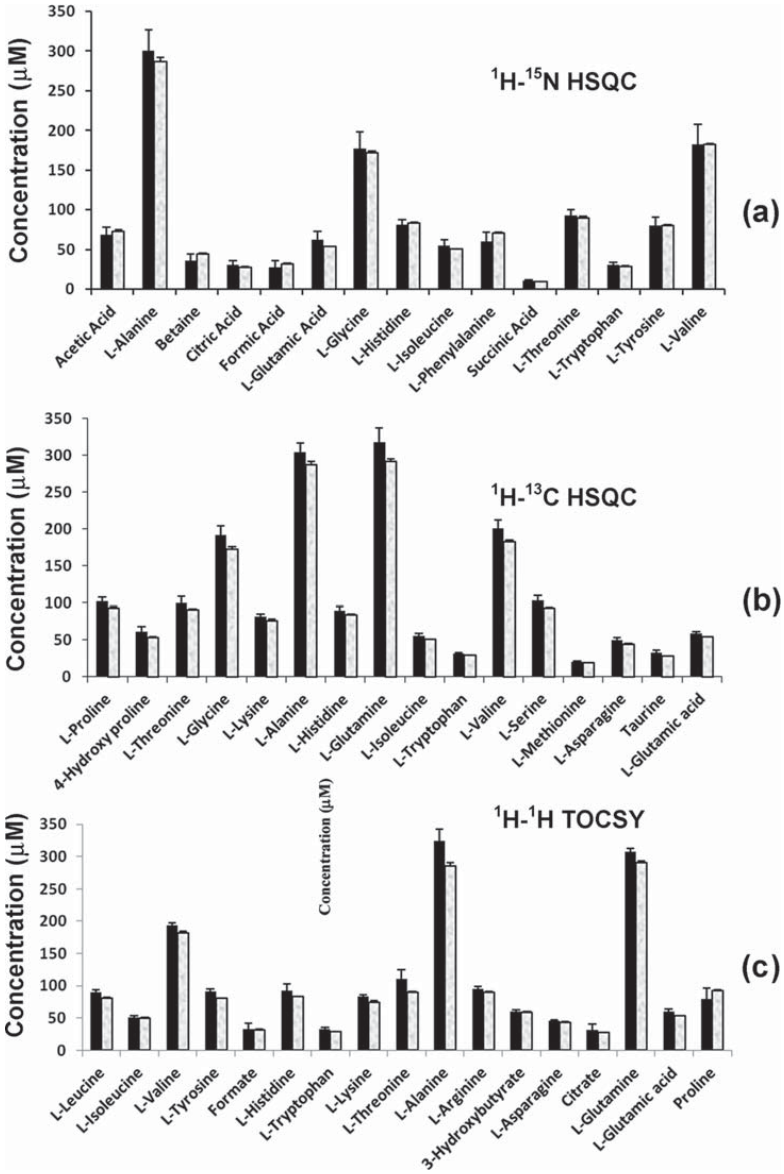


Figure 2. Concentration of 28 standard metabolites obtained by combining 2D NMR experiments with and without ^{15}N or ^{13}C tagging: (a) obtained from ^{1}H−^{15}N HSQC NMR after ^{15}N tagging; (b) obtained from ^{1}H−^{13}C HSQC NMR after ^{13}C tagging, and (c) obtained from ^{1}H−^{1}H TOCSY NMR of the neat mixture. The shaded bar on the right in each pair represents the actual concentration of the metabolite.

isotope tagged samples, sensitivity-enhanced ^{1}H−^{13}C 2D HSQC experiments were performed employing an INEPT transfer delay of 2.5 ms corresponding to a $^{1}J_{C-H}$ of 200 Hz. Spectral widths of approximately 10 kHz for the ^{1}H dimension and 600 Hz for ^{13}C were used at 800 MHz. For both ^{1}H−^{15}N HSQC and ^{1}H−^{13}C HSQC 2D experiments, 128 free induction decays were collected along the indirect (t_1) dimensions using four transients per increment and a 2 or 3 s recycle delay, resulting in a total acquisition time of 18 min for the ^{1}H−^{15}N HSQC and 28 min for the ^{1}H−^{13}C HSQC. Phase-sensitive data were obtained using echo-antiecho mode with nitrogen (for ^{1}H−^{15}N HSQC) or carbon (for ^{1}H−^{13}C HSQC) decoupling during acquisition (t_2 dimension) using the GARP (globally optimized alternating-phase rectangular pulses) sequence. ^{1}H−^{1}H 2D TOCSY experiments were performed on the neat (nonderivatized) samples with a spectral width of 6 kHz (500 MHz) or 12 kHz (800 MHz) in both the dimensions. The residual water signal was suppressed by presaturation. Free induction decays (400) were collected with t_1 increments using eight transients per increment and a 2 s recycle delay, resulting in a total acquisition time of 116 min (500 MHz) or 111 min (800 MHz).

All 1D data were Fourier transformed with a 0.3 Hz line broadening function. The 2D data were zero-filled to 1024 points in the t_1 dimension after forward linear prediction to 512 points and Fourier-transformed after multiplying by a squared sine-bell window function shifted typically by $\pi/4$ or $\pi/2$ along both the dimensions. All NMR data were processed with Bruker Topspin 2.0 on a Redhat Linux platform and Bruker XWINNMR 3.5 on a SGI/IRIX platform. An automatic baseline correction using a polynomial of degree 5 was used to correct the baseline in both 1D and 2D spectra.

Peaks in the 1D and 2D NMR spectra were assigned to various metabolites on the basis of literature reports.[17−19] Integrals for well resolved peaks in the 1D and 2D spectra were obtained with respect to the peak for the internal standard DSS, maleic acid or ethanolamine. Integral limits for each peak in the 2D spectra were selected such that the selected region encompassed the whole peak and that no other peak interfered with the selection. Once chosen for each type of 2D spectrum, the same sets of integral limits were used for all other samples. Concentrations of the plasma metabolites were determined by comparing the peak integrals from the spectra obtained with and without spiking with the synthetic analogues and also by directly comparing the peak integrals of the plasma metabolites with those from the standards. The accuracy, reproducibility, and errors were estimated from two to eight measurements, depending on the detection of the resolved peak for a particular metabolite in one or more types of spectra, for both synthetic mixtures and plasma samples. The ^{1}H−^{15}N HSQC, ^{1}H−^{13}C HSQC, and ^{1}H−^{1}H TOCSY experiments and the data analyses were performed by independent persons.

RESULTS

The standard solutions of the synthetic analogues of the plasma metabolites and internal standards (maleic acid and ethanolamine), prepared on the basis of their weights, were calibrated using 1D ^{1}H NMR. The actual concentration of the standard solutions prepared on the basis of the weight varies depending on the purity and hygroscopic nature of the compounds, and hence, it is important to calibrate the standard solutions especially for accurate quantitative analysis. The DSS solution, which was first calibrated using a primary stoichiometric standard, tris(hydroxymethyl)aminomethane, was used for calibrating all the standard solutions (Supplementary Table S2, Supporting Information). The difference between the concentrations determined on the basis of sample weight and calibration using 1D NMR varied as much as 10% for all but three metabolites, which varied up to nearly 20% (Supplementary Table S1, Supporting Information) due to hygroscopic nature of the metabolites or sample impurities.

Analysis of Synthetic Metabolite Mixture. A mixture of 28 metabolites was analyzed using both ^{15}N and ^{13}C isotope tagging approaches. Figure 1 shows 2D spectra of the mixture of 28 compounds with ^{15}N and ^{13}C isotope tagging, as well as without tagging. The integrated 2D peak volumes were obtained and then used to calculate the metabolite concentrations. Figure 2 shows the concentration of the compounds, thus, determined. As can be seen in the figure, an excellent match between the metabolite concentrations derived from NMR methods and the actual values was obtained. Further, as shown in Supplementary

Table 2. ^{1}H, ^{13}C, and ^{15}N Chemical Shifts of the Peaks Used in the Analysis of NIST Plasma Metabolites

experiment: ^{1}H−^{15}N HSQC

label	metabolite	^{1}H (ppm)	^{15}N (ppm)
2	acetate	8.01	120.93
3	L-alanine	8.23	113.70
5	citrate	8.04	122.62
8	L-glutamic acid	8.21	115.58
10	L-glycine	8.12	114.78
11	L-histidine	8.25	116.42
17	L-phenylalanine	8.13	117.19
19	L-threonine	8.28	117.52
20	L-tryptophan	8.03	116.90
21	L-tyrosine	8.18	117.25
22	L-valine	8.30	118.43
24	betaine	8.57	124.28
25	4-hydroxy proline	8.26	116.74

experiment: ^{1}H−^{13}C HSQC

label	metabolite	^{1}H (ppm)	^{13}C (ppm)
3	L-alanine	7.96	163.40
10	L-glycine	8.03	164.15
11	L-histidine	7.97	163.56
12	L-isoleucine	8.03	163.81
18	L-proline	8.10	163.10
19	L-threonine	8.12	164.12
20	L-tryptophan	7.88	163.51
22	L-valine	8.05	163.88
25	4-hydroxy proline	8.04	165.21
26	L-serine	8.07	163.92
27	L-asparagine	8.01	163.57
28	taurine	7.99	164.23

experiment: ^{1}H 1D NMR

label	metabolite	^{1}H (ppm)
1	3-hydroxybutyrate	1.19
2	acetate	1.91
3	L-alanine	1.47
7	formate	8.45
9	L-glutamine	2.13
11	L-histidine	7.06
12	L-isoleucine	1.01
16	L-methionine	2.13
17	L-phenylalanine	7.42
20	L-tryptophan	7.74
21	L-tyrosine	6.89
22	L-valine	1.03
29	glucose	5.23

experiment: ^{1}H−^{1}H TOCSY

label	metabolite	^{1}H (ppm) F_2 dimension	^{1}H (ppm) F_1 dimension
4	L-arginine	1.68	3.23
5	citrate	2.65	2.51
11	L-histidine	7.07	7.06
12	L-isoleucine	3.66	0.98
13	lactate	1.32	1.32
14	L-leucine	0.95	1.70
15	L-lysine	3.02	1.48
19	L-threonine	3.57	1.33
20	L-tryptophan	7.53	7.72
21	L-tyrosine	6.90	7.18
22	L-valine	3.60	2.27

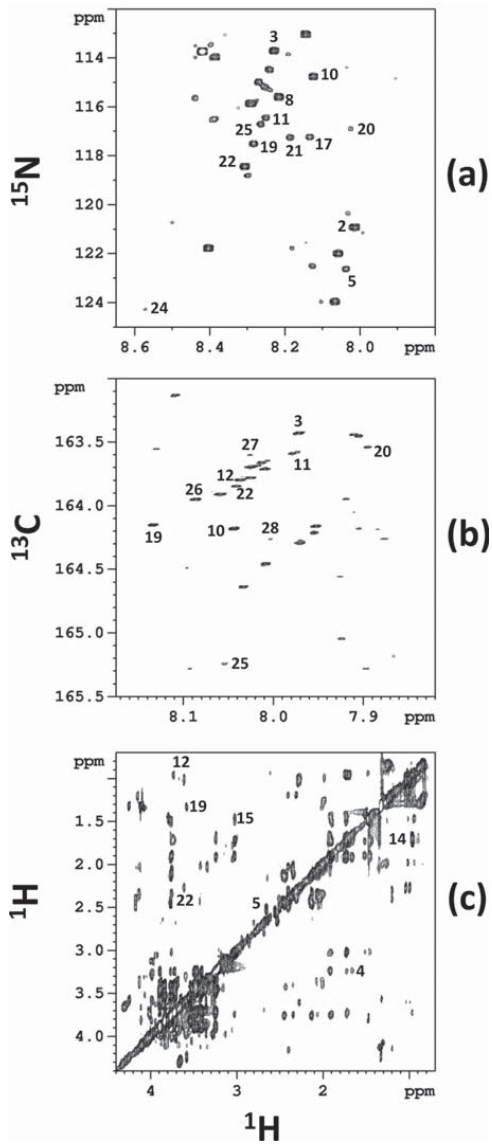


Figure 3. 2D spectra of NIST plasma obtained with and without isotope tagging: (a) $^{1}H-^{15}N$ HSQC spectrum obtained after ^{15}N tagging of carboxylic acids, (b) $^{1}H-^{13}C$ HSQC spectrum obtained after ^{13}C tagging of amines and amino acids, and (c) $^{1}H-^{1}H$ TOCSY spectrum of the neat mixture. All the spectra were obtained on an 800 MHz spectrometer. The labeled peaks correspond to the numbered metabolites in Table 1.

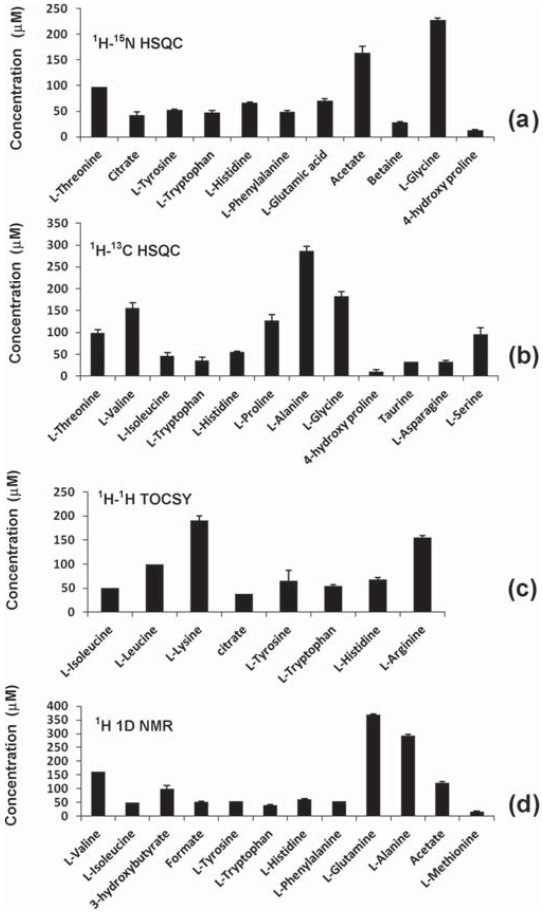


Figure 4. Concentrations of metabolites in the NIST plasma obtained using 1D/2D NMR experiments with and without isotope tagging: (a) obtained from $^{1}H-^{15}N$ HSQC NMR after ^{15}N tagging, (b) obtained from $^{1}H-^{13}C$ HSQC NMR after ^{13}C tagging, (c) obtained from $^{1}H-^{1}H$ TOCSY NMR of neat plasma, and (d) obtained from 1D NMR of the neat plasma sample.

Figure S1 (Supporting Information), a correlation of the NMR derived values with the expected values showed a very good agreement for all low and high concentration metabolites ($R^2 > 0.99$).

Quantitation of Plasma Metabolites. The 1D ^{1}H NMR spectrum of the plasma sample obtained without isotope labeling is highly complex, with only a relatively small number of metabolite signals being isolated from other signals (Supplementary Figure S2, Supporting Information). 2D HSQC spectra of plasma samples tagged with ^{15}N and ^{13}C isotopes provide resolved peaks for a much larger number of carboxylic acid and amine class of metabolites. Figure 3 shows 2D spectra of the plasma obtained with and without ^{15}N or ^{13}C tagging. The 2D TOCSY spectrum of the same plasma sample also provided a number of well resolved peaks (Figure 3c). However, unlike the HSQC spectra, the TOCSY spectrum showed a number of redundant peaks for the same metabolite, which increases the complexity of the spectrum.

Quantitation of the plasma metabolites followed an identical procedure used for the determination of the concentrations of metabolites in the synthetic mixture. The integrated peak areas/volumes in the 1D/2D spectra of the neat and the spiked plasma samples were obtained, and the metabolite concentrations were determined. Twenty-seven metabolites that were identified in human plasma were analyzed in duplicate measurements with and without ^{15}N and ^{13}C isotope tagging. Figure 4 depicts the

Table 3. NIST Plasma Metabolite Concentrations Obtained Using a Combination of NMR Experiments with or without Isotope Tagging

label	metabolite	NIST plasma concentration (μM)[a]	label	metabolite	NIST plasma concentration (μM)[a]
1	3-hydroxybutyrate[e]	99.3 ± 13.1	16	L-methionine[e]	16.6 ± 2.8
2	acetate[b,e]	142.0 ± 3.0	17	L-phenylalanine[b,e]	50.6 ± 3.8
3	L-alanine[b,c,e]	279.4 ± 18.9	18	L-proline[c]	127.6 ± 13.1
4	L-arginine[d]	155.9 ± 16.0	19	L-threonine[b,c,d]	107.3 ± 16.7
5	citrate[b,d]	40.1 ± 2.3	20	L-tryptophan[b,c,d,e]	45.9 ± 8.1
7	formate[e]	51.2 ± 2.1	21	L-tyrosine[b,d,e]	58.9 ± 7.2
8	L-glutamic acid[b]	69.3 ± 5.3	22	L-valine[b,c,d,e]	159.7 ± 11.6
9	L-glutamine[e]	368.5 ± 2.3	24	betaine[b]	27.3 ± 2.8
10	L-glycine[b,c]	204.5 ± 31.2	25	4-hydroxy proline[b,c]	11.5 ± 1.3
11	L-histidine[b,c,d,e]	63.1 ± 5.7	26	L-serine[c]	95.8 ± 15.0
12	L-isoleucine[c,d,e]	48.2 ± 2.4	27	L-asparagine[c]	33.4 ± 2.8
13	lactate[d]	2403.6 ± 127.6	28	taurine[c]	32.4 ± 0.8
14	L-leucine[d]	100.1 ± 0.1	29	glucose[e]	8778.5 ± 62.8
15	L-lysine[d]	190.8 ± 21.9			

[a] The errors are standard deviations. [b] Obtained from ^{1}H−^{15}N HSQC. [c] Obtained from ^{1}H−^{13}C HSQC. [d] Obtained from ^{1}H−^{1}H TOCSY. [e] Obtained from ^{1}H 1D NMR. Two samples were used for each type of experiment resulting in two, four, six, or eight independent measurements for each metabolite.

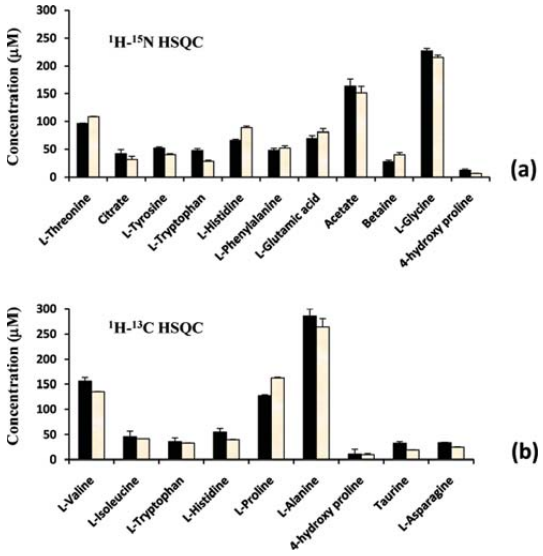


Figure 5. Comparison of the concentrations of (a) carboxylic acid and (b) amino group containing metabolites in the NIST plasma obtained with spiking (left bars) and without spiking (right bars) with the standard compounds.

concentration of the metabolites, thus, determined. The ^{1}H, ^{13}C, and ^{15}N chemical shifts for the blood plasma metabolites analyzed in this study are shown in Table 2, and the derived concentrations using a combination of four different NMR methods are shown in Table 3. Further, the concentration of the carboxylic acid and amine class of metabolites were also determined by directly comparing the 2D peak integrals with those for the corresponding standard compound. Comparison of the metabolites concentration determined with and without spiking is shown in Figure 5. Notably, the values determined from both approaches agree well.

DISCUSSION

^{1}H NMR spectroscopy is an attractive tool for the quantitative analysis of multiple metabolites from intact biological samples. Considering its ease of use, reproducibility, and high-throughput nature, 1D ^{1}H NMR spectroscopy is often used for metabolomics-based studies. However, it is challenging to analyze the 1D NMR spectrum of plasma in an absolute quantitative fashion as it contains a large number of overlapping signals due to hundreds of metabolites present at variable concentrations. The multiplicity of the signals due to *J*-coupling makes 1D ^{1}H NMR spectra of plasma particularly challenging. The interference from macromolecules such as proteins and lipids adds to the complexity and causes baseline distortions in the spectra. Such overlap and baseline issues substantially affect the accuracy of the quantitative analysis using 1D NMR. To offset such limitations, a majority of the studies that use 1D NMR resort to comparisons of the relative intensities of the 1D NMR signals between disease and healthy samples. While the use of relaxation edited techniques such as the Carr−Purcell−Meiboom−Gill (CPMG) experiment, serum/plasma deproteinization, and line fitting approaches significantly improves the analysis of metabolites,[20] such methods are not ideal. Diffusion-sensitized 1D NMR spectroscopy, which uses data from two separate 1D experiments, one obtained using low diffusion gradients and the other using high gradients to suppress macromolecular background signals effectively, was shown to be useful for the quantitative analysis of blood plasma metabolites.[21] However, spectral overlap still significantly limits the number of metabolites that can be analyzed.

2D NMR promises quantitative analysis of a large number of metabolites on a routine basis. An important requirement is that the cross-peaks in 2D spectra should be devoid of overlaps for reliable quantitative results; however, this criterion is not often met for a large number of metabolites by a single 2D experiment due to the extremely high complexity of plasma. The advantage of the new 2D NMR approaches used here is that the use chemoselective isotope tags greatly reduces the complexity of the spectra, since only a single peak is observed for the metabolites

(20) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. *Anal. Chem.* **2006**, *78*, 4430–4442.
(21) de Graaf, R. A.; Behar, K. L. *Anal. Chem.* **2003**, *75*, 2100–2104.

with a single functional group (see Figures 1 and 3). The reduced complexity of the spectra due to the absence of less interesting chemical signals is particularly important for the analysis of low-concentration metabolites (Figures 2 and 4). This method, however, does not work for the analysis of lipoproteins, which represents a major class of metabolites in blood plasma and which have been effectively analyzed using a multivariate deconvolution approach.[22,23]

An important criterion for the quantitative analysis method to be robust is that it does not require the use of spiking standards for each sample. To test this, we also determined the concentration of ^{15}N and ^{13}C isotope tagged metabolites in plasma by comparing the 2D peak integrals with those from the synthetic analogues. It may be interesting to note that, as shown in Figure 5, the values, thus, determined using both ^{15}N and ^{13}C isotope tagging agree well with those determined on the basis of spiking with synthetic analogues. Therefore, it is sufficient to obtain the integral for each synthetic analogue only once, which can be used for the analysis of any number of samples. Utilization of 2D HSQC experiments involving the isotopes has the additional advantage since both the magnitude of the coupling and the relaxation properties of the nuclear pairs (^{15}N/^{13}C and ^{1}H) do not appreciably vary across the metabolites of interest and, hence, provide the relative cross peak intensities that are less sensitive to instrumental settings.

In this study, we quantified 27 metabolites with an average CV of 2.4% for 17 metabolites and 5.6% when all the metabolites were considered. When the results from all the four NMR methods were combined for the same metabolites, the average CV's were 4.8% and 8.7%, respectively. We note that, as the metabolite library expands, we can quantify additional metabolites from the same and already acquired 2D data by comparison of the peak integrals with those from the standards. Mass spectrometry (MS), another very useful method for quantitative analysis, is highly sensitive and provides quantitative information on a larger number of metabolites. However, MS invariably involves the combination of a separation method such as gas chromatography or liquid chromatography for accurate analysis and often renders the obtained results to be sensitive to the specific column and separation parameters and especially the ionization conditions. In addition, a standard compound is needed for each quantified metabolite.

In conclusion, this investigation presents quantitative analysis of over 25 plasma metabolites using ^{15}N and ^{13}C isotope tagging methods. Carboxylic acids and amines represent a majority of the metabolites in body fluids, and their analysis by isotope tagging significantly enhances the detectable metabolic pool for biomarker discovery applications. The combination of improved sensitivity and resolution and the reduced time required when compared to natural abundance heteronuclear NMR methods are attractive for the routine and accurate analysis of metabolites in complex biological samples. Although, the isotope tagging methods use 2D NMR experiments, each 2D experiment requires only 30 min or less (<10 min with a cryoprobe), and hence, the approach can be useful for high throughput analysis of human plasma as well as other biological fluids. Further, combination of the isotope tagging approach with the latest advancements in NMR technology, such as detection using microcoil probes, for example, can significantly minimize the volume of biofluid samples required for routine analysis.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health Grants 1 R01GM085291-01 and 3R01GM085291-02S1. The authors would like to thank Dr. Art Castle at the NIH/NIDDK for useful suggestions regarding this work. D.R. is a member of the Purdue Center for Cancer Research and Oncological Sciences Center.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review July 22, 2010. Accepted September 9, 2010.

AC101938W

(22) Otvos, J. D.; Jeyarajah, E. J.; Bennett, D. W. *Clin Chem.* **1991**, *37* (3), 377–386.
(23) Otvos, J. D. *Clin Lab.* **2002**, *48*, 171–180.

analytical chemistry

Article

pubs.acs.org/ac

^{15}N-Cholamine—A Smart Isotope Tag for Combining NMR- and MS-Based Metabolite Profiling

Fariba Tayyari,[†] G. A. Nagana Gowda,[‡] Haiwei Gu,[‡] and Daniel Raftery*[,†,‡,§]

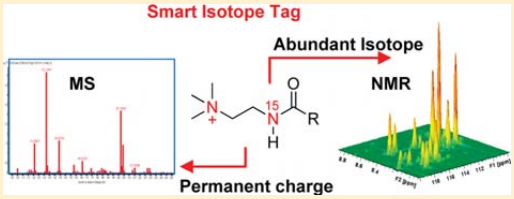
[†]Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States

[‡]Northwest Metabolomics Research Center, Anesthesiology and Pain Medicine, University of Washington, Seattle, Washington 98109, United States

[§]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States

Supporting Information

ABSTRACT: Recently, the enhanced resolution and sensitivity offered by chemoselective isotope tags have enabled new and enhanced methods for detecting hundreds of quantifiable metabolites in biofluids using nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry. However, the inability to effectively detect the same metabolites using both complementary analytical techniques has hindered the correlation of data derived from the two powerful platforms and thereby the maximization of their combined strengths for applications such as biomarker discovery and the identification of unknown metabolites. With the goal of alleviating this bottleneck, we describe a smart isotope tag, ^{15}N-cholamine, which possesses two important properties: an NMR sensitive isotope and a permanent charge for MS sensitivity. Using this tag, we demonstrate the detection of carboxyl group containing metabolites in both human serum and urine. By combining the individual strengths of the ^{15}N label and permanent charge, the smart isotope tag facilitates effective detection of the carboxyl-containing metabolome by both analytical methods. This study demonstrates a unique approach to exploit the combined strength of MS and NMR in the field of metabolomics.



The metabolomics field has witnessed exponential growth over the past decade due to its capabilities for systems biology research and potential applications in numerous disciplines including biomedicine, toxicology, food and nutrition, drug development, and environmental science.[1−5] Commonly used analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy and/or mass spectrometry (MS) have evolved in response to the high demand for resolving the complexity of biological mixtures and identifying the large pool of quantifiable metabolites. However, despite numerous advances, the biological complexity still often outweighs the capabilities of these advanced analytical methods; no single technique currently is capable of detecting all metabolites in a single experiment. Each analytical method is sensitive to certain classes of metabolites, and depending on the nature of the metabolites of interest, generally one or sometimes a combination of NMR or MS techniques are used to profile as many metabolites as possible and thereby derive the biological meaning. A major hurdle of such an approach is that the metabolite data obtained from NMR and LC-MS or GC-MS methods for the same or similar samples are often not directly comparable. The inability to compare and correlate data from different analytical techniques for the same or similar samples is a significant challenge that prevents drawing meaningful conclusions from the vast amount of metabolite data existing in the literature and exploiting the combined strength of NMR and MS for unknown metabolite identification. The main contributing factors for this bottleneck are the limited NMR sensitivity, complex spectral signatures, and variable MS ionization efficiency or suppression.

The use of chemo-selective tags provides an avenue to improve the sensitivity of metabolite detection by both NMR and MS methods. For example, the sensitivity of MS detection is shown to be enhanced by three orders of magnitude or more by tagging metabolites with chemoselective tags containing a permanent charge.[6−10] Because of the permanent charge, the tagged metabolites are effectively detected with high sensitivity and better quantitative accuracy, irrespective of the pH or nature of the solvents used to separate metabolites before detection by MS. Separately, based on differential dansylation using ^{12}C/^{13}C-dansyl chloride, absolute or relative quantitation of amine and phenol containing metabolites has been achieved with a sensitivity enhancement of three orders of magnitude.[11,12] Similarly, NMR-sensitive isotope based chemo-selective tags have been shown to detect many quantifiable metabolites with high sensitivity and resolution by NMR.[13−17] Using ^{15}N-ethanolamine as the tag, for example, over a hundred carboxyl-containing metabolites have been detected by ^{1}H−^{15}N two-dimensional NMR with high resolution and sensitivity.[13] However, while metabolites can be detected with high

Received: June 7, 2013
Accepted: August 9, 2013

   dx.doi.org/10.1021/ac401712a | Anal. Chem. XXXX, XXX, XXX−XXX

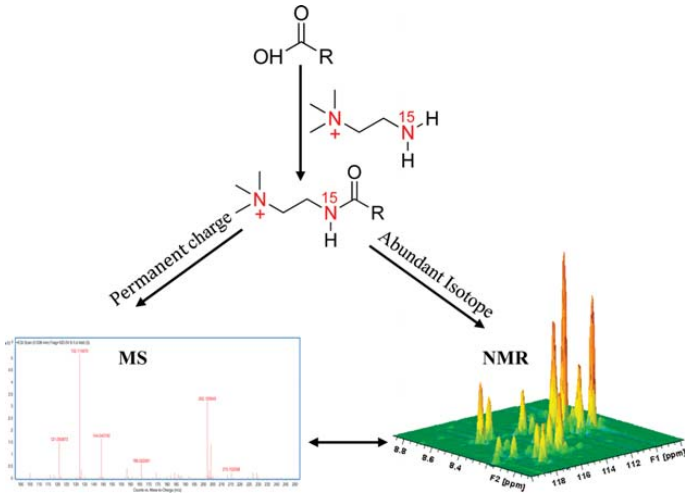


Figure 1. Schematic figure illustrating the "smart isotope tag" approach used to detect the same metabolites using NMR and MS with high sensitivity. Tagging carboxyl-containing metabolites with ^{15}N-cholamine enables their enhanced detection by both NMR and MS.

sensitivity by both MS and NMR separately using chemoselective tags, the inability to compare and correlate the data from the two methods is a major bottleneck in the metabolomics field.

