Discovery Based Yeast Metabolomic Analysis using Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry and Chemometrics

Rachel E. Mohler

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2007

Program Authorized to Offer Degree:
Department of Chemistry

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

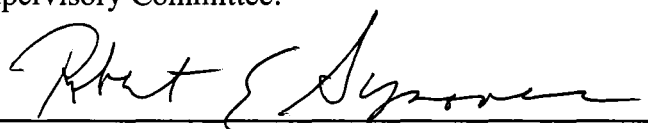
University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Rachel E. Mohler

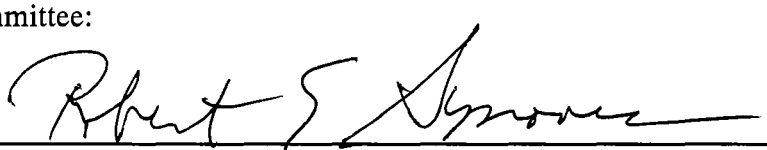
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

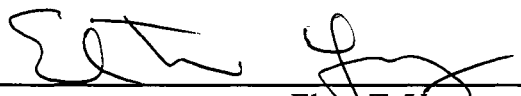


Robert E. Synovec

Reading Committee:



Robert E. Synovec

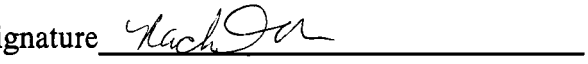


Elton T. Young

James Callis

Date: August 2 2007

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 8/2/2007

University of Washington

Abstract

Discovery Based Yeast Metabolomic Analysis using Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry and Chemometrics

Rachel E. Mohler

Chair of Supervisory Committee:
Professor Robert E. Synovec
Department of Chemistry

Comprehensive two-dimensional (2D) separations are highly beneficial in the characterization of complex samples. The research presented in this dissertation involves the analysis of the small polar molecules comprising the yeast metabolome using commercially available GCxGC-TOFMS instrumentation. The studies described herein are some of the first on metabolomic data using GCxGC-TOFMS and the analysis required a tremendous amount of procedural and software method development. A proof of principle experiment was performed using fermenting and respiring yeast cells. Initially, principal component analysis (PCA) was performed on three selective mass channels (m/z) to identify the metabolite locations exhibiting changes between sample types. Twenty-six metabolite peaks were reported. Following proof that the GCxGC-TOFMS is applicable to the study of yeast metabolite data, an extensive study was performed to determine the ability of GCxGC-TOFMS to detect metabolites and ultimately be utilized to distinguish classes. In this study, a newly developed in-house Fisher ratio method was applied, using the entire 3D data cube (*i.e.,* column 1 retention time, column 2 retention time, m/z), more than doubling the number of quantified metabolite peaks, relative to the PCA study. To obtain the confident identification and accurate metabolite quantification, parallel factor analysis (PARAFAC) was applied. The

statistical significance of the results was determined by applying a Student's *t*-test to the deconvoluted peak volumes. It was determined that 54 identified metabolite peaks were statistically different at the 95% confidence level.

Some cell strains exhibit robust oscillations in gene expression and molecular oxygen. A study with metabolites collected from cells showing periodicity in molecular dissolved oxygen was also performed to ascertain the changes in the metabolome. A new software method was developed to calculate the severity of the periodic pattern without signal bias or limiting the cycling frequency. This method is based on the raw signal intensities and, although it searches all *m/z,* only information from the three most selective *m/z* is output. Peak volumes were determined by PARAFAC and objective grouping of the metabolites with similar patterns was obtained with PCA. Metabolites were shown to cycle with four different patterns, with nineteen cycling with similar frequency but ~180° out of phase.

Table of Contents

List of Figures

List of Tables

Abbreviations

1D ..... one-dimensional

2D ..... two-dimensional

3D ..... three-dimensional

~ ..... approximately equal to

°C ..... degrees Celsius

%RSD ..... percent relative standard deviation

.csv ..... comma separated variable

ϕ ..... phase

µl ..... micro liter

ANOVA ..... analysis of variance

CE x CE ..... comprehensive two-dimensional capillary electrophoresis

d_f ..... film thickness

DIMS ..... direct infusion electrospray mass spectrometry

d-o-f ..... degrees of freedom

dO_2 ..... dissolved oxygen

DR ..... derepressed

FID ..... flame ionization detector

FT-IR ..... fourier transform infrared spectroscopy

FWHM ..... full width at half maximum

GC x GC ..... comprehensive two-dimensional gas chromatography

GC x GC-TOFMS ..... comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry

GRAM..... generalized rank annihilation method

GUI..... graphical user interface

Hz..... hertz

i.d..... inner diameter

ITP x CE..... comprehensive two-dimensional isotachophoresis capillary electrophoresis

kPa..... kilopascal

LC x CE ..... comprehensive two-dimensional liquid chromatography capillary electrophoresis

LC x GC..... comprehensive two-dimensional liquid chromatography gas chromatography

LC x LC ..... comprehensive two-dimensional liquid chromatography

LDI-MS..... laser desorption ionization mass spectrometry

LOD ..... limit-of-detection

min ..... minutes

mg/ml ..... milligrams per milliliter

ml ..... milliliter

ml/min..... milliliters per minute

mM..... millimolar

ms..... millisecond

MS..... mass spectrometer

MSD..... mass spectral detector

MV ..... mass spectral match value

m/z ..... mass to charge ratio or mass channel

N_{DR} ..... concentration in DR samples

N_{DR}/N_R ..... concentration ratio

ng/ml ..... nanograms per milliliter

NIST ..... national institute of standards and technology

NMR ..... nuclear magnetic resonance

N_R ..... concentration in R samples

OD ..... optical density

pA ..... picoamperes

PARAFAC ..... parallel factor analysis

PC ..... principal component

PCA ..... principal component analysis

P_M ..... modulation period

PW ..... pulse-width

R ..... repressed

s ..... seconds

SC ..... synthetic complete

SFC x GC ..... comprehensive two-dimensional supercritical fluid chromatography gas chromatography

SFC x SFC ..... comprehensive two-dimensional supercritical fluid chromatography

S/N ..... signal-to-noise

Sratio ..... signal ratio

TCA..... tricarboxylic acid cycle

TIC ..... total ion count

TMS ..... trimethylsilyl

TOFMS ..... time-of-flight mass spectrometer

t_{R1} ..... retention time on column 1

t_{R2} ..... retention time on column 2

tri-PLS..... trilinear partial least squares

Vratio ..... volume ratio

w_b ..... width at the base

Acknowledgements

For every achievement there is a host of people to thank. I would first like to thank my advisor, Professor Rob Synovec. His interest in research developments helped keep projects moving forward and maximized the learning experience.

My research with metabolomic samples would have been impossible without my collaborators. Thanks, Professor E. "Ted" Young for your support and Ken for growing and extracting the metabolites from fermenting and respiring yeast cells. I appreciate the many hours you put, not only, into growing and extracting the yeast metabolites, but also to answer all of my biochemistry questions. Ben, thanks for growing, extracting and shipping the cycling yeast metabolites from Texas. It is truly amazing how yeast cells can synchronize and then continuously produce and consume metabolites in order to stay alive.

I would also like to thank the past and present Synovec group members. Bryan and Gwen, you two were the instrumentation king and queen. No instrumentation problem was too difficult for you, and I hope I picked up some of your skills. Karisa, I enjoyed those "lab" talks we shared while walking/running around Green Lake. Jamin, all those Matlab discussions were much appreciated. You make programming look so easy.

Finally, I would like to thank my church family for all their support and prayers. Ron, thanks for sharing your dog and truck! I appreciate all you did to help me keep my perspective clear.

Dedication

To my wonderful parents, Robert and Sharon Mohler, who were always there for me and instilled in me the belief, that with determination, common sense and God's help all things are possible.

To my brothers (Bubba, Charlie, Ryan and Rodney) who were always 100% behind all of my endeavors. You are the best brothers any sister could have, although I must admit you all were a *little* overprotective growing up.

Chapter 1. Comprehensive Two-Dimensional Separations with Chemometrics*

1.1 Comprehensive Two-Dimensional Separations

The benefits of two-dimensional (2D) separations over one dimensional (1D) separations for complex sample analysis were noted back in the 1970s.[1] Erni and Frei reported that 1D separations do not provide the selectivity or peak capacity necessary for the study of complex samples such as plant extracts, but if two unrelated, *i.e.* orthogonal, separation mechanisms were utilized for the separation a significant increase in the number of compounds was obtained. As powerful as the 2D separations became, not all of the sample off the first separation dimension was subjected to the second separation mechanism. These coupled systems were either designed to collect column 1 effluent and at a later time inject each column 1 aliquot onto a second column (off-line analysis) or as a heart-cut technique where only a selected fraction of column 1 effluent was sent to column 2. About two decades later, it was realized that in order to obtain an accurate description of complex samples, the entire column 1 effluent must be representatively subjected to separation on the second column in an online manner in order to ideally achieve the multiplicative total of the peak capacities for the two separation column dimensions,[2] referred to as a comprehensive 2D separation.

In the early 1990s the first comprehensive 2D separation instrumentation was developed where the entire column 1 effluent was representatively subjected to separation on the second column while preserving the separation achieved on column 1.

---

* Large portions of this chapter have been submitted for publication:
Karisa M. Pierce, Jamin C. Hoggard, Rachel E. Mohler, Robert E. Synovec. *J. Chromatogr. A.*, 2007

This means all components in the injected sample were quantitatively analyzed while preserving the separation resolution on column 1.[3-6] A more in-depth discussion of comprehensive 2D separations can be found in Chapter 2. To date, a broad range of separation techniques have been interfaced with another separation technique to yield comprehensive 2D separations. Comprehensive 2D separations in order of year reported are LCxLC[3], LCxCE[4], GCxGC[5], SFCxGC[6], LCxGC[7], ITPxCE[8], SFCxSFC[9] and CExCE[10]. In addition to being implemented with traditional univariate detectors, *e.g.* flame ionization detector (FID) and single wavelength absorbance, comprehensive 2D instruments have also been interfaced with multivariate detectors, *e.g.* multi-channel optical absorbance and mass spectrometers.[11-13] One of the first tasks after collection of the 2D data is the visualization of the results. Detectors continually collect 1D data, thus requiring the researcher to reshape the data into the 2D separation space. This reshaping marked the first application of what is known as chemometrics to comprehensive 2D data.

1.2 Chemometrics

Chemometrics, which originated in the 1970s[14], has been loosely defined as "the development and use of mathematical techniques to extract useful information from data acquired through chemical analysis."[15-17] This very broad definition encompasses areas of visualization, method optimization, orthogonality, target analysis, calibration, classification and signal preprocessing. Because 2D separation instrumentation yields such large and complex datasets, human interpretation without the use of chemometric tools is extremely cumbersome if not impossible and the increase in information obtained

by comprehensive 2D separations is beneficial only if the results can be quickly compiled and understood by the users.

As mentioned previously, visualization of the data is a chemometric technique that is qualitative and extremely important. Due to the orthogonality of the two separation dimensions, comprehensive 2D separations yield structured chromatograms that have the potential to give group type information.[18-20] Chemometrics has also been utilized to determine the orthogonality of the two separation dimensions for comprehensive 2D separations.[21] Method optimization is another area which is well suited for chemometrics. Without the use of chemometrics it is extremely difficult to determine optimum instrumentation parameters since the large number of peaks detected may obscure the human ability to visually determine the optimum method. One method reported for optimizing a separation is to pick a few different instrumental conditions and columns and study the effects on the resolution of peaks.[22] A different method recently reported is the use of a closed-loop machine learning based on maximizing the peak count, average nearest neighbor peak distances, average signal to noise and minimizing run time.[23]

Although it is important to optimize a given separation method and obtain the maximum number of peaks with good resolution, it is becoming clear that it is not crucial (and indeed not practical) to obtain completely resolved peaks for all analytes of interest at the cost of increased run-time, especially for very complex samples. Target analysis using peak deconvolution tools have been developed to allow the mathematical resolution of overlapping peaks in order to allow accurate quantitative results while providing

analyte identification information. The generalized rank annihilation method (GRAM) was the first deconvolution technique applied to comprehensive 2D separations[24, 25] and more recently parallel factor analysis (PARAFAC) was used for the deconvolution of overlapping peaks in third order data.[26, 27] PARAFAC will be discussed in more detail in later chapters. In this same theme, the multivariate calibration technique trilinear partial least squares (tri-PLS) was applied to naphtha samples providing quantitative results for group type, even though the individual analytes were present at low chromatographic resolution.[28]

When studying a large number of samples, classification algorithms are frequently utilized to rapidly locate samples in the data set of similar origin as well as areas of the separation space showing class differences. There are two main types of classification algorithms, supervised and unsupervised. In supervised analysis the origin of samples must be known, but in unsupervised algorithms classification occurs without specification of sample origin. Johnson *et. al.* reported an Analysis of Variance (ANOVA) Fisher ratio method to locate differences in jet fuel of known origin. [29] The jet fuel data was also submitted to principal component analysis (PCA), one of the more common unsupervised classification tools, to determine classification.[29] Greater details on classification can be found in chapters 3 and 4.

Before qualitative or quantitative data analysis, however, preprocessing tools must be applied to the raw 2D data to correct for signal fluctuations due to random error, instrumentation fluctuations and detector noise. Types of work in this area include baseline subtraction, normalization, noise filtering and retention time alignment.

Reichenbach et. al. developed GC Image software that removes noise in an automated fashion and with good success.[30] A number of 1D retention time alignment algorithms have been developed, some of which have been adapted and then applied to comprehensive 2D separations [31-35] There have been some chemometric techniques developed for third order data, such as PARAFAC2, that are robust against retention shifting in one of the separation dimensions.[36] Although there have been advancements in chemometrics, there is still a gap between data collection and data interpretation for 2D separations especially in third order data (e.g. GCxGC-TOFMS), *ie.*, going from *data* to *useful information*. The bridge that continues to be built and broadened between the two is user-friendly, high-throughput chemometrics-based software that can readily be used in the pharmaceutical[37], environmental[38], food[39], petrochemical,[19] and biology[40] arenas.

1.3 Metabolome Analysis

One area of biology that has recently experienced an explosion in research is metabolome analysis. Metabolites are all polar molecules produced and consumed by metabolism and include compounds such as sugars, sugar phosphates, vitamins, cofactors, organic acids and amino acids. Although metabolite analysis has been around since the 1970s[41], the term metabolome was not coined until the late 1990s to connect the metabolite analysis with research involving the genome, transcriptome and proteome.[42, 43] Due to the complexity of the metabolome, the analysis has recently been divided into four categories: metabolomics, metabolite profiling, metabolite fingerprinting, and metabolite target analysis. [44-46] Metabolomics has been defined as the analysis of the entire metabolome under a certain set of conditions, but with the currently available

instrumentation this type of analysis is unattainable. A majority of current research is actually metabolite profiling, which is defined as the analysis of specific classes of compounds. Because there is no universal instrumentation system that has the ability to detect all metabolites, the instrumentation used for the study generally dictates which classes of compounds are studied. Metabolite fingerprinting is used to classify samples in the general sense. This analysis does not involve identification or quantification of specific metabolites. As the name suggests, metabolite target analysis involves the quantitative analysis of metabolites, believed to participate in specific parts of metabolism. Current analytical technologies that have been used for metabolome analysis include high performance liquid chromatography with mass spectrometry (HPLC-MS)[47], liquid chromatography with nuclear magnetic resonance (LC-NMR)[48], gas chromatography with mass spectrometry (GC-MS)[49], nuclear magnetic resonance (NMR)[50], direct infusion electrospray mass spectrometry (DIMS)[51], laser desorption ionization mass spectrometry (LDI-MS)[52], Fourier transform infrared spectroscopy (FT-IR)[53], capillary electrophoresis with mass spectrometry[54], and just within the last few years comprehensive 2D gas chromatography (GCxGC-MS).[55, 56]

The driving force behind metabolome analysis involves deciphering the functionality of genes. Many of the genes sequenced and included in the genome of different organisms have unknown function and it is hypothesized that the metabolome can give insight into the function since it is more closely connected to the observed phenotype, Figure 1.1. The importance of this area of research is readily noted by the formation of the Metabolomics Society and meetings that have formed to standardize the

reporting of metabolic data.[57] Also in 2005, a journal entitled *Metabolomics* was introduced that is devoted entirely to metabolomic studies.

1.4 Hypotheses

Comprehensive 2D gas chromatography is one of the most powerful analytical separation techniques, exhibiting both good selectivity and sensitivity. Valve-based comprehensive 2D GCxGC is one of the most compact, robust, and inexpensive GCxGC instrument designs. The major drawback of a valve-based modulation configuration lies in diminished detection sensitivity. This loss in sensitivity is because under typical operating conditions the fraction of the column 1 effluent transferred to the second column is likely to be ~ 5% to 10%. It was envisioned, however, that by simply blocking one of the appropriate ports of the high-speed six-port diaphragm valve that is used as the modulator between columns 1 and 2 an increase in sensitivity would result without sacrificing selectivity. This would not only be due to the injection of the entire sample, but also a focusing of column 1 effluent in the modulation device. To address this loss in sensitivity, we report in Chapter 2 the development of a unique total-transfer (i.e., 100%) valve-based GCxGC, without adding complexity to the instrumentation.

In 2000 the first reports of a GCxGC coupled to a time-of-flight mass spectrometer (TOFMS) were reported.[11, 12] This commercially available instrumentation is not as portable as the valve-based total transfer GCxGC due to the former's cryogenic modulator; however the mass spectrometer allows identification of compounds separated in the 2D chromatographic space. Also the cryogenic modulator does not have the temperature restraints of a valve-based modulator. In addition to the 2D separation data

collected from the GCxGC-FID instrument discussed in Chapter 2, the GCxGC-TOFMS instruments generate full mass spectra at each data point, making it ideal for the in-depth analysis of complex samples. The data cube collected for each sample injected is shown in Figure 1.2. Metabolome analysis was predicted to be a prime candidate for GCxGC-TOFMS and was an area that, at the time of the study, had not been exhaustively analyzed using comprehensive 2D gas chromatography.