The ability to more easily detect the same metabolites by both NMR and MS methods would offer new avenues to compare data between MS and NMR platforms and to exploit the combined strength of the complementary methods. Toward this goal, we introduce a new "smart isotope tag" approach, using ^{15}N-cholamine in this case, which possesses the characteristics of high NMR sensitivity and resolution through its isotope enrichment and high MS sensitivity through its permanent positive charge (see schematic Figure 1). The tag combines the strengths of individual chemoselective tags, demonstrated previously and separately for NMR and MS detection,[6,13] and offers news avenues to exploit the combined strength of these powerful and complementary techniques for areas such as metabolite profiling and unknown metabolite identification.

EXPERIMENTAL SECTION

Chemicals and Biofluids. A total of 48 carboxyl-containing metabolite standards (Table I), (2-bromoethyl)-trimethylammonium bromide, dimethylformamide (DMF), methanol, acetonitrile, acetone, hydrochloric acid (HCl), sodium hydroxide (NaOH) (all from Sigma-Aldrich, St. Louis, MO), 4-(4,6-dimethoxy[1,3,5]triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) (Acros Organic, Pittsburgh, PA), ^{15}N-phthalimide potassium, and deuterium oxide (Cambridge Isotope Laboratories, Andover, MA) were used without further purification. Human serum samples were obtained from Innovative Research, Inc., (Novi, MI) and urine from healthy volunteers, in accordance with the Internal Review Board at Purdue University. Deionized (DI) water was from in-house Synergy Ultrapure Water System from Millipore (Billerica, MA).

Design and Synthesis of the Smart Isotope Tag—^{15}N-Cholamine. Synthesis of ^{15}N-cholamine involved a two-step reaction and followed the Gabriel synthesis procedure with modifications as described below to yield the pure product.[18,19] The first step involved reacting potassium ^{15}N-phthalimide with (2-bromoethyl)trimethylammonium bromide in DMF to obtain the ^{15}N-substituted phthalimide intermediate (Scheme 1). The second step involved alkaline and acid hydrolyses of the ^{15}N-substituted phthalimide to yield the smart isotope tag, ^{15}N-cholamine (Scheme 2).

Briefly, for the synthesis of ^{15}N-substituted phthalimide (Scheme 1), (2-bromoethyl)trimethylammonium bromide (9.5 mmol, 2.35 g) was mixed with ^{15}N-phthalimide potassium (10 mmol, 1.86 g) in a 250 mL round-bottom flask and dry DMF (100 mL) was added under nitrogen gas. The mixture was then refluxed at 100 °C with stirring for 12 h. The supernatant from the reaction mixture was separated, and the solvent was removed using a rotary evaporator.[18] The resulting crude residue was washed thrice using acetonitrile (5 mL each time), twice with acetone (2 mL each time) followed by washing again once with acetonitrile (3 mL) to obtain the pure ^{15}N-substituted phthalimide. ^{1}H NMR spectra in D_2O at each step were monitored to assess the purity of the intermediate product. For the synthesis of ^{15}N-cholamine, in the second step, the ^{15}N-substituted phthalimide (538 mg) (Scheme 1) was dissolved in DI water (24 mL); 1 N NaOH (2.69 mL) was added to the solution, and the mixture was left at room temperature with stirring for 30 min to complete the alkaline hydrolysis (Scheme 2).[19] Subsequently, 12 N HCl (1.8 mL) was added to the solution, the temperature was raised to 70 °C, and left for 12 h with stirring to complete the acid hydrolysis (Scheme 2).[19] The solvent was then removed using a rotary evaporator. The resulting crude residue was washed thrice with acetonitrile (4 mL each time) followed by washing thrice with 25:75 water/acetone mixture (2 mL each time) to yield the pure product, ^{15}N-cholamine. ^{1}H NMR spectra in D_2O at each step were monitored to assess the purity of the final product.

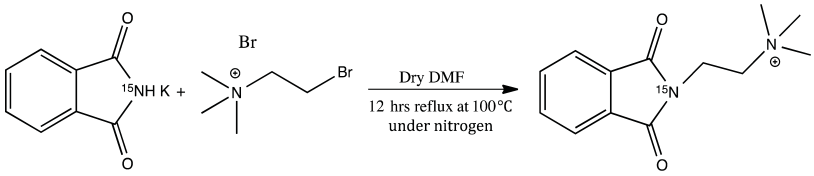
Tagging Metabolites Using the Smart Isotope Tag—^{15}N-Cholamine. ^{15}N-Cholamine (5 mg, 50 μmol) was added to 500 μL sample in an eppendorf tube, and the pH of the mixture was adjusted to 7.0 with 1 M HCl or NaOH. A 21 mg portion of DMTMM was added to initiate the

dx.doi.org/10.1021/ac401712a | *Anal. Chem.* XXXX, XXX, XXX–XXX

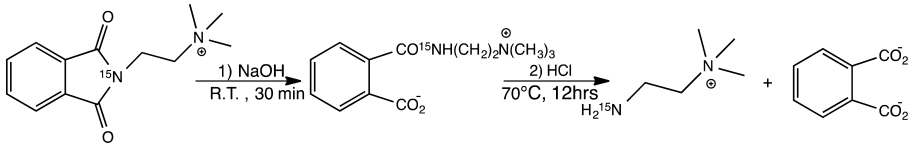
Table I. ^{1}H and ^{15}N-NMR Chemical Shifts for ^{15}N-Cholamine Tagged Carboxyl-Containing Metabolites That Were Measured with Reference to Smart Tagged Formic Acid

label	name	^{1}H (ppm)	^{15}N (ppm)	label	name	^{1}H (ppm)	^{15}N (ppm)
1	cis-aconitic acid	8.5	118.24	**23**	2-hydroxyisobutyric acid	7.95	117.51
		8.14	121.47	**24**	DL-isocitric acid	8.40	117.15
		8.06	119.49			8.11	120.77
		8.07	120.21	**25**	isoleucine	8.37	118.19
		8.23	116.00	**26**	isovaleric acid	8.07	121.92
		8.14	120.81	**27**	α-ketoglutaric acid	8.69	116.34
2	adipic acid	8.05	120.57			8.63	111.84
3	DL-alanine	8.30	114.39	**28**	lactic acid	8.23	114.18
4	4-aminobenzoic acid	8.25	111.35			8.49	114.45
5	arginine	8.34	115.96	**29**	leucine	8.34	115.74
6	asparagine	8.31	116.03	**30**	lysine	8.33	115.88
7	aspartic acid	8.15	120.01	**31**	maleic acid	8.39	120.39
		8.38	115.27	**32**	malic acid	8.28	122.83
		8.31	115.6			8.29	122.15
						8.08	115.14
		8.16	121.35	**33**	malonic acid	8.19	121.44
8	betaine	8.55	122.69	**34**	methionine	8.36	116.08
9	citric acid	8.20	121.46	**35**	oxalic acid	8.47	117.13
		8.07	123.95	**36**	oxaloacetic acid	8.35	112.67
		7.87	121.88			8.63	111.40
10	cysteine	8.35	115.93	**37**	L-phenylalanine	8.21	118.85
11	cystine	8.5	115.22	**38**	L-proline	8.35	115.58
12	formic acid	8.05	123.93	**39**	propionic acid	7.95	118.85
13	fumaric acid	8.42	122.68	**40**	pyroglutamic acid	8.29	115.88
		8.56	124.24	**41**	Pyruvic acid	8.63	111.39
14	glucuronic acid	8.38	119.54			8.35	112.72
15	glutamic acid	8.28	115.99	**42**	serine	8.17	117.63
		8.05	120.42	**43**	succinic acid	7.96	119.16
						8.01	119.64
16	glutamine	8.35	115.90	**44**	succinyl-COA	8.03	119.17
						8.11	119.67
17	glycine	8.2	115.45	**45**	L-threonine	8.34	117.79
18	glycolic acid	8.22	114.97	**46**	L-tryptophan	7.98	119.37
		8.37	115.19	**47**	tyrosine	8.27	118.05
19	hippuric acid	8.2	115.62	**48**	valine	8.38	118.20
20	histidine	8.36	116.60				
21	3-hydroxybutyric acid	8.07	122.20				
22	4-hydroxy-L-proline	8.5	115.89				
		8.36	117.62				

Scheme 1. Synthesis of ^{15}N-Substituted Phthalimide

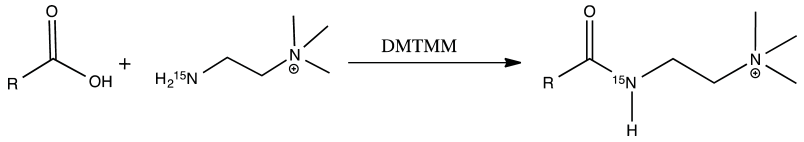


Scheme 2. Alkaline and Acid Hydrolyses of the ^{15}N-Substituted Phthalimide to Yield ^{15}N-Cholamine



dx.doi.org/10.1021/ac401712a | Anal. Chem. XXXX, XXX, XXX−XXX

Scheme 3. General Reaction for Tagging Carboxyl-Containing Metabolites with the Smart Isotope Tag ^{15}N-Cholamine



reaction.[13,20,21] The mixture was stirred at room temperature for 4 h to complete the reaction. The general reaction for tagging metabolites with the smart isotope tag is shown in Scheme 3. To maintain ^{15}N amide protonation, the pH was adjusted to 5.0 by adding 1 M HCl or 1 M NaOH, and the solution volume was adjusted to 580 μL by adding DI water. Serum was deproteinized using methanol prior to metabolite tagging and urine was used with no pretreatment.[13]

NMR Spectroscopy. For each sample, 580 μL was mixed with 30 μL D_2O and placed in a 5 mm NMR tube. NMR experiments were performed on a Bruker DRX 500 MHz or Avance III 800 spectrometer equipped with a room temperature probe or cryoprobe, respectively, suitable for ^{1}H inverse detection with Z-gradients at 298 K. A one pulse sequence with or without solvent signal suppression using presaturation was used for ^{1}H 1D NMR experiments. The sensitivity-enhanced ^{1}H−^{15}N 2D heteronuclear single quantum coherence (HSQC) experiments employed an INEPT transfer delay of 6 ms corresponding to the J_{NH} of 90 Hz. Spectral widths for the ^{1}H and ^{15}N dimensions were approximately 8 and 3 kHz, respectively. Here, 128 free induction decays of 1024 data points each were collected in the indirect (t_1) dimension with 1 or 4 transients per increment. Nitrogen decoupling during the direct acquisition (t_2 dimension) was achieved with the GARP (globally optimized alternating-phase rectangular pulses) sequence. The resulting 2D data were zero-filled to 1024 points in the t_1 dimension after forward linear prediction to 256 or 512 points. A 45° shifted sine-bell window function was applied to both dimensions before Fourier transformation. Chemical shifts were referenced to the ^{1}H signal of TSP for the 1D spectra or the derivatized formic acid signal (^{1}H 8.05 ppm; ^{15}N 123.93 ppm) in the HSQC spectra. Bruker Topspin versions 3.0 or 3.1 software packages were used for NMR data acquisition or processing.

Mass Spectrometry. LC-MS and LC-MS/MS experiments were performed using an Agilent 1200 SL-LC system coupled online with an Agilent 6520 Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA). The sample (8 μL) was injected onto an Agilent Poroshell 120 EC-C18 column (30 mm × 50 mm, 2.7 μm), which was heated to 50 °C. The flow rate was 0.5 mL/min. Mobile phase A was 5 mM ammonium acetate in water, and mobile phase B was 0.1% water in ACN. The mobile phase composition was initially kept isocratic at 3% B for 1 min, then increased to 90% B over 4 min; after another 4 min at 90% B, the mobile phase composition returned to 3% B. Electrospray ionization (ESI) was used in positive mode, and the voltage was 3.5 kV. The mass analyzer was scanned over a range of 50−1000 m/z. The collision energy for auto LC-MS/MS experiments was fixed at 10 V, targeting preselected compounds.

RESULTS AND DISCUSSION

The smart isotope tag, ^{15}N-cholamine, designed, developed, and used in this study retains all the characteristics of the ^{15}N-ethanolamine tag including the solubility of the tagged metabolites in aqueous media, large one-bond J-coupling between ^{1}H and ^{15}N of ∼90 Hz for efficient polarization transfer between ^{1}H and ^{15}N nuclei, and wide chemical shift dispersion for different metabolites in the resulting 2D NMR spectra.[13] In addition, and importantly, ^{15}N-cholamine possesses a permanent positive charge, which enables efficient positive mode detection of the same carboxyl-containing metabolites by MS, irrespective of the pH or solvent conditions of the eluting media, commonly used for chromatographic separation before detection by MS.[6]

Synthesis of ^{15}N-cholamine involved a two-step reaction and followed the Gabriel synthesis procedure with suitable modifications to yield the pure product.[18,19] As seen in the ^{1}H NMR spectrum (Supporting Information Figure S1), the pure intermediate compound, ^{15}N substituted phthalimide, was obtained after the first step of the synthesis. Hydrolysis of this compound yielded the ^{15}N-cholamine in pure form as can be ascertained from its ^{1}H NMR spectrum (Supporting Information Figure S2; peaks at 3.16; 3.48; 3.64 ppm). The accurate MS and MS/MS spectra for ^{15}N-cholamine, shown in Supporting Information Figure S3, help further verify the identity and purity of the synthesized smart isotope tag (m/z = 104.120).

The compound was then used to tag 48 metabolites that were selected for their prominence as carboxyl-containing metabolites in biofluids that represent a variety of metabolic pathways. The ^{1}H and ^{15}N chemical shift data derived from the 2D NMR experiments, after tagging with ^{15}N cholamine, are shown in Table I. Because the ^{15}N-cholamine and the previously used ^{15}N-ethanolamine differ only in their terminal group, the tagging efficiency, reproducibility and chemical shift values for metabolites with ^{15}N-cholamine tag were similar to those obtained using the ^{15}N-ethanolamine tag.[13]

Importantly, as anticipated based on the ^{15}N-ethanolamine tagging approach shown earlier in our laboratory,[13] the ^{15}N-cholamine tagging of metabolites in human serum provided a rich NMR spectrum due to the large number of carboxyl-containing metabolites normally present in blood (Figure 2). The low natural abundance of ^{15}N (0.37%) ensures that none of the nitrogen containing metabolites interferes with the detection of carboxyl-metabolites through the ^{15}N-cholamine tag. Each peak in the spectrum corresponds to different metabolite from the carboxylic acid class. However, metabolites, which contain more than one carboxyl group, provide additional peaks depending on the number of carboxyl groups and molecular symmetry. In addition, metabolites such as lactate, which possess α-hydroxyl groups, show more than one peak for the same metabolite as we described earlier using the ^{15}N-ethanolamine tag.[13] Some of the peaks assigned based on the chemical shift values for the standard compounds are marked with the corresponding number shown in Table I and Figure 2b. Similarly, tagging of metabolites in human urine with ^{15}N-cholamine also enabled the detection of peaks correspond-

dx.doi.org/10.1021/ac401712a | Anal. Chem. XXXX, XXX, XXX−XXX

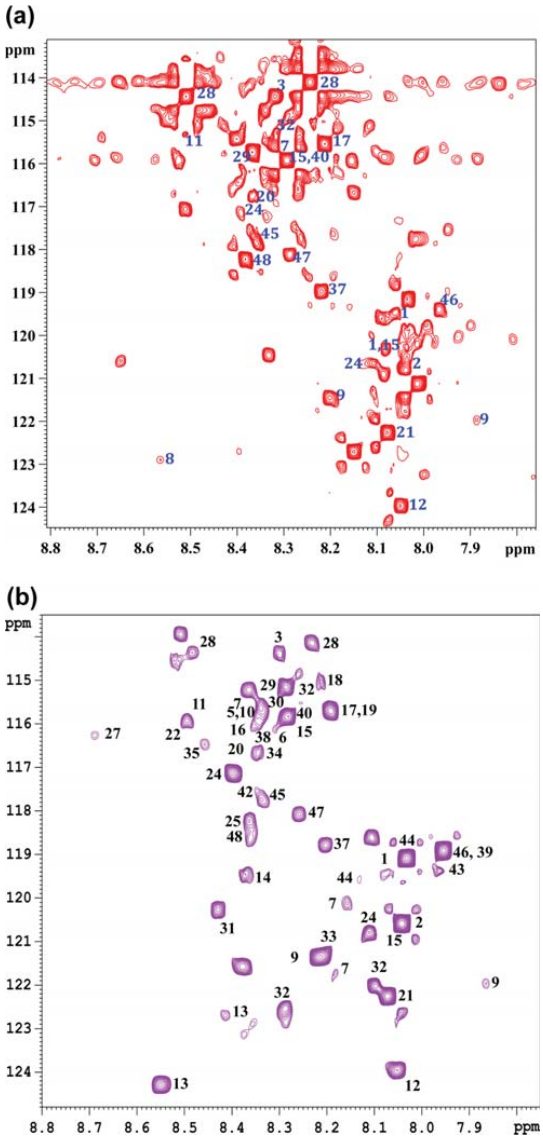


Figure 2. (a) Portion of the $^{1}H-^{15}N$ HSQC spectrum of human serum tagged with ^{15}N-cholamine: (1) aconitic acid; (2) adipic acid; (3) alanine; (7) aspartic acid; (8) betaine; (9) citric acid; (11) cystine; (12) formic acid; (15) glutamic acid; (17) glycine; (20) histidine; (21) 3-hydroxybutyric acid; (24) isocitric acid; (28) lactic acid; (29) leucine; (32) malic acid; (37) phenylalanine; (40) pyroglutamic acid; (45) threonine; (46) tryptophan; (47) tyrosine; (48) valine. (b) Portion of the $^{1}H-^{15}N$ HSQC spectrum of a mixture of standard compounds at various concentrations obtained after tagging with ^{15}N-cholamine. The peak numbers correspond to the compounds shown in Table I.

ing to well over a hundred carboxylic acid group containing metabolites (Figure 3). Peaks tentatively assigned based on the values for the standard compounds are marked by their numbers shown in Table I and Figure 2b.

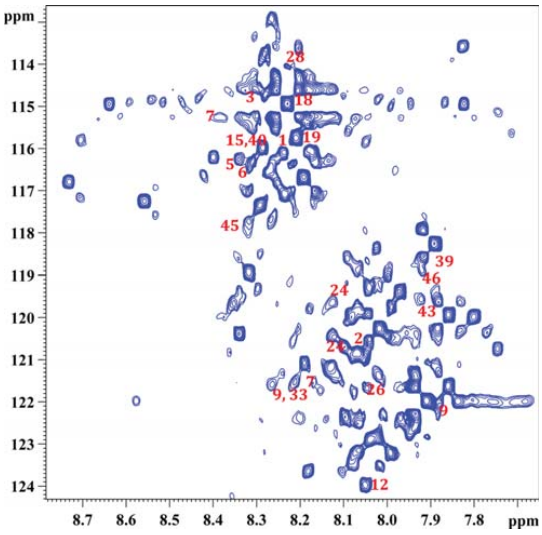


Figure 3. Portion of the $^{1}H-^{15}N$ HSQC spectrum of human urine tagged with ^{15}N-cholamine: (1) aconitic acid; (2) adipic acid; (3) alanine; (5) arginine; (6) asparagine; (7) aspartic acid; (9) citric acid; (12) formic acid; (15) glutamic acid; (18) glycolic acid; (19) hippuric acid; (24) isocitric acid; (28) lactic acid; (33) malonic acid; (39) propionic acid; (40) pyroglutamic acid; (43) succinic acid; (45) threonine; (46) tryptophan.

The ^{15}N-cholamine tagging of metabolites in aqueous media enabled a sensitivity enhancement of up to 3 orders of magnitude or more in the MS detection of carboxyl metabolites. The derivatized metabolites could be detected easily in positive ion mode as compared to the same metabolites detected in negative ion mode without the tag. For example, the sensitivity for pyruvic acid detected in positive ion mode after ^{15}N-cholamine tagging was enhanced by a factor of about 1500 when compared to that detected for the same metabolite without the ^{15}N-cholamine tag, in negative ion mode. Figure 4 shows typical mass spectra for formic acid and pyruvic acid after tagging with ^{15}N-cholamine. The enhancement in sensitivity is primarily due to the high ionization efficiency imparted by the permanent positive charge of the ^{15}N-cholamine and is in agreement with results by Smith and co-workers for fatty acid analysis using the heavy and light forms of cholamine.[6] In that study, reactions of metabolites with cholamine were made in organic solution in contrast to the aqueous media used here. The ^{15}N-cholamine derivatized serum samples were then analyzed by LC-MS. As anticipated, due to the presence of the permanent positive charge, tagged metabolites could be readily detected in positive ion mode with high sensitivity. Sensitivity enhancement by a factor of up to nearly 3000 could be achieved for tagged acids. The extracted ion chromatograms for a few typical carboxylic acids detected in serum with ^{15}N-cholamine tag are shown in the Supporting Information Figure S4.

One potential issue is the effect on chromatographic retention time caused by the addition of the cholamine tag. However, separation of the tagged metabolites using HILIC columns offers an opportunity to effectively separate before detection using MS. For example, the results of separation of a mixture of standard carboxylic and amino acids performed

dx.doi.org/10.1021/ac401712a | Anal. Chem. XXXX, XXX, XXX−XXX

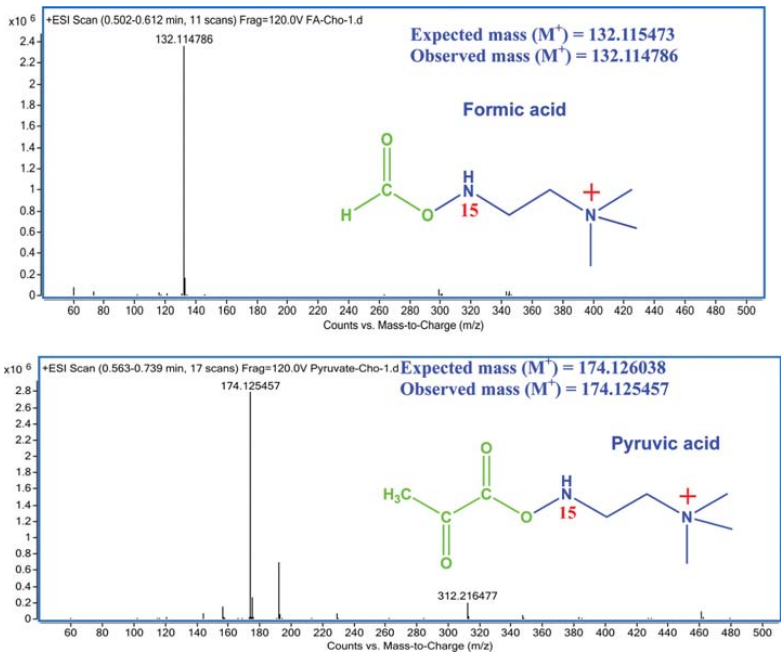


Figure 4. Typical LC-QTOF-MS spectra for formic acid and pyruvic acid obtained after tagging with the smart isotope tag, ^{15}N-cholamine. The permanent charge on the tagged metabolites enables their sensitive detection; the observed peak is from the intact tagged metabolite.

using a HILIC column, without attempting to optimize chromatography conditions, indicate that ^{15}N-cholamine tagged metabolites can be separated effectively (Supporting Information Figure S5). More broadly, we can contemplate the use of dual purpose smart tags for other NMR-MS combinations. For GC-MS, the addition of a charged species will likely cause problems related to reduced volatility; however, a different tag, such as ^{13}C or even ^{29}Si labeled silyl-type tags can be contemplated.[22] Another alternative is the use of smart tags for capillary electrophoresis (CE) coupled to MS, which is increasingly of interest in metabolomics.[23] In fact, positively charged derivatization agents (based on pyridinum containing compounds) have been demonstrated for the use of metabolite profiling of carboxylic acids by CE-MS.[24] Thus, the potential for the use of smart tags such as cholamine for CE-MS and NMR is quite promising.

In conclusion, we have developed a smart isotope tag, ^{15}N-cholamine, which possesses dual characteristics for metabolite profiling in complex biological mixtures using the powerful analytical techniques of NMR and MS. By combining the individual strengths of the ^{15}N label and permanent charge, the smart isotope tag facilitates detection of carboxyl-containing metabolome by both NMR and LC-MS techniques with high sensitivity. Detection of the same metabolites by both NMR and MS (Supporting Information Figure S6), effectively opens unique opportunities for identification of unknown metabolites and direct comparison of metabolite data from the two powerful analytical platforms.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S6.

This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: draftery@uw.edu. Tel.: (206) 543-9709.

Notes

The authors declare the following competing financial interest: D.R. is an officer at Matrix-Bio.

■ ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (2R01GM085291). We thank the reviewer who suggested we think more broadly about possible applications of smart tags. The authors thank Dr. David Thompson, Purdue University, for discussions and Agilent Technologies for the generous donation of a Poroshell 120 HILIC column.

■ REFERENCES

(1) Gowda, G. A. N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Mol. Diagn.* **2008**, *8* (5), 617−633.
(2) Shintu, L.; Banudin, R.; Navratil, V.; Prot, J. M.; Pontoizeau, C.; Defernez, M.; Blaise, B. J.; Domange, C.; Pèry, A. R.; Toulhoat, P.; Legallais, C.; Brochot, C.; Leclerc, E.; Dumas, M. E. *Anal. Chem.* **2012**, *84*, 1840−1848.
(3) Wishart, D. S. *Trends in Food Sci. Technol.* **2008**, *19*, 482−493.
(4) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Pharm. Res.* **2006**, *23*, 1075−1088.
(5) Veldhoen, N.; Ikonomou, M. G.; Helbing, C. C. *Ecotoxicol. Environ. Saf.* **2012**, *76*, 23−38.
(6) Lamos, S. M.; Shortreed, M. R.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. *Anal. Chem.* **2007**, *79*, 5143−5149.

(7) Yang, W. C.; Adamec, J.; Regnier, F. E. *Anal. Chem.* **2007**, *79*, 5150−5157.
(8) Yang, W. C.; Sedlak, M.; Regnier, F. E.; Mosier, N.; Ho, N.; Adamec, J. *Anal. Chem.* **2008**, *80*, 9508−9516.
(9) Yang, W. C.; Regnier, F. E.; Silva, D.; Adamec, J. *J. Chromatogr. B.* **2008**, *870*, 233−240.
(10) Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *J. Chromatogr. A.* **2010**, *1217*, 667−675.
(11) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919−3932.
(12) Wu, Y.; Li, L. *Anal. Chem.* **2013**, *85*, 5755−5763.
(13) Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009**, *81* (12), 4882−4888.
(14) Ye, T.; Zhang, S.; Mo, H.; Tayyari, F.; Nagana Gowda G, A.; Raftery, D. *Anal. Chem.* **2010**, *82* (6), 2303−2309.
(15) Shanaiah, N.; Desilva, M. A.; Nagana Gowda G, A.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11540−11544.
(16) DeSilva, M. A.; Shanaiah, N.; Gowda, G. A. N.; Rosa-Perez, K.; Hanson, B. A.; Raftery, D. *Magn. Reson. Chem.* **2009**, *47*, S74−S80.
(17) Gowda, G. A. N.; Tayyari, F.; Ye, T.; Suryani, Y.; Wei, S.; Shanaiah, N.; Raftery, D. *Anal. Chem.* **2010**, *82*, 8983−8990.
(18) Iida, K.; Ohtaka, K.; Kajiwara, M. *J. Label. Compd. Radiopharm.* **2007**, *50*, 251−253.
(19) Khan, M. N. *J. Org. Chem.* **1996**, *61* (23), 8063−8068.
(20) Kunishima, M.; Kawachi, C.; Monta, J.; Terao, K.; Iwasaki, F.; Tani, S. *Tetrahedron* **1999**, *55*, 13159−13170.
(21) Kunishima, M.; Kawachi, C.; Hioki, K.; Terao, R.; Tani, S. *Tetrahedron* **2001**, *57*, 1551−1558.
(22) Schraml, J. *Prog. NMR Spect.* **1990**, *22*, 289−348.
(23) Ramautar, R.; Somsen, G. W.; de Jong, G. J. *Electrophoresis* **2013**, *34* (1), 86−98.
(24) Yang, W.-C.; Regnier, F. E.; Adamec, J. *Electrophoresis* **2008**, *29* (22), 4549−4560.

analytical chemistry

ARTICLE

pubs.acs.org/ac

Ratio Analysis Nuclear Magnetic Resonance Spectroscopy for Selective Metabolite Identification in Complex Samples

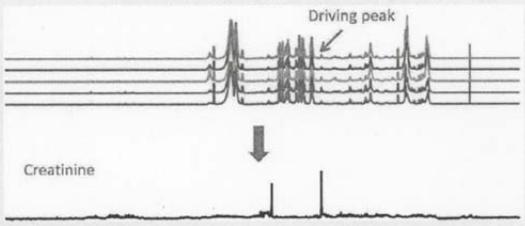
Siwei Wei,[†] Jian Zhang,[†] Lingyan Liu,[‡] Tao Ye,[†] G. A. Nagana Gowda,[†] Fariba Tayyari,[†] and Daniel Raftery*[,†]

[†]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, Indiana 47907, United States

[‡]Weldon School of Biomedical Engineering, Purdue University, 206 S. Martin Jischke Drive, West Lafayette, Indiana 47907, United States

Supporting Information

ABSTRACT: Metabolite identification in the complex NMR spectra of biological samples is a challenging task due to significant spectral overlap and limited signal-to-noise. In this study we present a new approach, RANSY (ratio analysis NMR spectroscopy), which identifies all the peaks of a specific metabolite on the basis of the ratios of peak heights or integrals. We show that the spectrum for an individual metabolite can be generated by exploiting the fact that the peak ratios for any metabolite in the NMR spectrum are fixed and proportional to the relative numbers of magnetically distinct protons. When the peak ratios are divided by their coefficients of variation derived from a set of NMR spectra, the generation of an individual metabolite spectrum is enabled. We first tested the performance of this approach using one-dimensional (1D) and two-dimensional (2D) NMR data of mixtures of synthetic analogues of common body fluid metabolites. Subsequently, the method was applied to ^{1}H NMR spectra of blood serum samples to demonstrate the selective identification of a number of metabolites. The RANSY approach, which does not need any additional NMR experiments for spectral simplification, is easy to perform and has the potential to aid in the identification of unknown metabolites using 1D or 2D NMR spectra in virtually any complex biological mixture.