All metabolomic research in this dissertation involves yeast species. Yeast has been deemed the model eukaryote for a number of reasons. The first is the ease and speed of yeast cell growth. This growth process is also inexpensive. The second reason is genetic analysis. Both the haploid and diploid strains can be studied, which means that mutations of essential genes that would otherwise be lethal are possible. The mating of genes from different strains is also possible for experimentation. The third reason is there has been extensive research on the biochemistry of yeast with reports dating back to the late nineteenth century when Louis Pasteur reported that yeast can exist under both fermenting and respiring conditions. A final reason yeast was deemed the model eukaryote is that yeast has a relatively small genome with approximately 6,000 genes of which only around 1,000 remain uncharacterized.[58] It has been determined that many functions in yeast, e.g., DNA replication, cell cycle regulation and protein synthesis, are similar to those of higher eukaryotes such as humans. Studies with yeast cells could lead to a better understanding of human diseases and possibly cures for those diseases.[59] The main focus of yeast research presented in this dissertation involves the hypothesis that GCxGC-TOFMS instrumentation is applicable for comprehensive analysis of metabolite

data. It was also predicted that the metabolome would support and enhance the conclusions drawn for gene expression. To reach these goals, cutting edge chemometric tools must be utilized and in some instances developed in-house.

Since it is known that the expression level of more than a quarter of yeast genes is glucose dependent,[60] the proof of principle metabolite study was performed on yeast cells undergoing the diauxic shift, i.e., shift from fermenting to respiring conditions. Metabolome analysis using GCxGC-TOFMS and commercially available unsupervised classification software is discussed in Chapter 3 and a more exhaustive study using an in-house developed supervised classification method in Chapter 4. A study involving yeast cells exhibiting oscillatory behavior in the transcriptome was performed with the hypothesis that the metabolome would oscillate in a manner analogous to the transcriptome and give insight into the origin of the oscillations. As stated in section 1.2 of this chapter the chemometric methods for analysis are still being developed. Since some of the metabolite concentrations are predicted to be very similar between different samples but show large concentration changes over the course of the cycle a new algorithm had to be developed as will be discussed in Chapter 5. Currently available algorithms do not perform optimally when different samples have similar concentrations even if overall there is a large change between most sample types.

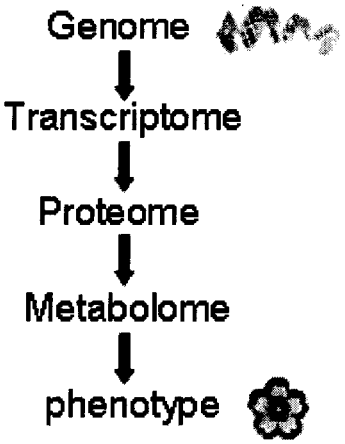


Figure 1.1: Connection of "–ome" series to observed phenotype.
The genome is far removed from the phenotype of a system which could explain the difficulty researchers encounter in determining the functionality of genes. Linking changes in metabolite levels to the specific genes might be a more practical option to determine gene functionality.

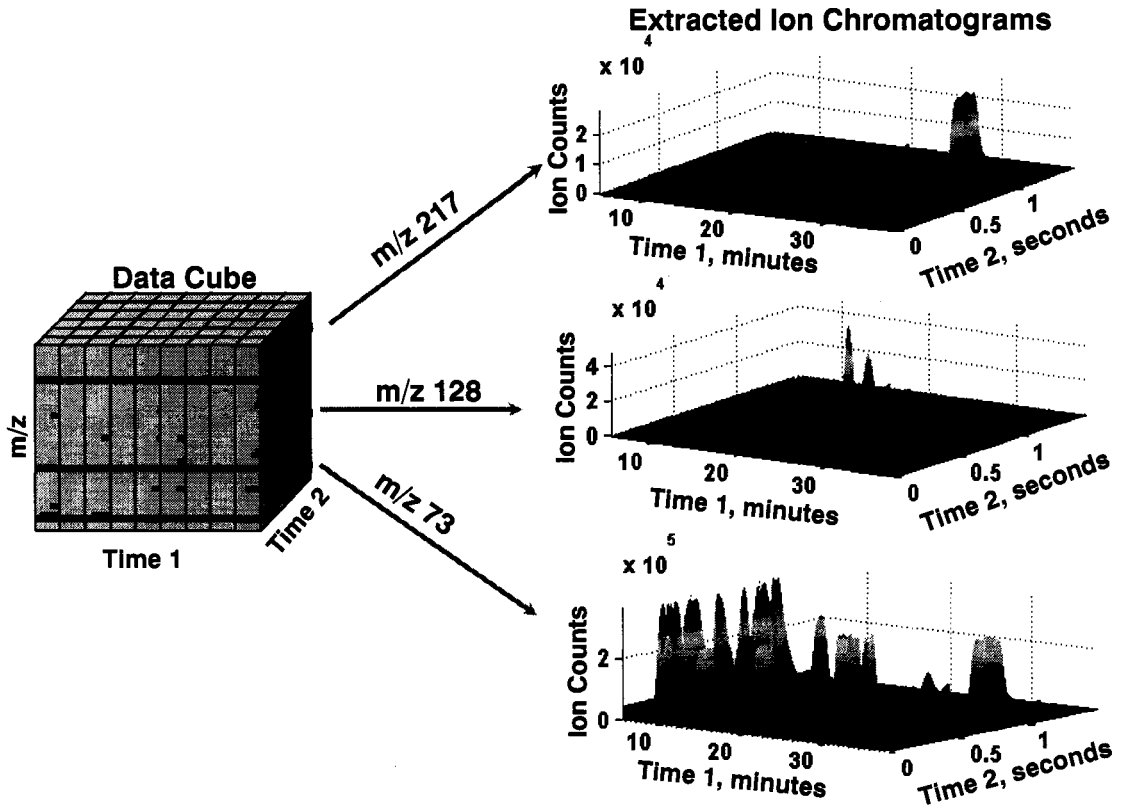


Figure 1.2: Illustration of third order GCxGC-TOFMS data.
A cube of data is collected for each sample injected into the GCxGC-TOFMS. Each *m/z* can be extracted and plotted to obtain relative concentration information.

Notes to Chapter 1

(1) Erni, F.; Frei, R. W. *J. Chromatogr.* **1978**, *149*, 561-569.

(2) Giddings, J. C. *Unified Separation Science*; Wiley-Interscience: New York, 1991.

(3) Bushey, M. M.; Jorgenson, J. W. *Anal. Chem.* **1990**, *62*, 161-167.

(4) Bushey, M. M.; Jorgenson, J. W. *Anal. Chem.* **1990**, *62*, 978-984.

(5) Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227-231.

(6) Liu, Z.; Ostrovsky, I.; Farnsworth, P. B.; Lee, M. L. *Chromatographia* **1993**, *35*, 567-573.

(7) Quigley, W. W. C.; Fraga, C. G.; Synovec, R. E. *J. Microcolumn Sep.* **2000**, *12*, 160-166.

(8) Chen, S.; Lee, M. L. *Anal. Chem.* **2000**, *72*, 816-820.

(9) Hirata, Y.; Hashiguchi, T.; Kawata, E. *J. Sep. Sci.* **2003**, *26*, 531-535.

(10) Michels, D. A.; Hu, S.; Dambrowitz, K. A.; Eggertson, M. J.; Lauterbach, K.; Dovichi, N. J. *Electrophoresis* **2004**, *25*, 3098-3105.

(11) Dimandja, J.-M. D.; Grainger, J.; Jr., D. G. P.; Turner, W. E.; Needham, L. L. *J. Exposure Analysis and Environmental Epidemiology* **2000**, *10*, 761-768.

(12) van Deursen, M.; Beens, J.; Reijenga, J.; Lipman, P.; Cramers, C.; Blomberg, J. *J. High Resol. Chromatogr.* **2000**, *23*, 507-510.

(13) de Koning, S.; Janssen, H.-G.; Deursen, M. v.; Brinkman, U. A. T. *J. Sep. Sci.* **2004**, *27*, 397-409.

(14) Kowalski, B. R. *J. Chem. Inf. Comp. Sci.* **1975**, *15*, 201-203.

(15) Kramer, R. *Chemometric Techniques for Quantitative Analysis*; Marcel Dekker: New York, 1998.

(16) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; Wiley-Interscience: New York, 1998.

(17) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*; Wiley: New York, 1986.

(18) Janssen, H.-G.; Boers, W.; Steenbergen, H.; Horsten, R.; Floter, E. *J. Chromatogr. A* **2003**, *1000*, 385-400.

(19) Blomberg, J.; Schoenmakers, P. J.; Beens, J.; Tijssen, R. *J. High Resol. Chromatogr.* **1997**, *20*, 539-544.

(20) Dugo, P.; Skerikova, V.; Kumm, T.; Trozzi, A.; Jandera, P.; Mondello, L. *Anal. Chem.* **2006**, *78*, 7743-7750.

(21) van Gyseghem, E.; Dejaegher, B.; Put, R.; Forlay-Frick, P.; Elkihel, A.; Daszykowski, M.; Heberger, K.; Massart, D. L.; Heyden, Y. V. *J. Pharm. Biomed. Anal.* **2006**, *41*, 141-151.

(22) Beens, J.; Blomberg, J.; Schoenmakers, P. J. *J. High Resol. Chromatogr.* **2000**, *23*, 182-188.

(23) O'Hagan, S.; Dunn, W. B.; Knowles, J. D.; Broadhurst, D.; Williams, R.; Ashworth, J. J.; Cameron, M.; Kell, D. B. *Anal. Chem.* **2007**, *79*, 464-476.

(24) Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **1998**, *70*, 2796-2804.

(25) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **2000**, *72*, 4154-4162.

(26) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1056*, 145-154.

(27) Sinha, A. E.; Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1027*, 269-277.

(28) Prazen, B. J.; Johnson, K. J.; Weber, A.; Synovec, R. E. *Anal. Chem.* **2001**, *73*, 5677-5682.

(29) Johnson, K. J.; Synovec, R. E. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 225-237.

(30) Reichenbach, S. E.; Ni, M.; Zhang, D.; Jr., E. B. L. *J. Chromatogr. A* **2003**, *985*, 47-56.

(31) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **2000**, *72*, 4154-4162.

(32) Johnson, K. J.; Prazen, B. J.; Young, D. C.; Synovec, R. E. *J. Sep. Sci.* **2004**, *27*, 410-416.

(33) Micyus, N. J.; Seeley, S. K.; Seeley, J. V. *J. Chromatogr. A* **2005**, *1086*, 171-174.

(34) Krebs, M. D.; Kang, J.-M.; Cohen, S. J.; Lozow, J. B.; Tingley, R. D.; Davis, C. E. *Sensors and Actuators B* **2006**, *119*, 475-482.

(35) Pierce, K. M.; Wood, L. F.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2005**, *77*, 7735-7743.

(36) Bro, R.; Andersson, C. A.; Kiers, H. A. L. *J. Chemometrics* **1999**, *13*, 295-309.

(37) Guttman, A.; Varoglu, M.; Khandurina, J. *Drug Discovery Today* **2004**, *9*, 136-144.

(38) Paníc, O.; Górecki, T. *Anal. Bioanal. Chem.* **2006**, *386*, 1013-1023.

(39) Tranchida, P. Q.; Dugo, P.; Dugo, G.; Mondello, L. *J. Chromatogr. A* **2004**, *1054*, 3-16.

(40) Evans, C. R.; Jorgenson, J. W. *Anal. Bioanal. Chem.* **2004**, *378*, 1952-1961.

(41) Horning, E. C.; Horning, M. G. *J. Chromatogr. Sci.* **1971**, *9*, 129-140.

(42) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends Biotechnology* **1998**, *16*, 373-378.

(43) Tweeddale, H.; Notley-McRobb, L.; Ferenci, T. *J. Bacteriology* **1998**, *180*, 5109-5116.

(44) Fiehn, O. *Comparative and Functional Genomics* **2001**, *2*, 155-168.

(45) Nielsen, J.; Oliver, S. *Trends Biotechnology* **2005**, *23*, 544-546.

(46) Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. *Analyst* **2005**, *130*, 606-625.

(47) Buchholz, A.; Takors, R.; Wandrey, C. *Anal. Biochem.* **2001**, *295*, 129-137.

(48) Wolfender, J. L.; Ndjoko, K.; Hostettmann, K. *J. Chromatogr. A* **2003**, *1000*, 437-455.

(49) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nature Biotechnology* **2000**, *18*, 1157-1161.

(50) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181-1189.

(51) Vaidyanathan, S.; Kell, D. B.; Goodacre, R. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 118-128.

(52) Vaidyanathan, S.; Jones, D.; Broadhurst, D. I.; Ellis, J.; Jenkins, T.; Dunn, W. B.; Hayes, A.; Burton, N.; Oliver, S. G.; Kell, D. B.; Goodacre, R. *Metabolomics* **2005**, *1*, 243-250.

(53) Harrigan, G. G.; LaPlante, R. H.; Cosma, G. N.; Cockerell, G.; Goodacre, R.; Maddox, J. F.; Luyendyk, J. P.; Ganey, P. E.; Roth, R. A. *Toxicology Letters* **2004**, *146*, 197-205.

(54) Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. *Anal. Chem.* **2002**, *74*, 2233-2239.

(55) Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Lidstrom, M. E.; Synovec, R. E. *Talanta* **2005**, *65*, 380-388.

(56) Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O. *Metabolomics* **2005**, *1*, 65-73.

(57) Castle, A. L.; Fiehn, O.; Kaddurah-Daouk, R.; Lindon, J. C. *Briefings Bioinformatics* **2006**, *7*, 159-165.

(58) Pena-Castrillo, L.; Hughes, T. R. *Genetics* **2007**, *176*, 7-14.

(59) Williams, N. *Science* **1996**, *272*, 481.

(60) Young, E. T.; Dombek, K. M.; Tachibana, C.; Ideker, T. *J. Biol. Chem.* **2003**, *278*, 26146-26158.

Chapter 2. Total-Transfer, Valve-Based Comprehensive Two-Dimensional Gas Chromatography[†]

2.1 Introduction

Comprehensive two-dimensional gas chromatography (GCxGC) separates analytes in the gas phase using two columns that provide complementary separations, which are as orthogonal (independent) as possible.[1] Aliquots of effluent from the first column, referred herein as column 1, are repeatedly transferred to the second column, *i.e.*, column 2, at a rapid synchronized interval called the modulation period, P_M, for separation on column 2. Each peak eluting from column 1 is sampled on column 2 multiple times. The separations leaving column 2 are generally sent to a flame ionization detector (FID) or mass spectral detector (MSD). The resulting separation data, which can be viewed as a two-dimensional (2D) separation, has increased peak capacity over either "standard" single column GC or heart-cutting GC separations. In order to increase the peak capacity from what would be achieved in standard one-dimensional (1D) GC, the GCxGC separation provided by each column must be as independent as possible of the other, and the separation information on column 1 must be preserved.[2] The separation on column 1 is often provided by a non-polar stationary phase and based primarily on boiling point, while the separation on column 2 often utilizes a stationary phase of greater polarity and separates with more sensitivity to chemical functional groups. The modulator interface, which is placed between the two columns, collects and injects narrow plugs of the column 1 effluent onto column 2. Preservation of the column 1

[†] Large portions of this chapter have been published:
Rachel E. Mohler, Bryan J. Prazen, Robert E. Synovec Analytica Chimica Acta, 2006, 555, 68-74.

separation (*i.e.*, information) occurs when the column 1 effluent is sampled a minimum of three to four times over the course of the elution of any given analyte peak.[3]

Since the introduction of GCxGC in 1991,[5] various configurations have evolved with most varying at the modulator interface between columns 1 and 2. The initial GCxGC pioneered by Phillips and Liu was configured using a thermal modulator.[5] The second modulator configuration developed, and currently the most widely used, is the cryogenic thermal interface.[6, 7] A third interface configuration involves a valve-based modulator. In this configuration a diaphragm valve is generally used and allows independent control of flows on columns 1 and 2.[8-11] Injection from column 1 onto column 2 occurs by actuating the valve. Under the most sensitive yet restrictive operating conditions, approximately 80% of the sample injected into the head of column 1 reaches the detector in diaphragm valve-based GCxGC.[9] Typically, however, ~ 5% to 10% of the sample passing through column 1 reaches the detector. The separations can still be comprehensive at this low sample transfer efficiency, since the modulation period can be chosen such that all analytes eluting from column 1 are adequately sampled.[3,8] This incomplete sampling from column 1 to column 2 at the modulator interface leads to a lower sensitivity than either the thermal or cryogenic modulators, but the valve-based system(s) have the attributes of being relatively inexpensive, compact and have the potential to be more readily used in a portable instrument. In related work, a recent valve configuration has been developed that is analogous to a Dean's switch.[12] A three-way solenoid valve outside the sample path controls the direction of sample flow allowing 100% transfer from column 1 to column 2.

Recently, there has been increased interest in stopping the flow in GC instruments to obtain better selectivity and separation of components. Sacks and co-workers developed an in-series coupled column ensemble where the flow was periodically stopped on column 1 to improve the selectivity and resolution of components.[13,14] It was shown that stopping the flow on column 1 up to 10 s could be used without major increases in band broadening. Also Górecki and co-workers developed a stop-flow GCxGC-TOFMS configuration that uses both a six-port valve and a cryogenic trap.[15] The focus was to extend the period of time for the second dimension separation, *i.e.*, a programmable modulation period, without sacrificing the separation in the primary dimension. This configuration allowed a longer column 2 separation without under-sampling the column 1 separation.