Nuclear magnetic resonance (NMR) based metabolomics is increasingly used in numerous studies and applications that include drug response, early disease diagnosis, toxicity and nutritional studies, and basic systems biology using a variety of complex biological samples.[1−10] A number of reviews have been published that describe advancements in the field.[11−16] NMR spectroscopy is an important analytical technique in metabolomics primarily because of its quantitative nature and high reproducibility. Due to the complexity of the NMR spectra of biological samples, however, obtaining information on low-concentration metabolites is challenging due to the high degree of spectral overlap. Nevertheless, these species are important to analyze because their changing concentrations may distinguish biological status, such as health and disease.

Numerous advances in NMR methods and databases have so far been made to aid the analysis of complex NMR spectra, as well as to identify specific metabolites and pathways associated with the onset of various diseases.[17−20] Some important approaches for peak assignments in the complex NMR spectra using NMR databases include targeted profiling of complex NMR spectra using mathematical modeling of the pure compound spectra to fit the experimental data,[21] the COLMAR approach that screens chemical shift lists or cross-sections of multidimensional NMR spectra,[22] and annotation of metabolites using a statistical index called spin-assign *p*-value.[23]

Methods such as statistical total correlation spectroscopy (STOCSY),[24] selective TOCSY,[25,26] and isotope enhanced methods[27−32] offer approaches to simplify NMR spectra of complex samples such as biofluids and cells. In addition, spectral simplification in terms of increased resolution along the indirect dimension can be achieved through covariance NMR spectroscopy, introduced by Zhang and Bruschweiler.[33−35]

Among the methods applicable to metabolomics, STOCSY[24] has proven to be a very useful and general data analysis method to identify related metabolite peaks from complex spectra. The autocorrelation matrix of a set of NMR spectra identifies peaks from the same metabolite because these peaks show high correlation, which can simplify assignments. In addition, from the STOCSY analysis, peaks from other metabolites in the same or different pathways that are also correlated appear in the spectrum, which provides benefits but also complicates the analysis. STOCSY has successfully been extended to analyze spectra obtained from different nuclei[36] and even different analytical techniques, such as NMR, cryoflow LC-NMR, and UPLC-MS.[37−39] Correlation was also used to group variables in complex spectra and prove interpretability of latent variables for

Received: February 25, 2011
Accepted: September 6, 2011
Published: September 06, 2011

metabolic biomarker recovery.[40] However, due to the often substantial number of high correlation values, it can be difficult to find the meaningful correlations without choosing a threshold empirically. Clean spectra for individual metabolites are sometimes difficult to achieve.

Here we propose an alternative method to detect metabolite signals selectively from the complex one-dimensional (1D) and two-dimensional (2D) spectra. The new approach involves an analysis of the ratio between a selected peak and all of the other peaks in the spectrum. The mean of each ratio is calculated across the multiple spectra in the data set and is then divided by the standard deviation of this ratio to generate a spectrum that clearly illustrates the relationship among the peaks from the same molecular species. The newly generated spectrum contains clean peaks with good signal-to-noise; identification can then be easily performed by comparing the ratio spectrum with a standard spectrum from a database. The RANSY method therefore allows easier identification of individual molecules in overlapped spectral regions without the need for additional experiments. The same methodology can also be applied to simplify 2D spectra. For example, in the spectra generated by isotope tagging methods, metabolites with multiple functional groups can be identified easily.[30] Because of intrinsic characteristics of the calculated ratio, no data normalization is needed.

■ THEORY

RANSY is based on the fact that the ratio between two NMR peak intensities or areas from the same metabolite will be equal to the relative number of magnetically nonequivalent spins and will be constant across all spectra; and therefore the standard deviation of this ratio will in theory be zero, or very close to zero. Although spectral noise contributes to some variation in this ratio, the noise contribution is still relatively small. Of course, in highly overlapped spectra, it is the contributions to the peaks in the spectra from a particular metabolite that will have fixed ratios. On the other hand, if the two peaks originate from different metabolites, their ratio should vary across different spectra. Thus, the standard deviation of the ratio of peaks from separate molecules across all of the spectra will be typically large, except in those rare cases where these metabolites are very highly correlated.

To obtain the ratio spectrum of an unknown metabolite with only one known peak, the ratios between all of the other points in each spectrum and the known peak are calculated. The mean of each ratio is then divided by its standard deviation to get its reciprocal coefficient of variation (1/CV). Since the ratio's standard deviation is used as the denominator, small standard deviations will produce a very large reciprocal value, generating a peak, but large standard deviations will give small numbers, similar to noise values. The reciprocal of the CV is used instead of the standard deviation because CV is dimensionless and standard deviations can only be compared in the context of the mean value.

The RANSY algorithm is as follows. We denote the ith spectrum of a set of n spectra as a vector $\mathbf{X}_i$. The jth data point of m total points in that spectrum is denoted as $\mathbf{X}_{i,j}$. We designate one peak of interest from a given metabolite (the kth point in spectrum i, $\mathbf{X}_{i,k}$) as the "driving" peak. The first step is to calculate all of the ratios between all of the other points and the driving peak; we can denote the ratio matrix as $\mathbf{D}$, which is an $n \times m$ matrix. Each point in that matrix is defined as follows:

$$\mathbf{D}_{i,j} = \frac{\mathbf{X}_{i,j}}{\mathbf{X}_{i,k}} \tag{1}$$

Next, to generate a ratio spectrum, which is an m-element vector $\mathbf{R}$, the quotient of means and standard deviations across columns of $\mathbf{D}$ are calculated. Thus, the jth element of the 1D RANSY spectrum (vector $\mathbf{R}$) is calculated as follows:

$$\mathbf{R}_j = \frac{(1/n)\sum_{i=1}^{n} \mathbf{D}_{i,j}}{\sqrt{(1/n)\sum_{i=1}^{n}\left(\mathbf{D}_{i,j} - (1/n)\sum_{i=1}^{n}\mathbf{D}_{i,j}\right)^2}} \tag{2}$$

However, since the standard deviation is zero for the driving peak itself, the corrected ratio for the driving peak $\mathbf{X}_k$ is infinite. To not sacrifice the driving peak, we assign the corrected ratio of that peak equal to the maximum of $\mathbf{R}$. So the final RANSY spectrum is given by eq 2 except for the driving peak $\mathbf{R}_k$, which is given by

$$\mathbf{R}_k = \max(\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_m)$$

All ratio analyses were performed on a personal computer using R version 2.10.0 software.

■ EXPERIMENTAL METHODS

Chemicals. Thirty-three synthetic analogues of human serum metabolites (see Supporting Information Table S-1) and the NMR standard, TSP (trimethylsilylpropionic acid-d_4, sodium salt), were purchased from Sigma-Aldrich (St. Louis, MO). Deuterium oxide (D_2O, 99.9% D) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). Chemicals for ^{15}N-ethanolamine tagging, 4-(4,6-dimethoxy[1,3,5]triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM), and ^{15}N-ethanolamine were purchased from Cambridge Isotope, and Isotech (Miamisburg, OH), respectively. All chemicals were used without further purification.

Serum Samples. 100 serum samples for 1D NMR were obtained from Innovative Research, Inc. (Novi, MI), and 550 μL of each sample was used for the Carr−Purcell−Meiboom−Gill (CPMG) experiment.

For 1D RANSY experiments, 15 standard samples were prepared by mixing 15 μL of TSP (5 mM in D_2O) with stock solutions of the 33 metabolites (20 mM each). Volumes for each of the standards were chosen randomly within a range, and their final concentrations can be found in Supporting Information Table S-1. Concentrations of the standards were varied by nearly 2 orders of magnitude (0.03−2 mM) and were chosen to match roughly with their physiological concentrations in blood or urine.[41] A 285 μL solution of 0.5 M phosphate buffer in D_2O (pH = 7.4) was added to each standard mixture to minimize peak shifts. After mixing, 550 μL of each sample was used to acquire the NMR spectra. A second set of 10 sample mixtures containing 15 different metabolites was also prepared in a similar manner, except that the concentration ranges for each of these metabolites was varied from 0.03 to 2 mM (see Supporting Information Table S-2).

For 2D RANSY experiments, 15 standard mixtures were prepared by mixing 100 μL of water with stock solutions of the 33 metabolites (20 mM each) to obtain the final concentrations. These 15 samples were derivatized with ^{15}N-ethanolamine,

according to the procedure described previously.[30] Briefly, 3 μL of ^{15}N-ethanolamine (50 μM) was added to each sample. After adjustment to pH 7.0 with 1 M HCl, 21 mg of DMT-MM was added to the solution. The reaction mixture was kept at room temperature for 4 h with stirring to complete the reactions. To maintain ^{15}N amide protonation, the pH was adjusted to 5.0 and the solution volume was adjusted to 600 μL by adding water prior to 2D NMR experiments, of which 550 μL was used for NMR experiments. A blank sample was also prepared by the same procedure as above but without adding any metabolites.

NMR Spectroscopy. One-dimensional NMR experiments for the 15 standard mixture samples were performed on a Bruker Avance-III-800 spectrometer equipped with a room-temperature ^{1}H inverse detection Z-gradient probe. NMR data were acquired using the 1D NOESY pulse sequence with water presaturation. A set of 128 scans with 16 k time domain data points was collected using a spectral width of 12 800 Hz. An exponential weighting function corresponding to 1.0 Hz line broadening was applied to the free induction decay (FID) before Fourier transformation. The spectra were then phased, baseline corrected, and referenced to TSP (δ = 0.000 ppm) using Bruker Topspin 3.0 software.

^{1}H NMR experiments were obtained for 100 commercial serum samples using a standard 1D CPMG pulse sequence coupled with water presaturation on a Bruker Avance-500 spectrometer equipped with a TXI gradient cryoprobe. A set of 128 scans with 16 k time domain data points was collected using a spectral width of 6000 Hz. An exponential weighting function corresponding to 1.0 Hz line broadening was applied to the FIDs before Fourier transformation. The acquired spectra were then phased, baseline-corrected, and referenced to alanine (δ = 1.479 ppm) using Bruker's Topspin 3.0 software.

Two-dimensional experiments for standard samples were performed at 298 K on a Bruker Avance-III-800 spectrometer as well. ^{1}H−^{15}N 2D heteronuclear single quantum coherence (HSQC) experiments employed an INEPT transfer delay of 5.5 ms corresponding to a $^1J_{NH}$ coupling of 90 Hz. Spectral widths of approximately 10 kHz for the ^{1}H dimension and 5 kHz for ^{15}N were used at 800 MHz. A total of 128 free induction decays of 2048 data points each were collected in the indirect (t_1) dimension with 32 transients per increment. ^{15}N decoupling in the direct detection dimension (t_2) was achieved using globally optimized alternating-phase rectangular pulses (GARP). The resulting 2D data were zero-filled to 1024 points in the t_1 dimension after forward linear prediction to 512 points. The spectra were then phased, baseline-corrected, and integrated using Bruker Topspin 3.0 software as described earlier.[30,42]

■ RESULTS AND DISCUSSION

The RANSY approach was first applied to analyze a set of 15 standard metabolite mixtures, and some of these results are shown in Figure 1. The 1D NMR spectrum of a mixture of 33 standard compounds can be seen in Figure 1a. Some regions in this spectrum are quite crowded even in this relatively simple mixture of standard compounds. RANSY was applied to the collection of 15 samples by selecting one of the isolated metabolite peaks as the "driving peak". For example, in Figure 1b the driving peak is at 3.26 ppm and corresponds to the upfield peak of the methyl singlet of betaine. The other singlet at 3.91 ppm also appears in the RANSY spectrum at the correct position. In Figure 1c, the driving peak is at 3.18 ppm and corresponds to the downfield triplet of β-alanine. The other triplet appears in the RANSY spectrum at the expected position (2.56 ppm) and intensity. Both betaine and β-alanine are quite low in concentration, and their resonances sit in relatively crowded regions, as shown in the expanded region of Figure 1a. Nevertheless, one can easily generate their spectra by RANSY. The example shown in Figure 1d is hippuric acid, which has several peaks in the aromatic region, and a doublet at 3.96 ppm. Three multiplets could be found in the aromatic region, centered at 7.81, 7.64, and 7.54 ppm, corresponding to hydrogens located on the phenyl ring. Using the center peak at 7.64 ppm as the driving peak, the RANSY spectrum reveals the other two peaks plus a doublet centered at 3.96 ppm, which is due to hippuric acid's CH_2 group, and also had its correct chemical shift. In the spectra of Figure 1, the noise levels are significantly lower when compared with the peaks generated by RANSY, and the RANSY spectra are very similar to standard spectra of these metabolites. The relative peak heights are also close to those for the standards, indicating that the standard deviations for the ratios from the same compounds are small and very similar to one another. As will be shown below, using a larger number of samples will further improve the signal-to-noise ratio (vide infra). To investigate RANSY on mixtures in which the signals vary over a larger dynamic range, we also record spectra of 10 mixtures of 15 metabolites where the concentrations all varied over a (nonbiological) range of 0.03−2 mM (Table S-2). The RANSY spectra of the same three metabolites (betaine, β-alanine, and hippuric acid) were calculated by the driving peaks mentioned above and shown in Supplorting Information Figure S-1. Similar RANSY results were obtained with this larger concentration range, with the one difference being that the noise appeared to be more evenly distributed in this case.

RANSY was then applied to a set of 100 serum samples. In Figure 2a the CPMG spectrum of serum is shown prior to the application of RANSY. Subsequently, RANSY was applied to the upfield peaks of creatinine and creatine centered at 3.03 and 3.02 ppm, respectively, as shown in Figure 2a. These two metabolites were chosen because their upfield peaks are very close to one another. They are somewhat difficult to identify in serum since they both often have overlapping singlets, and these signals are somewhat weak due to their low concentrations in serum (~30 μM for adults). The RANSY spectra for both metabolites are shown in Figure 2b,c; it is very clear that the low-field singlet centered at 4.05 ppm is associated with creatinine, and the singlet at 3.92 ppm is associated with creatine. While the RANSY spectra for both metabolites are clear, there is some noise generated by the large number of small peaks in the baseline of the serum spectra. One can, however, distinguish the RANSY peaks from baseline noise by applying a minimum threshold value across the spectrum. A threshold of R = 6 for the calculated ratio of eq 2 can be applied to the spectrum of Figure 2b,c, and indicated by the dashed line. It can be seen that the all of the noise peaks in both figures are below that threshold.

To test RANSY further in serum samples, the method was also applied to identify resonances from valine and leucine. Both of the upfield peaks for valine and leucine come from methyl groups that overlap with isoleucine. In addition, leucine's upfield triplet centered at 0.95 ppm overlaps with the tail of a strong, broad lipid peak. The chosen driving peaks were at 1.02 ppm for valine and 0.94 ppm for leucine. Correlation spectra were also calculated using the same driving peak for comparison. The valine RANSY spectrum shown in Figure 3b shares two upfield doublets from the two methyl groups, and another doublet at around 3.6 ppm,

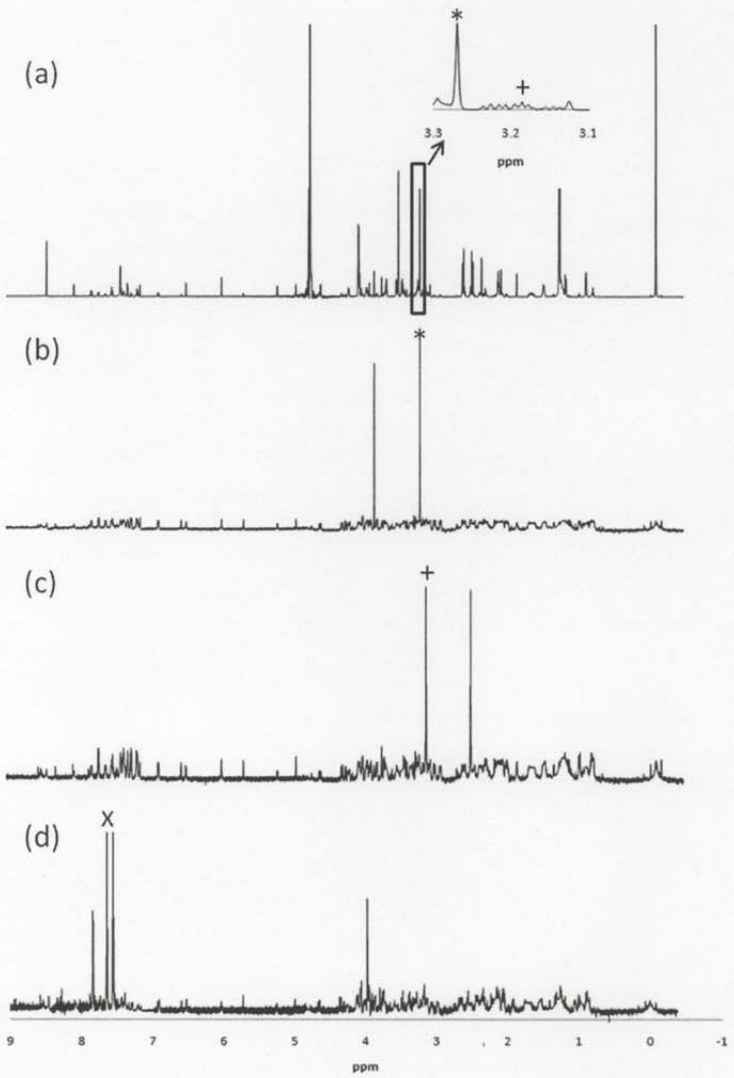


Figure 1. (a) ^{1}H NMR spectrum of a mixture of 33 standard compounds. (b–d) Selective detection of (b) betaine, (c) β-alanine, and (d) hippuric acid by the RANSY method. Panels b–d were generated from 15 ^{1}H NMR spectra of mixtures, and the driving peaks are indicated by an asterisk, plus sign, and letter X, respectively. The inset in a shows the crowded region for driving peaks of betaine and β-alanine.

which was from the α-CH proton. The RANSY method is not perfect, as the multiplet at 2.25 ppm ($R = 3.1$) was missing due to its low intensity and some peak shifting observed in the CPMG spectra. For the RANSY spectrum calculated for leucine, we can see in Figure 3d that another multiplet centered at 1.70 ppm appears for leucine, which comes from the β-CH_2 and γ-CH protons. However, the multiplet centered at 3.71 ppm was missing ($R = 0.69$) compared with its standard spectrum, because of the very strong signal intensity and overlap from glucose. To compare RANSY with the conventional statistical correlation method, correlation spectra based on Pearson correlation coefficients were calculated using the same driving peaks for valine and leucine. Peaks in the spectra as shown in Figure 3c,e thus represent correlation coefficients between the vector of the driving peak and those of all other peaks. While the RANSY results are not perfect in these cases, they are still much better than the correlation method calculated and shown in Figure 3c,e, in which most of the peaks can hardly be seen.

In general, to calculate the RANSY spectra, it is necessary to use multiple spectra to calculate the relative CV values.

dx.doi.org/10.1021/ac201625f |*Anal. Chem.* 2011, 83, 7616–7623

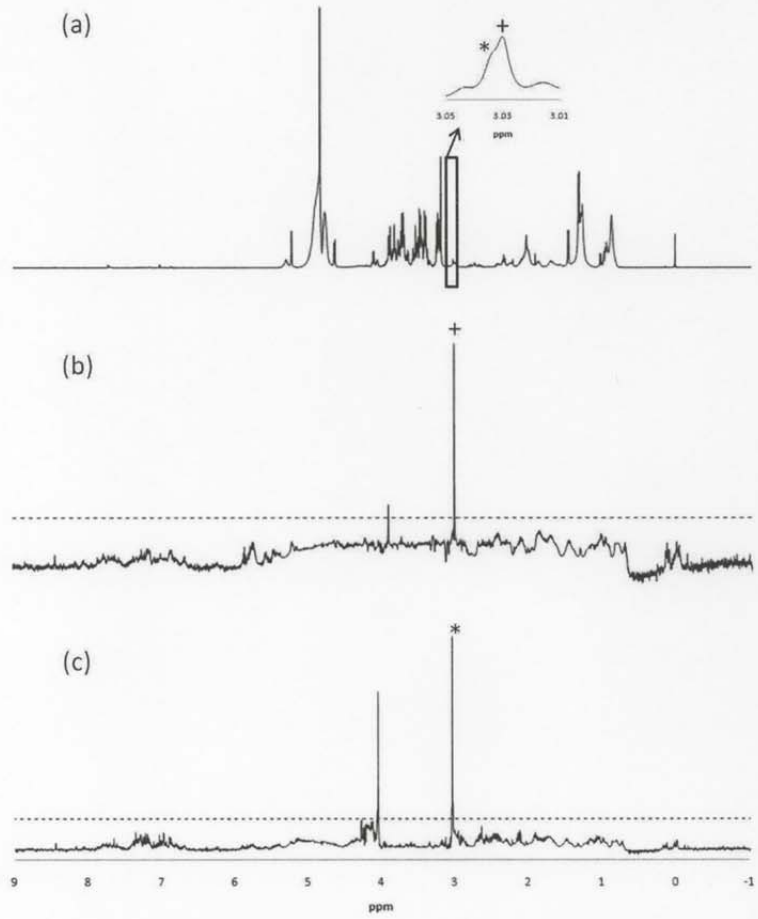


Figure 2. (a) ^{1}H NMR spectrum of serum, obtained using the CPMG pulse sequence. (b) Selective detection of creatine by the RANSY method. The driving peak is at 3.02 ppm (indicated by the plus sign). (c) Selective detection of creatinine driven by the singlet at 3.03 ppm (indicated by the asterisk). The noise threshold value of $R = 6$ is indicated by the dashed horizontal line in both b and c. The inset in a shows the two very closely spaced peaks for creatine and creatinine used as driving peaks.

To evaluate the effect of limited sample numbers, an analysis of the RANSY spectra of serum was performed using different numbers of spectra. The RANSY spectrum for valine was calculated by averaging the results from using each of the four peaks from the two valine doublets located at 1.04 and 0.98 ppm as the driving peaks. First, five samples were randomly picked from the 100 commercial serum samples. Next, additional and randomly chosen samples were added to the original five samples and the RANSY spectrum was recalculated. The results of applying RANSY to different numbers of spectra, from 5 to 80 are shown in Supporting Information Figure S-2. When only five samples were used, the two doublets were not of the same height and at least one peak was distorted. The addition of two spectra in the RANSY calculation (a total of seven spectra) resulted in much better resolution for the two doublet peaks and the peak shape also improved. When 20 samples or more were used, the peaks became well-shaped and showed no additional changes when additional samples were added to the RANSY calculation. Overall, the differences in the RANSY spectra between 20 and 80 samples were minor and not easy to identify. From this example, we expect that 20 samples would often be sufficient to deconvolute metabolites well even in complex samples such as serum.

RANSY was also applied to 2D HSQC spectra of isotope-tagged metabolite mixtures. Fifteen mixtures each containing different concentrations of the same 33 metabolites were tagged with ^{15}N-ethanolamine, as described in Experimental Methods. Of these 33 metabolites, three metabolites (isocitric acid, citric acid, and *cis*-aconitic acid) have three carboxyl groups, and three metabolites (L-glutamic acid, oxalic acid, and malic acid) have two carboxyl groups, while the remaining metabolites have only one carboxyl group each. Compared with the library,[30] all peaks were observed as expected except for one peak each for L-glutamic acid and oxalic acid. A typical isotope-enhanced 2D

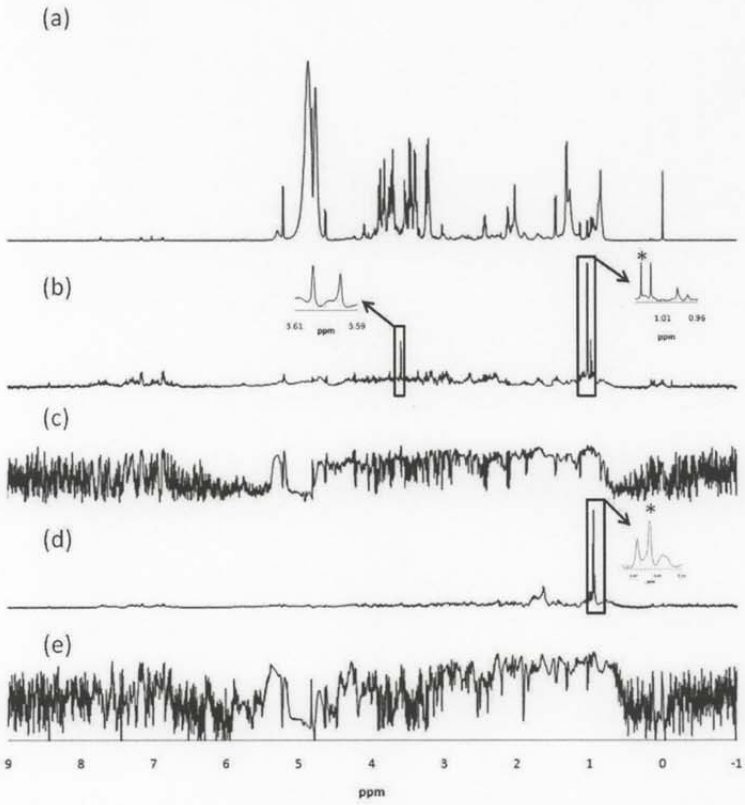


Figure 3. ^{1}H NMR spectrum of serum (a), obtained using the CPMG pulse sequence. (b) Selective detection of valine from the serum ^{1}H NMR spectrum by RANSY. The identified peaks around 3.6 ppm are from the α-CH proton. (c) Selective detection of valine from the serum ^{1}H NMR spectrum by statistical correlation. (d) Selective detection of leucine from the complex serum ^{1}H NMR spectrum by RANSY. The identified multiplet around 1.70 ppm is from the β-CH_2 and γ-CH protons. (e) Selective detection of leucine from the complex serum ^{1}H NMR spectrum by statistical correlation. The driving peaks are indicated by asterisks.