Certainly, there is a growing need for a GCxGC that provides the advantages of a valve-based GCxGC along with optimum detection sensitivity acheieved by 100% mass transfer from column 1 to column 2. While these attributes have been provided to a significant extent by the previously mentioned Dean's switch-type modulation configuration,[12] the same attributes can be readily obtained based on the diaphragm valve used in previously reported valve-based GCxGC,[8-11] and with narrower peaks and thus better detection sensitivity. This configuration, as reported herein, will be referred to as total-transfer valve-based GCxGC. The basic idea of the total-transfer valve-based GCxGC is to have all the mass that enters column 1 go through column 2 and to detection, and operates as follows. The valve-based GCxGC is used with one of the high-speed diaphragm valve ports appropriately blocked, with the head pressure on

column 1 higher than the head pressure on column 2 (see Figure 2.1). As such, during each modulation period, effluent from column 1 is compressed in the valve loop prior to injection onto column 2. It is during this compression that material is moving through column 1, and the column 1 separation occurs. The length of the modulation period P_M then determines the extent of effluent compression prior to injection onto column 2. This configuration allows one to utilize the compact features of valve-based GCxGC while achieving 100% effluent transfer from column 1 to 2. Performance of this instrumental design was evaluated using test mixtures of standard compounds. Quantitative precision, peak-widths at half height and limit of detection (LOD) are reported. The instrument was also evaluated using natural complex samples (gasoline and an essential oil).

2.2 Experimental

2.2.1 Valve configuration

A diagram of the heart of the total-transfer valve-based GCxGC is shown in Figure 2.1. The total-transfer valve-based GCxGC, consisted of a HP 6890 Gas Chromatograph equipped with an HP 7673A automatic sampler (Agilent, Palo Alto, CA, USA) and a high-speed six-port diaphragm valve (VICI, Valco, Houston, TX, USA) that was face mounted on the oven so that the six ports were the only portion of the valve inside the oven. This valve mounting allows separation temperatures up to at least 275 °C without degradation of the valve's three temperature sensitive O-rings.[10] Column 1 in Figure 2.1 was a 10 m x 180 μm i.d. x 0.25 μm capillary column with a 5% diphenyl/95% dimethylpolysiloxane film (DB-5 J&W; Scientific/Agilent Technologies, USA), and column 2 was a 3 m x 250 μm i.d. x 0.25 μm capillary column with a

polyethylene glycol film (DB-WAX; J&W Scientific/Agilent Technologies, USA). Hydrogen gas was used as the carrier gas for both columns with the column 2 controlled by an auxiliary flow (uncoupled from column 1).

Upon injection onto column 1 using the automatic sampler, sample eventually travels through column 1 and into the modulator (diaphragm valve) sample loop. At regular intervals the modulator valve interfacing the two columns was actuated and the contents of the sample loop were directed into column 2 using the hydrogen gas auxiliary flow. Using the six port valve configuration, the head pressure on column 1 must be higher than the head pressure on column 2, ensuring compression instead of diffusion of column 1 effluent into the sample loop. The column head pressures for each experiment will be specified. An in-house designed control box was used to signal valve actuation, which defined the modulation period, P_M. The time length of the injection pulse onto column 2 is defined as the pulse-width, PW. The analytes were detected using a FID and data collected at an acquisition rate of 200 Hz. For data collection, the P_M defined the length of time for each column 2 separation and it included the PW. The PW was kept short to minimize backup into column 1 but long enough to ensure all the loop contents were injected. The difference between the P_M and PW defines the time that column 1 effluent was compressed into the sample loop.

One hypothesis was that the P_M would control the operation of the instrument, with a longer P_M providing higher detection sensitivity, with fewer P_M required to sample an analyte peak leaving column 1. Furthermore, if the PW was sufficiently long so the sample loop contents were fully injected onto column 2, it was hypothesized that

instrument operation should be independent of PW. These hypotheses were investigated experimentally (second study described below).

2.2.2 Instrumental conditions

In the first study, we present proof-of-principle that the total-transfer valve-based GCxGC operates successfully, and relate the data to a single-column GC separation, in particular, the separation on only column 1 of the GCxGC at the same linear flow velocity. For this study, the GCxGC was set to an isothermal temperature of 120 °C using a column 1 head pressure of 280 kPa and a column 2 head pressure of 97 kPa. The P_M was set to 1.0 s with a PW of 20 ms. A fourteen component mixture was used: n-hexane, n-heptane, octane, nonane, decane, methyl ethyl ketone, 2-pentanone, methanol, ethyl alcohol, 1-propanol, benzene, toluene, ethylbenzene, and propylbenzene. All standard components except for methanol, benzene (Fisher Scientific, Fairlawn, NJ, USA), methyl ethyl ketone and toluene (J.T.Baker, Phillipsburg, NJ, USA) were obtained from Sigma-Aldrich (St. Louis, MO, USA). For the single-column GC separation, the same retention times were achieved as in the column 1 portion of the GCxGC separation, using a head pressure of 11 kPa with the same DB-5 column used with the GCxGC mentioned previously, but with column 1 going directly to the FID.

In the second study, the dependence of the total-transfer valve-based GCxGC upon the P_M, PW and the difference in pressure from column 1 to column 2 was evaluated. The P_M was investigated using the following values: 1.5, 2, 3 or 6 s. However, only the results from a P_M of 1.5 s and 6 s will be shown for brevity, with the conclusions drawn representative of remaining results. In this study, a PW of 20, 40 and

60 ms were used. Column 1 was operated under a constant pressure of 450 kPa. Column 2 had a constant pressure of 210 kPa for all modulation periods. The difference in column 1 to column 2 pressure was also studied by keeping PW, P_M and the column 2 pressure (210 kPa) constant and lowering the column 1 pressure from 450 to 340 kPa. The inlet septum purge was delayed for 12 s after the splitless injection. The initial oven temperature was set to 40 °C and then increased at 5 °C/min to a final temperature of 150 °C. A split-less 1-μl injection of the following ten components was made: 1-pentanol, 2-pentanone, 2-hexanone, 2-heptanone, 3-octanone, octane, nonane, decane, toluene and ethylbenzene. All the standard components except toluene were purchased from Sigma-Aldrich (St. Louis, MO, USA). Toluene was purchased from J.T. Baker (Phillipsburg, NJ, USA). All components were injected at a concentration of 500 ng/μl except 2-pentanol and 2-hexanone, which had concentrations of 800 ng/μl.

2.2.3 Natural Samples

A gasoline sample and Eucalyptus Oil were two natural samples used to demonstrate this new GCxGC configuration performance. The gasoline sample was run with a 10:1 split, a P_M of 3 s, and PW of 40 ms. The head pressure was held constant at 480 kPa on column 1 and 210 kPa on column 2. The temperature increased from 40-200°C at 5°/min. Eucalyptus oil was run with a 25:1 split ratio, a P_M of 1.5 s, and PW of 40 ms. Column 1 had an inlet pressure of 410 kPa and a column 2 inlet pressure of 210 kPa. The temperature was held at 70°C for 8 minutes and then ramped at 5°C/min to a final temperature of 200°C where it was held for 6 minutes.

2.3 Results and Discussion

2.3.1 Proof-of-principle

Unlike the previous valve-based GCxGC design(s), with a constant outlet flow velocity for column 1, the total-transfer valve-based GCxGC has a non-constant column 1 outlet flow velocity. A schematic showing the conceptualized flow velocity profiles at the end of column 1 and head of column 2 are shown in Figures 2.2A and B, respectively. There are two valve configurations: load and inject. Assume for the moment that effluent from column 1 is loaded in the valve sample loop at a pressure higher than the head pressure on column 2. When the valve is then turned to the inject mode, there will be a spike in the flow velocity at the head of column 2 as depicted in Figure 2.2B. This spike occurs as sample is pushed out of the valve loop onto column 2. Following injection onto column 2 at the specified pulse width PW, the valve is switched back to the load mode, for a time period equal to P_M minus PW and there is a spike then decay in the flow at the end of column 1 as depicted in Figure 2.2A. Effluent leaving column 1 (and simultaneously separated by column 1) is compressed into the valve loop, while the separation of the previous injection occurs on column 2. With a six port valve configuration, the longer the P_M, the greater the decay in flow velocity, as column 1 effluent is compressed into the sample loop to a greater extent. It is important to note, however, that this flow into the sample loop rapidly approaches zero as shown in Figure 2.2A, as gas is constantly being compressed into the sample loop. The sample loop is never empty when it changes from the inject mode to the load mode since there is the column 2 head pressure in the loop, isolated from the column 2 outlet at that juncture,

that the column 1 effluent must compress. The reduction of the column 1 flow rate results in the subsequent increased sensitivity in detection at the end of column 2, since the quantity of effluent compressed is directly proportional to the pressure at constant sample loop volume and constant temperature. Given this conceptual understanding of how the instrument should function, the ideas were evaluated by systematically changing PW, P_M, and the difference in column 1 to column 2 pressure.

First, we will demonstrate that the total-transfer valve-based GCxGC provides a separation, and compare the separation provided by column 1 with the GCxGC to the separation provided by the same column in a single-column GC, as outlined in first study in the Experimental Section. The single-column GC separation was run at the same flow velocity as the net velocity on column 1 of the total-transfer GCxGC separation. In this proof-of-principle study, the flow velocity on column 1 was suboptimal, serving primarily to compare the pressures needed for equivalent retention times on column 1 in the total-transfer GCxGC mode versus the single-column GC mode. For the total-transfer valve-based GCxGC, Figure 2.3 shows a contour plot of the 14-component mixture consisting of alkanes, alcohols, alkyl benzenes and ketones. As in the typical GCxGC isothermal separation, the homologous series of various compound classes follow along lines. This 2D separation was summed onto the first dimension for comparison to a standard 1D GC separation as shown in Figures 2.4A and B. As mentioned, the separations were run at the same net flow velocity on column 1. To achieve the same net flow velocity on column 1, the total-transfer GCxGC required column 1 with a head pressure of 280 kPa, combined with a column 2 head pressure of 97

kPa, and the single-column GC required a pressure of 11 kPa. The conclusion from Figure 2.4 is that the new GCxGC configuration provides a 2D separation, and when summed onto the column 1 axis can be compared to the single-column GC separation run at the same net flow rate (i.e., the pressure requirements can be compared).

2.3.2 Pulse width and modulation period

For the second study of the instrument dependence on PW and P_M, a ten-component mixture consisting of alkanes, ketones, alkyl benzenes and 2-pentanol was used. These components were chosen to represent a variety of functional groups and to show the elution trends within each functional group type. A contour plot showing the ten components separated by total-transfer valve-based GCxGC is shown in Figure 2.5. Theoretically, at a column 2 head pressure of 210 kPa, the sample loop should empty in 4.4 ms, which means that the sample loop should empty during every column 2 injection with PW > 4.4 ms, and the amount of sample injected should be independent of the PW. This is what is seen experimentally. Varying the PW from 20 to 60 ms while keeping the P_M constant at 1.5 s yielded similar separations, essentially eight modulation periods ("slices") across each analyte leaving column 1, as shown in Figures 2.6A and B, respectively. The number of first dimension slices is the same and the integrated signal is approximately the same with a difference of less than 10%. With the particular valve used, lengthening the PW does not yield any significant increase of the amount of sample injected into column 2 since all effluent leaving column 1 enters the valve loop and is subsequently injected onto column 2. Note, one would expect the longer PW would result in slightly longer retention times on column 1, and this is observed. A related

experiment was performed in which a PW of 20 ms, a P_M of 1.5 s and column 2 pressure of 210 kPa were held constant as in Figure 2.6A, but the column 1 pressure was decreased from 450 kPa to 340 kPa. At this lower column 1 pressure (i.e., closer to the column 2 pressure), decane and 3-octanone required more modulation periods, with the peaks eluting about 2 minutes later on column 1 as expected based upon the mechanism of coupling the two columns together in the total-transfer mode with the particular valve used for modulation (figure not shown for brevity).

However, the amount of sample sent to column 2 is very dependent on the P_M as shown in Figures 2.7A and B. Figure 2.7A shows that at a constant PW of 40 ms and P_M of 1.5 s (and column pressures as in Figures 2.6A and B) there are approximately seven slices of the decane peak sent to column 2 with the most intense slice at approximately 800 pA. When the P_M is increased to 6.0 s and the PW kept constant, the number of slices of the decane peak decreases by about half and the most intense slice has an intensity of approximately 1300 pA as shown in Figure 2.7B. As stated previously there is an increased S/N at the longer P_M due to a greater amount of sample being introduced into column 2 as well as a higher temperature at the time of the second column injection. This results in a lower LOD.

Even though the number of slices decreases in going from the shorter to longer P_M the quantitative accuracy and precision is maintained. Table 2.1 gives the integrated signal at a P_M of 1.5 s and 6 s for one compound from each of the classes sampled relative to the internal standard octane. There is good quantitative agreement between the two P_M, for a given compound. The results for 2-pentanol varied the most, likely due to a

strong interaction between the alcohol and the highly polar column 2 and possibly adsorption in the valve modulator, with the peak tailing impacting quantification. Other important figures-of-merit for this total-transfer configuration at a P_M of 1.5 s and 6 s are given in Tables 2.2 and 2.3, respectively. With an increase in the P_M from 1.5 s to 6 s the full width at half maximum (FWHM) decreases from an average of 59 ms at a P_M of 1.5 s, to an average of 29 ms at a P_M of 6 s. This is a little over a factor of two decrease which leads to an improved signal intensity. The LOD decreased by a factor of two from an average of 0.21 to 0.098 ng/μl. Another observation is that the longer P_M results in a longer run time on column 1. Certainly, the results are impetus for further study of the various instrument configurations.

2.3.3 Natural sample separations

In order to further demonstrate the utility of the total-transfer valve-based GCxGC, two natural samples were separated. These two samples were gasoline and Eucalyptus oil, in Figure 2.8A and B, respectively. The results of these separations were satisfactory, with the alkane, alkyl benzenes and substituted aromatics homologous series clearly visible in the gasoline sample. The gasoline separation in Figure 2.8A was run with a P_M of 3 s, a PW of 40 ms with all components eluting in 20 minutes. Although the use of the primary separation space is not as efficient, the separation and run time are comparable to that obtained by Harynuk and Górecki[15] where a valve was used in combination with a cryogenic modulator. An acceptable orthogonal separation of an Eucalyptus oil was also obtained in Figure 2.8B, with a P_M of 1.5 s, a PW of 40 ms. All components eluted in 15 minutes.

2.4 Conclusions

A total-transfer valve-based GCxGC was developed and the overall performance tested. The major advantage to increasing the modulation period is the improved LOD, and thus improved sensitivity of valve-based GCxGC. A LOD of 0.03 ng/μL was obtained for octane at a PW of 40 ms and a P_M of 6.0 s (Table 2.3). Another advantage is that columns 1 and 2 are fully decoupled, so column 2 can be operated at its optimum linear flow velocity, which is not possible with non-stop-flow systems. Further study of this advantage is warranted. While the present system employed a six port valve, a simpler approach could implement a three-way valve, thus relaxing the requirement with the current system that the column 1 pressure be higher than the column 2 pressure. With this new GCxGC configuration, a number of possibilities for applications requiring more sensitive, relatively inexpensive and portable GCxGC instrumentation are feasible.

Table 2.1: Dependence of peak volumes on P_M.
Signal ratios are the volume of the analytes (sum of all baseline corrected peaks for a given analyte) relative to an internal standard (octane) as a function of the P_M indicated, at constant PW of 40 ms from data shown in Figures 2.7A and B.

Compound	Volumn Ratio (P_M 1.5 s)	Volumn Ratio (P_M 6 s)
2-pentanol	0.77 ± 0.008	0.68 ± 0.016
toluene	0.59 ± 0.003	0.56 ± 0.004
decane	0.61 ± 0.003	0.62 ± 0.004
3-octanone	0.45 ± 0.003	0.46 ± 0.007

Table 2.2: Figures-of-merit at P_M 1.5.

The figures-of-merit for a P_M of 1.5 s and a PW of 40 ms, for data as in Figure 2.7A. The full-width-at-half-maximum (FWHM) and limit of detection (LOD) are reported and the standard deviation based on four replicate injections. The LOD was determined based upon finding the standard deviation of the noise adjacent to the peaks, multiplying by 3 and setting up a ratio using the signal of a known concentration.

Compound	Column 2 FWHM (ms)	LOD (ng/μl)
2-pentanone	45.13 ± 0.25	0.25 ± 0.07
2-pentanol	86.88 ± 0.63	0.34 ± 0.06
toluene	50.25 ± 0.29	0.14 ± 0.04
octane	23.50 ± 0.41	0.03 ± 0.00
2-hexanone	59.25 ± 0.65	0.34 ± 0.13
ethylbenzene	57.88 ± 0.63	0.13 ± 0.03
nonane	25.50 ± 0.00	0.08 ± 0.02
2-heptanone	71.00 ± 2.80	0.28 ± 0.03
decane	27.50 ± 0.41	0.12 ± 0.05
3-octanone	68.38 ± 1.18	0.42 ± 0.19

Table 2.3: Figures-of-merit at P_M 6.0.
The figures-of-merit at a P_M of 6.0 s and a PW of 40 ms, for data as in Figure 2.7B. The full-width-at-half-maximum (FWHM) and limit of detection (LOD) are reported and the standard deviation was based on four replicate injections. The LOD was determined as described in Table 2.2.

Compound	Column 2 FWHM (ms)	LOD (ng/µl)
2-pentanone	28.31 ± 0.69	0.24 ± 0.09
2-pentanol	38.69 ± 0.94	0.15 ± 0.04
toluene	28.88 ± 0.25	0.07 ± 0.01
octane	22.06 ± 0.43	0.03 ± 0.01
2-hexanone	31.56 ± 0.52	0.15 ± 0.04
ethylbenzene	29.81 ± 0.47	0.06 ± 0.02
nonane	22.50 ± 0.41	0.03 ± 0.01
2-heptanone	33.94 ± 0.72	0.08 ± 0.02
decane	23.75 ± 0.29	0.05 ± 0.02
3-octanone	32.81 ± 0.24	0.12 ± 0.04

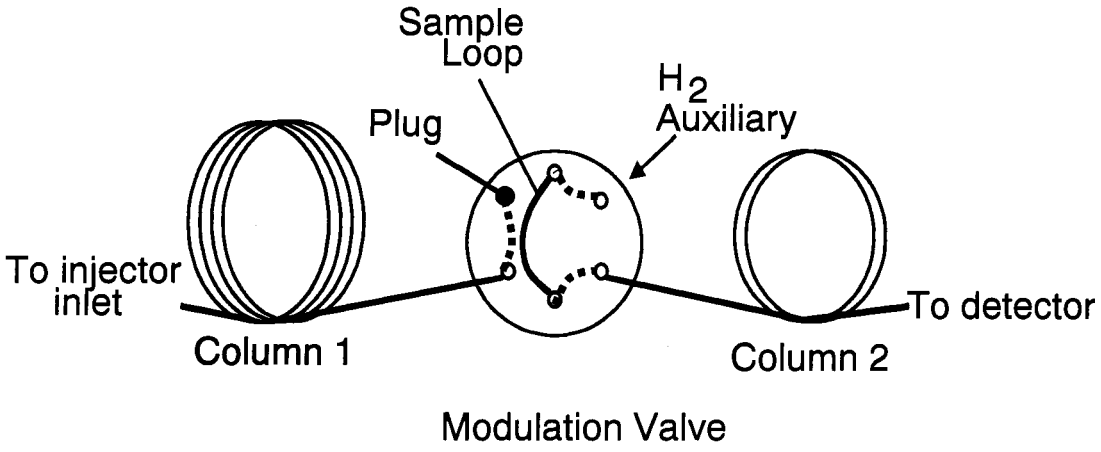


Figure 2.1: Diagram of the valve modulator.
The dotted lines show the port connection in the column 2 inject mode for total transfer valve-based GCxGC. Operation is explained in the text. The main requirement is that the head pressure on column 1 is sufficiently higher than the head pressure on column 2 (hydrogen gas reference).

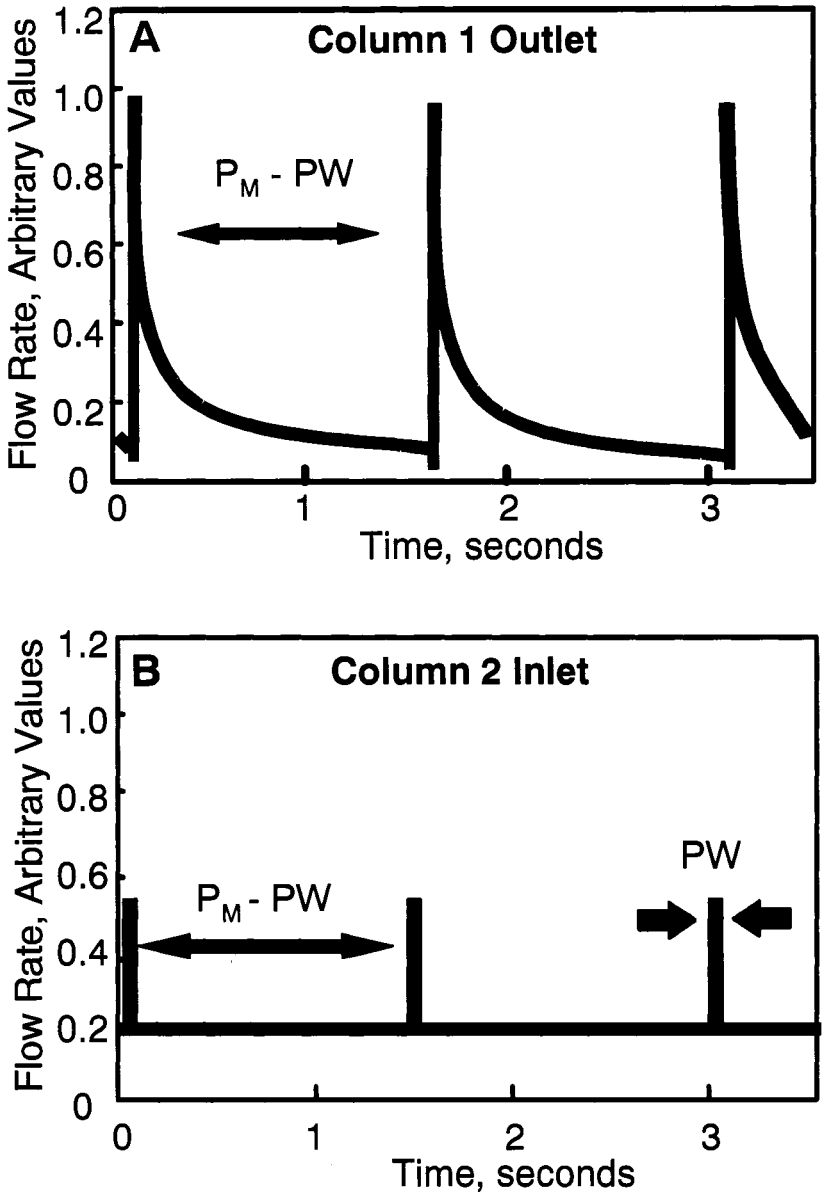


Figure 2.2: Flow profile schematic.

(A) Schematic of the conceptual flow velocity profile at the end of column 1, i.e., in the sample loop as explained in the text. (B) Schematic of the conceptual flow profile at the head of column 2 as explained in the text. Note that the areas under the curve for (A) and (B) are for conceptual illustration purposes only and should not be quantitatively compared.

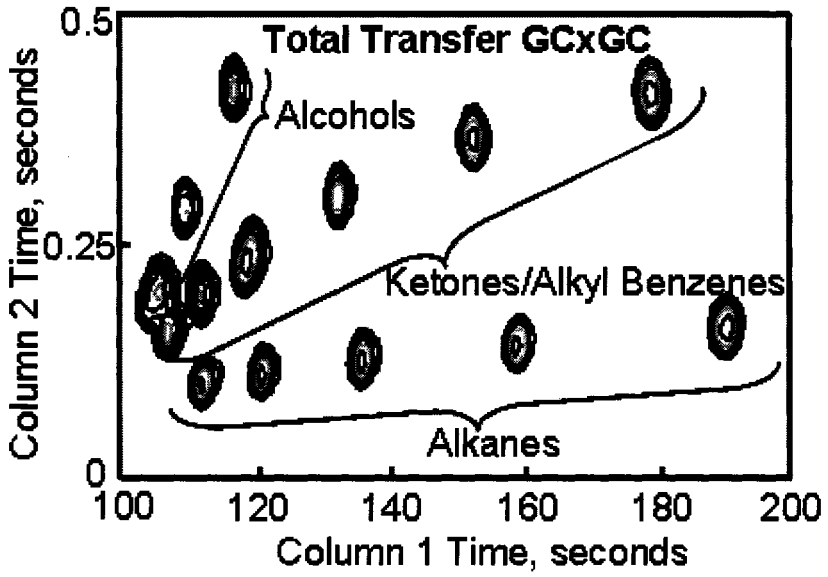


Figure 2.3: A contour plot of the 14 component mixture.

Contour plot of a 14-component mixture separated using the total-transfer valve-based GCxGC at P_M of 1.0 s and PW of 20 ms. Column 1 was at a head pressure of 280 kPa and column 2 at a head pressure of 97 kPa.

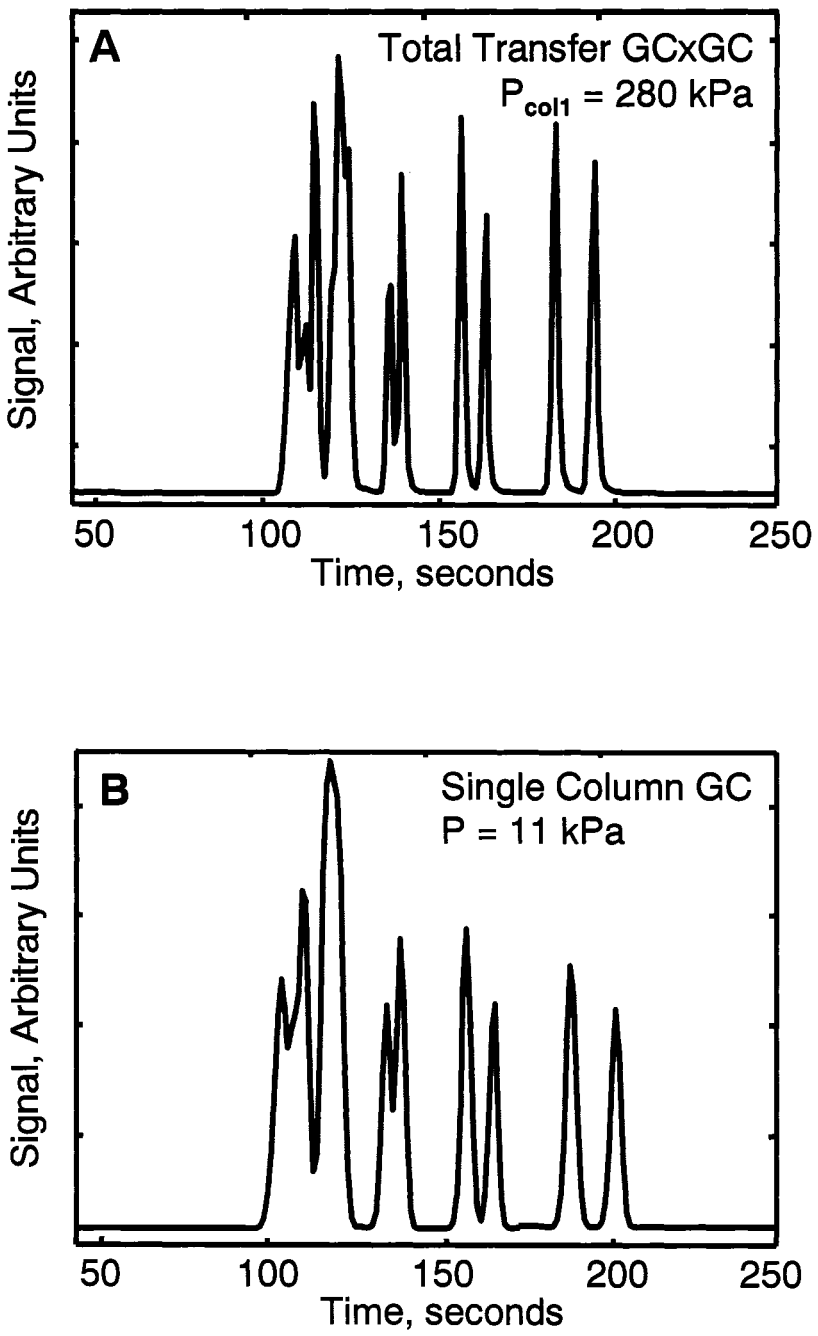


Figure 2.4: Comparison of conventional GCxGC and total transfer GCxGC.

(A) One-dimensional view of the GCxGC separation in Figure 2.3 with column 2 separations summed onto the column 1 time axis. (B) Single-column GC separation of the same 14-component mixture as in (A) at the same average flow velocity of column 1 of the GCxGC. A single-column head pressure of 11 kPa was required.

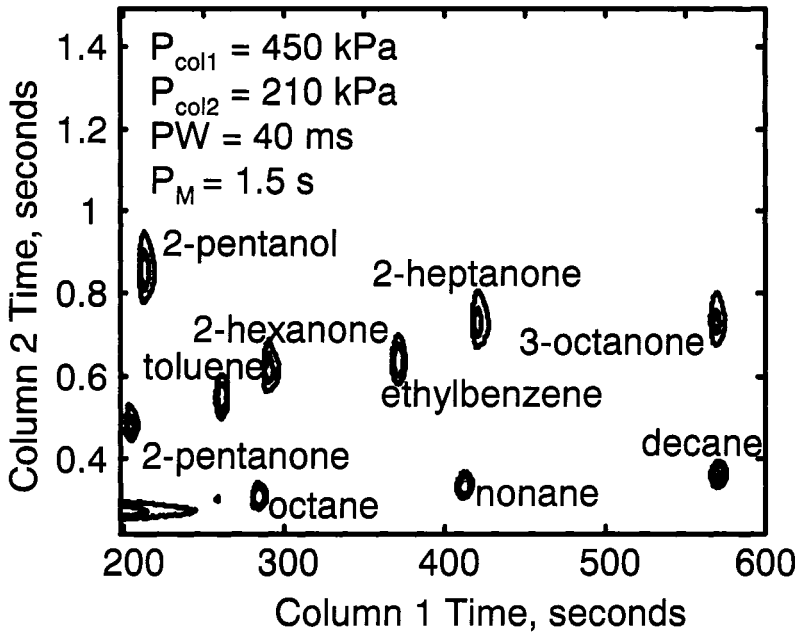


Figure 2.5: Representative contour plot of the 10 component 2D separation
Representative contour of total-transfer valve-based GCxGC separation showing the 10 components used for the PW and P_M study, with results presented in Figures 2.6 and 2.7, and Tables 2.1,2.2 and 2.3. Column 1 was operated under a constant pressure of 450 kPa and column 2 had a constant pressure of 210 kPa.

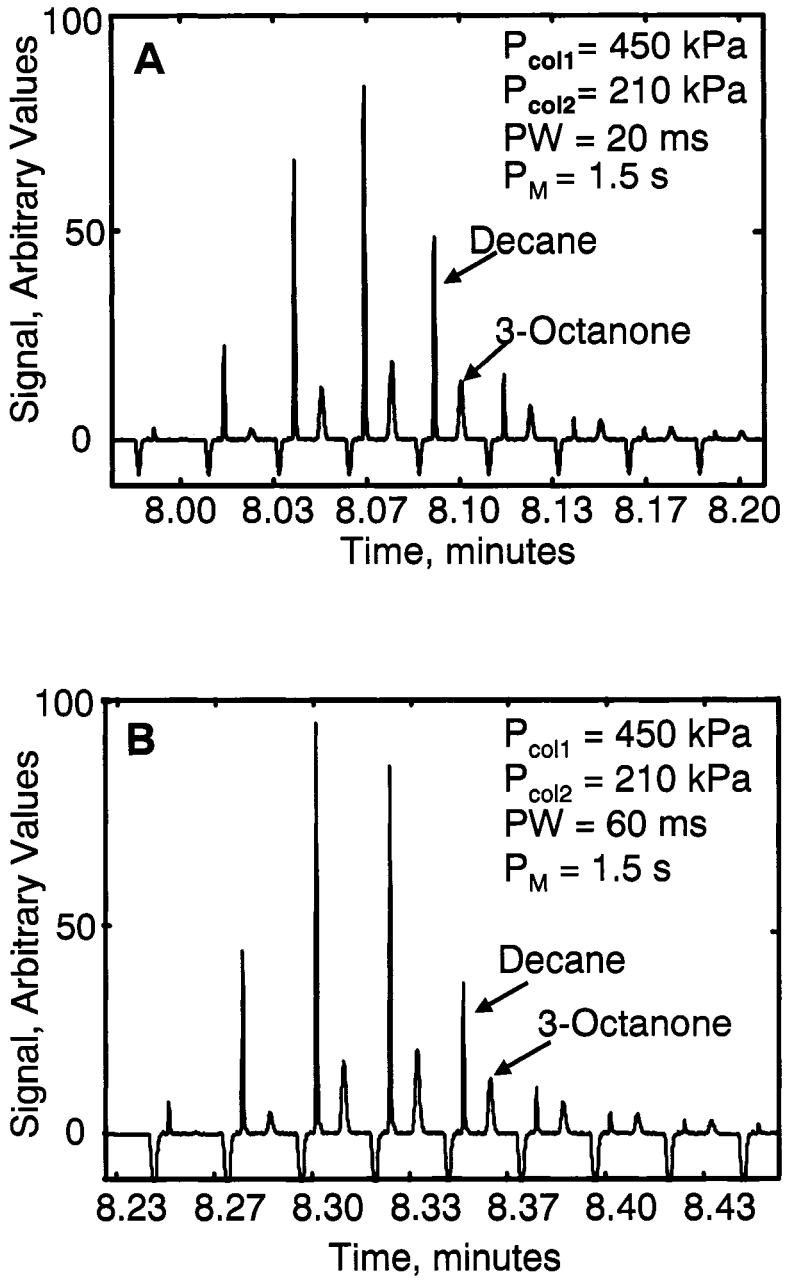


Figure 2.6: Pulse-width (PW) dependence.
Dependence of the total-transfer valve-based GCxGC on changing PW is investigated using the separation shown in Figure 2.5. At a constant P_M of 1.5 s the number of slices of the analytes shown (decane and 3-octanone) eluting from column 1 remains the same for a PW of (A) 20 ms and (B) 60 ms, about 8 slices for each, where a slice is a modulation period involved in separating a given analyte.