HSQC spectrum is shown in Figure 4a. A few minor peaks can also be seen, which we determined to be due to the impurities present in the starting materials, or in a few cases, incomplete derivatization of standard compounds with multiple functional groups (especially OH groups). Similar issues are sometimes present in GC-MS analysis employing derivatization as well. All 47 metabolite peaks detected in every spectrum were integrated, and the RANSY analysis was performed for all pairs of peaks by an in-house-developed R code (see the Supporting Information). Whether the two peaks are associated with the same metabolite can be determined by the RANSY method. The algorithm used to calculate individual metabolite spectra is as follows: if two peaks showed a RANSY *R* value above the threshold mutually, we regard them as associated with a single metabolite. And a third peak will be added to that metabolite if and only if it showed high *R* values when using both peaks as the driving peak. Following this approach, 2D spectra of individual metabolites could then be generated. In a few examples as shown in Figure 4b−i, the peaks labeled as 1, 2, and 47 in Figure 4a were identified as coming from lactic acid; peaks 3 and 4 from glucuronic acid; peaks 5−7 from citric acid; peaks 11−14 from isocitric acid (including its OH group that was also derivatized); peaks 17 and 18 from malic acid; and peaks 20−22 from *cis*-aconitic acid. All of these peaks were detected with a threshold value, $R \geq 6.5$. Two examples of single-peak metabolites, leucine (peak 15) and betaine (peak 24), are also shown here. Interestingly, the ratio threshold of 6.5 is close to what we have used for 1D RANSY, which indicates at least some generality of this threshold value. Overall, 30 out of 33 metabolites can be assigned correctly from this approach (Supporting Information Table S-3). The metabolites not detected correctly were oxalic acid, L-glutamic acid, and 4-hydroxyl-L-proline. The first two metabolites only showed one peak in some of the spectra due to low concentration and incomplete reaction. One of the two peaks of 4-hydroxyl-L-proline (peak 10 in Figure 4a) overlapped with a background peak and became distorted. The other peaks thus identified from RANSY match very well with the peak assignments for the metabolites using the ^{15}N tagging.[30] By this method, the spectra can be substantially simplified, leading to better assignments. Considering that 20 samples or more improve the RANSY analysis, as discussed before, a larger sample set will improve the discrimination between peaks from the same metabolite and

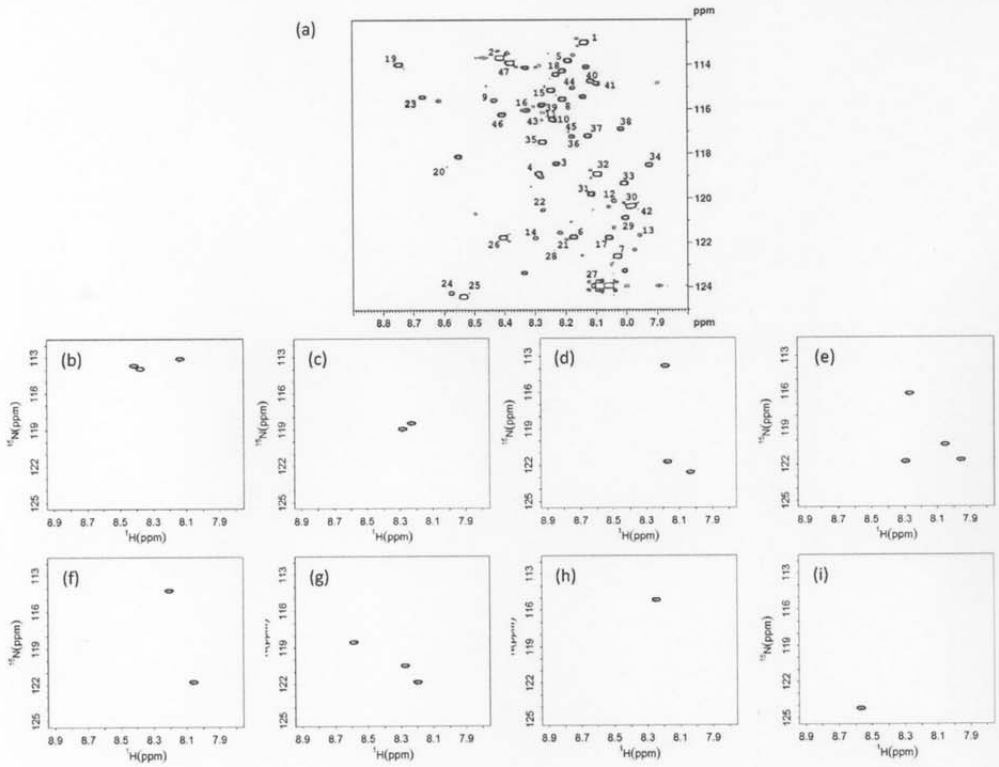


Figure 4. (a) $^{1}H-^{15}N$ HSQC spectrum of a standard mixture containing 33 metabolites with identifications and concentrations given in Table S-1. Examples of metabolites detected by RANSY: (b) lactic acid (peaks 1, 2, and 47); (c) glucuronic acid (peaks 3 and 4); (d) citric acid (peaks 5–7); (e) isocitric acid (peaks 11–14); (f) malic acid (peaks 17 and 18); (g) *cis*-aconitic acid (peaks 20–22); (h) leucine (peak 15); and (i) betaine (peak 24). Individual HSQC spectra b–i were generated by thresholding the RANSY ratio of peaks above a value, $R \geq 6.5$.

background peaks, and thus metabolites with multiple functional groups will be easier to identify.

■ CONCLUSIONS

NMR spectra of complex biosamples can pose difficulties in assignment and analysis. On the basis of the relationship of spin-correlated peaks, RANSY can provide a useful approach to help solve this problem. Using peak ratios to analyze the relationship between intrinsically correlated signals is a new approach that can be applied to complex spectral analysis problems. While a variety of correlation methods have been proved to be quite useful in this effort, the RANSY approach is quite complementary and has some advantages. For example, no additional experiments are needed to apply the RANSY method; it is computationally efficient and does not require enormous numbers of samples. The clearly visible spectra generated by RANSY compare well with correlation calculations. No normalization is needed before analysis. And, in principle, RANSY should be applicable to the identification of unknown species, even those which are not present in spectral libraries. As shown here, RANSY analysis results in an efficient recovery of metabolite peak information from complex 1D and 2D spectra of standard mixtures as well as blood samples. Peak identification can be facilitated with the development of RANSY.

■ ASSOCIATED CONTENT

Supporting Information. Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: (765) 494-6070. Fax: (765) 494-0239. E-mail: raftery@purdue.edu.

■ ACKNOWLEDGMENT

This work is supported by the NIH/NIGMS (Grant 1R01 GM085291). The authors also thank Dr. Shanaiah Murthy at Matrix-Bio, Inc., and the Purdue Interdepartmental NMR Facility (PINMRF) for their assistance with NMR experiments.

■ REFERENCES

(1) Clayton, T. A.; Lindon, J. C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Provost, J. P.; Le Net, J. L.; Baker, D.; Walley, R. J.; Everett, J. R.; Nicholson, J. K. *Nature* **2006**, *440*, 1073–1077.

Anal. Chem. **2010**, *82*, 2303–2309

^{13}C-Formylation for Improved Nuclear Magnetic Resonance Profiling of Amino Metabolites in Biofluids

Tao Ye,[†] Shucha Zhang,[†] Huaping Mo,[‡,§] Fariba Tayyari,[†] G. A. Nagana Gowda,[†] and Daniel Raftery*[,†]

Department of Chemistry, Purdue Inter-Departmental NMR Facility, and Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47907

An increased interest in metabolite profiling is driving the need for improved analytical techniques with greater performance for a variety of important applications. Despite their limited sensitivity, nuclear magnetic resonance (NMR) methods are attractive because of their simplicity, reproducibility, quantitative nature, and wide applicability. The use of chemoselective isotopic tags has the potential to advance the application of NMR for analyzing metabolites in complex biofluids by allowing detection of metabolites down to the low micromoalr level with high resolution and specificity. Here, we report a new ^{13}C-tagging method using ^{13}C-formic acid that delivers high sensitivity, good quantitation, and excellent resolution for ^{1}H−^{13}C 2D NMR profiling of amino metabolites. High reproducibility (coefficient of variation (CV) = 2%) was observed for metabolites in urine with concentrations down to 10 μM. As amino compounds comprise an important class of metabolites and small molecules of biological roles, this new method therefore should be amenable to a variety of applications.

The study of the metabolite profile or metabolome can reveal the status of biological systems from a variety of perspectives, including insights into the normal and abnormal (or disease) states of an organism. For example, concentrations and fluxes of metabolites are constrained to a certain extent by their metabolic pathways and related enzymes under homeostasis, but these metabolite quantities are also subject to stresses such as environmental factors and disease. A quantitative measurement of the metabolome therefore provides an important snapshot of the ongoing normal and abnormal processes in complex biological systems.[1−7] A more accurate, precise, and rapid profile of the metabolome in an organism will lead to a better understanding of the systems biology and potentially a series of biomarkers that can be used for a variety of practical purposes, including diagnostics, drug development, nutrition, and environmental studies.[8−17] As the importance of metabolic profiling has grown, the need for advanced analytical methods that deliver higher sensitivity, resolution, and throughput has become recognized.

It is believed that most eukaryotic organisms possess at least 3000−5000 metabolites with very diversified molecular structures and physical/chemical properties.[5] The quantitative determination of so many compounds in a single analysis remains out of reach for current technologies. Only a small fraction, normally the most abundant species, of the metabolites can currently be accurately and precisely detected. The information derived from this small sampling of metabolites is very often too superficial and nonspecific to reveal enough biochemical detail. Alternatively, one can measure a larger number of metabolites semiquantitatively; however, this approach is often limiting as well.

The primary analytical methods used for metabolic profiling are mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) due to their ability to provide rich information of complex mixtures at high throughput.[18] MS is very attractive because of its high sensitivity, experimental flexibility, and ability to determine unknown molecules. While a high number of metabolites can be detected and identified by MS (typically a few hundred), far fewer can be quantified to better than 10% coefficient of variation (CV) on a routine basis because of ion suppression and other matrix effects. Compared to MS, NMR is much less sensitive but produces data that is more easily reproduced and quantified using a single internal standard. The

* To whom correspondence should be addressed. E-mail: raftery@purdue.edu. Phone: (765) 494-6070. Fax: (765) 494-0239.
[†] Department of Chemistry.
[‡] Purdue Inter-Departmental NMR Facility.
[§] Department of Medicinal Chemistry and Molecular Pharmacology.

(1) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–1189.
(2) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157–1161.
(3) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.
(4) Kell, D. B. *Curr. Opin. Microbiol.* **2004**, *7*, 296–307.
(5) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 763–769.
(6) van der Greef, J.; Smilde, A. K. *J. Chemom.* **2005**, *19*, 376–386.
(7) Van Dien, S.; Schilling, C. H. *Mol. Syst. Biol.* **2006**, *2*, 2006.0035.
(8) Serkova, N. J.; Niemann, C. U. *Expert Rev. Mol. Diagn.* **2006**, *6*, 717–731.
(9) Griffin, J. L. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2006**, *361*, 147–161.
(10) Gowda, G. A. N.; Zhang, S. C.; Gu, H. W.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Mol. Diagn.* **2008**, *8*, 617–633.
(11) Zhang, S.; Gowda, G. A. N.; Asiago, V.; Shanaiah, N.; Barbas, C.; Raftery, D. *Anal. Biochem.* **2008**, *383*, 76–84.
(12) Bedair, M.; Sumer, L. W. *TrAC, Trends Anal. Chem.* **2008**, *27*, 238–250.
(13) Bundy, J. G.; Davey, M. P.; Viant, M. R. *Metabolomics* **2009**, *5*, 3–21.
(14) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *FEBS J.* **2007**, *274*, 1140–1151.
(15) Robertson, D. G. *Toxicol. Sci.* **2005**, *85*, 809–822.
(16) Rezzi, S.; Ramadan, Z.; Fay, L. B.; Kochhar, S. *J. Proteome Res.* **2007**, *6*, 513–525.
(17) Viant, M. R. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943–948.
(18) Pan, Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, *387*, 525–527.

10.1021/ac9024818
Published on Web 02/24/2010

same nuclei (usually ^{1}H) are detected with the same sensitivity in one NMR experiment, irrespective of molecular origin.

Nevertheless, serious impediments to more widespread use of NMR for quantitative metabolite profiling, especially for lower concentration species, are the severe spectral overlap and complexity. The signals of metabolites with low concentrations are often buried under nearby strong signals due to insufficient resolution. These species become nondetectable unless selective methods[19,20] or multidimensional experiments[21,22] are used. Methodological innovations that significantly improve the resolving power of NMR would reduce a current bottleneck and greatly advance the application of NMR in metabolic profiling.

Toward this goal, we recently introduced several methods for in vitro chemical tagging of biofluids.[23−25] A molecular tag containing an enriched or high-abundance isotope such as $^{13}C/^{15}N$ or ^{31}P can be introduced to the biofluid; it selectively reacts with metabolites with a certain functional group and therefore tags these metabolites with that isotope. The subsequent heteronuclear 2D NMR experiment selectively detects the tags and provides a simple spectrum free of the background signals from the rest of the tagged molecule as well as the untagged molecules. Instead of the crowded and coupled signals from all the protons into one dimension of ~10 ppm, one can easily resolve a smaller number of signals (each metabolite has only one cross-peak unless multiply tagged) in two dimensions. Since every metabolite molecule has at least one functional group in order to function in its involved biological processes, one can in principle use different isotopic tags to profile all metabolites in selective classes. The improved resolving power comes from the high-resolution heteronuclear 2D NMR, the chemoselective tagging, and the reduced complexity and number of signals. It should be noted that chemoselective tags including the isotopic variant tags have also been successfully adopted for metabolic or proteomic profiling with MS for separation or quantification purposes.[26−30] Previously we have reported the ^{13}C-tagging of amino metabolites with ^{13}C-acetic anhydride for $^{1}H-^{13}C$ heteronuclear single quantum coherence (HSQC) 2D NMR profiling,[23] ^{15}N-tagging of carboxyl-containing metabolites with ^{15}N-ethanolamine for $^{1}H-^{15}N$ HSQC 2D NMR profiling,[24] and ^{31}P-tagging of active hydrogen moieties in lipid extracts with 2-chloro-4,4,5,5,-tetramethyldioxaphospholane for ^{31}P NMR analysis.[25] Because of the closeness of the tag to the metabolite molecule and the relatively strong $^{1}H-^{15}N$ J-coupling, the use of ^{15}N-ethanolamine has been especially successful for profiling over 100 carboxyl-containing metabolites in biofluids with 2D NMR. Here, ^{13}C-formic acid is applied as the isotopic tag to further improve the performance of 2D NMR in profiling amino metabolites. Amino acids are the most important and common amino metabolites; they are not only the building blocks for proteins but also the precursors for nucleotides,[31] and they provide an energy source through transamination, the urea cycle, the citric acid cycle, and gluconeogenesis.[32−34] Other common amino metabolites include derivatives of amino acids, taurine, dimethylamine, methylamine, and many neurotransmitters such as dopamine, serotonin, and histamine. Drugs such as amphetamine, procaine, rimantadine, and their metabolites also belong to this group of compounds.

EXPERIMENTAL SECTION

Chemicals and Biological Samples. All metabolite standards, *N,N*-dicyclohexylcarbodiimide (DCC), *N*-hydroxysuccinimide (HOSu) (Sigma-Aldrich), and ^{13}C-formic acid (Cambridge Isotope Laboratories) were purchased and used without further purification. Human serum and urine samples were obtained either from commercial sources (Innovative Research, Novi, MI) or from healthy volunteers in accordance with the Institutional Review Board at Purdue University.

General Procedure for ^{13}C-Formylation. ^{13}C-formic acid (2 μL, 0.05 mmol) and *N*-hydroxysuccinimide (5 mg, 0.04 mmol) were dissolved in tetrahydrofuran (100 μL). *N,N*-Dicyclohexylcarbodiimide (9 mg, 0.04 mmol) in tetrahydrofuran (50 μL) was added to the mixture and stirred at room temperature.[35] After 15 min, the reaction mixture was centrifuged to remove the insoluble urea; the supernatant containing $^{13}C-N$-formyloxysuccinimide was then added to the biofluid sample (500 μL) along with 2 M $NaHCO_3$ (50 μL, 0.1 mmol) aqueous solution. The reaction mixture was stirred at room temperature for 4 h and dried under vacuum. After redispersing the dried product mixture in D_2O (500 μL), the pH of the solution was adjusted to 7.0 by adding 1 M HCl and then transferred to a standard 5 mm NMR tube for analysis. $^{13}C-N$-formyloxysuccinimide can also be purified by recrystallization in ethanol[31] and used for the tagging reaction instead of in situ generation.

NMR Spectroscopy and Data Processing. NMR experiments were carried out at 298 K on a Bruker Avance-III-800 spectrometer equipped with a room temperature ^{1}H inverse detection Z-gradient probe. ^{1}H NMR spectra were obtained using the water Pre-SAT180 sequence.[36] The sensitivity-enhanced $^{1}H-^{13}C$ 2D HSQC experiments employed an INEPT transfer delay of 2.5 ms corresponding to a $^{1}J_{C-H}$ of 200 Hz. Spectral widths were approximately 10 kHz for the ^{1}H dimension and 3 kHz for ^{13}C. A total of 128 or 256 free induction decays of 2 048 data points each were collected in the indirect (t_1) dimension using 4 or 8 transients per increment. ^{13}C

(19) Sandusky, P.; Raftery, D. *Anal. Chem.* **2005**, *77*, 7717–7723.
(20) Sandusky, P.; Raftery, D. *Anal. Chem.* **2005**, *77*, 2455–2463.
(21) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385–9390.
(22) Chikayama, E.; Suto, M.; Nishihara, T.; Shinozaki, K.; Kikuchi, J. *PLoS One* **2008**, *3*, e3805.
(23) Shanaiah, N.; Desilva, M. A.; Gowda, G. A. N.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11540–11544.
(24) Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009**, *81*, 4882–4888.
(25) DeSilva, M. A.; Shanaiah, N.; Gowda, G. A. N.; Rosa-Pérez, K.; Hanson, B. A.; Raftery, D. *Magn. Reson. Chem.* **2009**, *47*, S74–S80.
(26) Carlson, E. E.; Cravatt, B. F. *Nat. Methods* **2007**, *4*, 429–435.
(27) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660–666.
(28) Brittain, S. M.; Ficarro, S. B.; Brock, A.; Peters, E. C. *Nat. Biotechnol.* **2005**, *23*, 463–468.
(29) Lamos, S. M.; Shortreed, M. R.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. *Anal. Chem.* **2007**, *79*, 5143–5149.
(30) Huang, X.; Regnier, F. E. *Anal. Chem.* **2008**, *80*, 107–114.
(31) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; W. H. Freeman & Co.: New York, 2002; pp 693–698.
(32) Sakami, W.; Harrington, H. *Annu. Rev. Biochem.* **1963**, *32*, 355–398.
(33) Brosnan, J. T. *J. Nutr.* **2000**, *130*, 988S–990S.
(34) Young, V. R.; Ajami, A. M. *J. Nutr.* **2001**, *131*, 2449S–2459S.
(35) Hecht, S. M.; Werner, D. *J. Chem. Soc., Perkin Trans. 1* **1973**, 1903–1906.
(36) Mo, H.; Raftery, D. *J. Magn. Reson.* **2008**, *190*, 1–6.

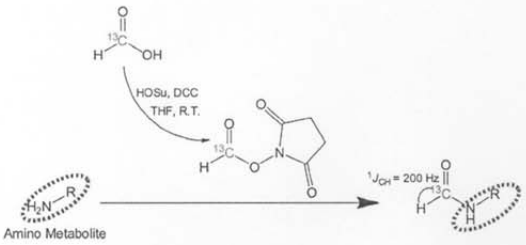


Figure 1. Amino metabolites in biofluids are isotopically tagged by ^{13}C-formylation. The tagged metabolites are subsequently detected using sensitivity-enhanced 2D ^{1}H−^{13}C correlation NMR.

decoupling during the direct detection dimension (t_2) was achieved with the globally optimized alternating-phase rectangular pulses (GARP) sequence. The resulting 2D data were zero-filled to 1 024 points in the t_1 dimension after forward linear prediction to 512 points. A 45°-shifted sinebell window function was then applied to both dimensions before Fourier transformation. NMR data were processed using Bruker Topspin 3.0 spectrometer software on a Redhat Linux platform. The reproducibility of the method was evaluated using peak volumes generated through the automatic peak picking routine of Topspin; the 2D signals in three spectra were then compared individually after normalization with respect to the most intense signal (L-alanine) in order to reduce errors from volumetric measurements and probe tuning.

Calibration. Stock solutions of nine standard compounds were made individually at 50 mM with their precise concentrations determined using 1D ^{1}H NMR against a calibrated sodium 2,2-dimethyl-2-silapentane-5-sulfonate-d_6 (DSS) solution. The individual stock solutions were then mixed at equal volumes and diluted with water to make a mixed stock solution at 4 mM, which was further diluted to make a calibration series ranging from 5 μM to 2 mM. A fixed amount of ethanolamine (0.5 μmol, 1 mM internal reference) was added to 500 μL of each calibration solution to make the final calibration samples, which were then subjected to ^{13}C-formylation and ^{1}H−^{13}C HSQC analysis. The 2D peaks were integrated and referenced to the integration volume of ^{13}C-formylated ethanolamine and plotted against the concentrations previously determined by 1D ^{1}H NMR.

RESULTS

After conversion to the active NHS (*N*-hydroxysuccinimidyl) ester, ^{13}C-formic acid readily reacted with amino metabolites and introduced a ^{1}H−^{13}C pair with a one-bond *J*-coupling of 200 Hz into the metabolite molecule as shown in Figure 1. The newly introduced HCO-moiety was highly polar and therefore retained good solubility in water. An HSQC experiment was then applied to provide a highly resolved spectrum of the ^{1}H−^{13}C pairs in the tagged metabolite molecules. The stuctural variations of the metabolites led to a good dispersion of the 2D signals in both ^{1}H and ^{13}C dimensions. At the same time, the *J*-couplings and relaxation rates do not vary significantly across different metabolites, which is essential for unbiased detection of all tagged amino metabolites.

The performance of the method was first evaluated with standard compounds that are commonly seen in biofluids. A

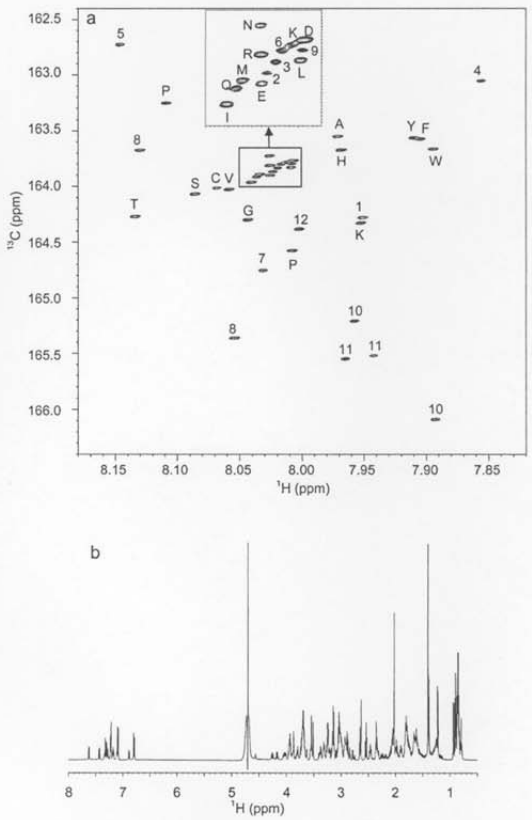


Figure 2. (a) The ^{1}H−^{13}C HSQC spectrum of a mixture of 32 standard amino compounds after ^{13}C-formylation (pH 7.0, 298 K). Inset shows the expanded view of the boxed region (^{1}H, 8.00−8.05 ppm; ^{13}C, 163.65−164.00 ppm). (b) 1D ^{1}H NMR spectrum of the same mixture in water before ^{13}C-formylation (pH 7.0, 298 K).

mixture of 32 standard compounds with amino groups was isotopically tagged by ^{13}C-formylation and analyzed using a ^{1}H−^{13}C HSQC experiment at 800 MHz. The resulting spectrum is shown in Figure 2 with peak assignments, and the chemical shifts for signals of the tagged standard compounds are listed in Table 1. Each compound has one signal in the ^{1}H−^{13}C 2D spectrum except for metabolites such as lysine and spermidine that have more than one unique amino group. Proline, 4-hydroxyproline, and sarcosine also have two signals due to the cis- and trans-isomers formed by ^{13}C-formylation. The signals of the ^{13}C-fomylated compounds are dispersed over 4 ppm in the ^{13}C dimension and over 0.3 ppm in the ^{1}H dimension with good spectral separation.[23] By comparison, the 1D ^{1}H NMR detection of the same mixture before ^{13}C-tagging produces a complex spectrum with significant signal overlap (Figure 2b).

Because of the reproducibility and good sensitivity due to the strong *J*-coupling between ^{1}H and ^{13}C, HSQC detection of the ^{13}C-formylated amines is quantitative as shown in Figure 3. Good linearity ($R^2 > 0.995$) and quantitation (average deviation from the trend line for the 9 standard compounds was 6% for concentrations above 10 μM and 12% for concentrations below

Table 1. Chemical Shifts for 32 ^{13}C-Formylated Amino Metabolites Standard Compounds (D_2O, pH 7.0, 298 K)

label	name	^{1}H (ppm)	^{13}C (ppm)	label	name	^{1}H (ppm)	^{13}C (ppm)
A	L-alanine	7.972	163.521	L	L-leucine	8.009	163.803
1	β-alanine	7.951	164.252	K	L-lysine	8.013	163.754
2	L-2-aminoadipic acid	8.023	163.841			7.953	164.299
3	L-2-aminobutyric acid	8.020	163.807	M	L-methionine	8.034	163.862
4	L-2-aminoisobutyric acid	7.856	163.025	9	L-norleucine	8.008	163.771
5	4-aminophenol	8.147	162.699	F	L-phenylalanine	7.905	163.544
R	L-arginine	8.026	163.783	P	L-proline	8.109	163.224
N	L-asparagine	8.026	163.695			8.008	164.545
D	L-aspartic acid	8.007	163.738	10	sarcosine	7.958	165.178
6	L-citrulline	8.017	163.772			7.893	166.057
C	L-cysteine	8.069	163.986	S	L-serine	8.086	164.042
7	ethanolamine	8.032	164.723	11	spermidine	7.965	165.521
E	L-glutamic acid	8.026	163.872			7.942	165.494
Q	L-glutamine	8.037	163.887	12	taurine	8.002	164.353
G	glycine	8.044	164.272	T	L-threonine	8.134	164.24
H	L-histidine	7.964	163.644	W	L-tryptophan	7.895	163.635
8	4-hydroxy-L-proline	8.130	163.645	Y	L-tyrosine	7.910	163.534
		8.054	165.331	V	L-valine	8.059	164.000
I	L-isoleucine	8.041	163.935				

10 μM) was observed using the integrated 2D signal intensities for the standard compounds over a large range of concentrations. The strong *J*-coupling also ensures sensitive detection of trace amounts of metabolites. The ^{1}H−^{13}C HSQC experiment run at 800 MHz can detect ^{13}C-formylated alanine with a concentration down to 4 μM (SNR ~4) within 8 min (1 scan, 128 increments). Though conventional 1D ^{1}H NMR can detect substances at low micromolar levels, its performance is often reduced by ^{1}H−^{1}H couplings and spectral overlap in complex samples, as well as residual solvent signals. As a result, ^{1}H NMR is normally utilized for profiling metabolites above 100 μM. The improved sensitivity and excellent detection linearity for the ^{13}C-formylated amines result in part from the relatively large 200 Hz J_{C-H} coupling between the ^{1}H and ^{13}C of the formyl tag.

^{13}C-formylation was then applied to profile amino metabolites in human biofluid samples. More than 40 signals, mostly of amino acids, were clearly seen and well resolved from a human serum sample as shown in Figure 4a. The urine spectrum (Figure 4b) showed not only the presence of most amino acids that were found in serum at different levels but also additional signals representing unidentified amino metabolites. However, as a collection of biowastes and byproducts rather than a circulating body fluid, urine has a much greater variation in chemical content compared to blood.[37] Very often the urinary metabolites are reflections of what has been introduced into the

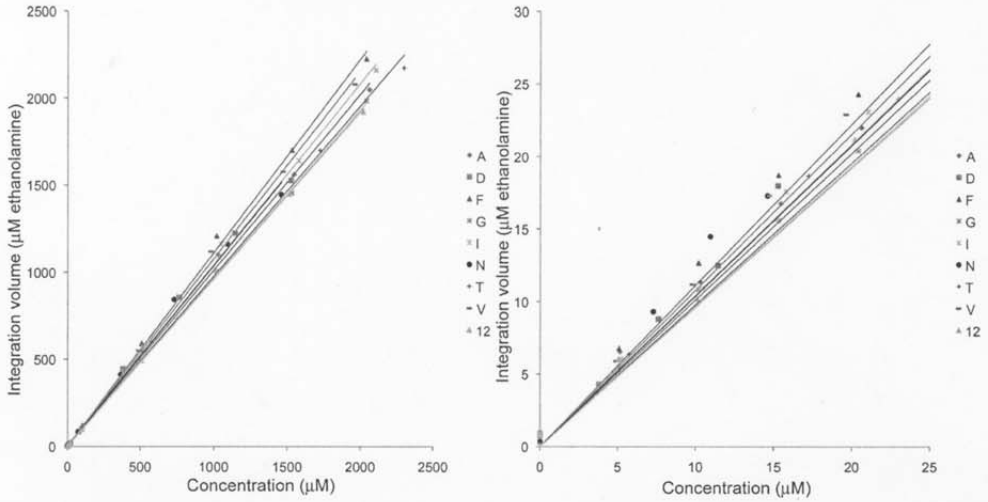


Figure 3. Calibration curves for nine standard amino compounds detected by ^{1}H−^{13}C HSQC after ^{13}C-formylation, with an expanded view of the 0−25 μM concentration range shown in the right panel. The 2D peaks were integrated and referenced to the integration volume of ^{13}C-formylated ethanolamine (internal standard) and plotted against the concentrations previously determined by 1D ^{1}H NMR. The integration volume of ^{13}C-formylated 1 μM ethanolamine serves as the unit for the vertical axis. The trend lines were generated by linear regression; R^2 values were greater than 0.995 for all 9 compounds. Good quantitation was observed: the average deviation from the calibration trend line is 6% for concentrations above 10 μM and 12% for concentrations below 10 μM.