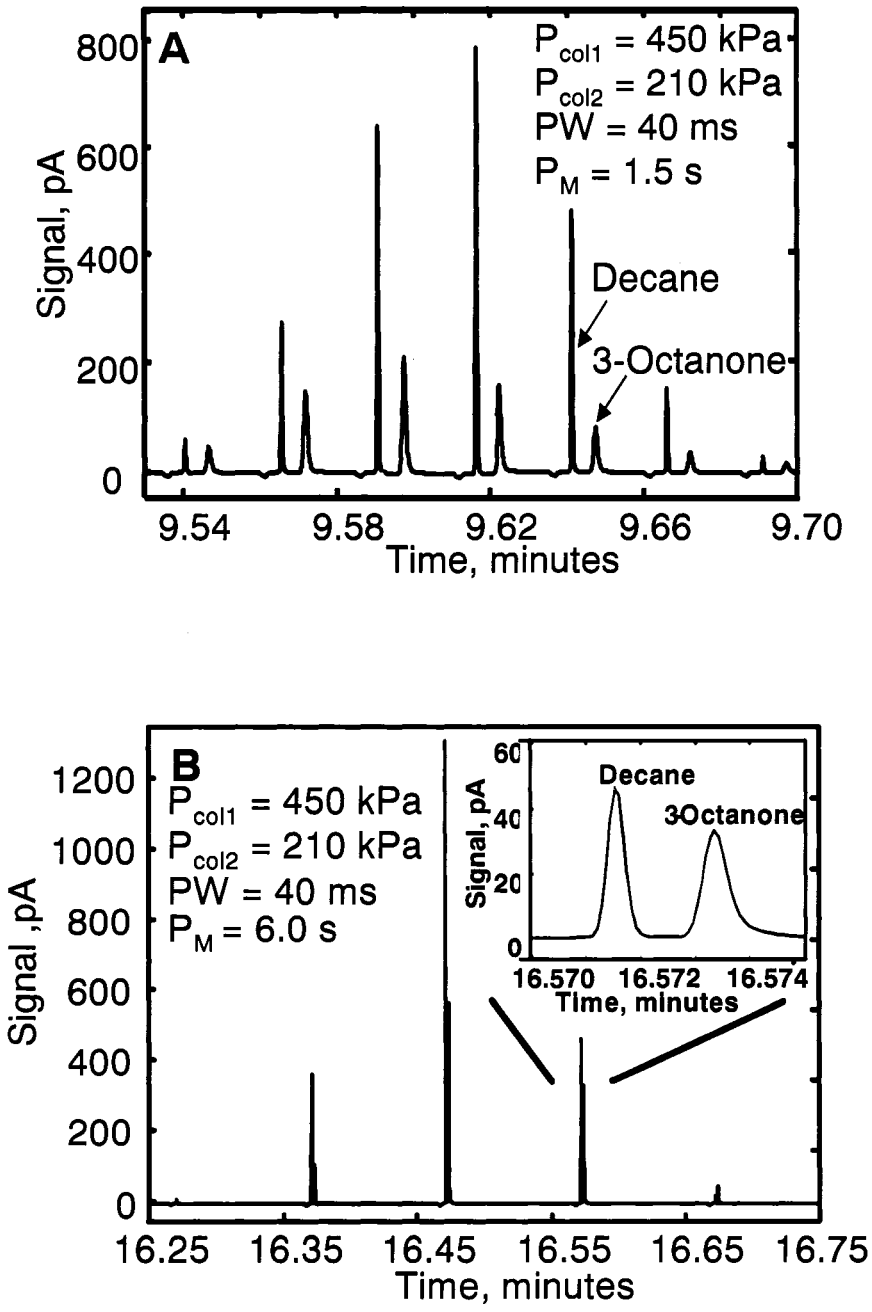


Figure 2.7: Modulation period (P_M) dependence.
Dependence of the total-transfer valve-based GCxGC on changing P_M is investigated using the separation shown in Figure 2.5. At a constant PW of 40 ms the number of slices of the analytes shown (decane and 3-octanone) eluting from column 1 changes as P_M is changed from (A) 1.5 s to (B) 6.0 s. Specifically, the number of second column slices is cut in half and there is a two-fold S/N increase for this P_M increase.

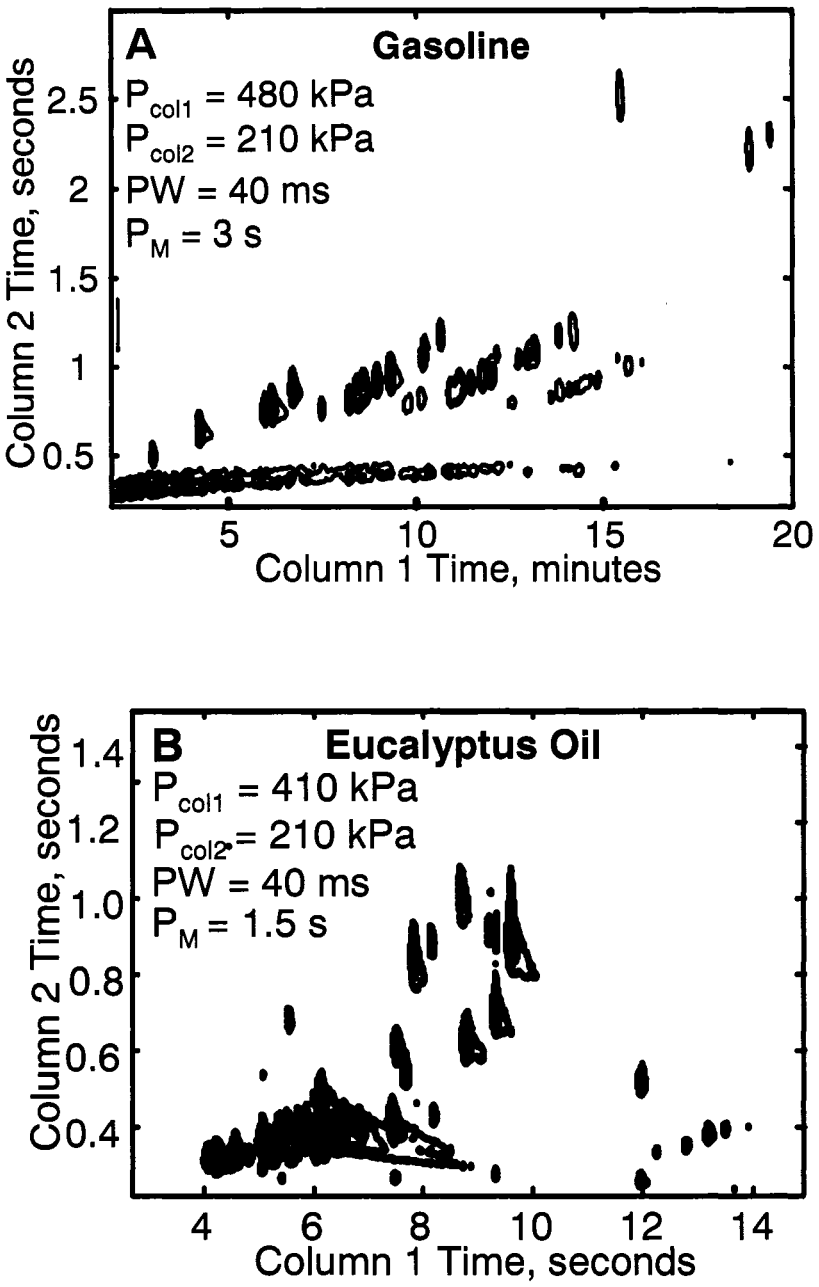


Figure 2.8: Total-transfer valve-based GCxGC of complex samples.
Separations of a (A) gasoline sample and (B) Eucalyptus oil. Gasoline was separated at P_M of 3 s and PW of 40 ms with the column 1 head pressure at 480 kPa and column 2 head pressure at 210 kPa. The Eucalyptus oil was separated at a P_M of 1.5 s and PW of 40 ms with column 1 at 410 kPa and column 2 at 210 kPa.

Notes to Chapter 2

(1) Górecki, T.; Harynuk, J.; Panić, O. *J. Sep. Sci.* **2004**, *27*, 359-379.

(2) Giddings, J. C. *Anal. Chem.* **1984**, *6*, 1258A-1270A.

(3) Murphy R.; Schure, M.; Foley, J. *Anal. Chem.* **1998**, *70*, 1585-1594.

(4) Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227-231.

(5) Phillips, J. B; Ledford, E. B. *Field Anal. Chem. Tech.* **1996**, *1*, 23-29.

(6) Kinghorn, R. M.; Marriott, P. J. *J. High Resol. Chromatogr.* **1998**, *21*, 620-622.

(7) Beens, J.; Adahchour, M.; Vreuls, R. J. J.; van Altena, K.; Brinkman, U. A. Th. *J. Chromatogr. A* **2001**, *919*, 127-132.

(8) Bruckner, C. A; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **1998**, *70*, 2796-2804.

(9) Seeley, J. V.; Kramp, F.; Hicks, C. J. *Anal. Chem.* **2000**, *72*, 4346-4352.

(10) Sinha, A. E.; Johnson, K. J.; Prazen, B.J.; Lucas, S. V.; Fraga, C.; Synovec, R. E. *J. Chromatogr. A* **2003**, *983*, 195-204.

(11) Sinha, A. E.; Prazen, B.; Fraga C.; Synovec, R. E. *J. Chromatogr. A* **2003**, *1019*, 79-87.

(12) Bueno Jr., B. A.; Seeley, J. V. *J. Chromatogr. A* **2004**, *1027*, 3-10.

(13) Veriotti, T.; Sacks, R. *Anal. Chem.* **2001**, *73*, 3045-3050.

(14) Wittrig, R. E.; Dorman, F. L.; English C. M.; Sacks, R. D. *J. Chromatogr. A* **2004**, *1027*, 75-82.

(15) Harynuk, J.; Górecki, T. *J. Sep. Sci.* **2004**, *27*, 431-441.

Chapter 3. Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells‡

3.1 Introduction

Two-dimensional (2D) comprehensive gas chromatography coupled with time-of-flight mass spectrometry (GCxGC-TOFMS) has been used for the analysis of a number of complex mixtures such as jet fuels,[1] environmental samples,[2] drug screening,[3] cigarette smoke condensate[4] and pesticides.[5] The major advantage of a time-of-flight mass spectrometer (TOFMS) over other GCxGC detectors is its ability to aid in the identification of a large number of compounds in very complex mixtures. Obtaining useful information from the analysis of these mixtures, which may contain literally thousands of different compounds, requires a reliable and relatively rapid procedure to identify compounds of interest and to deconvolute any overlapping mass spectra to yield pure chromatographic peak profiles for accurate quantification.

One way to glean information from GCxGC-TOFMS data is to use chemometric analysis and take advantage of the multivariate selectivity in the third order data.[6] An analysis tool that is commonly used to locate distinguishing regions of signal in data sets is principal component analysis (PCA). PCA recognizes patterns in complex data sets based upon the amount of variance in the data sets. The chemical compounds exhibiting the greatest variance are captured on principal component 1, PC1. The next greatest variance in the data set is captured on PC2, and so on. The two outputs from PCA are a

‡ Large portions of this chapter have been previously published:
Rachel E. Mohler, Kenneth M. Dombek, Jamin C. Hoggard, Elton T. Young and Robert E. Synovec, Analytical Chemistry 2006, 78, 2700-2709

scores plot and a loadings plot. The scores plot indicates the degree of similarity between the individual chromatographic profiles and groups them accordingly. The loadings plot identifies the locations in the 2D separation plane of the chemical compounds responsible for these groupings. The loadings values correspond with quantitative differences between classes for each differentiating chromatographic feature. A larger absolute loadings value is associated with a greater class-to-class difference.

Parallel Factor Analysis (PARAFAC) is a quantitative multivariate tool that can be applied to three-dimensional (3D) GCxGC-TOFMS data structures.[6] PARAFAC uses the GC column 1, GC column 2 and the mass spectral signal in one sample to obtain pure component profiles (i.e., pure chromatographic and mass spectral peak profiles) by removing the background noise and overlapping peaks. The deconvoluted 3D peak can then be reconstructed by taking the outer product between the first and second dimension chromatographic profiles. This deconvoluted analyte peak can then be integrated to give a total peak volume that is proportional to analyte concentration.

Recently, there has been interest in using GCxGC-TOFMS for metabolite analysis.[7-11] Several methods of identifying differences in metabolomic data obtained by gas chromatography have been reported in the literature.[8-13] Each of these methods employs different tools to achieve the goal of high-throughput analysis. One method of analysis involves the use of chemometric and multivariate techniques to analyze complex metabolic data.[8,9,12,13] Recently, the Synovec lab reported the "DotMap" algorithm, which uses the dot product of the scaled, weighted and normalized library mass spectrum and the observed mass spectrum along each point in the GCxGC-TOFMS data cube to

locate analytes of interest.[8,9] Using the DotMap algorithm, the analytes of interest must be known prior to analysis. Jonsson and co-workers used multivariate curve resolution (MCR) followed by PCA or PLS-DA on GC-TOFMS data to identify and quantify metabolites of interest.[12,13] Using this method all samples were smoothed, baseline subtracted, aligned, resolved using MCR, submitted to a multivariate analysis and exported to a mass spectral search library. The PARAFAC algorithm has also been used to obtain pure mass spectra from GCxGC-TOFMS data derived from a complex derivatized plant metabolite extract.[9] This study used the entire mass spectral range collected to obtain pure component peak profiles. Shellie and co-workers used direct chromatogram comparisons and difference plots of the total ion chromatogram (TIC) for GCxGC-TOFMS data to locate differences in mice metabolite data.[10-11]

In this chapter, we examined the suitability of PCA for normalized GCxGC-TOFMS chromatographic data followed by the use of ChromaTOF and PARAFAC software for locating, identifying and quantifying metabolites of interest in complex yeast metabolite extracts. First, an optimization study was performed to determine the amount of yeast cells to collect for the extraction. Once the optimum number of cells to use in the extraction was approximated, small molecules were extracted from yeast cells either metabolizing glucose via fermentation or metabolizing ethanol by respiration. Growth on the carbon source ethanol (derepressed, DR) should lead to the accumulation within cells of different pools of metabolites than growth on the carbon source glucose (repressed, R). In this study, nine extracts from cells growing on each carbon source were prepared and a total of 70 injections on the GCxGC-TOFMS were processed for comparison. Since the

compound classes of interest were known, i.e., derivatized metabolites, three selective mass channels (*m/z* 73, 205 and 387) were submitted to PCA, thereby utilizing more of the 3D data cube than if only one *m/z* was examined to locate a significant number of analytes responsible for much of the variability between the sample extracts. The compounds with the highest loadings values were then identified but not quantified using ChromaTOF software. This resulted in the discovery of key analytes that differed in concentration between fermenting (R) and respiring cells (DR). The concentration ratios of these metabolites in the two types of sample extracts were obtained using PARAFAC. Thus, ChromaTOF, PCA and PARAFAC software are used in a complementary way in the reported methodology. The quantitative precision obtained for the metabolites that vary most between growth conditions is also reported along with the mass spectral match values (MV).

3.2 Experimental

3.2.1 Growth, extraction and derivatization conditions: optimization study

The yeast strain used in this study was W303-1a (*MATa ade2 can1-100 his3-11,15 leu2-13,112 trp1-1 ura3-1*). Cells were grown at 30 °C in YPD (yeast extract, peptone, dextrose) medium containing 5% glucose followed by incubation for 6 hours. Three aliquots containing 4.5 x 10^7 cells were removed from the yeast culture. Six serial dilutions were performed with each sample containing half the number of cells of the previous sample, Figure 3.1. Each aliquot from each culture was diluted into 4 volumes of quenching buffer (10 mM tricine, pH 7.4, in 60% methanol) at –40 °C. The metabolic activity of the yeast cells was quenched and small polar metabolites extracted using the

method reported by Castrillo et al.[14] as a guide. Each cell suspension was spun at 1000 g in a Sorvall RC-5B Plus centrifuge at –20 °C for 3 min and the resulting cell pellet was washed once with 1 ml of the quenching buffer at –40 °C. Each cell pellet was suspended in 1 ml of extraction buffer (0.5 mM tricine, pH 7.4, in 75% ethanol) at 80 °C and held at this temperature for approximately 3 min. Extracted cell suspensions were cooled on ice for 5 min and then spun twice as described above to pellet large cellular debris. The resulting ethanolic metabolite extracts were dried in a SpeedVac at room temperature and stored at –80 °C under argon prior to GCxGC-TOFMS analysis. The extracted metabolites were methoximated (20 mg/ml methoxyamine in pyridine) and then trimethylsilylated (BSTFA: TMCS, 99:1) thus replacing the active hydrogens with the trimethylsilyl (TMS) adducts.[16] The samples were warmed at 30 °C for 90 minutes after the addition of 30 μL methoxime solution. Following oximation, 70 μL of TMS reagent was added and the sample was heated at 60 °C for 60 minutes.

3.2.2 Growth, extraction and derivatization conditions: repressed and derepressed study

The same strain of yeast mentioned above was grown at 30 °C in synthetic complete medium (SC)[15] containing either 5% glucose to prepare fermenting cells (R) or 3% ethanol and 0.05% glucose followed by incubation for 6 hours to prepare cells metabolizing ethanol by respiration (DR). Three R and three DR cultures were grown to obtain a measure of biological variability, Figure 3.2. At late log phase a volume of each culture containing $1x10^7$ cells was diluted into 4 volumes of quenching buffer (10 mM tricine, pH 7.4, in 60% methanol) at –40 °C. Three aliquots were taken from each culture

for extraction, Figure 3.2, and extraction of metabolites was performed as stated in the optimization study. Derivatization also followed the above procedure.

3.2.3 Instrumental Parameters

An Agilent 6890N gas chromatograph equipped with an Agilent 7683 auto-injector (Agilent Technologies, Palo Alto, CA, USA) coupled with a LECO Pegasus III time-of-flight mass spectrometer with the commercially available 4D thermal modulator upgrade (LECO, St. Joseph, MI, USA) was used to analyze the yeast extracts. Column 1 was a 20 m x 250 μm i.d. x 0.5 μm RTX-5MS film (Restek, Bellefonte, PA, USA) and column 2 a 2 m x 180 μm i.d. x 0.2 μm RTX-200MS film (Restek, Bellefonte, PA, USA). Injections of 1 μl were made in split-less mode with each extract having 4 replicate injections except A1R and A1DR (nomenclature given in Figure 3.2), which were injected in triplicate. This resulted in 70 injections (GC x GC-TOFMS chromatograms) for the analysis. The GC inlet and transfer line were set at 280 °C. Column 1 was held at 60 °C for 0.25 minutes and then increased at 8°/min to 280 °C where it was held for 10 minutes. Column 2 was initially set at 70 °C and followed the same temperature program as column 1 giving a total run time of 37.8 minutes. The modulator was kept 40 °C higher than column 1 and the modulation period was 1.5 seconds. A constant flow rate of 1 ml/min was held at the head of column 1. The ion source was set to 200 °C. Mass Channels m/z 40-600 were collected at 100 spectra /second after a 5 minute solvent delay.