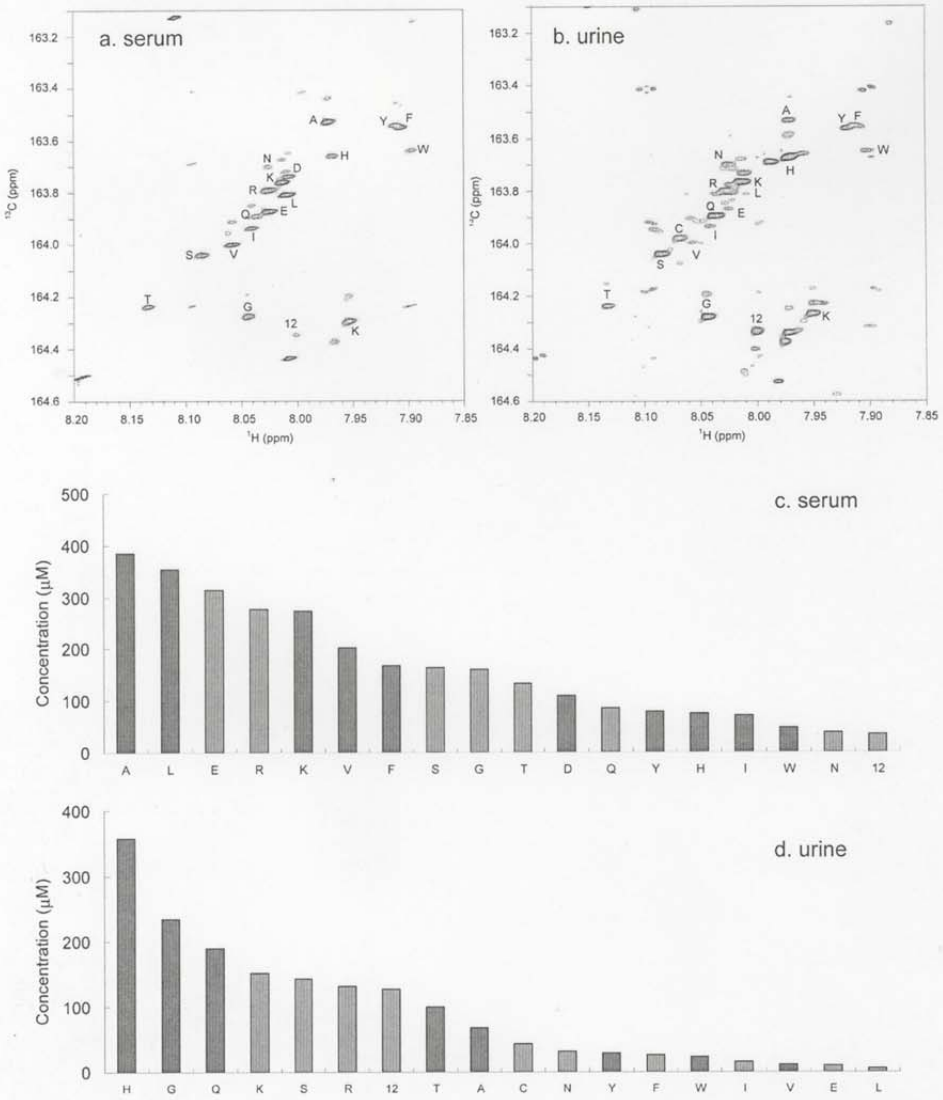


Figure 4. ^{1}H−^{13}C HSQC spectra of a healthy human serum sample (a) and a healthy human urine sample (b) after ^{13}C-isotopic tagging with ^{13}C-formic acid. A total of 18 metabolites were indentified and quantified in each spectrum. Concentrations for amino metabolites identified and quantified using the 2D ^{1}H−^{13}C HSQC spectra for the serum sample (c) and the urine sample (d) after ^{13}C-formylation. Metabolites that have resolved 1D ^{1}H signals are marked in red and others are marked in light blue.

human body. Therefore, the extra signals in the studied urine spectrum could arise from amino compounds generated via the metabolism of nitrogen-containing dietary intake.

The assignment of 2D ^{1}H−^{13}C signals could be accomplished by simply comparing the ^{1}H and ^{13}C chemical shifts of a particular peak with the library of 32 compounds (Table 1). This is in contrast to the sometimes challenging and often time-consuming assignment of low level 1D ^{1}H NMR signals in complex samples that involves checking chemical shifts, coupling constants, integrations, or even spiking standard compounds. Quantification of these amino metabolites is straightforward as the integration volume of a 2D NMR signal is proportional to the number of protons. A total of 18 amino metabolites were identified and quantified in both serum and urine samples with their concentrations shown in parts c and d of Figure 4, respectively. Taurine and 16 amino acids were identified from both samples; in addition, aspartic acid was quantified in the serum and cysteine was quantified in the urine. In general, the observed metabolite concentrations in serum were within the healthy concentration ranges appearing in the Human

(37) Saude, E. J.; Adamko, D.; Rowe, B. H.; Marrie, T.; Sykes, B. D. *Metabolomics* **2007**, *3*, 439–451.

Metabolome Database.[38] Profiling of amino metabolites using 1D ^{1}H spectra was challenging especially for metabolites of low concentrations that appeared in crowded spectral regions. As a result, more than half of the amino-metabolites profiled using the ^{13}C tagging and 2D detection approach could not be identified and quantified using 1D ^{1}H detection (parts c and d of Figure 4 and Figures S1 and S2 in the Supporting Information). Most of the signals that were clearly resolved and quantifiable in the 1D ^{1}H spectrum had relatively high intensities and appeared either in the methyl region (~1−2 ppm) where signals had narrow line widths and simple splitting patterns or the aromatic region (~6−8 ppm) where fewer signals were present.

The reproducibility of the method was tested in triplicate experiments on a split sample of the same urine. As shown in Figures S3 and S4 in the Supporting Information, the three spectra are very similar. Peak integrations indicated that the average coefficient of variation (CV) was 2% for 12 metabolites above 10 μM and 11% for 6 metabolites below 10 μM (signal to noise ratios (SNR) are less than 8 for 3 metabolites below 4 μM). Again, only 7 of the 18 amino metabolites had recognizable 1D ^{1}H NMR signals (Figure S5 in the Supporting Information). This high reproducibility results from the quantitative nature of both the NMR detection and the ^{13}C-tagging reaction, which is essential for metabolic profiling.

DISCUSSION

Successful metabolic profiling methods rely on the sensitive and unbiased detection of numerous metabolites. The combination of ^{13}C-formyl tagging and 2D ^{1}H−^{13}C HSQC detection provides a sensitive, unbiased method with excellent quantitation and resolution. In the ^{13}C-formylated metabolite molecule, the large one-bond *J*-coupling (200 Hz) between the labeled ^{13}C and its attached ^{1}H significantly improves polarization transfer that is essential for the HSQC experiments. The Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) delay of 2.5 ms in the current method is significantly shorter than the 83 ms delay (corresponding to the $^2J_{C-H}$ of 6 Hz) that was used in our previous work involving acetic anhydride tagging.[23] The faster polarization transfer significantly reduces polarization losses due to relaxation or competing transfer pathways involving $^2J_{C-H}$ or $^3J_{C-H}$ of 2−4 Hz and therefore leads to better detection sensitivity, reproducibility, and linearity.

To further improve the resolving ability of ^{1}H−^{13}C HSQC 2D NMR for metabolic profiling, one can improve the performance of the spectrometer (i.e., using higher magnetic fields, more increments in the indirect ^{13}C dimension), or make the 2D signals disperse in a wider region of chemical shifts across different metabolites. Chemical shifts originate from the impact of varying local electronic environments on the resonance frequency of the nuclear spin. For the ^{13}C-tagged molecules, the chemical shifts of the 2D ^{1}H−^{13}C signals vary because of the structural differences among metabolites. The close distance between the isotope labeled ^{1}H−^{13}C nuclear pair and the rest of the metabolite leads to an increase in the chemical shift sensitivity of the tag, a better dispersion of signals in two dimensions, and less chance for spectral overlap. In addition, the tagging reaction is now carried out in a simple mix-and-stir procedure that is easy to repeat using the NHS ester. The reaction time for this mild tagging reaction is longer than the previous acetylation using acetic anhydride; however, no acetic acid is generated during the reaction and there is no need to control the pH by addition of NaOH throughout the reaction. Therefore both the time and labor for setting up one reaction are reduced. Moreover ^{13}C-enriched formic acid is relatively inexpensive and has a long shelf life, making the method economically attractive.

Though the assignments of 2D signals for known metabolites can be easily accomplished by comparing chemical shifts to the library, the identification of unknown signal remains challenging. Efforts are currently underway to expand this library and to identify unassigned peaks in the spectrum using high-performance liquid chromatography (HPLC)−NMR and microcoil NMR approaches.[39−41] While 1D ^{1}H NMR metabolic profiling primarily emphasizes signals from highly concentrated (millimolar) species that are often too nonspecific to be used as biomarkers, the new method allows the quantitative NMR profiling of metabolites at lower levels (micromolar) that should lead to new findings in future applications. The limit of detection may be further improved using preconcentration methods and microcoil NMR,[39−41] or potentially new technologies such as dynamic nuclear polarization[42] and para-hydrogen induced polarization.[43] Though 2D NMR experiments can require longer times than simple 1D NMR experiments, the experiments here could be conducted in only 8 min per sample. In addition, advanced data collection and processing algorithms such as ultrafast 2D NMR,[44] covariance NMR,[45] nonlinear sampling, and forward maximum entropy reconstruction[46] can greatly reduce the acquisition time for 2D NMR experiments and allow high-throughput operation.

In conclusion, a new method to incorporate ^{13}C nuclei into metabolites in vitro has been introduced for conducting quantitative 2D NMR metabolic profiling with improved resolution and sensitivity stemming from the structure of ^{13}C-formic acid. The performance of method has been demonstrated with a mixture of 32 standard compounds, human urine, and serum. High reproducibility and detection linearity were observed for metabolites down to 10 μM; therefore, we anticipate that this method will be a useful tool for quantitative profiling of low-concentration amino metabolites in complex mixtures. As amino-containing compounds constitute an important class of

(38) Wishart, D.; Knox, C.; Guo, A.; et al. *Nucleic Acids Res.* **2009**, *37*, D603–610.

(39) Djukovic, D.; Liu, S.; Henry, I.; Tobias, B.; Raftery, D. *Anal. Chem.* **2006**, *78*, 7154–7160.

(40) Djukovic, E.; Appiah-Amponsah, E.; Shanaiah, N.; Gowda, G. A. N.; Henry, I.; Everly, M.; Tobias, B.; Raftery, D. *J. Pharm. Biomed. Anal.* **2008**, *47*, 328–334.

(41) Kc, R.; Henry, I.; Park, G. H. J.; Raftery, D. *J. Magn. Reson.* **2009**, *197*, 186–192.

(42) Wilson, D. M.; Hurd, R. E.; Keshari, K.; Van Criekinge, M.; Chen, A. P.; Nelson, S. J.; Vigneron, D. B.; Kurhanewicz, J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5503–5507.

(43) Adams, R. W.; Aguilar, J. A.; Atkinson, K. D.; Cowley, M. J.; Elliott, P. I. P.; Duckett, S. B.; Green, G. G. R.; Khazal, I. G.; López-Serrano, J.; Williamson, D. C. *Science* **2009**, *323*, 1708–1711.

(44) Frydman, L.; Blazina, D. *Nat. Phys.* **2007**, *3*, 415–419.

(45) Zhang, F. L.; Brüschweiler, R. *J. Am. Chem. Soc.* **2004**, *126*, 13180–13181.

(46) Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luithardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K.; Falchuk, K. H.; Halperin, J. A.; Wagner, G. *J. Am. Chem. Soc.* **2007**, *129*, 5108–5116.

molecules in biological processes, this new method is expected to find a number of applications in the quantitative analysis of these molecules.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health Grant 1 R01GM085291-01. D.R. is a member of the Purdue Center for Cancer Research and Oncological Sciences Center.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review November 1, 2009. Accepted February 11, 2010.

AC9024818

MOLECULAR CARCINOGENESIS

Altered Glucose Metabolism in Harvey-*ras* Transformed MCF10A Cells

Wei Zheng,[1] Fariba Tayyari,[2] G.A. Nagana Gowda,[2] Daniel Raftery,[2] Eric S. McLamore,[3] D. Marshall Porterfield,[3,4,5,6] Shawn S. Donkin,[1] Brian Bequette,[7] and Dorothy Teegarden[1]*

[1]*Interdepartmental Nutrition Program, Purdue University, West Lafayette, Indiana 47906*
[2]*Department of Chemistry, Purdue University, West Lafayette, Indiana 47906*
[3]*Birck-Bindley Physiological Sensing Facility, Purdue University, West Lafayette, Indiana 47906*
[4]*Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47906*
[5]*Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana 47906*
[6]*Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana 47906*
[7]*Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742*

Metabolic reprogramming that alters the utilization of glucose including the "Warburg effect" is critical in the development of a tumorigenic phenotype. However, the effects of the Harvey-*ras* (H-*ras*) oncogene on cellular energy metabolism during mammary carcinogenesis are not known. The purpose of this study was to determine the effect of H-*ras* transformation on glucose metabolism using the untransformed MCF10A and H-*ras* oncogene transfected (MCF10A-*ras*) human breast epithelial cells, a model for early breast cancer progression. We measured the metabolite fluxes at the cell membrane by a selective micro-biosensor, [$^{13}C_6$]glucose flux by ^{13}C-mass isotopomer distribution analysis of media metabolites, intracellular metabolite levels by NMR, and gene expression of glucose metabolism enzymes by quantitative PCR. Results from these studies indicated that MCF10A-*ras* cells exhibited enhanced glycolytic activity and lactate production, decreased glucose flux through the tricarboxylic acid (TCA) cycle, as well as an increase in the utilization of glucose in the pentose phosphate pathway (PPP). These results provide evidence for a role of H-*ras* oncogene in the metabolic reprogramming of MCF10A cells during early mammary carcinogenesis.

Key words: glucose; breast cancer; ras; metabolism

INTRODUCTION

The increased glycolytic metabolism and pyruvate oxidative phosphorylation noted in tumors termed the "Warburg effect" was described decades ago [1]. This shift in cellular metabolism describes an increased glucose uptake and a shift of the pyruvate oxidative phosphorylation in the mitochondria towards a more rapid aerobic glycolysis even in a normoxic environment, and increased conversion of pyruvate to lactate [1]. The resulting lactate may also serve as an energy source for tumor cells [2]. It is now known that this metabolic reprogramming also occurs in proliferating cells [3]. These increased rates of glucose uptake and metabolism provide an advantage to proliferating and cancer cells by favoring utilization of the most abundant energy and carbon sources. For example, glucose metabolism yields ribose for nucleic acid synthesis and NADPH through the pentose phosphate pathway (PPP) while greater glycolysis provides intermediates to maintain anaplerosis and supply biosynthetic intermediates [3]. The biological importance of this metabolic shift is supported by the high rate of glycolysis and an over expression of glucose transporters and glycolytic enzymes in many types of solid tumors [4]. Deprivation of glucose can induce oxidative stress and other defects in metabolism which leads to cancer cell apoptosis [5,6]. Further, inhibitors of glucose transporters and glycolytic inhibitors have been implemented as effective anticancer treatments and can also sensitize the tumor cells to other chemotherapeutic drugs [7,8]. Unlike in the normal proliferating cells, such metabolic reprogramming in cancer cells is controlled by oncogenes which lead to the growth factor independent, chronic activation of the proliferative pathways [3].

The Ras subfamily of proteins is a group of small GTPases which serve as an important effector essential for the signal transduction induced by

---

The present address of G.A. Nagana Gowda and Daniel Raftery are Mitochondria and Metabolism Center, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA 98109.

Grant sponsor: National Institutes of Health, National Cancer Institute; Grant number: R25CA128770; Grant sponsor: Cancer Prevention Internship Program; Grant sponsor: Oncological Sciences Center; Grant sponsor: Discovery Learning Research Center at Purdue University; Grant sponsor: NIH; Grant number: 1R01GM085291

*Correspondence to: Interdepartmental Nutrition Program, Purdue University, West Lafayette, IN 47906.

Received 14 December 2012; Revised 27 April 2013; Accepted 26 July 2013

DOI 10.1002/mc.22079

Published online in Wiley Online Library (wileyonlinelibrary.com).

numerous growth factors to stimulate cell proliferation. The *ras* proto-oncogene is frequently mutated in cancers [9,10] and affects a variety of processes involved in cancer progression. The oncogenic *ras* drives cellular proliferation in the transformed cells by promoting pro-growth and inhibiting anti-growth signals in a growth factor independent manner [9]. Although mutations in the *ras* gene are not common in breast cancers [11,12], Ras may be pathologically activated in breast cancer by overexpression of growth factor receptors signaling through Ras such as the ErbB2 receptor, which is activated in 30% of breast cancers [13,14]. Harvey-*ras* (H-*ras*)-induced tumors are characterized by activation of mitogen-activated protein kinase signaling [15] and is associated with early neoplasia and poor prognosis [13,16]. Although K-*ras* transfection has been shown to alter cellular metabolism in fibroblast cells [17], the impact of H-*ras* in epithelial cells in models representative of early progression has not been studied.

The purpose of the current study was to determine the effect of the Harvey-*ras* oncogene (H-*ras*) on cellular energy metabolism in untransformed MCF10A and H-*ras* transfected MCF10A (MCF10A-*ras*) human breast epithelial cells, which serve as a model for studying early mammary carcinogenesis. The hypothesis of the study is that MCF10A-*ras* cells have increased glycolytic activity and lactate production as well as reduced flux through the tricarboxylic acid (TCA) cycle. These results will contribute to understanding the effect of H-*ras* on the regulation of cellular energy metabolism during early breast cancer progression.

MATERIALS AND METHODS

Chemicals and Reagents

Dulbecco's modified Eagle medium (DMEM/F12), horse serum, trypsin and penicillin/streptomycin were obtained from Life Technologies, Gibco-BRL (Rockville, MD). Cholera toxin was purchased from Calbiochem (Darmstadt, Germany). Protein assay reagents were obtained from Pierce (Rockford, IL). Protease inhibitors cocktail, trypan blue, insulin, epidermal growth factor, and hydrocortisone were purchased from Sigma (St. Louis, MO). All reagents for gas chromatography-mass spectrometry (GC–MS) analyses were from Pierce. D-[$^{13}C_6$]Glucose was purchased from Cambridge Isotope labs (Woburn, MA). Mass spectrometry analysis confirmed its chemical and isotopic purity (92.7% [$^{13}C_6$]glucose and 6.9% [$^{13}C_5$]glucose).

Cell Culture

MCF10A human breast epithelial cells and MCF10A-*ras* cells were a gift from Dr. Michael Kinch, Purdue University. The phenotypes of the two cell lines which were originally derived from human fibrocystic mammary tissue have been well characterized in the literature. The MCF10A cells are spontaneously immortalized but otherwise normal, which do not form colonies in soft agar or grow in immunocompromised mice [18], but undergo a well-defined program of proliferation and differentiation in three-dimensional (3-D) reconstituted basement membrane culture, forming acinar structures that recapitulate many aspects of mammary architecture *in vivo* [19]. The MCF10A-*ras* cells were premalignant breast epithelial cells generated by transfecting the MCF10A cells with constitutively active T24 Harvey-*ras* oncogene. They can form complex multi-acinar structures that produce a basement membrane but undergo delayed cell cycle arrest and have incomplete luminal development when grown in 3-D culture [19]. Therefore, these two cell lines with the same genetic background serve as a unique model to represent early breast cancer progression. The MCF10A and MCF10A-*ras* cells were cultured in DMEM/F12 (1:1) containing 5% horse serum and supplemented with 10 mg/L insulin, 20 μg/L epidermal growth factor, 50 μg/L cholera toxin, 50 mg/L hydrocortisone, 100 units/mL penicillin, and 0.1 mg/mL streptomycin in a humidified environment at 37°C with 5% CO_2. Cells were maintained in fresh media changed every 24 h for 4 d before measurement or harvest.

RNA Isolation and Analysis

RNA was isolated with TriReagent (Molecular Research Center, Cincinnati, OH) following the manufacturer's instructions. Reverse transcription of total RNA was performed using MMLV reverse transcriptase (Promega, Madison, WI). Real-time quantitative PCR was performed using the Brilliant II SYBR Green QPCR Master Mix (Agilent, Santa Clara, CA). The mRNA abundances of enzymes involved in glucose metabolism were determined from the threshold cycle (Ct) value. The mRNA expression was normalized to 18S expression and results were expressed as arbitrary units. The primers used are shown in Table 1.

Metabolomics

Cells were washed with calcium and magnesium free-phosphate buffer saline (CMF-PBS) and were harvested on ice into doubly distilled water and the intracellular metabolites were extracted following freeze-thaw procedure specially optimized for mammalian cell cultures [20]. Cell debris was pelleted by centrifugation at 12,000 RPM for 2 min at 4°C. The supernatant was collected for metabolite profiling analysis using nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) [21–23]. For metabolites analysis using NMR, water was removed by freeze-drying and the resulting residue reconstituted in 100 mM phosphate buffer (pH 7.4) prepared using deuterated water. Metabolite levels were normalized to protein content, which was determined by protein assay (Pierce).

Table 1. Primers Used in QPCR Analysis of Gene Expression

Genes	Primer information
GLUT1	Forward: 5′-TATCGTCAACACGGCCTTCACTGT-3′
	Reverse: 5′-CACAAAGCCAAAGATGGCCACGAT-3′
SGLT1	Forward: 5′-GCTCATGATTGCCGGAAGGTTGTT-3′
	Reverse: 5′-AATGGGTGGTCCCAAGTAACTGGT-3′
HK2	Forward: 5′-CTGCAGCGCATCAAGGAGAACAAA-3′
	Reverse: 5′-ACGGTCTTATGTAGACGCTTGGCA-3′
PGK1	Forward: 5′-TCACTCGGGCTAAGCAGATTGTGT-3′
	Reverse: 5′-CGTGTTCCATTTGGCACAGCAAGT-3′
PKM1	Forward: 5′-AGAACTTGTGCGAGCCTCAAGTCA-3′
	Reverse: 5′- CATTCATGGCAAAGTTCACCCGGA-3′
PKM2	Forward: 5′-ATTATTTGAGGAACTCCGCCGCCT-3′
	Reverse: 5′-CATTCATGGCAAAGTTCACCCGGA-3′
LDHA	Forward: 5′-TGGTCCAGCGTAACGTGAACATCT-3′
	Reverse: 5′-TTGCAACCGCTTCCAATAACACGG-3′
PDK1	Forward: 5′-TCATGTCACGCTGGGTAATGAGGA-3′
	Reverse: 5′-AACACGAGGTCTTGGTGCAGTTGA-3′
PEPCK	Forward: 5′-AGATCATCTCCTTTGGCAGTGGGT-3′
	Reverse: 5′-GTGCGTCAAACTTCATCCAGGCAA-3′
G6PD	Forward: 5′-TGCCTTCCATCAGTCGGATACACA-3′
	Reverse: 5′-GCATAGCCCACGATGAAGGTGTTT-3′
18S	Forward: 5′-TTAGAGTGTTCAAAGCAGGCCCGA-3′
	Reverse: 5′-TCTTGGCAAATGCTTTCGCTCTGG-3′

^{13}C-Metabolite Flux Analysis

Two hours before cell harvest, media were changed to fresh media containing equal concentrations of unlabelled and labeled glucose, and media collected after incubation for 2 h and stored at −80°C. Subsequently, media was used to monitor the ^{13}C-mass isotopomer distribution in metabolites using GC–MS. To 1 mL of media was added 0.2 mL of sulfosalicylic acid (50% w:v). The acid-supernatant was desalted by cation (AG 50W-X8, H+ form) exchange, and amino acids, pyruvate and lactate eluted with 2 mol/L NH_4OH followed by water. The frozen eluate was lyophilized to dryness, and analytes converted to their *t*-butyldimethylsilyl derivative prior to GCMS (HP 5973N Mass Selective Detector, Agilent, Palo Alto, CA). Fragment ions containing all carbons of an analyte (lactate, pyruvate, aspartate, and glutamate) were monitored under electron impact mode. Normalized crude ion abundances of the enriched analytes were corrected for the measured natural abundance of stable isotopes present in the original molecule and that contributed by the derivative using the matrix approach [24].

Flux calculations were based on tracer:tracee ratios (TTR) in the form mol ^{13}C-isotopomer (M+n) per 100 mol ^{12}C analyte (M+0), where n equals the number of ^{13}C-labelled carbons in the analyte, for example [M+1], [M+2] and [M+3]pyruvate. Catabolism of [$^{13}C_6$]glucose via the glycolytic pathway results in distinctive ^{13}C-labelling patterns in metabolites that provide information on the contributions of glucose to pathway fluxes and the activity of the enzymatic pathways through which the ^{13}C-skeleton traversed [25]. Catabolism of [$^{13}C_6$]glucose via the glycolytic pathway leads to the synthesis of [M+3] phosphoglycerate, thus [M+3]serine. It is important to note that the DMEM media contains serine, thus results are expressed as relative flux. Catabolism of [$^{13}C_6$]glucose leads to [M+3]pyruvate (and [M+3] lactate) which is readily released into the media after synthesis. It is important to note that the DMEM/F12 media does not contain pyruvate and lactate, thus the appearance and ^{13}C-labelling of these metabolites in media directly measures their activity in the intracellular pool. Thus, the contribution of glucose to the flux of pyruvate (and lactate) can be assessed from the ratios [M+3]pyruvate to [M+6]glucose. For measurement of pyruvate dehydrogenase (PDH) activity, we took advantage of the unique labeling patterns that result when [M+3]pyruvate is metabolized in the tricarboxylic acid (TCA). First, metabolism of the [M+3]pyruvate isotopomer via pyruvate carboxylase (PC) introduces the [M+3]oxaloacetate isotopomer into the TCA cycle and that this [M+3]oxaloacetate eventually leads to formation of [M+3]α-ketoglutarate. Second, the [M+3]pyruvate isotopomer can also be metabolized via PDH to yield [M+2]acetyl-CoA and thence [M+2]α-ketoglutarate. However, the [M+2]α-ketoglutarate isotopomer can also arise as a consequence of the equilibrium reaction between oxaloacetate and fumarate. This metabolic cycle yields an equal mixture of two positional isotopomers of [M+3]oxaloacetate, one labeled in carbons 1–3 and the other in carbons 2–4. In consequence, because the decarboxylation step between citrate and α-ketoglutarate leads to the loss of carbon 1 of oxaloacetate (i.e., half of [M+3]oxaloacetate contributes to [M+2]α-ketoglutarate enrichment), a correction must be made to the [M+2]α-ketoglutarate enrichment [26]. Direct measurement of intracellular oxaloacetate and α-ketoglutarate enrichments is technically challenging, particularly in the current study with cells in culture. As an alternative, we measured [M+2], [M+3]aspartate and [M+2], [M+3]glutamate in media since these isotopomers can only arise from intracellular synthesis from oxaloacetate and α-ketoglutarate, respectively. And, even though the DMEM/F12 media contained unlabelled aspartate and glutamate, the dilution of the ^{13}C-isotopomers of these amino acids will not alter the relative labeling of the [M+2] and [M+3]isotopomers. In consequence, the relative contribution of [M+3]pyruvate to [M+2]acetyl-CoA, that is, PDH activity, can be assessed by the ratio of [M+2]acetyl-CoA to [M+3]pyruvate [26].

Membrane Metabolite Fluxes

A sensitive and selective enzyme-based microbiosensor decorated with platinum nanoparticle was employed in self-referencing mode to measure real-

time physiological glucose, oxygen and lactate flux across the cell membrane [27]. Self-referencing involves oscillation of a single microsensor via computer-controlled stepper. This non-invasive technique provides direct measurement of trans-membrane analyte flux, reviewed in detail by McLamore and Porterfield [28].

Statistical Analysis

Data were analyzed by ANOVA to account for the effects of treatment and experiment replication. Values are presented as means and standard errors (SEM). Means were compared using the Student's *t*-test and by analysis of variance (ANOVA) and means were considered different when $P < 0.05$.

RESULTS

MCF10A-*ras* Cells Have Greater Aerobic Glycolysis

The impact of the activated H-*ras* gene on glucose uptake and glycolytic activity was investigated in the MCF10A and MCF10A-*ras* cells. The flux of glucose into the glycolytic pathway was assessed by the flux contribution of $^{13}C_6$-labeled glucose to the glycolytic intermediates 3-phosphoglycerate and pyruvate. Results showed that glucose flux into 3-phosphoglycerate was increased 94% in MCF10A-*ras* cells compared to MCF10A cells (Figure 1A), while pyruvate flux from glucose was not different in the two cell lines (Figure 1A), suggesting an increased flux of glucose through the glycolytic pathway in the H-*ras* transformed MCF10A cells.

Basal glucose influx at the cell membrane from the media was measured by the nano-biosensor as described previously. Glucose influx at the cell membrane was more than two-fold greater in the MCF10A-*ras* cells (328 ± 17 pmol/cm^2/sec) than in the MCF10A cells (156 ± 19 pmol/cm^2/sec, $P < 0.01$; Figure 1B), suggesting an increase in glucose uptake in *ras* transformed MCF10A cells, a hallmark of the Warburg effect during cancer progression. However, mRNA expression of the glucose transporter 1 (GLUT1) gene, the major glucose transporter in mammalian cells, was 35% lower in the MCF10A-*ras* cells (Figure 1C). In addition, expression of the sodium dependent glucose transporter (SGLT1) gene in the MCF10A-*ras* cells was also significantly lower than in the MCF10A cells (Figure 1C). These results suggest that the increase in glucose uptake in the MCF10A-*ras* cells is not due to an induction of the expression of these two glucose transporters by the *ras* oncogene.