3.2.4 Data Analysis

The LECO ChromaTOF software v.2.21 (LECO, St. Joseph, MI, USA) was used to collect the data and obtain raw mass spectral match values (MV). Three mass spectral libraries were searched against: the NIST main library, an in-house metabolite library and the metabolite library obtained from the Max Planck Institute of Molecular and Plant Physiology (http://www.mpimp-golm.mpg.de/mms-library/index-e.html). Data from *m/z* 73, 205 and 387 were exported as comma separated value (.csv) files to Matlab v.7.0.4 (Mathworks, Natick, MA, USA) where principal component analysis, PCA, was performed, using the PLS toolbox version 3.51 (Eigenvector Research, Manson, WA, USA). All samples were normalized to the summed TIC from regions 430.5 s to 1125 s and 1145 s to the end of the run and mean centered. The regions omitted contained overloaded reagent artifacts that were similar in both the R and DR 2D chromatograms. Only data from *m/z* 387 required retention time alignment for the greatest amount of class-to-class variance to be captured on PC1. The data was shifted slightly in both separation dimensions, but it was found that only alignment on column 2 was necessary to assist the data analyis, so the peak match one-dimensional retention time alignment algorithm was used.[17] The peak match algorithm was sufficient since the magnitude of the shifting on column 2 was less than the typical peak width at the base. The retention time alignment algorithm creates a list of where each peak is located in the specified target chromatogram and all sample chromatograms, where each chromatogram is represented by a single row vector and peaks are defined as local maxima with signal greater than five times the standard deviation of the chromatographic noise. Each peak in

a sample chromatogram is then shifted to match the retention time of the nearest peak in the target chromatogram using linear interpolation. The *m/z* 387 chromatograms were aligned to A1R injection 1 (see Figure 3.2 for sample extract label nomenclature) and the shifting threshold was set to 20. This alignment algorithm did not mismatch peaks due to the scarcity of peaks in the 2D chromatogram at *m/z* 387. A PARAFAC graphical user interface (GUI) developed in-house and implementing the N-way toolbox was used for the quantification of the metabolites identified by PCA as exhibiting the most variability between R and DR sample extracts. The PARAFAC algorithm in this GUI was implemented from the following website: http://www.models.kvl.dk/courses/.[18] The GUI automatically imports raw ChromaTOF data and converts LECO format chromatograms into Matlab variables. The GUI allows the user to quickly specify the sub region of the chromatograms for the analysis as well as the *m/z* to use for the deconvolution of overlapping mass spectra. The major benefits to using a PARAFAC GUI are the speed and ease of analysis.

3.3 Results and Discussion

3.3.1 Metabolite response to number of cells extracted

As stated in Chapter 1, one of the challenges of yeast metabolome analysis is the wide range of metabolite concentrations present at each point in time. Even with the 2D separation space, metabolites that overload the detector and column 2 can potentially obscure metabolites of lower concentration. The optimum number of extracted cells would minimize the overloading while enabling detection of lower concentrated metabolites. To quickly approximate an optimum number of cells to extract, the signal

intensity at the most intense mass channel (*m/z* 73) as a function of cells extracted was plotted for four selected metabolites, Figure 3.3A-D. Samples one and four were not included in this study due to problems encountered during sample preparation. The four metabolites studied encompassed peak volumes over three orders of magnitude. Based on the linear region of the graphs and the extract to extract percent relative standard deviation (%RSD), it was decided that the optimum number of cells to process was between 5×10^6 and 1×10^7. The extract to extract %RSD jumped from an average of 26% over the four metabolites at 1.1×10^7 cells to 63% at 2.81×10^6 cells. The upper limit (1×10^7) was chosen to minimize the chance of losing metabolites of low concentration.

3.3.2 Sample classification

GCxGC-TOFMS is suitable for the analysis of the very complex methoximated and trimethylsilylated yeast extracts, Figures 3.4A and B. It is important to note that even though the modulation period is short, 1.5 s, and some wraparound did occur, sufficient resolution of the eluting components was obtained along with three or more modulation periods across each peak to assist the chemometric analysis. Since the trimethylsilyl (TMS) group ($-Si(CH_3)_3$) has a *m/z* of 73, any chemical species in a sample extract that is tagged with one or more TMS group will have a peak at m/z 73. While providing some useful information, the major drawback to using this single m/z for identifying differences in the samples is the complexity of the chromatogram at m/z 73. Of the more than 2500 peaks recognized by ChromaTOF as present in the 2D chromatogram, many are reagent artifacts along with column bleed. To exclude these

artifacts from the analysis and at the same time investigate a certain class of compounds, selected m/z were chosen that would be specific for a known class of compounds of interest, as demonstrated in Figures 3.4C and D. Trimethylsilyl carbohydrates have a distinguishing mass fragment at m/z 205 (Figure 3.4C). A majority of TMS sugar phosphates have a distinguishing fragment at m/z 387 (Figure 3.4D). These m/z showed the metabolites of interest and significantly lowered the complexity of the 2D chromatographic profiles. Only the chromatograms for the R samples are shown, but the DR samples show a similar number of peaks.

The complexity of the chromatograms shown in Figure 3.4 made it difficult to visually observe differences between the R and DR chromatograms. A rapid and objective method for determining the difference between two or more classes of samples is to use PCA on the entire data set of samples. Prior to submission of all 70 chromatograms to PCA, samples were normalized to the sum of the TIC thus correcting for injection volume discrepancies and aligned only if the class-to-class variance was not captured on PC1. The summed TIC signals for the R and DR samples were compared to ensure they fell in the same range and they did. The mean TIC signal for the R samples was $9.0x10^8$ with a standard deviation of $1.1x10^8$ and the mean signal for the DR samples was $9.7x10^8$ with a standard deviation of $1.6x10^7$. Thus, since the standard deviations encompassed the means, the TIC could be objectively used to correct injection volume variation. In the future, it may be prudent to add internal standards from multiple compound classes to correct for variable losses during extract preparation.[19] In addition to the major mass channel (m/z 73), PCA was performed individually on m/z 205 and

387 to ascertain the extent of class discrimination followed by the use of the loadings plots to locate differentiating metabolites. Since only two experimental classes, R and DR, were sampled, the class-to-class variance was captured primarily on PC1 as is shown in the scores plots in Figures 3.5A-C. The 70 injections mentioned previously in the Experimental section are labeled by a chromatogram number. The chromatogram number does not reflect the order in which the samples were derivatized or injected, but rather the order in which the samples were subjected to PCA. At m/z 73 approximately 47% of the variance was captured in PC1 (Figure 3.5A). This low captured variance is reasonable given that reagent artifacts and column bleed are present at this m/z. The m/z 205 data (Figure 3.5B) resulted in the greatest amount of variance captured on PC1 (72.1%) and, like m/z 73, was also successful in separating the R and DR samples. The class-to-class variance of the sugar phosphates at m/z 387 was initially obscured by run-to-run retention time variations on column 2, resulting in the R and DR differences being captured on PC2 with 25% of the variance, instead of separation on PC1. Potential reasons for the dependence of the m/z 387 mass channel on alignment are the lower S/N of the eluting compounds as well as smaller differences between the R and DR samples. At m/z 73 and 205 the differentiating peaks were of much higher S/N and with much larger differences between the R and DR samples, so slight retention time shifting was not detrimental to the class-to-class separation. Lower S/N for the sugar phosphate peaks was possibly caused by poor solubility in the derivatization solvent, i.e., pyridine. The scores plot for the aligned m/z 387 data is shown in Figure 3.5C. Even after alignment and normalization, sample extracts B3R and C3R (chromatograms numbered 31-34 and

55-58) overlapped with sample extracts of the DR cultures and the variance captured by PC1 was only 47%. Although the captured variance was somewhat low, successful location of distinguishing sugar phosphates was also achieved as will be shown.

Following distinction of the R from DR samples on the PC1 scores plot, the PC1 loadings plots were reconfigured into the 2D chromatographic space giving the retention time location of the metabolites showing the biggest difference between the R and DR samples as shown in Figures 3.6A-C. In order to obtain only the most useful information, a threshold was thoughtfully, yet empirically, set for each of the mass channels to include only metabolites of interest for the proof-of-principle demonstration of this analytical methodology, thus excluding artifacts or peaks of intensity too low to be definitively identified. The loading plot thresholds were set at the absolute values of 0.025, 0.002 and 0.01 for m/z 73, 205 and 387, respectively. Note that m/z 73 had the highest threshold so only the largest intensity differences between the two classes of samples would be discovered. The largest number of compounds that distinguish the DR samples from the R samples were found on the PC1 loadings at m/z 73 as shown in Figure 3.6A. This was expected due to the large number of peaks at this m/z. The PC1 loadings for m/z 205 revealed an additional three metabolites that were not discovered at m/z 73 (Table 3.1), one of which (erythrose) was obscured by the larger lysine peak at m/z 73 (Figure 3.6B). The loadings for m/z 387 revealed nine sugar phosphates that were not identified at either m/z 73 or 205. Examining the selective mass channels proved beneficial as it increased the number of compounds that served to distinguish R and DR sample extracts. Indeed, it

is likely that a more exhaustive evaluation of additional m/z would yield a greater amount of information, but further analysis was beyond the scope of the current study.

Table 3.1 lists the differentiating chemical components identified by PC1 in order of increasing column 1 retention time. The mass spectral match value (MV) is a measure of the similarity between the collected mass spectra and the library mass spectra. A perfect MV is 999, but for the purpose of this study, any MV greater than 750 was considered a positive identification. Approximately 77% of the metabolites have spectra that match the library spectra with a MV greater than 750. Some of the lower MV values are a result of the low S/N of the peak or of overlapping chromatographic peaks. Sugar phosphates resulted in two peaks, one each from the α and β isomers, which have been reported previously.[20] The loadings values listed in column 6 of Table 3.1 are the values obtained from PC1 at m/z 73, 205 or 387. As the absolute value of the loadings increases, the difference between the amounts of the metabolite in the two classes of samples also increases. The negative loadings values correspond to the metabolites that have higher relative concentration in DR cells than in R cells, and *vice versa*.

3.3.3 Quantification

The 26 metabolites identified by PCA and the raw mass spectral information (Table 3.1) was then further analyzed to obtain quantitative information. We sought to obtain the concentration ratio of a given metabolite in the DR samples relative to in the R samples. The concentration ratios for each of the 26 metabolites were obtained by dividing the average amount of each metabolite in the DR sample extracts (N_{DR}), by the average amount of the same metabolite in the R sample extracts (N_R). One of the issues

with using peak volumes from the raw chromatograms as a measure of the amount of a given metabolite is that co-eluting compounds would bias the quantitative information. Overlapping compounds could significantly contribute to the strength of the chromatographic signal at the selected m/z and interfere with the accurate quantification of the metabolite of interest. Thus, to obtain useful quantitative information, the PARAFAC algorithm was applied since it analyzes the entire 2D peak simultaneously.

Citric acid was chosen to illustrate the use of PARAFAC to obtain accurate quantitative information, in particular in the presence of overlapping compounds. Figure 3.7A shows the library mass spectrum for citric acid. When the raw mass spectrum, i.e., the unprocessed spectrum output from the ChromaTOF software, is compared to the library mass spectrum, a MV of 752 is obtained (Figure 3.7B). The raw mass spectrum has significant ion fragments at m/z 157 and 256 that are not present in the library spectrum leading to the low MV that would also result in an inaccurate peak volume. When the m/z 44 and 73-480 are submitted to the PARAFAC GUI, a MV of 858 is obtained and the ion fragments at m/z 157 and 256 are dramatically reduced from the raw spectrum (Figure 3.7C). This yielded fewer interfering ions and a more accurate concentration ratio. On the other hand, one issue with submitting such a large range of m/z to PARAFAC is the potentially excessive computation time and memory required. This issue can be adequately addressed if only a few selective analyte mass channels are chosen and submitted to the PARAFAC GUI instead of submitting the entire mass range. When submitting only a selected number of mass channels to PARAFAC, the user must realize that some of the signal is being ignored. However, if a concentration ratio is used,

as in this report, it is independent of the number of mass channels and can be compared between analytes. Since we seek the concentration ratio of the metabolites in the DR to R sample extracts, a unique set of mass channels can be applied to each metabolite, as will be further described below. Keep in mind that the PARAFAC analysis is complemented by using the commercially available ChromaTOF software to determine the identity of the analyte. The number of mass channels submitted to PARAFAC depends on the mass spectrum of the analyte of interest and the presence of overlapping mass spectra from co-eluting compounds. Generally, the 10 most sensitive mass channels for a given metabolite are sufficient for PARAFAC, however more mass channels can be added if the overlapping components have similar mass spectra or if the analyte of interest has a low S/N. The benefit of using PARAFAC to minimize the impact of overlapping interferences is illustrated in Figure 3.8. Figures 3.8A and B show the resulting first and second dimension chromatographic peak profiles for citric acid in a R sample extract after only seven selective m/z were submitted to PARAFAC. Additionally, the resulting deconvoluted mass spectrum is shown in Figure 3.8C. Fewer than 10 m/z were selected for citric acid to optimize the selectivity and sensitivity of the deconvolution of citric acid from the other interfering species. The m/z 44 data was included to separate the baseline from citric acid. Interfering species overlap the peak profiles of citric acid on both the first and second GC dimensions, i.e. GC column 1 and 2, yet PARAFAC was able to obtain a pure citric acid chromatographic peak profile using the selected m/z. The mass spectral pattern in Figure 3.8C is in good agreement with the library spectrum in Figure 3.7A, and the MV increased significantly due to the

exclusion of noisy mass channels from the analysis. Using PARAFAC results for citric acid in Figure 3.8, the reconstructed 2D GC peak profile was obtained from simple linear algebra (outer product of column 1 and 2 peak profiles), and the peak volumes from the 3D peak determined, i.e., the N values. Thus, the peak volume corresponds to the sum of the individual column 2 peaks for a given analyte, obtained from all modulation periods for the elution of the analyte from column 1. For citric acid, on average, N_R was 4.9 x 10^{-4} (13%RSD) and N_{DR} was 6.4 x 10^{-3} (26%RSD), resulting in a concentration ratio N_{DR}/N_R of 13. Even though a subset of m/z was applied with the PARAFAC GUI, the concentration ratio, N_{DR}/N_R, is independent of the number of mass channels, since the same number of mass channels is used for both the DR and R sample extracts for a given metabolite. Accordingly, the adjusted MV and the m/z used to obtain the different spectra for all 26 metabolites, which were ultimately used for determining N_{DR} and N_R as in the citric acid example, are listed in Table 3.2. Note that the standard deviation in the R MV for citric acid decreased by a factor of approximately 20 between the raw MV, Table 3.1, and post PARAFAC value, Table 3.2. This provides additional evidence that PARAFAC has accurately obtained the citric acid mass spectrum.

Before going into the details for the quantification of the 26 metabolites, we now turn our attention to a detailed evaluation of the injection-to-injection, extract-to-extract and culture-to-culture precision in the determination of N_{DR} and N_R for four metabolites: citric acid, ornithine, malic acid, and glucose 6-phosphate (Figures 3.9A-D). The injection precision, encompassing instrumental and analytical variation, averaged 8% for the DR sample extracts and 9% for the R sample extracts after normalization. The

sample extract variability can be assessed by comparing the values obtained from the three samples of each culture. For future studies, extraction precision could be enhanced by the use of appropriate internal standards. The extraction precision averaged 11% for the DR sample extracts and 31% for the R sample extracts. The culture-to-culture variability or biological precision for these four metabolites averaged 9% for DR sample extracts and 25% for R sample extracts. The R and DR cells were grown, harvested, extracted and derivatized separately. The larger variation in the R samples is likely due to the extraction and derivatization steps and could also arise from the R cells growing faster than the DR cells. Citric acid, ornithine, and malic acid all showed insignificant extraction and culture variation (Figures 3.9 A-C). For these metabolites no concentration overlap was seen between the DR and R sample extracts. Glucose-6-phosphate was more variable between cultures for the reasons described previously (Figure 3.9D).

The PARAFAC quantitative results (N_{DR} and N_R) and the ratio of N_{DR} divided by N_R (labeled N_{DR}/N_R) for the 26 most variable metabolites identified by PCA were calculated using PARAFAC on selected mass channels to determine the biological variability due to the cell growth conditions (Table 3.3). To save analysis time, only one of the replicate sample extracts for each culture was analyzed by PARAFAC. Thus, the average biological relative standard deviation, %RSD, for the R sample extracts was approximately 40% and for DR was 21%. The %RSD values are, on average, almost double those calculated above in the detailed precision study of four of the metabolites (Figure 3.9), likely due to using fewer of the sample extracts (and their replicates). For

these 26 differentiating metabolites the %RSD's obtained did not hinder the class separation, and the use of fewer replicates was justified. One third of the identified metabolites were detected only in the R sample extracts. These values for biological variability are similar to previous studies of plant metabolites analyzed using GC/MS.[21] When numerically defined (neither zero nor infinite), the N_{DR}/N_R concentration ratio for the 26 components listed in Table 3.3 ranged from 0.02 for glucose to 67 for trehalose. All analytes that were located and quantified in both classes of samples were not overloaded, so the quantification was accurate in this regard. Metabolites of lower concentration and exhibiting smaller differences between the R and DR sample extracts were identified in the more selective 205 and 387 mass channels. The results suggest that the use of selective mass channels can facilitate the identification and quantitation of the less abundant metabolites, giving impetus for future study.

3.3.4 Biological interpretation

Although our main focus in the present study is to demonstrate the feasibility of GCxGC-TOFMS combined with chemometric analysis for the separation, identification and initial quantification of metabolites in yeast extracts, it is important to note that the biological results presented in Table 3.3 are consistent with the known regulation of the metabolic pathways operative in yeast cells growing in the two conditions tested. When a yeast cell is fermenting glucose, the glycolytic pathway is highly active. Glucose is phosphorylated to glucose-6-phosphate and then converted to fructose 1,6 diphosphate. Glucose-6-phosphate and fructose 1,6 diphosphate are both elevated in the extracts from R cells. Trehalose, a storage carbohydrate, is produced when yeast cells are grown in

stressed conditions where growth rate is slow and a carbon source is abundant.[22] These conditions are present when cells are metabolizing ethanol by respiration. Trehalose was found to be more abundant in the DR cells metabolizing ethanol. Two intermediates in the tricarboxylic acid (TCA) cycle, malic acid and citric acid were also found to be more abundant in the DR cells. This is consistent with the increased expression of many of the genes encoding the enzymes of the TCA cycle in accord with a more active TCA cycle in these cells.[23,24] Thus, the analytical method presented identified and quantified metabolites that differentiate fermenting and respiring cells and which are consistent with the known regulation of the metabolic pathways active in these types of cells.