Expression of genes for key enzymes in the glycolytic pathway were also measured in both cell lines. The expression of hexokinase 2 (HK2), the enzyme mediating the first step of phosphorylation of glucose during glycolysis, was not significantly different in the MCF10A-*ras* and MCF10A cells (Figure 1D). Phosphoglycerate kinase 1 (PGK1) catalyzes the seventh step of glycolysis, where 1,3-bisphosphoglycerate is converted to 3-phosphoglycerate with the formation of one ATP molecule. The MCF10A-*ras* cells have 20% greater expression of PGK1 (Figure 1D), consistent with the increase of flux into 3-phosphoglycerate from glucose (Figure 1A). Pyruvate kinase M1 (PKM1) and M2 catalyze conversion of phosphoenolpyruvate (PEP) to pyruvate, the rate-limiting final step of glycolysis. PKM2 is the predominant isoform expressed in both MCF10A and MCF10A-*ras* cells, and the switch of PKM1 to PKM2 has been shown to be important for the shift in cellular metabolism to aerobic glycolysis which promotes tumor growth [29]. There was a 47% greater expression of PKM1 but not PKM2 in the MCF10A-*ras* cells (Figure 1D). In addition, the enzyme activity assay of total pyruvate kinase (PK) showed that PK activity was not different between the two cell types (Figure 1E). Moreover, metabolic profiling showed that the intracellular PEP level was not different in the two cell types (Figure 1F), suggesting that the increased glucose flux into glycolysis may not be due to an increased conversion of PEP to pyruvate in the MCF10A-*ras* cells.

MCF10A-*ras* Cells Have Greater Lactate Production

One of the results of the Warburg effect in cells during cancer progression is the increased conversion of pyruvate to lactate. To determine whether there is an increase in lactate production in the H-*ras* transformed cells, the mRNA expression of lactate dehydrogenase A (LDHA), the enzyme which converts pyruvate to lactate, was examined in both cell types. There was a 34% increase in LDHA gene expression in the MCF10A-*ras* cells compared to the MCF10A cells (Figure 2A). Consistent with the increased LDHA expression in MCF10A-*ras* cells, metabolic profiling of the cells showed that intracellular lactate level was 2.4-fold higher in MCF10A-*ras* than in MCF10A cells (Figure 2B).

MCF10A-*ras* Cells Have Reduced TCA Cycle Activity

Since our results suggest that H-*ras* transformed cells have increased aerobic glycolysis and lactate production in progression to cancer, the activity of tricarboxylic acid (TCA) cycle was examined in both MCF10A and MCF10A-*ras* cells. The flux of glucose into the TCA cycle was assessed by the flux contribution of $^{13}C_6$-glucose to the intermediates in the TCA cycle. Although there was no significant reduction in the $^{13}C_6$-glucose flux to TCA cycle intermediates such as acetyl-CoA and oxaloacetate, the enzyme activity of pyruvate dehydrogenase (PDH), the mitochondrial enzyme complex converting pyruvate to acetyl-CoA for entering the TCA cycle, was reduced by 12% in MCF10A-*ras* cells compared to MCF10A cells as assessed by the $^{13}C_6$-glucose tracer kinetics (Figure 3A), suggesting a reduction in TCA cycle flux from glucose. Furthermore, the mRNA expression of pyruvate dehydrogenase kinase 1 (PDK1), which

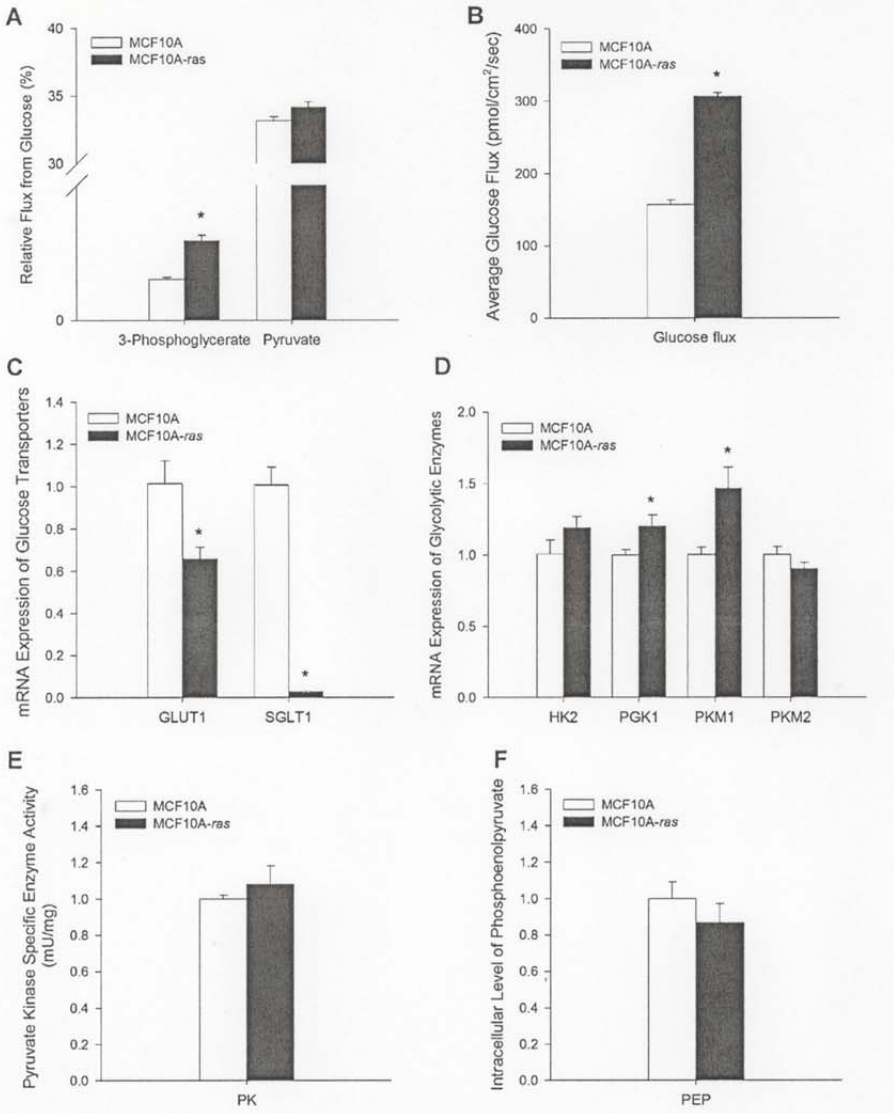


Figure 1. Assessment of glycolysis in MCF10A and MCF10A-*ras* cells. MCF10A and MCF10A-*ras* cells were cultured for 4 d before measurement or harvest. (A) Relative flux contributions of $^{13}C_6$-labeled glucose to 3-phosphoglycerate and pyruvate shown in percent metabolite flux from glucose (mean ± SEM, $n = 4$). (B) Basal glucose influx at the cell membrane (pmol/cm^2/sec) in normal culture conditions. Results are expressed as mean ± SEM ($n = 3$). (C) mRNA expressions of glucose transporters GLUT1 and SGLT1 are expressed relative to mRNA expression in the MCF10A cells as mean ± SEM ($n = 3$). (D) mRNA expression of glycolytic enzymes are expressed relative to mRNA expression in the MCF10A cells as mean ± SEM ($n = 3$). (E) Specific activity of pyruvate kinase are shown in mU enzyme activity per mg total protein (mean ± SEM, $n = 3$). (F) Intracellular level of phosphoenolpyruvate (PEP) relative to that in the MCF10A cells (mean ± SEM, $n = 4$). *Significant difference between the two cell types ($P < 0.05$).

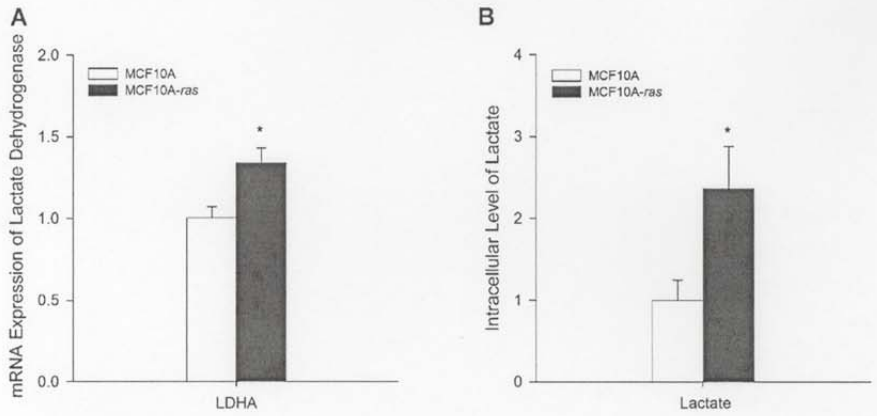


Figure 2. Lactate production in MCF10A and MCF10A-*ras* cells. MCF10A and MCF10A-*ras* cells were cultured for 4 d before harvest. (A) mRNA expression of lactate dehydrogenase A (LDHA) is expressed relative to mRNA expression in the MCF10A cells as mean ± SEM ($n = 3$). (B) Intracellular level of lactate relative to that in the MCF10A cells (mean ± SEM, $n = 4$). *Significant difference between the two cell types ($P < 0.05$).

acts to inactivate PDH by phosphorylation, was 54% higher in MCF10A-*ras* cells (Figure 3B), further supporting a reduced flux of glucose into the TCA cycle through the PDH.

MCF10A-*ras* Cells Have Greater G6PD Expression

One explanation for the shift to aerobic glycolysis is to provide metabolic intermediates as precursors and NADPH as reducing equivalent for the synthesis of fatty acids, protein and nucleic acids for rapid cell proliferation [3]. The PPP is an anabolic alternative to glycolysis, which produces NADPH and ribose-5-phosphate used in the synthesis of nucleotides. To determine the impact of the *ras* oncogene on the PPP, the expression of glucose-6-phosphate dehydrogenase (G6PD) was examined. The gene expression of G6PD was 45% higher in the MCF10A-*ras* cells (Figure 4). Since G6PD is the rate limiting enzyme in the PPP and is also important in maintaining NADPH levels against oxidative damage [30], this result indicates that the *ras* transformed cells may have increased PPP activity and thus potentially increased

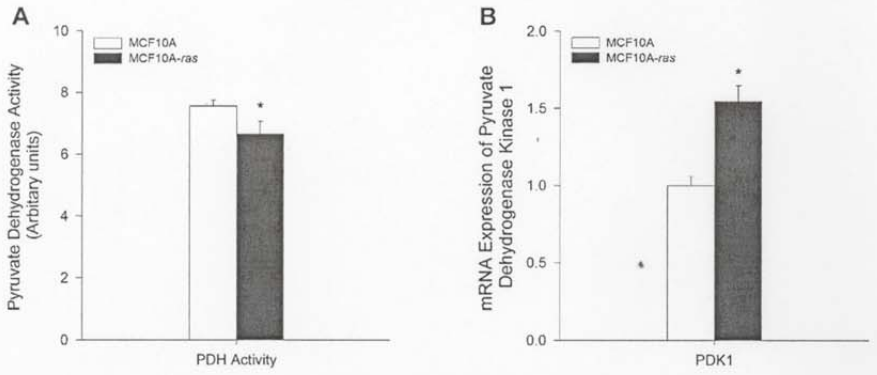


Figure 3. TCA cycle activity in MCF10A and MCF10A-*ras* cells. MCF10A and MCF10A-*ras* cells were cultured for 4 d before harvest. (A) Relative enzyme activity of pyruvate dehydrogenase (PDH) in arbitrary units, determined from flux contribution of $^{13}C_6$-labeled glucose (mean ± SEM, $n = 4$). (B) mRNA expression of pyruvate dehydrogenase kinase 1 (PDK1) is expressed relative to mRNA expression in the MCF10A cells as mean ± SEM ($n = 3$). *Significant difference between the two cell types ($P < 0.05$).

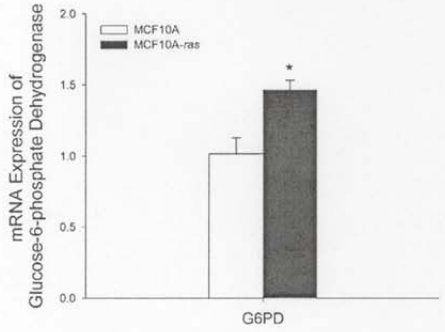


Figure 4. G6PD expression in MCF10A and MCF10A-*ras* cells. MCF10A and MCF10A-*ras* cells were cultured for 4 d before harvest. mRNA expression of glucose-6-phosphate dehydrogenase (G6PD) is expressed relative to mRNA expression in the MCF10A cells as mean ± SEM ($n = 3$). *Significant difference between the two cell types ($P < 0.05$).

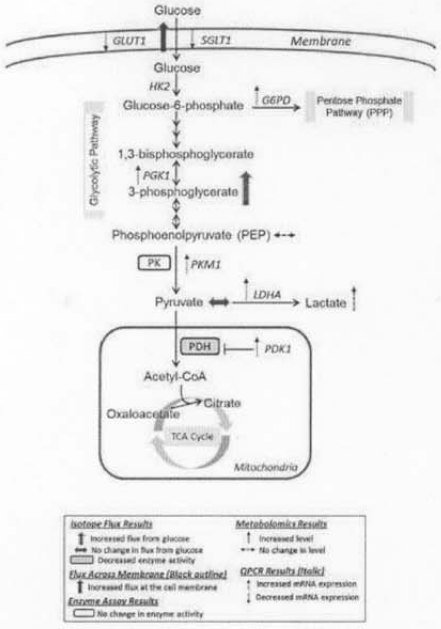


Figure 5. Summary of altered glucose metabolism in MCF10A-*ras* cells. MCF10A-*ras* cells exhibited enhanced glycolytic activity and lactate production, decreased glucose flux through the tricarboxylic acid (TCA) cycle, as well as a potential increase in glucose utilization in the pentose phosphate pathway. GLUT1, glucose transporter 1; SGLT1, sodium dependent glucose transporter1; HK2, hexokinase 2; PGK1, phosphoglycerate kinase1; PK, pyruvate kinase; PKM1, pyruvate kinase isoform M1; LDHA, lactate dehydrogenase A; PDH, pyruvate dehydrogenase; PDK1, pyruvate dehydrogenase kinase 1; G6PD, glucose-6-phosphate dehydrogenase.

nucleotides and NADPH synthesis, which promotes the survival and proliferation of these cells in early progression to cancer.

DISCUSSION

Alteration in cellular energy metabolism, especially glucose metabolism, is a signature characteristic of cancer cells. These alterations drive cell proliferation through increasing bioenergetics and cellular biosynthesis, maintaining anaplerosis and redox potential, as well as through initiating signal transductions which are controlled by changes in cellular metabolism [3,31,32]. Interventions targeting metabolic pathways are now emerging as potential preventive or therapeutic approaches for cancer [33,34]. In the current study, these data support the hypothesis that compared to the untransformed MCF10A cells, MCF10A-*ras* cells have greater glycolytic activity and glucose uptake, increased lactate, as well as reduced TCA cycle flux from glucose at an early stage of cancer progression. These results together suggest that MCF10A-*ras* cells, which are in the early stage of cancer progression, have a dramatic alteration in energy metabolism compared to the untransformed MCF10A cells. To our knowledge, these results are the first evidence of an early shift in energy metabolism mediated by a single H-*ras* oncogene in epithelial cells.

Our results are consistent with previous literature that an increase in glycolysis may be mediated by only the activity of the activated K-Ras gene in mouse and human cells [17], similar to cancer cells [3,4]. Previous literature demonstrates that the increase in glycolysis in the Warburg effect in cancer cells is mediated by increases in multiple enzymes in the glycolytic pathway, including GLUT1, HK2, PGK1, and PKM2 [4,29,35–41] and an increase in oxygen flux [42]. Similarly, the results of the current study, summarized in Figure 5, demonstrate the reprogramming of glucose metabolism that occurs in the MCF10A-*ras* breast epithelial cells. In the MCF10A-*ras* cells, glucose influx to the cells is upregulated (Figure 5, solid black arrow). The flux of glucose to the glycolytic intermediate 3-phosphoglycerate was increased (Figure 5, solid grey arrow), and consistent with this, the expression of glycolytic enzyme PGK1 was increased (Figure 5, dashed upward arrows). However, PKM1 is increased in the *ras*-transformed MCF10A cells, whereas the PKM2 isoform is predominantly expressed in the MCF10A and MCF10A-*ras* cells, and it is the switch of PKM1 to PKM2 that has been shown to associate with the shift in cellular metabolism to aerobic glycolysis which promotes tumor growth [3,29]. However, the enzyme activity of total PK was not different between the two cell types (Figure 5, open box). Therefore, the current results support an increase in glycolytic activity, with a potential difference in the mechanism underlying this metabolic reprogramming.

An underlying mechanism by which glucose uptake is increased in cancer cells is by an increase in expression of the GLUT1 membrane transporter [43]. In contrast, in the current study the expression of major glucose transporters in mammary epithelial cells GLUT1 and SGLT1 are decreased significantly in the *ras*-transformed cells (Figure 5, dashed downward arrows), suggesting that the increased glucose uptake is not a result of increased expression of glucose transporters but may be due to increased glucose flux into the glycolytic pathway. It is possible that with the increased rate of glycolysis in the MCF10A-*ras* cells, which was reflected by increased expression of glycolytic enzymes (Figure 1D), and the increased glucose flux to 3-phosphoglycerate (Figure 1A), there may be increased activity of the glucose transporters to accommodate the increased glycolytic activity downstream. More work is needed to investigate the molecular mechanism such as how *ras* oncogene may regulate the GLUT1 activity, to explain the dramatic increase in glucose uptake in the MCF10A-*ras* cells.

Consistent with the classic Warburg effect that the cancer cells have increased production of lactate as a result of the increased glycolysis [4], our results showed increases in both the expression of LDHA and the intracellular lactate level (Figure 5, dashed upward arrows) in the MCF10A-*ras* cells, indicating a similar effect in lactate production mediated by the single *ras* gene activation and that in the cancer cells. Interestingly, the proliferation rates of the MCF10A and MCF10A-*ras* cells are not different (data not shown), suggesting that the *ras*-oncogene mediated shifts in glucose metabolism occur at earlier stage during tumor progression before the increase in cell proliferation [43,44].

Another key feature of the classic Warburg effect is a decrease in glucose flux into the TCA cycle, which is controlled by the key enzyme PDH and its inhibitory kinase PDK1 [45,46]. The current results show that PDH activity is lower (Figure 5, solid box), concomitant with a substantially increased PDK1 expression (Figure 5, dashed upward arrow) in the MCF10A-*ras* cells, suggesting a decrease in glucose utilization through the TCA cycle. This decrease is consistent with previous studies showing a significantly lower mitochondrial Complex I activity in K-Ras transformed mouse and human cells [17,47,48].

Evidence suggests that the transcription factors HIF-1 and Myc, both of which are the key transcription factors in the Warburg effect and tumorigenesis, regulate the expression of the enzymes involved in the classic Warburg effect, including GLUT1, HK2, PGK1, LDHA, PDK1, and PK [45,46,49]. The results of the current study suggest that the increase flux of glucose through the glycolytic pathway and lactate production may involve the increased expression of PGK1, PKM1, and LDHA, and the decreased flux of glucose into the TCA cycle may involve the increased expression of PDK1, consistent with regulation by HIF-1 and Myc. However, the expression of GLUT1 and HK2 were not altered in the MCF10A-*ras* cells, which suggests there may be a difference in regulators of the effect compared to more advanced stage cancer cells. Therefore, these results suggest that the *ras*-oncogene may mediate the shift in glucose metabolism at least in part through the transcription regulation by HIF-1 and Myc, but may also involve other regulatory elements.

The expression of G6PD is also increased in the MCF10A-*ras* cells (Figure 5, dashed upward arrow), which is consistent with the literature that G6PD was particularly overexpressed in some human cancer cell lines, and its overexpression can result in the neoplastic transformation [50]. Since G6PD is the rate limiting enzyme in the PPP and plays an important role in cell growth and proliferation [30], these result may suggest a *ras*-oncogene mediated increase in the activity of the PPP and potentially an increased production of reducing equivalents (NADPH) and nucleotides to promote the survival and proliferation of these cells in early progression to cancer.

In conclusion, Harvey-*ras* transformed MCF10A human breast epithelial cells have altered glucose metabolism including increased glycolytic activity, lactate production, as well as reduced glucose flux through the TCA cycle. These alterations in glucose metabolism, consistent with the classic Warburg effect, may promote cell proliferation and/or survival during early breast cancer progression mediated by the initiating event of the activation of the H-*ras* gene [3]. Further research is needed to determine the mechanism by which H-*ras* oncogene regulates the shifts in glucose metabolism. The results of this study may aid in the identification of targeting points in the metabolic pathways to contribute to the development of effective agents for breast cancer prevention.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health, National Cancer Institute R25CA128770 (D. Teegarden) Cancer Prevention Internship Program (W. Zheng) administered by the Oncological Sciences Center and the Discovery Learning Research Center at Purdue University and by NIH 1R01GM085291 (D. Raftery).

REFERENCES

1. Warburg O. On the origin of cancer cells. Science 1956;123:309–314.
2. Feron O. Pyruvate into lactate and back: From the Warburg effect to symbiotic energy fuel exchange in cancer cells. Radiother Oncol 2009;92:329–333.
3. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: The metabolic requirements of cell proliferation. Science 2009;324:1029–1033.

4. Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? Nat Rev Cancer 2004;4:891–899.
5. Spitz DR, Sim JE, Ridnour LA, Galoforo SS, Lee YJ. Glucose deprivation-induced oxidative stress in human tumor cells. A fundamental defect in metabolism? Ann N Y Acad Sci 2000;899:349–362.
6. Caro-Maldonado A, Tait SW, Ramirez-Peinado S, et al. Glucose deprivation induces an atypical form of apoptosis mediated by caspase-8 in Bax-, Bak-deficient cells. Cell Death Differ 2010;17:1335–1344.
7. Cao X, Fang L, Gibbs S, et al. Glucose uptake inhibitor sensitizes cancer cells to daunorubicin and overcomes drug resistance in hypoxia. Cancer Chemother Pharmacol 2007;59:495–505.
8. Pelicano H, Martin DS, Xu RH, Huang P. Glycolysis inhibition for anticancer treatment. Oncogene 2006;25:4633–4646.
9. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: Weaving a tumorigenic web. Nat Rev Cancer 2011;11:761–774.
10. Kiaris H, Spandidos D. Mutations of ras genes in human tumors (review). Int J Oncol 1995;7:413–421.
11. Khleif SN, Abrams SI, Hamilton JM, et al. A phase I vaccine trial with peptides reflecting ras oncogene mutations of solid tumors. J Immunother 1999;22:155–165.
12. Eckert LB, Repasky GA, Ulku AS, et al. Involvement of Ras activation in human breast cancer cell signaling, invasion, and anoikis. Cancer Res 2004;64:4585–4592.
13. von Lintig FC, Dreilinger AD, Varki NM, Wallace AM, Casteel DE, Boss GR. Ras activation in human breast cancer. Breast Cancer Res Treat 2000;62:51–62.
14. Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 1989;244:707–712.
15. Dunn KL, Espino PS, Drobic B, He S, Davie JR. The Ras-MAPK signal transduction pathway, cancer and chromatin remodeling. Biochem Cell Biol 2005;83:1–14.
16. Malaney S, Daly RJ. The ras signaling pathway in mammary tumorigenesis and metastasis. J Mammary Gland Biol Neoplasia 2001;6:101–113.
17. Gaglio D, Metallo CM, Gameiro PA, et al. Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. Mol Syst Biol 2011;7:523.
18. Heppner GH, Wolman SR. MCF-10AT: A model for human breast cancer development. Breast J 1999;5:122–129.
19. Imbalzano KM, Tatarkova I, Imbalzano AN, Nickerson JA. Increasingly transformed MCF-10A cells have a progressively tumor-like phenotype in three-dimensional basement membrane culture. Cancer Cell Int 2009;9:7.
20. Bradley SA, Ouyang A, Purdie J, Smitka TA, Wang T, Kaerner A. Fermentanomics: Monitoring mammalian cell cultures with NMR spectroscopy. J Am Chem Soc 2010;132:9531–9533.
21. Asiago VM, Alvarado LZ, Shanaiah N, et al. Early detection of recurrent breast cancer using metabolite profiling. Cancer Res 2010;70:8309–8318.
22. Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. Expert Rev Mol Diagn 2008;8:617–633.
23. Zhang S, Liu L, Steffen D, Ye T, Raftery D. Metabolic profiling of gender: SPME/GC–MS and ^{1}H NMR analysis of urine. Metabolomics 2012;8:323–334.
24. Brauman JI. Least square analysis and simplification of multi-isotope mass spectra. Anal Chem 1966; 607–610.
25. Bequette BJ, Sunny NE, El-Kadi SW, Owens SL. Application of stable isotopes and mass isotopomer distribution analysis to the study of intermediary metabolism of nutrients. J Anim Sci 2006;84:E50–E59.
26. Berthold HK, Wykes LJ, Jahoor F, Klein PD, Reeds PJ. The use of uniformly labelled substrates and mass isotopomer analysis to study intermediary metabolism. Proc Nutr Soc 1994;53:345–354.
27. McLamore ES, Shi J, Jaroch D, et al. A self referencing platinum nanoparticle decorated enzyme-based microbiosensor for real time measurement of physiological glucose transport. Biosens Bioelectron 2010;26:2237–2245.
28. McLamore ES, Porterfield DM. Non-invasive tools for measuring metabolism and biophysical analyte transport: Self-referencing physiological sensing. Chem Soc Rev 2011;40:5308–5320.
29. Christofk HR, Vander Heiden MG, Harris MH, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature 2008;452:230–233.
30. Tian WN, Braunstein LD, Pang J, et al. Importance of glucose-6-phosphate dehydrogenase activity for cell growth. J Biol Chem 1998;273:10609–10617.
31. Metallo CM, Vander Heiden MG. Metabolism strikes back: Metabolic flux regulates cell signaling. Genes Dev 2010;24:2717–2722.
32. Locasale JW, Cantley LC. Metabolic flux and the regulation of mammalian cell growth. Cell Metab 2011;14:443–451.
33. Michelakis ED, Sutendra G, Dromparis P, et al. Metabolic modulation of glioblastoma with dichloroacetate. Sci Transl Med 2010;2:31–34.
34. Vander Heiden MG, Christofk HR, Schuman E, et al. Identification of small molecule inhibitors of pyruvate kinase M2. Biochem Pharmacol 2010;79:1118–1124.
35. Chen C, Pore N, Behrooz A, Ismail-Beigi F, Maity A. Regulation of glut1 mRNA by hypoxia-inducible factor-1. Interaction between H-ras and hypoxia. J Biol Chem 2001;276:9519–9525.
36. Mathupala SP, Heese C, Pedersen PL. Glucose catabolism in cancer cells. The type II hexokinase promoter contains functionally active response elements for the tumor suppressor p53. J Biol Chem 1997;272:22776–22780.
37. Noguchi Y, Okamoto T, Marat D, et al. Expression of facilitative glucose transporter 1 mRNA in colon cancer was not regulated by k-ras. Cancer Lett 2000;154:137–142.
38. Osthus RC, Shim H, Kim S, et al. Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. J Biol Chem 2000;275:21797–21800.
39. Finley LW, Carracedo A, Lee J, et al. SIRT3 opposes reprogramming of cancer cell metabolism through HIF1alpha destabilization. Cancer Cell 2011;19:416–428.
40. Yeung SJ, Pan J, Lee MH. Roles of p53, MYC and HIF-1 in regulating glycolysis—The seventh hallmark of cancer. Cell Mol Life Sci 2008;65:3981–3999.
41. Hwang TL, Liang Y, Chien KY, Yu JS. Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. Proteomics 2006;6:2259–2272.
42. Cook CC, Kim A, Terao S, Gotoh A, Higuchi M. Consumption of oxygen: A mitochondrial-generated progression signal of advanced cancer. Cell Death Dis 2012;3:e258.
43. Gillies RJ, Robey I, Gatenby RA. Causes and consequences of increased glucose metabolism of cancers. J Nucl Med 2008;49:24S–42S.
44. Robey IF, Stephen RM, Brown KS, Baggett BK, Gatenby RA, Gillies RJ. Regulation of the Warburg effect in early-passage breast cancer cells. Neoplasia 2008;10:745–756.
45. Jones RG, Thompson CB. Tumor suppressors and cell metabolism: A recipe for cancer growth. Genes Dev 2009;23:537–548.
46. Kim JW, Dang CV. Cancer's molecular sweet tooth and the Warburg effect. Cancer Res 2006;66:8927–8930.

47. Ishikawa K, Takenaga K, Akimoto M, et al. ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. Science 2008;320:661–664.
48. Baracca A, Chiaradonna F, Sgarbi G, Solaini G, Alberghina L, Lenaz G. Mitochondrial Complex I decrease is responsible for bioenergetic dysfunction in K-ras transformed cells. Biochim Biophys Acta 2010;1797:314–323.
49. Denko NC. Hypoxia, HIF1 and glucose metabolism in the solid tumour. Nat Rev Cancer 2008;8:705–713.
50. Kuo W, Lin J, Tang TK. Human glucose-6-phosphate dehydrogenase (G6PD) gene transforms NIH 3T3 cells and induces tumors in nude mice. Int J Cancer 2000;85:857–864.