3.4 Conclusions

GCxGC-TOFMS analysis coupled with chemometrics software has the ability to objectively and unambiguously identify and quantitate metabolite differences between yeast cells growing in different media with straightforward and rapid preprocessing of the data. In addition to the major derivatization fragment mass channel at m/z 73, higher mass channels (e.g., 205 and 387) can be used to glean selective information. In the future, it would be highly advantageous to use of all mass channels to determine class differences, hence, a methodology to that uses Fisher ratios to do so is currently under investigation and will be the subject of Chapter 4. As we show here, multiple mass channels are especially useful in deconvoluting complex mass spectra to resolve individual components and to determine their relative abundance. The use of third order data, two retention times and ion currents at selective masses, is especially important for working with extremely complex samples such as metabolite extracts. Utilizing these

methods of analysis to obtain and interpret metabolomic data will narrow the gap between cellular genotype and phenotype. The lack of ability to differentiate the C6 sugar phosphates in Table 3.1 begs for the compilation of a metabolite standard library collected under the same instrumental conditions, thus yielding retention time and mass spectral information.

Table 3.1: Metabolites with the highest loadings on PC1.

Identification of the most highly loaded metabolites discovered from Figures 3.4 and 3.5 along with the column retention times and raw match values (MV) for the repressed (R) and derepressed (DR) samples. The metabolites listed in column 1 were detected as the methoximated and trimethylsilyl derivatives. At least five different classes of compounds are included in this list, including organic acids, amino acids, carbohydrates, sugar phosphates, and a sugar alcohol. The number following the compound name corresponds to anomers or different conformations of the same species. The MV for the C6 sugar phosphates was chosen from the first match to a C6 sugar phosphate in the LECO software. Unless otherwise stated, the loadings values are from figure 3.6A, m/z 73. * loadings values from figure 3.6B, m/z 205, † loadings values from figure 3.6C, m/z 387. [a] not detected in sample, [b] no match value available.

Metabolites	t_{R1}, min	t_{R2}, s	MV, R	MV, DR	Load Value
Glycolic Acid	7.5	0.68	893 ± 74	951 ± 8	- 0.0024*
Glycerol	11	0.43	881 ± 4	894 ± 1	0.1
Threonine	12.9	0.58	945 ± 7	939 ± 11	- 0.062
Arabino hexos-2-ulose	13.3	0.51	648 ± 2	651 ± 3	- 0.006*
Malic Acid	14.5	0.72	896 ± 16	925 ± 4	- 0.061
5-Oxoproline	15.1	1.36	917 ± 2	934 ± 2	- 0.068
Glutaric Acid	15.9	0.59	681 ± 10	--[a]	0.059
Glutamic Acid	16.4	0.74	861 ± 7	853 ± 5	- 0.035
α-Glycerophosphoric Acid	18.5	0.89	782 ± 16	869 ± 9	- 0.013 †
3-Phosphoglycerate	19.1	1.04	749 ± 48	813 ± 57	- 0.031 †
Ornithine	19.2	0.57	863 ± 5	844 ± 6	- 0.075
Citric Acid	19.3	0.66	695 ± 117	911 ± 8	- 0.068
Glucopyranose 1	20.3	0.42	899 ± 13	--[a]	0.099
Glucose 1	20.4	0.42	839 ± 8	845 ± 13	0.15
Erythrose	20.5	0.42	733 ± 5	--[a]	0.06*
Lysine	20.5	0.59	901 ± 13	871 ± 27	- 0.073
Glucose 2	20.6	0.46	837 ± 13	--[a]	0.14
Glucopyranose 2	21.3	0.41	899 ± 21	--[a]	0.11
C6 Sugar Phosphate 1	23.9	0.76	579 ± 99	--[a]	0.014†
Glucose-6-Phosphate 1	25.3	0.58	928 ± 19	881 ± 29	0.29†
Glucose-6-Phosphate 2	25.5	0.63	829 ± 39	638 ± 64	0.066†
C6 Sugar Phosphate 2	25.9	0.62	669 ± 117	520 ± 108	0.087†
C6 Sugar Phosphate 3	26.4	0.645	705 ± 66	--[a]	0.051†
Fructose Diphosphate 1	29.3	1.15	810 ± 29	--[a]	0.077†
Fructose Diphosphate 2	29.4	1.22	760 ± 37	--[a]	0.12†
Trehalose	29.7	0.67	920 ± 27	953 ± 3	-0.11

Table 3.2: PARAFAC mass spectral match values.

PARAFAC match values (MV) for the derivatized metabolites repressed (R) and derepressed (DR) cells using the selected mass channels. The number of mass channels depended on the selectiveness of the fragments and overlapping components for each metabolite. [a] not detected in sample, [b] no match value available.

Metabolite	MV, R	MV, DR	Mass Channels
Glycolic Acid	985 ± 18	994 ± 4	43,45,66,73,74,75,147,148,177,205
Glycerol	940 ± 8	948 ± 1	45,59,73,75,103,117,129,131,133,147,177,193,238 191,205,206,218,263,293,44,219,220,299,314
Threonine	994 ± 6	990 ± 4	73,205,147,103,117,218,133,206,75,45,44 159,203,86,291,320,321
Arabino hexos-2-ulose	869 ± 9	894 ± 11	45,73,74,75,89,103,117,147,205,234
Malic Acid	982 ± 1	984 ± 4	73,147,45,233,75,55,245,133,74,148,44 217,101,307, 335,265,117,175,189,190
5-Oxoproline	967 ± 12	988 ± 4	76,156,45,230,258,147
Glutaric Acid	928 ± 15	--[a]	73,147,185,259,349,133,44
Glutamic Acid	899 ± 14	884 ± 9	44,73,75,84,128,147,156,174,230,246,247 248,348, 363
α-Glycerophosphoric	638 ± 16	643 ± 35	73,299,357,147,445,315,44
3-Phosphoglycerate	904 ± 25	920 ± 26	73,357,299,147,103,358,101,75,315,445,44,387
Ornithine	948 ± 12	941 ± 3	73,142,174,258,420, 300
Citric Acid	991 ± 6	995 ± 1	73,147,273,347,211,465,44
Glucopyranose 1	963 ± 15	--[a]	73,74,75,147,191,192,204,205,206,217
Glucose 1	948 ± 6	930 ± 10	45,73,103,129,147,157,160,205,217,319
Erythrose	975 ± 12	--[a]	45,73,74,75,103,117,147,161,205,206,363
Lysine	976 ± 9	970 ± 9	44,73,128,156,174,317,434
Glucose 2	877 ± 9	--[a]	45,73,103,129,147,157,160,205,217,319
Glucopyranose 2	877 ± 38	--[a]	44,73,74,75,147,191,192,205,206,217 291,319,345,435
C6 Sugar Phosphate 1	803 ± 55	--[a]	45,73,147,191,217,299,315,343,357,387
Glucose-6-Phosphate 1	895 ± 17	880 ± 8	73,387,299,147,388,160,315,357,217,471,44,129
Glucose-6-Phosphate 2	879 ± 19	858 ± 28	73,387,299,147,388,160,315,357,217,471,44,129
C6 Sugar Phosphate 2	944 ± 21	920 ± 10	45,73,74,75,129,133,147,160,299,387
C6 Sugar Phosphate 3	991 ± 1	--[a]	44,73,74,129,147,204,205,299,357,387,388
Fructose Diphosphate 1	997 ± 2	--[a]	44,45,73,74,75,129,147,217,299,315,357,387,459
Fructose Diphosphate 2	994 ± 5	--[a]	44,45,73,74,75,129,147,217,299,315,357,387,459
Trehalose	931 ± 20	974 ± 3	73,147,191,103,217,129,169,204,361,271,44

Table 3.3: Quantitative results for metabolites with highest PC1 loadings. Quantification of highest loaded derivatized metabolites discovered in figures 3.4 and 3.5. The relative concentrations (N_R and N_{DR}) are the average of the peak volumes in one sample extract from cultures A, B and C for both the repressed (R) and (DR) cells. These N values were calculated using the mass channels given in table 3.2. [a] not detected in sample, [b] undefined.

Metabolite	N_R	N_{DR}	%RSD, R	%RSD, DR	N_{DR}/N_R
Glycolic Acid	4.24E-04	1.12E-03	13.0	4.0	2.6
Glycerol	1.99E-02	2.58E-03	10.7	13.5	0.1
Threonine	2.89E-03	9.60E-03	52.4	13.7	3.3
Arabino hexos-2-ulose	5.00E-03	6.46E-03	27.7	11.6	1.3
Malic Acid	4.94E-04	3.89E-03	34.8	15.3	7.9
5-Oxoproline	6.35E-03	1.54E-02	27.3	16.7	2.4
Glutaric Acid	2.3E-03	--[a]	65.6	--[a]	--[b]
Glutamic Acid	1.41E-03	4.40E-03	33.7	12.7	3.1
α-Glycerophosphoric Acid	1.2E-04	3.3E-04	18.7	18.1	2.8
3-Phosphoglycerate	1.10E-04	2.00E-04	22.6	33.3	1.8
Ornithine	1.90E-03	7.91E-03	24.3	33.6	4.2
Citric Acid	4.90E-04	6.39E-03	13.3	26.3	13.0
Glucopyranose 1	5.96E-03	--[a]	58.8	--[a]	--[b]
Glucose 1	3.28E-02	6.36E-04	33.7	106.2	0.02
Erythrose	1.09E-03	--[a]	59	--[a]	--[b]
Lysine	3.39E-03	1.09E-02	88.4	14.6	3.2
Glucose 2	1.35E-02	--[a]	38.9	--[a]	--[b]
Glucopyranose 2	5.88E-03	--[a]	70	--[a]	--[b]
C6 Sugar Phosphate 1	5.46E-05	--[a]	66.3	--[a]	--[b]
Glucose-6-Phosphate 1	4.85E-04	1.85E-04	27.7	14.6	0.4
Glucose-6-Phosphate 2	9.71E-05	3.16E-05	22	8	0.3
C6 Sugar Phosphate 2	8.42E-05	2.97E-05	87	29	0.4
C6 Sugar Phosphate 3	6.50E-05	--[a]	13	--[a]	--[b]
Fructose Diphosphate 1	1.97E-04	--[a]	12.8	--[a]	--[b]
Fructose Diphosphate 2	3.02E-04	--[a]	12.5	--[a]	--[b]
Trehalose	1.04E-03	6.92E-02	41.3	1.5	66.8

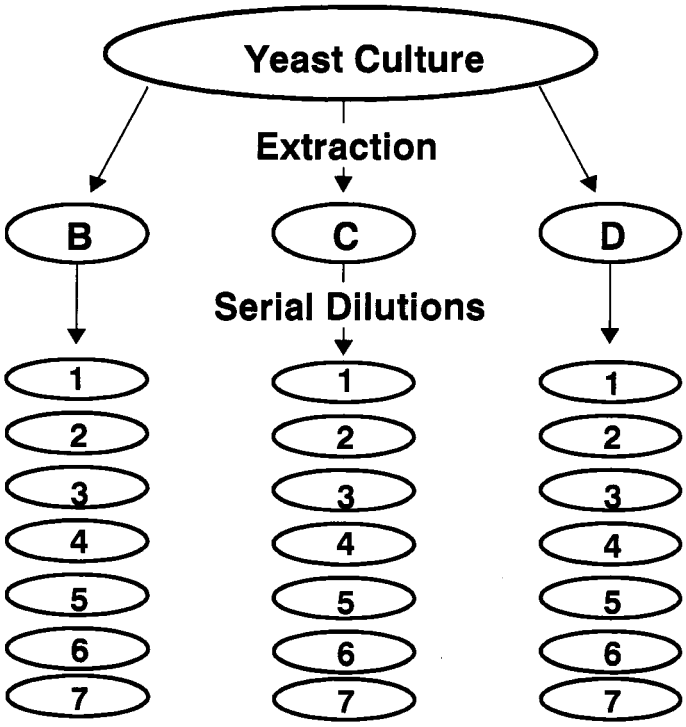


Figure 3.1: Schematic of optimization study.
Three aliquots from the yeast culture were collected for serial dilutions. Each sample contains half the number of cells of the previous sample. Sample 1 contains 4.5×10^7 cells and sample 7 contains 7.0×10^5 cells.

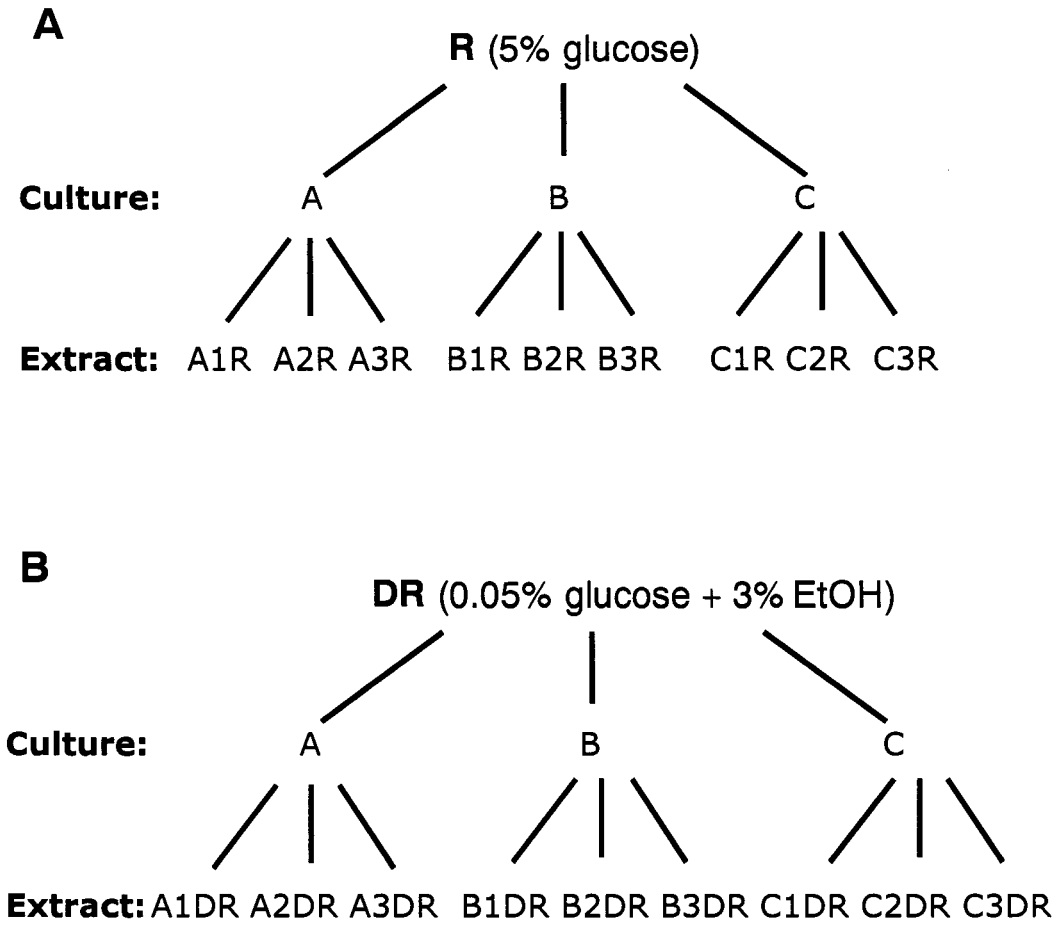


Figure 3.2: Experimental design and nomenclature for R and DR samples. Cultures are labeled as A, B or C. Sample extracts have the aliquot number (1, 2 or 3) in addition to the culture label. An R follows the sample extract number when referring to a fermenting (repressed) sample (A) and DR when referring to a respiring (derepressed) sample (B). The injection number for each extract is written out. For example: repressed sample extract, culture A, extract 2 injection 3 would be written as A2R injection 3.

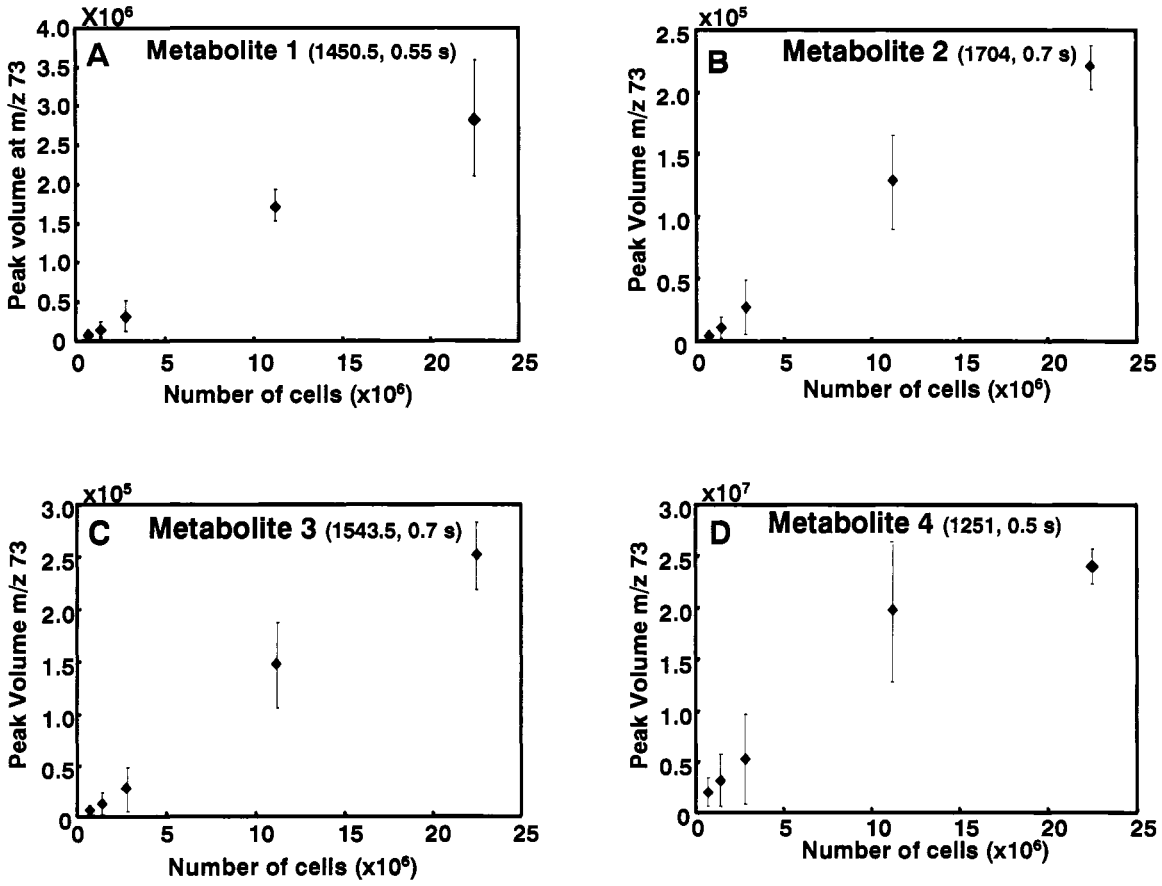


Figure 3.3: Response curves.
(A) Compiled results for metabolite 1, (B) metabolite 2, (C) metabolite 3 and (D) metabolite 4. The peak volumes cover metabolites over three orders of magnitude. The bars represent the error in extraction replicates.