Journal of Steroid Biochemistry & Molecular Biology 138 (2013) 81–89

Contents lists available at SciVerse ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/jsbmb

1,25-Dihydroxyvitamin D regulation of glucose metabolism in Harvey-*ras* transformed MCF10A human breast epithelial cells

Wei Zheng[a], Fariba Tayyari[b], G.A. Nagana Gowda[b,1], Daniel Raftery[b,1], Eric S. McLamore[c], Jin Shi[c,d], D. Marshall Porterfield[c,d,e,f], Shawn S. Donkin[a], Brian Bequette[g], Dorothy Teegarden[a,*]

[a] *Interdepartmental Nutrition Program, Purdue University, West Lafayette, IN 47906, United States*
[b] *Department of Chemistry, Purdue University, West Lafayette, IN 47906, United States*
[c] *Birck-Bindley Physiological Sensing Facility, Purdue University, West Lafayette, IN 47906, United States*
[d] *Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47906, United States*
[e] *Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47906, United States*
[f] *Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47906, United States*
[g] *Department of Animal and Avian Sciences, University of Maryland, College Park, MD, United States*

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form 22 March 2013
Accepted 24 March 2013

Keywords:
Vitamin D
Cancer prevention
Breast cancer
Glucose
Energy metabolism
ras
1,25-Dihydroxyvitamin D

ABSTRACT

This study was designed to investigate the impact of 1,25-dihydroxyvitamin D (1,25$(OH)_2$D) on glucose metabolism during early cancer progression. Untransformed and *ras*-oncogene transfected (*ras*) MCF10A human breast epithelial cells were employed to model early breast cancer progression. 1,25$(OH)_2$D modified the response of the *ras* cells to glucose restriction, suggesting 1,25$(OH)_2$D may reduce the *ras* cell glucose addiction noted in cancer cells. To understand the 1,25$(OH)_2$D regulation of glucose metabolism, following four-day 1,25$(OH)_2$D treatment, metabolite fluxes at the cell membrane were measured by a nanoprobe biosensor, [$^{13}C_6$]glucose flux by ^{13}C-mass isotopomer distribution analysis of media metabolites, intracellular metabolite levels by NMR, and gene expression of related enzymes was assessed. Treatment with 1,25$(OH)_2$D reduced glycolysis as flux of glucose to 3-phosphoglycerate was reduced by 15% ($P = 0.017$) and 32% ($P < 0.003$) in MCF10A and *ras* cells respectively. In the *ras* cells, 1,25$(OH)_2$D reduced lactate dehydrogenase activity by 15% ($P < 0.05$) with a concomitant 10% reduction in the flux of glucose to lactate ($P = 0.006$), and reduction in the level of intracellular lactate by 55% ($P = 0.029$). Treatment with 1,25$(OH)_2$D reduced flux of glucose to acetyl-coA 24% ($P = 0.002$) and 41% ($P < 0.001$), and flux to oxaloacetate 33% ($P = 0.003$) and 34% ($P = 0.027$) in the MCF10A and *ras* cells, respectively, suggesting a reduction in tricarboxylic acid (TCA) cycle activity. The results suggest a novel mechanism involving the regulation of glucose metabolism by which 1,25$(OH)_2$D may prevent breast cancer progression.

1. Introduction

Breast cancer is the second leading cancer among women in the US, with devastating consequences physically, emotionally and financially. Approximately 15% of cancer deaths in women result from breast cancer [1]. A growing body of evidence suggests that vitamin D may play a role in preventing the development of breast cancer [2–6]. For example, areas with higher latitudes and lower solar radiation, which leads to lower vitamin D synthesis in the skin, have increased mortality from breast cancer [3]. Substantial literature also supports that better vitamin D status is associated with reduced risk of breast cancer [2,4,5], but the mechanism is not clear.

The major circulating form of vitamin D, 25(OH)D, produced in the liver, is hydroxylated by 1α-hydroxylase in the kidney to the bioactive form of vitamin D, 1α,25-dihydroxyvitamin D (1,25$(OH)_2$D). Research supports that 1,25$(OH)_2$D has antineoplastic effects in colon, prostate, ovarian and breast cancer [2,7].

Abbreviations: 1,25$(OH)_2$D, 1,25-dihydroxyvitamin D; VDR, vitamin D receptor; TCA cycle, tricarboxylic acid cycle; LDH, lactate dehydrogenase; PDH, pyruvate dehydrogenase; PDK, pyruvate dehydrogenase kinase; GLUT1, glucose transporter 1; HK2, hexokinase 2; PGK1, phosphoglycerate kinase1; PEP, phosphoenolpyruvate; PK, pyruvate kinase; PBS, phosphate buffered saline; MTT, 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide; NMR, nuclear magnetic resonance spectroscopy; MS, mass spectrometry.

* Corresponding author at: Purdue University, 700 W. State Street, West Lafayette 47907, United States. Tel.: +1 765 494 8246; fax: +1 765 494 0906.
E-mail addresses: zheng38@purdue.edu (W. Zheng), ftayyari@purdue.edu (F. Tayyari), ngowda@uw.edu (G.A.N. Gowda), draftery@uw.edu (D. Raftery), emclamor@ufl.edu (E.S. McLamore), porterf@purdue.edu (D.M. Porterfield), sdonkin@purdue.edu (S.S. Donkin), bbequett@umd.edu (B. Bequette), teegarden@purdue.edu (D. Teegarden).
[1] Present address: Mitochondria and Metabolism Center, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA 98109, United States.

0960-0760/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jsbmb.2013.03.012

$1,25(OH)_2D$ is proposed to prevent cancer progression through the modulation of expression of many genes involved in cell growth, apoptosis, angiogenesis and immune responses [8,9]. However, the mechanisms in progression of breast cancer in particular and the effect of oncogenes on the action of $1,25(OH)_2D$ in early cancer progression are not fully understood.

One of the critical shifts in progression to tumorigenesis is in cellular energy metabolism [10]. The metabolic switch that occurs during carcinogenesis (including the Warburg effect) is a general characteristic of proliferating cells [10,11], which may lead to increased glucose metabolism and the dependence of cells on glucose (addiction) [12]. Proliferating cells not only require energy, but also nutrients in amounts greater than their bioenergetic needs in order to provide biosynthetic precursors, such as lipids, proteins and nucleic acids, for continued cell proliferation [11]. In nonproliferating cells, most of the pyruvate generated by glycolysis can be completely metabolized through the TCA cycle to produce large amount of ATPs in the presence of oxygen. In contrast, in rapidly proliferating cells and cancer cells, there is an increased glucose uptake and a shift of the pyruvate oxidative phosphorylation in the mitochondria toward a more rapid aerobic glycolysis, even in a normoxic environment, as described in the classic Warburg effect [10,11]. This decrease in glucose flux into the TCA cycle, which is controlled by the key enzyme pyruvate dehydrogenase (PDH) and its inhibitory kinase pyruvate dehydrogenase kinase (PDK), is another key feature of the classic Warburg effect [13,14]. These alterations in glucose metabolism result in dramatically reduced production of ATP and increased conversion of pyruvate to lactate, which allows glycolysis to continue by regenerating NAD+. The resulting lactate may also serve as an energy source for tumor cells [15]. Further, the rate of glucose metabolism in this metabolic switch increases dramatically. In fact, cancer cells divert about 10% of the glucose into biosynthetic pathways upstream of pyruvate production [11], which provides an advantage to cancer cells by favoring utilization of the most abundant energy and carbon sources. For example, glucose metabolism yields ribose for nucleic acid synthesis and NADPH through pentose phosphate pathway (PPP) while greater glycolysis provides intermediates to maintain anaplerosis and supply of biosynthetic intermediates [11] for the synthesis of fatty acids, nucleic acids and amino acids for continued cellular growth and replication. Further, the high rate of NADPH generation also aids in the anti-oxidant defense mechanisms of the cancer cells. Thus, deprivation of glucose can induce oxidative stress and other defects in metabolism which leads to cancer cell apoptosis [16,17]. Inhibitors of glucose transporters and glycolysis have been implemented as effective anticancer treatments and can also sensitize the tumor cells to other chemotherapeutic drugs [18,19]. Unlike in untransformed proliferating cells, the metabolic reprogramming in cancer cells is controlled by growth factor independent, chronic activation of the proliferative pathways [11]. Previous studies demonstrate that the increase in glycolysis in the Warburg effect in cancer cells is at least in part mediated by increases in multiple enzymes in the glycolytic pathway [19]. In order to establish strategies for cancer prevention by vitamin D, it is critical to determine whether and through which mechanism it regulates energy metabolism in normal cells and prevents the metabolic switch in cells containing oncogenes such as the activated *ras* gene.

The *ras* proto-oncogene is frequently mutated in cancer and affects a variety of tumorigenic processes including proliferation [20,21]. It encodes four distinct RAS proteins (HRAS, NRAS, KRAS4A and KRAS4B) which are small GTPases essential for the signal transduction induced by numerous growth factors to stimulate cell proliferation. The oncogenic RAS promotes both pro-growth and inhibits anti-growth signals in a growth factor independent manner [20]. The oncogenic RAS may also aid in metabolic reprogramming toward glycolysis in transformed cells. Previous studies show that K-*ras* transformed fibroblast cells have increased glycolytic activity and altered cellular glucose metabolism [22]. Furthermore, research supports that vitamin D receptor (VDR) transcriptional activity is down-regulated in the presence of *ras* oncogene [23–25], potentially disrupting the effect of $1,25(OH)_2D$ to inhibit tumorigenesis. Therefore, it is important to study the effect of $1,25(OH)_2D$ on cellular energy metabolism in *ras* oncogene transformed cells.

The effect of $1,25(OH)_2D$ on cellular glucose metabolism and its biological outcomes in early breast cancer progression have not been studied. The purpose of the current study was to investigate the effect of $1,25(OH)_2D$ regulation of cellular glucose energy metabolism in human breast epithelial cells with and without the Harvey-*ras* oncogene. The hypothesis of the current study is that $1,25(OH)_2D$ shifts glucose utilization toward reduced glycolysis and lactate production as well as reduced flux through the TCA cycle in Harvey-*ras* transfected breast epithelial cells but not in untransformed cells. These results will contribute to the understanding of $1,25(OH)_2D$ action on breast tissue during mammary carcinogenesis.

2. Materials and methods

2.1. Chemicals and reagents

The$1,25(OH)_2D$ was purchased from Biomol (Plymouth Meeting, PA). Dulbecco's modified Eagle medium: Nutrient Mixture F-12 (DMEM/F12) media, horse serum, trypsin and penicillin/streptomycin were obtained from Life Technologies, Gibco-BRL (Rockville, MD). Cholera toxin was purchased from Calbiochem (Darmstadt, Germany). Bicinchoninic acid (BCA) protein assay reagents were obtained from Pierce (Rockford, IL). Protease inhibitor cocktail, trypan blue, insulin, epidermal growth factor, and hydrocortisone were from Sigma (St. Lois, MO). Tris–HCl Bio-Rad Ready Gels were purchased from Bio-Rad Laboratories (Hercules, CA). The QuantiChrom Lactate Dehydrogenase Kit was from BioAssay Systems (Hayward, CA). All reagents for gas chromatography–mass spectrometry (GC–MS) analyses were from Pierce (Rockford, IL). D-[$^{13}C_6$]Glucose was purchased from Cambridge Isotope labs (Woburn, MA). Mass spectrometry analysis confirmed its chemical and isotopic purity (92.7% [$^{13}C_6$]glucose and 6.9% [$^{13}C_5$]glucose).

2.2. Cell culture

MCF10A human breast epithelial cells and MCF10A cells stably transfected with the Harvey-*ras* oncogene (MCF10A-*ras* cells) were a gift from Dr. Michael Kinch, Purdue University. MCF10A and MCF10A-*ras* cells were cultured in the standard media recommended for these cells [26], the Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12, 1:1) containing 17 mM glucose, and supplemented with 5% horse serum, 10 mg/L insulin, 20 μg/L epidermal growth factor, 50 μg/L cholera toxin, 50 mg/L hydrocortisone, 100 units/ml penicillin, and 0.1 mg/mL streptomycin in a humidified environment at 37 °C with 5% CO_2. DMEM/F12 (1:1) containing 17 mM glucose was used in all assays except for the MTT and flow cytometry analysis as indicated in Fig. 1. Cells were maintained in fresh media changed every 24 h during the 4-day treatment period. The $1,25(OH)_2D$ treatment was delivered to cells in 100% ethanol at a final ethanol concentration <1%.

2.3. MTT cell proliferation assay

Cells were cultured in media containing 5 mM glucose and the media changed that containing different glucose levels (0.1, 1, 5, and 17 mM) for the last 24 h. Relative viable cell levels were

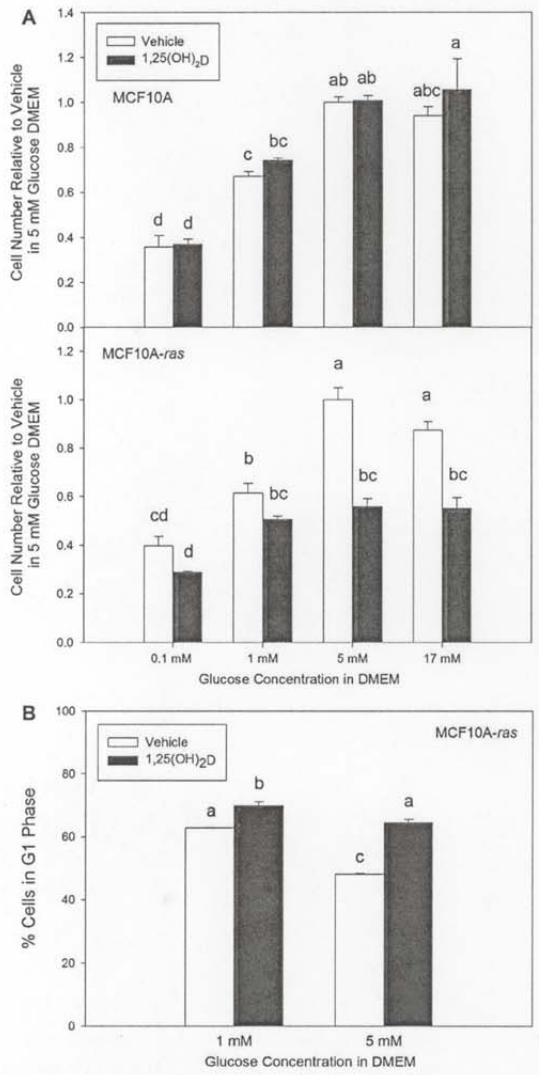


Fig. 1. 1,25$(OH)_2$D reduces glucose addiction. MCF10A and MCF10A-*ras* cells were treated with vehicle or 1,25$(OH)_2$D (10 nM) for 4 days before analysis. Cells were switched from media containing 5 mM glucose to the indicated glucose levels for the last 24 h. (A) Relative amount of MCF10A and MCF10A-*ras* cells compared to vehicle treated cells in 5 mM glucose media assessed by MTT assay. (B) Percentage of MCF10A-*ras* cells arrested in G1 cell cycle, analyzed by flow cytometry. Groups with the same letters are not significantly different ($P < 0.05$).

determined by the MTT assay according to the manufacturer's recommendations. Results were expressed as the relative absorbance compared to that in the vehicle treated cells in media containing 5 mM glucose.

2.4. Cell cycle analysis

From each sample, 1×10^6 cells were harvested with phosphate buffered saline (PBS) in single cell suspension. Cells were fixed with ice cold ethanol, pretreated with 0.2 mg/ml Rnase A, and stained with 10 ug/ml propidium iodide. Flow cytometry analysis was performed with a Beckman Coulter FC500 flow cytometer equipped with 488 nm laser. The results were analyzed using FlowJo (Tree Star, Inc., Ashland, OR). Results were expressed as percentage of total cells arrested in G1 phase in the cell cycle.

2.5. LDH assay

Cells were washed with calcium and magnesium free-phosphate buffered saline (CMF-PBS) and harvested on ice into buffer containing 100 mmol/L potassium phosphate (pH 7.0), 2 mmol/L EDTA, and 1% protease inhibitor cocktail and phosphatase inhibitor cocktail. Cells were briefly sonicated and cell debris was removed by centrifugation at 12,000 RPM for 15 min at 4 °C. Colorimetric kinetic determination of lactate dehydrogenase (LDH) activity in the cell lysate was measured using the QuantiChrom Lactate Dehydrogenase Kit (DLDH-100, BioAssay Systems). Values are expressed as LDH activity/total protein (specific activity).

2.6. Metabolomics

Cells were washed with CMF-PBS and harvested on ice into doubly distilled water. Cell lysates were obtained by freezing the cells in dry ice/ethanol bath for 5 min, and thawing them in 37 °C water bath for 1 min. Cell debris was pelleted by centrifugation at 12,000 RPM for 2 min at 4 °C. The supernatant was collected for metabolite profiling analysis using nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) as described previously [27–29]. Metabolite levels were normalized to protein content.

2.7. ^{13}C-metabolite flux analysis

Two hours prior to cell harvest, media were changed to fresh media with equal concentrations of unlabelled and $^{13}C_6$-labeled glucose, and collected after incubation for two hrs. Cells were rinsed with CMF-PBS, harvested on ice into lysis buffer and briefly sonicated. The cell lysates were saved for protein and DNA analysis, and media was used to determine the ^{13}C-mass isotopomer distribution analysis of metabolites and amino acids using GC–MS. Briefly, 0.2 mL of sulfosalicylic acid (50%, w:v) was added to 1 mL of media. The acid-supernatant was desalted by cation (AG 50W-X8, H+ form) exchange, and amino acids and lactate eluted with 2 mol/L NH_4OH followed by water. The frozen eluate was lyophilized to dryness, and amino acids converted to their *t*-butyldimethylsilyl derivative prior to GC–MS (HP 5973N Mass Selective Detector, Agilent, Palo Alto, CA). Fragment ions containing all carbons of an analyte (lactate, pyruvate, serine, aspartate and glutamate) were monitored under electron impact mode. Normalized crude ion abundances of the enriched analytes were corrected for the measured natural abundance of stable isotopes present in the original molecule and that contributed by the derivative using the matrix approach [30].

Flux calculations were based on tracer:tracee ratios (TTR) in the form mol ^{13}C-isotopomer ($M+n$) per 100 mol ^{12}C analyte ($M+0$), where n equals the number of ^{13}C-labeled carbons in the analyte, e.g. [$M+1$], [$M+2$] and [$M+3$]lactate. Catabolism of [$^{13}C_6$]glucose via the glycolytic pathway results in distinctive ^{13}C-labeling patterns in metabolites that provide information on the contributions of glucose to pathways fluxes and the activity of the enzymatic pathways through which the ^{13}C-skeleton traversed [31]. Under steady-state conditions, catabolism of [$^{13}C_6$]glucose leads to [$M+3$]pyruvate and then [$M+3$]lactate. Thus, the contribution of glucose to the flux of pyruvate and lactate can be assessed from the ratios [$M+3$]pyruvate to [$M+6$]glucose and [$M+3$]lactate to [$M+6$]glucose, respectively. Further, since lactate derives from cytosolic pyruvate, appearance of [$M+1$] and [$M+2$]lactate

represents metabolism of [$^{13}C_6$]glucose via the PPP. Thus, the ratio [M+3]lactate to [M+1]lactate provides a crude measure of the relative activities of glycolysis versus the PPP.

Metabolism of the [M+3]pyruvate isotopomer via pyruvate carboxylase (PC) introduces the [M+3]oxaloacetate isotopomer into the Krebs cycle which, subsequently, reaches a metabolic equilibrium with its transamination partner [M+3]aspartate [32]. Similarly, [M+3]oxaloacetate eventually leads to formation of [M+3]α-ketoglutarate, which is in metabolic equilibrium with its transamination partner [M+3]glutamate. Alternatively, the [M+3]pyruvate isotopomer can be metabolized via pyruvate dehydrogenase (PDH) to yield [M+2]acetyl-CoA and thence [M+2]α-ketoglutarate and [M+2]glutamate. However, the [M+2]glutamate also arises as a consequence of the equilibrium reaction between oxaloacetate and fumarate. This metabolic cycle yields an equal mixture of 2 positional isotopomers of [M+3]oxaloacetate, one labeled in carbons 1–3 and the other in carbons 2–4. In consequence, because the decarboxylation step between citrate and α-ketoglutarate leads to the loss of carbon 1 of oxaloacetate (i.e. half of [M+3]oxaloacetate contributes to [M+2]α-ketoglutarate enrichment), a correction to the [M+2]glutamate enrichment was made. In consequence, the contribution of glucose to oxaloacetate and acetyl-CoA fluxes can be assessed by the ratios [M+3]aspartate to [M+6]glucose and corrected [M+2]glutamate to [M+6]glucose, respectively. Furthermore, the relative activities of PDH vs. PC can be assessed from the ratio of [M+2]glutamate to [M+3]glutamate.

2.8. Membrane metabolite fluxes

A highly sensitive and selective glucose oxidase-based micro biosensor decorated with platinum nanoparticle was employed in self-referencing mode to measure real-time physiological glucose flux across the cell membrane [33]. Self-referencing involves oscillation of a single microsensor via computer-controlled stepper motors within the concentration boundary layer near cells/tissues. This non-invasive technique provides direct measurement of transmembrane analyte flux, and is reviewed in detail by McLamore and Porterfield [34].

2.9. RNA isolation and analysis

RNA was isolated with TriReagent (Molecular Research Center, Cincinnati, OH) following the manufacturer's instructions. Reverse transcription of total RNA was performed using MMLV reverse transcriptase (Promega, Madison, WI). Real-time quantitative PCR was performed using the Brilliant II SYBR Green QPCR Master Mix (Agilent, Santa Clara, CA). The mRNA abundances of enzymes involved in glucose metabolism were determined from the threshold cycle (Ct) value. The mRNA expression was normalized to 18S expression and results were expressed as arbitrary units. The primers used are shown Table 1.

2.10. Statistical analysis

Values are presented as mean ± SEM. Results are expressed compared to the vehicle within the same cell line, by the Student's *t*-tests (LSD), or by analysis of variance (ANOVA), with $P < 0.05$ considered statistically significant.

3. Results

3.1. 1,25(OH)$_2$D reduces glucose addiction

The impact of 1,25(OH)$_2$D treatment on glucose addiction of the cells was examined by MTT cell proliferation assay and flow cytometry analysis. 1,25(OH)$_2$D reduced the number of MCF10A-*ras* cells but not MCF10A cells by 44% and 37% in media containing 5 and 17 mM glucose, respectively (Fig. 1A). Glucose restriction at 0.1 and 1 mM reduced the number of MCF10A-*ras* cell by 60% and 39%, respectively (Fig. 1A), indicating the cell dependence on glucose (addiction). However, 1,25(OH)$_2$D prevented the reduction in cell number at 1 mM glucose, suggesting 1,25(OH)$_2$D may reduce the cell glucose addiction of MCF10A-*ras* cells. Consistent with these results, cell cycle analysis by flow cytometry showed that 1,25(OH)$_2$D increased the percentage of MCF10A-*ras* cells in G1 phase from 63% to 70% and from 48% to 64% in 1 and 5 mM glucose, respectively (Fig. 1B). Glucose restriction at 1 mM caused 30% more *ras* cells in G1 phase, but 1,25(OH)$_2$D treatment prevented the increase in G1 arrest (Fig. 1B), which suggest that 1,25(OH)$_2$D may reduce glucose addiction. These results suggest 1,25(OH)$_2$D reduces the MCF10A-*ras* cell glucose addiction, supporting the regulation of glucose metabolism by 1,25(OH)$_2$D during early cancer progression.

Table 1
Primers used in QPCR analysis of gene expression.

Genes	Primer information
GLUT1	Forward: 5′-TATCGTCAACACGGCCTTCACTGT-3′ Reverse: 5′-CACAAAGCCAAAGATGGCCACGAT-3′
HK2	Forward: 5′-CTGCAGCGCATCAAGGAGAACAAA-3′ Reverse: 5′-ACGGTCTTATGTAGACGCTTGGCA-3′
PGK1	Forward: 5′-TCACTCGGGCTAAGCAGATTGTGT-3′ Reverse: 5′-CGTGTTCCATTTGGCACAGCAAGT-3′
PKM2	Forward: 5′-ATTATTTGAGGAACTCCGCCGCCT-3′ Reverse: 5′-CATTCATGGCAAAGTTCACCCGGA-3′
PDK1	Forward: 5′-TCATGTCACGCTGGGTAATGAGGA-3′ Reverse: 5′-AACACGAGGTCTTGGTGCAGTTGA-3′
18S	Forward: 5′-TTAGAGTGTTCAAAGCAGGCCCGA-3′ Reverse: 5′-TCTTGGCAAATGCTTTCGCTCTGG-3′

3.2. 1,25(OH)$_2$D reduces glucose uptake and aerobic glycolysis

To investigate the impact of 1,25(OH)$_2$D on glucose uptake and glycolytic activity, MCF10A and MCF10A-*ras* cells were treated with vehicle or 1,25(OH)$_2$D for 4 days. Glucose influx across the cell membrane in response to increasing concentrations of glucose in the media was measured employing a bio-nanosensor [33]. There was a glucose concentration dependent increase in glucose influx for both 1,25(OH)$_2$D and vehicle treated MCF10A-*ras* cells (Fig. 2A). At high glucose concentration (15 mM), glucose influx was reduced in cells treated with 1,25(OH)$_2$D compared to cells treated with vehicle ($P < 0.05$). The effects of 1,25(OH)$_2$D on the expression of proteins and enzymes in the glycolytic pathway were also examined. Glucose transporter 1 (GLUT1) is the major glucose transporter expressed in the human breast epithelial cells. Results showed that mRNA expression of GLUT1 was reduced by 28% in the 1,25(OH)$_2$D treated MCF10A cells, but not for the MCF10A-*ras* cells compared to vehicle treated cells (Fig. 2B), suggesting that the reduction of glucose uptake by 1,25(OH)$_2$D treated MCF10A-*ras* cells may not be mediated through a decrease in expression of GLUT1. In contrast, mRNA expression of hexokinase 2 (HK2), the enzyme mediating the first step of phosphorylation of glucose during glycolysis, was induced by 23% by 1,25(OH)$_2$D in MCF10A cells, but not MCF10A-*ras* cells (Fig. 2C). Phosphoglycerate kinase 1 (PGK1) catalyzes the seventh step of glycolysis, where 1,3-bisphosphoglycerate is converted to 3-phosphoglycerate. 1,25(OH)$_2$D reduced the expression of PGK1 by 13% in MCF10A cells but not in MCF10A-*ras* cells (Fig. 2D). Consistent with the latter, results from the analysis of the flux

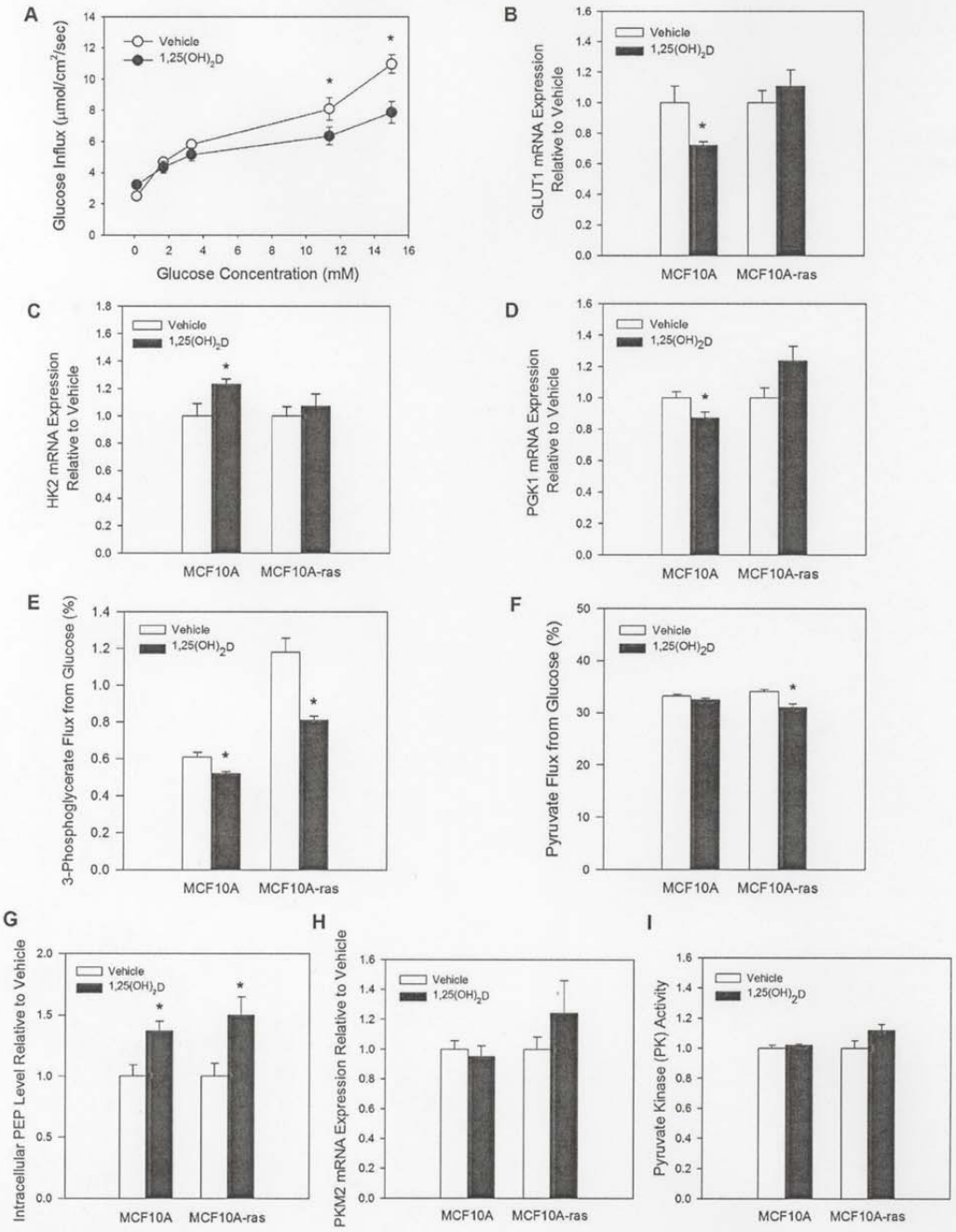


Fig. 2. 1,25(OH)$_2$D reduces glucose uptake and aerobic glycolysis. MCF10A and MCF10A-*ras* cells were treated with vehicle or 1,25(OH)$_2$D (10 nM) for 4 days before measurement or harvest. (A) Glucose influx at the cell membrane (μmol/cm^2/s) in response to increasing doses of added glucose in the media in MCF10A-*ras* cells ($n = 4$). The mRNA expression is shown for GLUT1 (B), HK2 (C), and PGK1 (D) relative to vehicle in each cell type ($n = 3$). Flux contributions of $^{13}C_6$-labeled glucose to 3-phosphoglycerate (E) and pyruvate (F) are shown in percent metabolite flux from glucose ($n = 4$). (G) Intracellular levels of phosphoenolpyruvate (PEP) relative to vehicle treatment in each cell type ($n = 4$). (H) The mRNA expression of PKM2 relative to vehicle in each cell type ($n = 3$). (I) The enzyme activity of total PK relative to vehicle in each cell type ($n = 3$). Results are expressed as mean ± SEM. An asterisk (*) indicates a significant difference between vehicle and 1,25(OH)$_2$D treatments within the same cell type ($P < 0.05$).

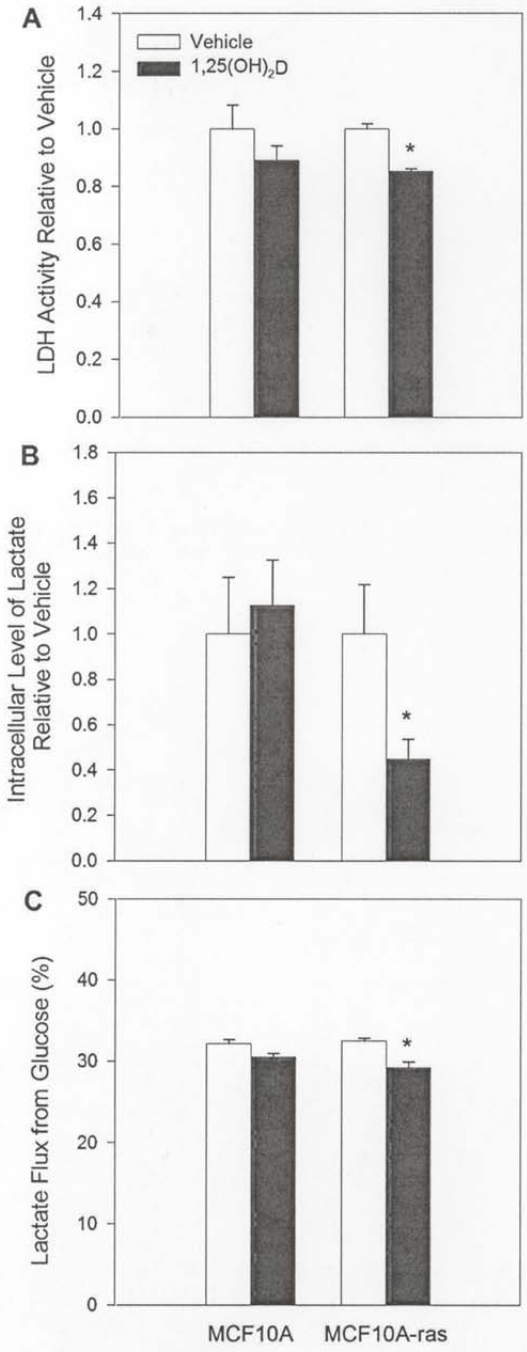


Fig. 3. 1,25$(OH)_2$D reduces lactate production. MCF10A and MCF10A-*ras* cells were treated with vehicle or 1,25$(OH)_2$D (10 nM) for 4 days before harvest. (A) Lactate dehydrogenase (LDH) activity relative to vehicle treatment in the same cell type ($n = 3$). (B) Intracellular levels of lactate relative to vehicle treatment in each cell type ($n = 4$). (C) Flux contributions of $^{13}C_6$-labeled glucose to lactate shown in percent metabolite flux from glucose ($n = 4$). Results are expressed as mean ± SEM. An asterisk (*) indicates a significant difference between vehicle and 1,25$(OH)_2$D treatments within the same cell type ($P < 0.05$).

contribution of $^{13}C_6$-labeled glucose to glycolytic intermediates showed that following 1,25$(OH)_2$D treatment, the contribution of glucose to 3-phosphoglycerate flux was reduced by 15% and 32% compared to vehicle in MCF10A and MCF10A-*ras* cells, respectively (Fig. 2E). Moreover, the proportion of pyruvate flux from glucose was reduced by 9% in MCF10A-*ras* cells, but was not altered in MCF10A cells (Fig. 2F), suggesting that glycolysis is reduced by 1,25$(OH)_2$D only in the MCF10A-*ras* cells.

Phosphoenolpyruvate (PEP) is a glycolytic intermediate converted to pyruvate via pyruvate kinase (PK) in the rate-limiting final step of glycolysis. But metabolic profiling of MCF10A and MCF10A-*ras* cells showed that compared to vehicle treated cells, intracellular levels of PEP was ~1.5-fold higher in MCF10A and MCF10A-*ras* cells treated with 1,25$(OH)_2$D (Fig. 2G). However, the mRNA expression of pyruvate kinase M2 (PKM2) (Fig. 2H), the predominant isoform of PK expressed in the MCF10A cells that promotes metabolic programming for tumor growth [35], was not different in vehicle and 1,25$(OH)_2$D treated MCF10A and MCF10A-*ras* (Fig. 2H). Further, total activity of PK was not affected by 1,25$(OH)_2$D treatment (Fig. 2I), suggesting that the accumulation of PEP is not a result of reduced PK activity by 1,25$(OH)_2$D. Collectively, these data support that 1,25$(OH)_2$D may reduce glucose uptake and glycolytic activity in MCF10A-*ras* cells at an early stage of cancer progression.

3.3. 1,25$(OH)_2$D reduces lactate production

One of the components of the Warburg effect in tumor cells is the increased conversion of pyruvate to lactate. To determine whether 1,25$(OH)_2$D impacts lactate production, the activity of lactate dehydrogenase (LDH) was examined in MCF10A and MCF10A-*ras* cells following 4 days of 1,25$(OH)_2$D treatment. There was a 15% reduction in LDH activity by 1,25$(OH)_2$D in MCF10A-*ras* cells, but not in MCF10A cells, compared to vehicle (Fig. 3A). Consistent with the reduction in LDH activity, 1,25$(OH)_2$D reduced the intracellular lactate by 55% in MCF10A-*ras* cells, but not in MCF10A cells (Fig. 3B). $^{13}C_6$-Glucose kinetics also showed that 1,25$(OH)_2$D reduced the contribution of glucose to lactate flux by 10% in MCF10A-*ras* cells, but not in MCF10A cells (Fig. 3C). These results suggest that 1,25$(OH)_2$D may reduce lactate production in MCF10A-*ras* cells during cancer progression, but not in untransformed MCF10A cells.

3.4. 1,25$(OH)_2$D reduces TCA cycle activity

The impact of 1,25$(OH)_2$D on TCA cycle activity in cells in progression to tumor were examined. The contribution of glucose to TCA cycle intermediate fluxes was assessed. Following 4 days of 1,25$(OH)_2$D treatment, there were 24% and 41% reductions in the contribution of glucose to acetyl-CoA flux in MCF10A and MCF10A-*ras* cells, respectively, compared to vehicle (Fig. 4A). Further, 1,25$(OH)_2$D treatment reduced the contribution of glucose to oxaloacetate flux by 33% and 34% (Fig. 4B) in MCF10A cells and MCF10A-*ras* cells, respectively, suggesting an overall reduction of glucose metabolism in the TCA cycle. Consistent with these results, the ^{13}C-tracer kinetics indicated a reduction in pyruvate dehydrogenase (PDH) activity by 22% and 24% in MCF10A and MCF10A-*ras* cells treated with 1,25$(OH)_2$D (Fig. 4C). In addition, metabolic profiling showed there was a 29% reduction in the intracellular level of succinate, an intermediate of the TCA cycle, by 1,25$(OH)_2$D treatment in the MCF10A-*ras* cells, but not in MCF10A cells, compared to vehicle (Fig. 4D). The latter provides further support of the overall reduction in TCA cycle activity in response to 1,25$(OH)_2$D in cells during early progression to tumors. In contrast, mRNA expression

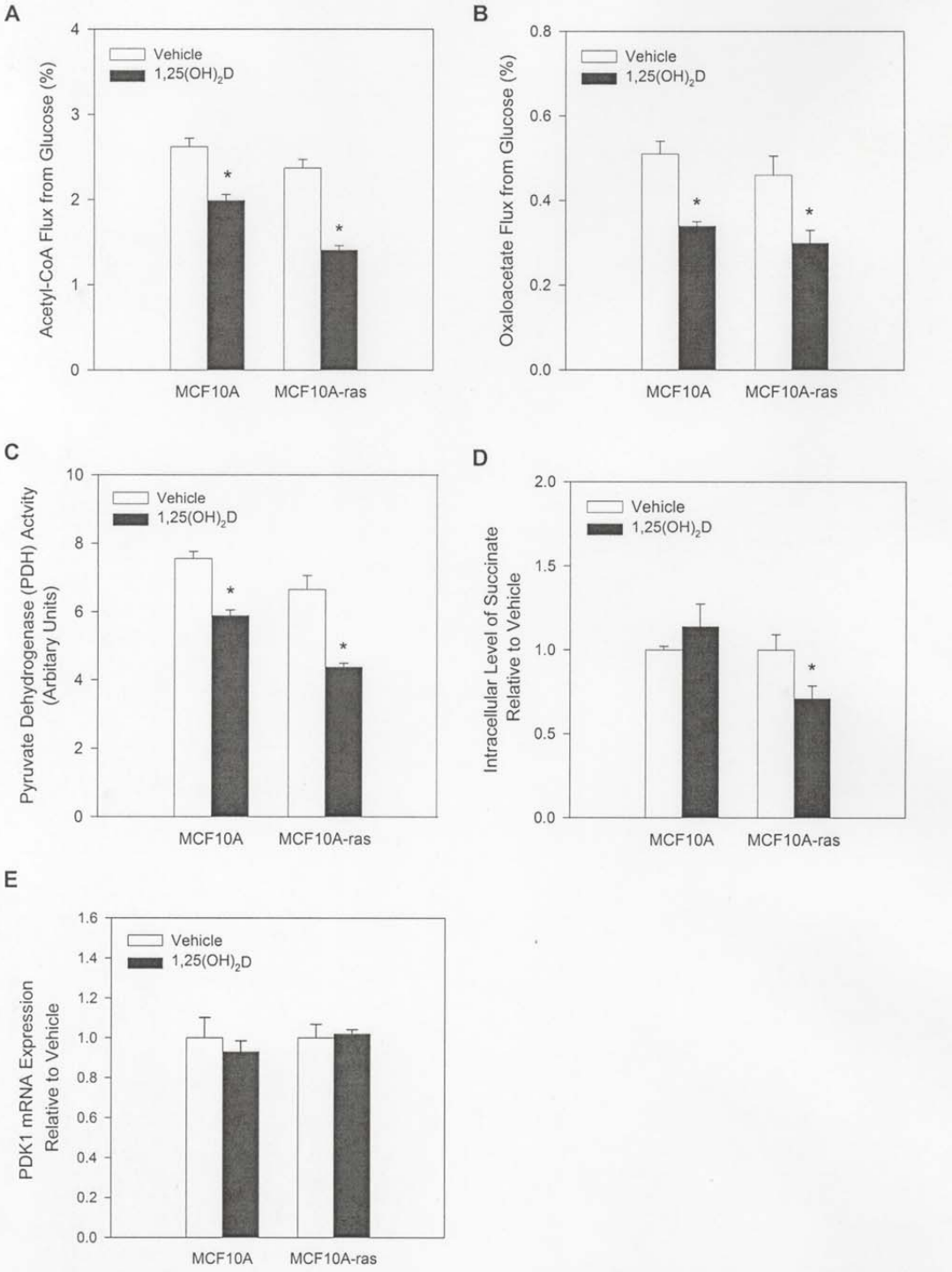


Fig. 4. $1,25(OH)_2D$ reduces TCA cycle activity. MCF10A and MCF10A-*ras* cells were treated with vehicle or $1,25(OH)_2D$ (10 nM) for 4 days before measurement or harvest. Flux contributions of $^{13}C_6$-labeled glucose to (A) acetyl-CoA and (B) oxaloacetate shown in percent metabolite flux from glucose. (C) Pyruvate dehydrogenase activity shown in arbitrary units ($n = 4$). (D) Intracellular levels of succinate relative vehicle in each cell type ($n = 4$). (E) The mRNA expression of PDK1 relative to vehicle in each cell type ($n = 3$). Results are expressed as mean ± SEM. An asterisk (*) indicates a significant difference between vehicle and $1,25(OH)_2D$ treatments within the same cell type ($P < 0.05$).

of pyruvate dehydrogenase kinase 1 (PDK1) (Fig. 4E), the inhibiting kinase of PDH that regulates the flux of pyruvate into the TCA cycle, was not altered by $1,25(OH)_2D$ in MCF10A or MCF10A-*ras* cells, suggesting that the reduced PDH activity and flux of glucose into the TCA cycle induced by $1,25(OH)_2D$ may not be mediated through upregulation of PDK1.

4. Discussion

Alteration in cellular glucose metabolism is a signature characteristic of tumor cells which drives cell proliferation by increasing bioenergetics and biosynthesis, maintaining redox potential, and via initiation of signal transduction controlled by changes in cellular metabolism [11,36,37]. Interventions that target metabolic pathways are now emerging as potential preventive or therapeutic approaches for the treatment of cancers [18,19,38,39]. In the current study, the effects of $1,25(OH)_2D$ on cellular energy metabolism were explored in untransformed and H-*ras* oncogene transfected MCF10A cells, a human breast epithelial cell model for studying early mammary carcinogenesis. The results support the hypothesis that $1,25(OH)_2D$ reduces the glucose addiction of cells in progression to cancer, as a consequence of the shift in glucose metabolism toward reduced glycolysis and lactate production (reversal of the classic Warburg effect) as well as reduced TCA cycle activity in H-*ras* transfected MCF10A cells, suggesting a preventive effect of $1,25(OH)_2D$ on glucose utilization for rapid cell proliferation during breast cancer progression. To our knowledge, these results are the first to show that $1,25(OH)_2D$ regulates cellular glucose metabolism which may be a potential mechanism for preventing early breast cancer progression.

Previous studies suggested that activity of the activated K-*ras* gene alone may lead to an increase in glycolysis in mouse and human cells [22], similar to that seen in cancer cells [11,40]. Results from our laboratory support these results. The MCF10A-*ras* cells, which represent an initiation stage of tumor progression, show changes in several energy status parameters that are characteristic of tumors, including the increased glucose uptake (by 2.1 fold, $P<0.01$) [41] and lactate accumulation (by 2.4 fold, $P<0.05$, data not shown), which is consistent with the Warburg effect [40,42], supporting the hypothesis that very early changes in energy status may occur during cancer progression in the presence of the H-*ras* oncogene. The current study suggests that $1,25(OH)_2D$ may reverse the alterations in glucose metabolism during early breast cancer progression mediated by the H-*ras* oncogene. One of the outcomes for the altered glucose metabolism by $1,25(OH)_2D$ was the reduced glucose addiction via reducing G1 cell cycle arrest at glucose restriction. It has been suggested that part of the chemopreventive effects of $1,25(OH)_2D$ against cancer is mediated by G1 cell-cycle arrest, as a result of the upregulation of proteins suppressing cyclin-dependent kinase activity [43–45]. Consistently, the current study showed that $1,25(OH)_2D$ induced the G1 cell cycle arrest of MCF10A-*ras* cells but prevented the increase in G1 arrest (Fig. 1B) at glucose restriction, suggesting $1,25(OH)_2D$ may reduce the glucose addiction of MCF10A-*ras* cells through the regulation of G1 cell cycle arrest.

Results suggesting a reduction in glycolysis by $1,25(OH)_2D$ are supported by the reduced flux of glucose to 3-phosphoglycerate by $1,25(OH)_2D$ in MCF10A and MCF10A-*ras* cells respectively. In the MCF10A-*ras* cells, $1,25(OH)_2D$ was shown to reduce lactate production, and a reduction of intracellular lactate levels, as well as a reduction in lactate dehydrogenase activity. $1,25(OH)_2D$-induced reduction in TCA cycle activity was observed as reduced glucose flux to acetyl-coA and to oxaloacetate in the MCF10A and MCF10A-*ras* cells, respectively. It is also intriguing that one of the glycolytic intermediates, PEP, was increased by $1,25(OH)_2D$ in the MCF10A-*ras* cells. One mechanism that may contribute to the accumulation of PEP may be due to the reduced activity of the enzyme downstream of this metabolite in the glycolytic pathway, PK. PK regulates the rate-limiting and final step of glycolysis, the conversion of PEP to pyruvate. The M2 isoform (PKM2) is a critical enzyme expressed predominantly in tumor tissues that promotes aerobic glycolysis [35]. The accumulation of PEP suggests a potentially decreased activity of PKM2 by $1,25(OH)_2D$. However, the mRNA expression of PKM2 was not different in vehicle and $1,25(OH)_2D$ treated MCF10A and MCF10A-*ras* cells (Fig. 2H); and neither was the total activity of PK enzyme (Fig. 2I), suggesting that accumulation of PEP is not a result of reduced PK activity by $1,25(OH)_2D$. Regulation of the activity of other enzymes in hexogenesis, glycolysis or flux into the TCA cycle may contribute to the increased accumulation of PEP mediated by $1,25(OH)_2D$. There are differential effects of $1,25(OH)_2D$ on energy metabolism and cell growth in the untransformed and *ras* oncogene transformed MCF10A cells. Overall, our results suggest that $1,25(OH)_2D$ has a greater effect on metabolic parameters in the MCF10A-*ras* cells than in MCF10A cells, which may contribute to the growth inhibitory effects in different glucose concentrations (Fig. 1A lower panel). Although there are changes in some of the metabolic parameters by $1,25(OH)_2D$ in MCF10A cells (Figs. 2 and 4), there are no changes in the biological consequences (growth) in these cells (Fig. 1A, upper panel). These differences in quantitative impact of $1,25(OH)_2D$ may suggest that the effects in MCF10A cells were not great enough to lead to a change with respect to growth in different glucose concentrations. Further investigations are needed to understand how the changes in metabolic parameters contribute to the biological consequences such as growth and glucose dependence.

Overall, the results of the current study support our hypothesis that $1,25(OH)_2D$ regulates the metabolic reprogramming during early breast cancer progression. Specifically, $1,25(OH)_2D$ reduces cell glucose addiction, shifts glucose utilization toward reduced glycolysis and lactate production as well as reduced flux through the TCA cycle in H-*ras* transformed MCF10A breast epithelial cells. To our knowledge, this is the first study to demonstrate that $1,25(OH)_2D$ regulates cellular energy metabolism in a model of early breast cancer progression. These results indicate that vitamin D, as a potential chemopreventive agent, has multiple functions during cancer progression. Acquiring a better understanding of vitamin D regulation of glucose metabolism during cancer progression should allow the identification of its targeted regulatory points in the metabolic pathways and its interaction with oncogenes such as *ras*, so as to contribute to the development of effective strategies for breast cancer prevention.

Conflict of interest

No potential conflicts of interest were disclosed by the authors.

Acknowledgments

This work was supported by the National Institutes of Health, National Cancer Institute R25CA128770 (D. Teegarden) Cancer Prevention Internship Program (W. Zheng) administered by the Oncological Sciences Center and the Discovery Learning Research Center at Purdue University and by NIH 1R01GM085291 (D. Raftery).

References

[1] J. Stebbing, S. Ngan, Breast cancer (metastatic), Clin Evid (Online) 2010 (2010).
[2] C.F. Garland, E.D. Gorham, S.B. Mohr, W.B. Grant, E.L. Giovannucci, M. Lipkin, H. Newmark, M.F. Holick, F.C. Garland, Vitamin D and prevention of breast cancer:

pooled analysis, The Journal of Steroid Biochemistry and Molecular Biology 103 (2007) 708–711.
[3] F.C. Garland, C.F. Garland, E.D. Gorham, J.F. Young, Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation, Preventive Medicine 19 (1990) 614–622.
[4] R.A. Hiatt, N. Krieger, B. Lobaugh, M.K. Drezner, J.H. Vogelman, N. Orentreich, Prediagnostic serum vitamin D and breast cancer, Journal of the National Cancer Institute 90 (1998) 461–463.
[5] E.C. Janowsky, G.E. Lester, C.R. Weinberg, R.C. Millikan, J.M. Schildkraut, P.A. Garrett, B.S. Hulka, Association between low levels of 1,25-dihydroxyvitamin D and breast cancer risk, Public Health Nutrition 2 (1999) 283–291.
[6] J.M. Lappe, D. Travers-Gustafson, K.M. Davies, R.R. Recker, R.P. Heaney, Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial, American Journal of Clinical Nutrition 85 (2007) 1586–1591.
[7] C.F. Garland, F.C. Garland, E.D. Gorham, M. Lipkin, H. Newmark, S.B. Mohr, M.F. Holick, The role of vitamin D in cancer prevention, American Journal of Public Health 96 (2006) 252–261.
[8] J.C. Fleet, M. DeSmet, R. Johnson, Y. Li, Vitamin D and cancer: a review of molecular mechanisms, Biochemical Journal 441 (2012) 61–76.
[9] A.V. Krishnan, D. Feldman, Mechanisms of the anti-cancer and anti-inflammatory actions of vitamin D, Annual Review of Pharmacology and Toxicology 51 (2011) 311–336.
[10] O. Warburg, On the origin of cancer cells, Science 123 (1956) 309–314.
[11] M.G. Vander Heiden, L.C. Cantley, C.B. Thompson, Understanding the Warburg effect: the metabolic requirements of cell proliferation, Science 324 (2009) 1029–1033.
[12] R.J. Gillies, I. Robey, R.A. Gatenby, Causes and consequences of increased glucose metabolism of cancers, Journal of Nuclear Materials 49 (Suppl 2) (2008) 24S–42S.
[13] R.G. Jones, C.B. Thompson, Tumor suppressors and cell metabolism: a recipe for cancer growth, Genes and Development 23 (2009) 537–548.
[14] J.W. Kim, C.V. Dang, Cancer's molecular sweet tooth and the Warburg effect, Cancer Research 66 (2006) 8927–8930.
[15] O. Feron, Pyruvate into lactate and back: from the Warburg effect to symbiotic energy fuel exchange in cancer cells, Radiotherapy and Oncology 92 (2009) 329–333.
[16] D.R. Spitz, J.E. Sim, L.A. Ridnour, S.S. Galoforo, Y.J. Lee, Glucose deprivation-induced oxidative stress in human tumor cells. A fundamental defect in metabolism? Annals of the New York Academy of Sciences 899 (2000) 349–362.
[17] A. Caro-Maldonado, S.W. Tait, S. Ramirez-Peinado, J.E. Ricci, I. Fabregat, D.R. Green, C. Munoz-Pinedo, Glucose deprivation induces an atypical form of apoptosis mediated by caspase-8 in Bax-, Bak-deficient cells, Cell Death & Differentiation 17 (2010) 1335–1344.
[18] X. Cao, L. Fang, S. Gibbs, Y. Huang, Z. Dai, P. Wen, X. Zheng, W. Sadee, D. Sun, Glucose uptake inhibitor sensitizes cancer cells to daunorubicin and overcomes drug resistance in hypoxia, Cancer Chemotherapy and Pharmacology 59 (2007) 495–505.
[19] H. Pelicano, D.S. Martin, R.H. Xu, P. Huang, Glycolysis inhibition for anticancer treatment, Oncogene 25 (2006) 4633–4646.
[20] Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, RAS oncogenes: weaving a tumorigenic web, Nature Reviews Cancer 11 (2011) 761–774.
[21] H. Kiaris, D. Spandidos, Mutations of ras genes in human tumors (review), International Journal of Oncology 7 (1995) 413–421.
[22] D. Gaglio, C.M. Metallo, P.A. Gameiro, K. Hiller, L.S. Danna, C. Balestrieri, L. Alberghina, G. Stephanopoulos, F. Chiaradonna, Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth, Molecular Systems Biology 7 (2011) 523.
[23] L.M. Taber, L.S. Adams, D. Teegarden, Mechanisms of nuclear vitamin D receptor resistance in Harvey-ras-transfected cells, Journal of Nutritional Biochemistry (2008).
[24] C. Solomon, J.H. White, R. Kremer, Mitogen-activated protein kinase inhibits 1,25-dihydroxyvitamin D3-dependent signal transduction by phosphorylating human retinoid X receptor alpha, Journal of Clinical Investigation 103 (1999) 1729–1735.
[25] C. Solomon, R. Kremer, J.H. White, J.S. Rhim, Vitamin D resistance in RAS-transformed keratinocytes: mechanism and reversal strategies, Radiation Research 155 (2001) 156–162.
[26] H.D. Soule, T.M. Maloney, S.R. Wolman, W.D. Peterson Jr., R. Brenz, C.M. McGrath, J. Russo, R.J. Pauley, R.F. Jones, S.C. Brooks, Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10, Cancer Research 50 (1990) 6075–6086.
[27] V.M. Asiago, L.Z. Alvarado, N. Shanaiah, G.A. Gowda, K. Owusu-Sarfo, R.A. Ballas, D. Raftery, Early detection of recurrent breast cancer using metabolite profiling, Cancer Research 70 (2010) 8309–8318.
[28] G.A. Gowda, S. Zhang, H. Gu, V. Asiago, N. Shanaiah, D. Raftery, Metabolomics-based methods for early disease diagnostics, Expert Review of Molecular Diagnostics 8 (2008) 617–633.
[29] S. Zhang, L. Liu, D. Steffen, T. Ye, D. Raftery, Metabolic Profiling of Gender: SPME/GC-MS and 1H NMR Analysis of Urine, Metabolomics 8 (2012) 323–334.
[30] J.I. Brauman, Least square analysis and simplification of multi-isotope mass spectra, Analytical Chemistry (1966) 607–610.
[31] B.J. Bequette, N.E. Sunny, S.W. El-Kadi, S.L. Owens, Application of stable isotopes and mass isotopomer distribution analysis to the study of intermediary metabolism of nutrients, Journal of Animal Science 84 (Suppl.) (2006) E50–E59.
[32] C.A. Fernandez, R.C. Des, Modeling of liver citric acid cycle and gluconeogenesis based on 13C mass isotopomer distribution analysis of intermediates, Journal of Biological Chemistry 270 (1995) 10037–10042.
[33] E.S. McLamore, J. Shi, D. Jaroch, J.C. Claussen, A. Uchida, Y. Jiang, W. Zhang, S.S. Donkin, M.K. Banks, K.K. Buhman, D. Teegarden, J.L. Rickus, D.M. Porterfield, A self referencing platinum nanoparticle decorated enzyme-based microbiosensor for real time measurement of physiological glucose transport, Biosensors and Bioelectronics 26 (2010) 2237–2245.
[34] E.S. McLamore, D.M. Porterfield, Non-invasive tools for measuring metabolism and biophysical analyte transport: self-referencing physiological sensing, Chemical Society Reviews 40 (2011) 5308–5320.
[35] H.R. Christofk, M.G. Vander Heiden, M.H. Harris, A. Ramanathan, R.E. Gerszten, R. Wei, M.D. Fleming, S.L. Schreiber, L.C. Cantley, The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth, Nature 452 (2008) 230–233.
[36] C.M. Metallo, M.G. Vander Heiden, Metabolism strikes back: metabolic flux regulates cell signaling, Genes and Development 24 (2010) 2717–2722.
[37] J.W. Locasale, L.C. Cantley, Metabolic flux and the regulation of mammalian cell growth, Cell Metabolism 14 (2011) 443–451.
[38] E.D. Michelakis, G. Sutendra, P. Dromparis, L. Webster, A. Haromy, E. Niven, C. Maguire, T.L. Gammer, J.R. Mackey, D. Fulton, B. Abdulkarim, M.S. McMurtry, K.C. Petruk, Metabolic modulation of glioblastoma with dichloroacetate, Science Translational Medicine 2 (2010) 31–34.
[39] M.G. Vander Heiden, H.R. Christofk, E. Schuman, A.O. Subtelny, H. Sharfi, E.E. Harlow, J. Xian, L.C. Cantley, Identification of small molecule inhibitors of pyruvate kinase M2, Biochemical Pharmacology 79 (2010) 1118–1124.
[40] R.A. Gatenby, R.J. Gillies, Why do cancers have high aerobic glycolysis? Nature Reviews Cancer 4 (2004) 891–899.
[41] W. Zheng, E.S. McLamore, D.M. Porterfield, S.S. Donkin, D. Teegarden, Effect of 1,25-dihydroxyvitamin D on energy metabolism in MCF10A breast epithelial cells, FASEB Journal 25 (235.1) (2011) (Ref type: Abstract).
[42] C.C. Cook, A. Kim, S. Terao, A. Gotoh, M. Higuchi, Consumption of oxygen: a mitochondrial-generated progression signal of advanced cancer, Cell Death and Disease 3 (2012) e258.
[43] B.A. Scaglione-Sewell, M. Bissonnette, S. Skarosi, C. Abraham, T.A. Brasitus, A vitamin D3 analog induces a G1-phase arrest in CaCo-2 cells by inhibiting cdk2 and cdk6: roles of cyclin E, p21Waf1, and p27Kip1, Endocrinology 141 (2000) 3931–3939.
[44] S.S. Jensen, M.W. Madsen, J. Lukas, L. Binderup, J. Bartek, Inhibitory effects of 1alpha, 25-dihydroxyvitamin D(3) on the G(1)-S phase-controlling machinery, Molecular Endocrinology 15 (2001) 1370–1380.
[45] S.A. Lamprecht, M. Lipkin, Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms, Nature Reviews Cancer 3 (2003) 601–614.