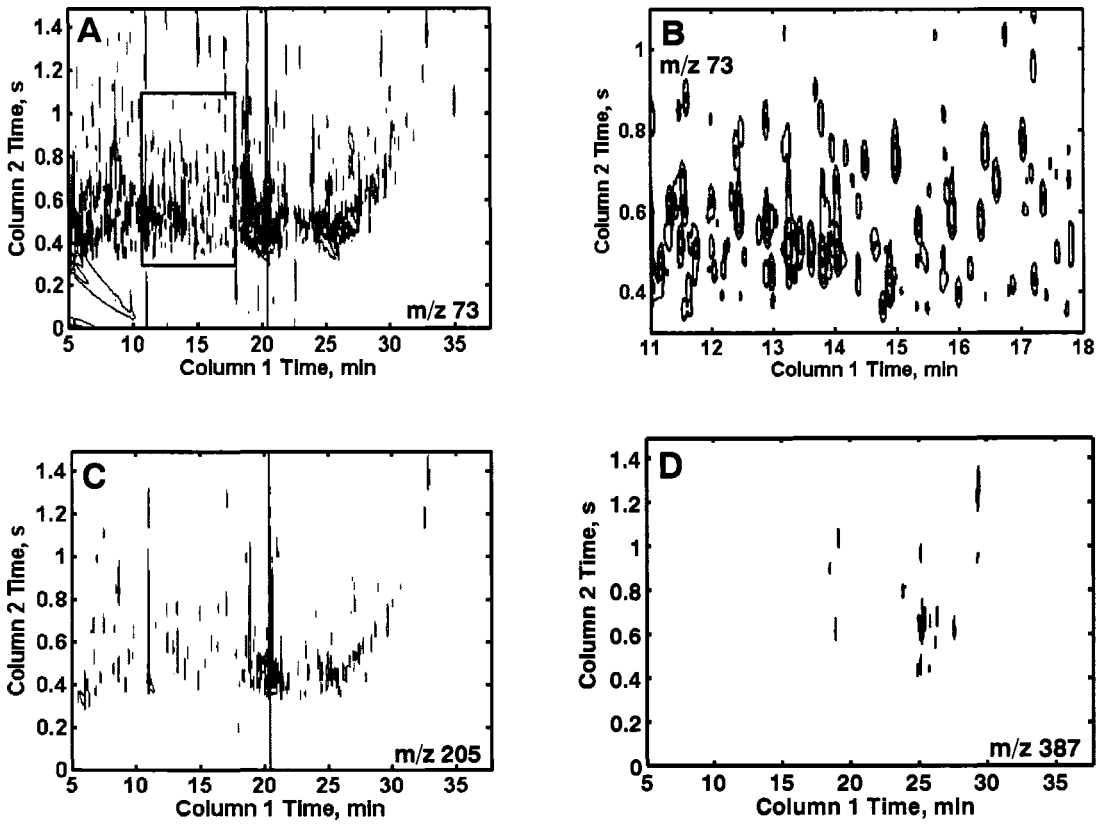


Figure 3.4: Two-dimensional plots of a R sample at selected mass channels. (A) Mass channel (*m/z*) 73, (B) boxed region of (A) emphasizing the separation resolution achieved. (C) Mass channel 205, which is fairly selective for carbohydrates and (D) mass channel 387, which is selective for sugar phosphates. The selective mass channels significantly reduce the complexity of the 2D chromatograms while providing information of interest.

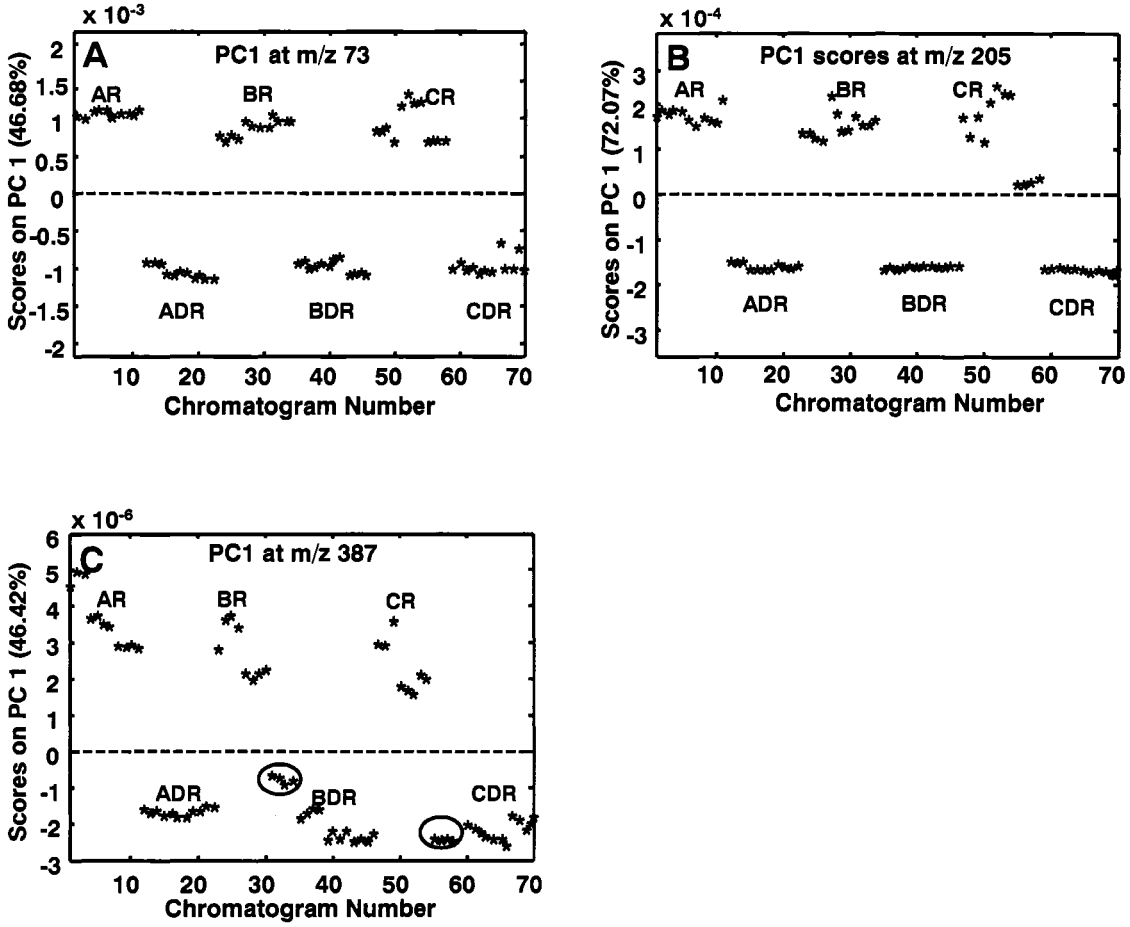


Figure 3.5: PC1 scores plots.
Results after submission of all 70 chromatograms to PCA following normalization to the TIC. PC1 scores plots at (A) *m/z* 73, (B) *m/z* 205 and (C) *m/z*387. Mass channel 387 required alignment in addition to normalization. The circled B3R and C3R sample extracts in (C) overlap with the DR sample extracts.

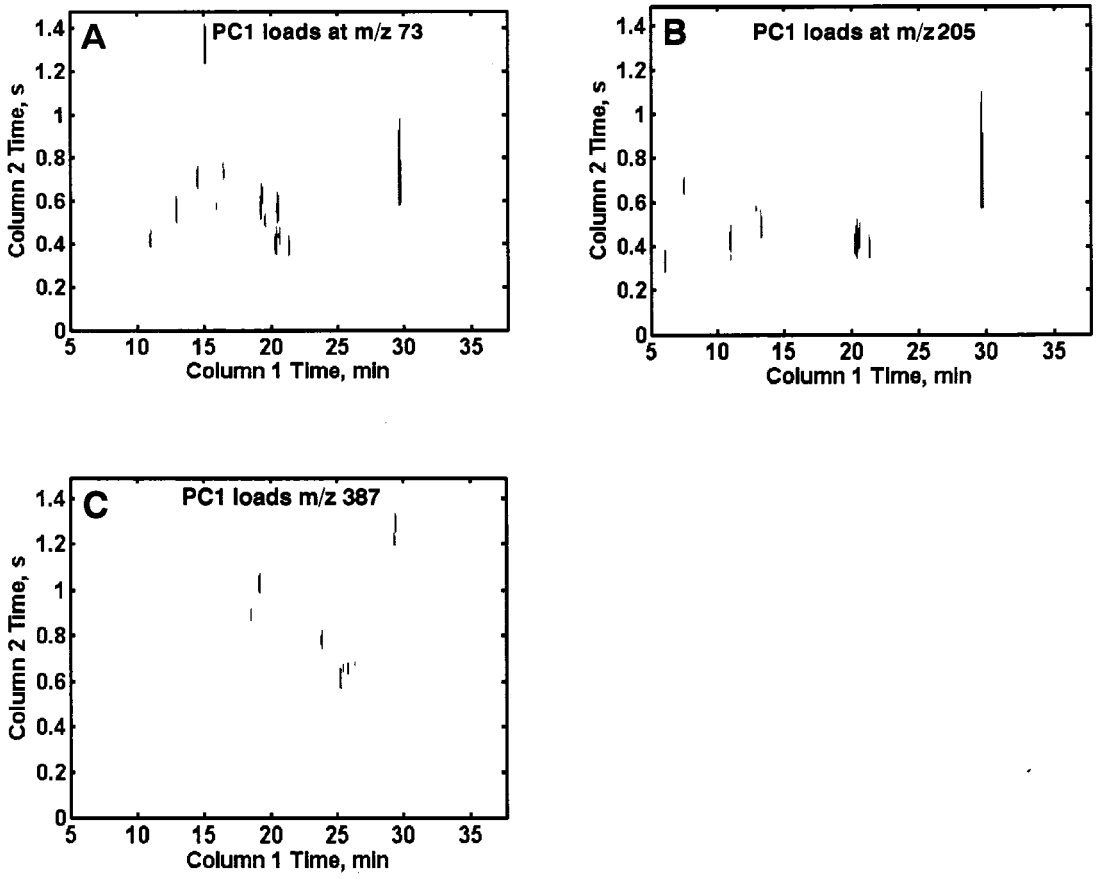


Figure 3.6: PC1 loadings plots.
Loadings results after submission of all 70 chromatograms to PCA. The thresholds were empirically set to include only metabolites and not reagent artifacts, see text for values. PC1 loadings plot for (A) m/z 73, (B) m/z 205 and (C) m/z 387. Fifteen differentiating compounds were found at m/z 73 (A). Mass channel 205, (B) obtained an additional three metabolites and m/z 387 (C) obtained nine chromatographic peaks not found at m/z 73.

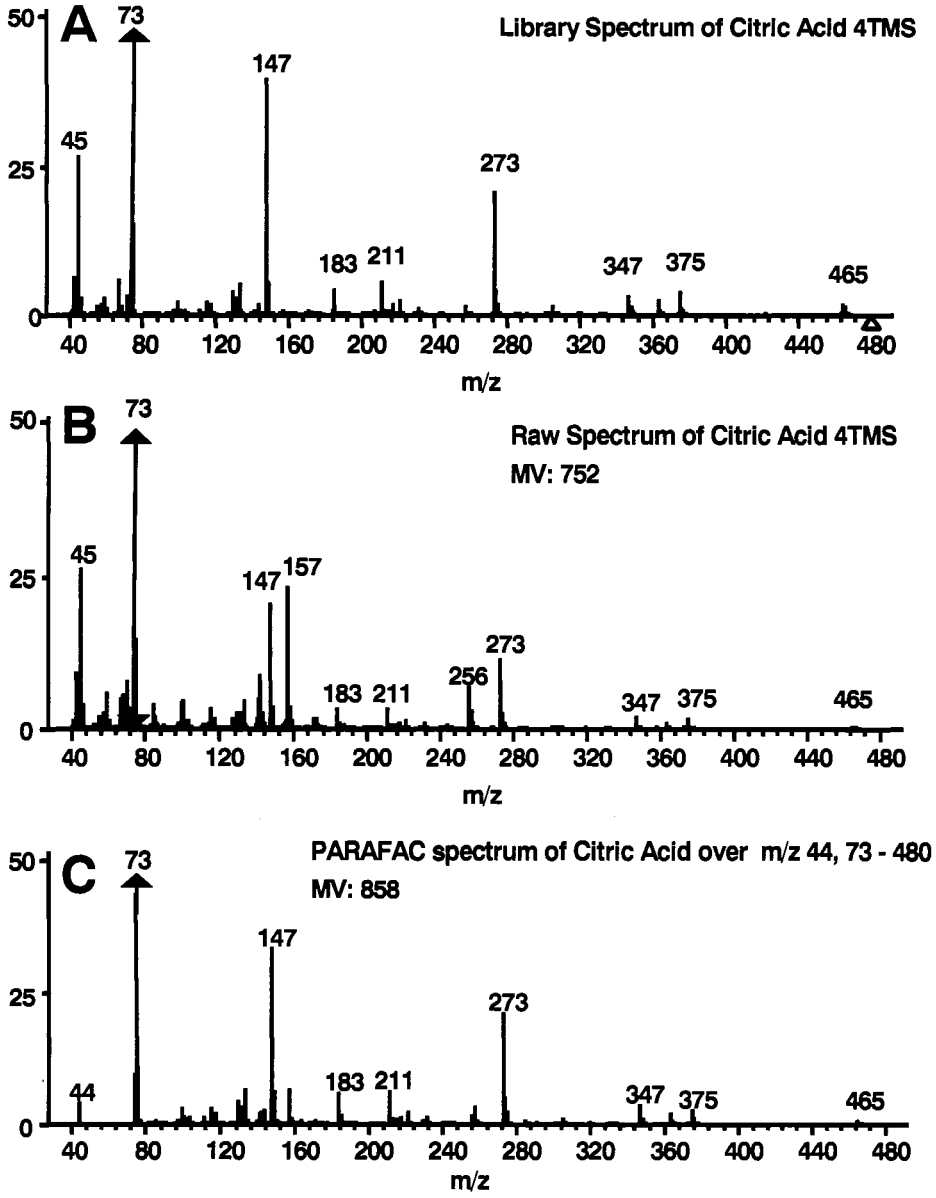


Figure 3.7: PARAFAC mass spectral purification demonstration.
An example of PARAFAC as a deconvolution and quantification tool for citric acid 4TMS in 2D chromatographic region. Mass spectra from (**A**) NIST main library, (**B**) raw and (**C**) PARAFAC of pure component are shown. The raw spectrum is defined as the spectrum obtained using the ChromaTOF software after baseline correction. The similarity match values (MV) for the raw (B) and PARAFAC (C) spectra are given. The raw spectrum (B) has significant extraneous fragments at 157 and 256.

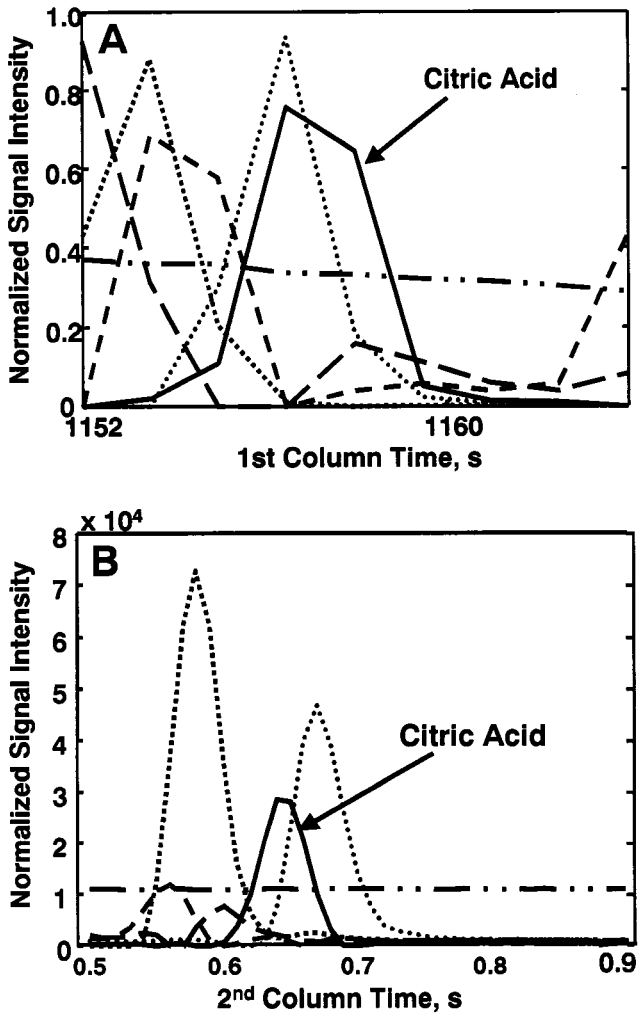


Figure 3.8: PARAFAC using selected mass channels for quantification.
Using only seven mass channels, the pure component peak profiles for citric acid 4TMS on column 1 (A), and column 2 (B) was separated from five overlapping interferents. The ratio of the fragment intensities (C) is comparable to those in the library spectrum. The protocol for choosing these 7 mass channels is described in the text.

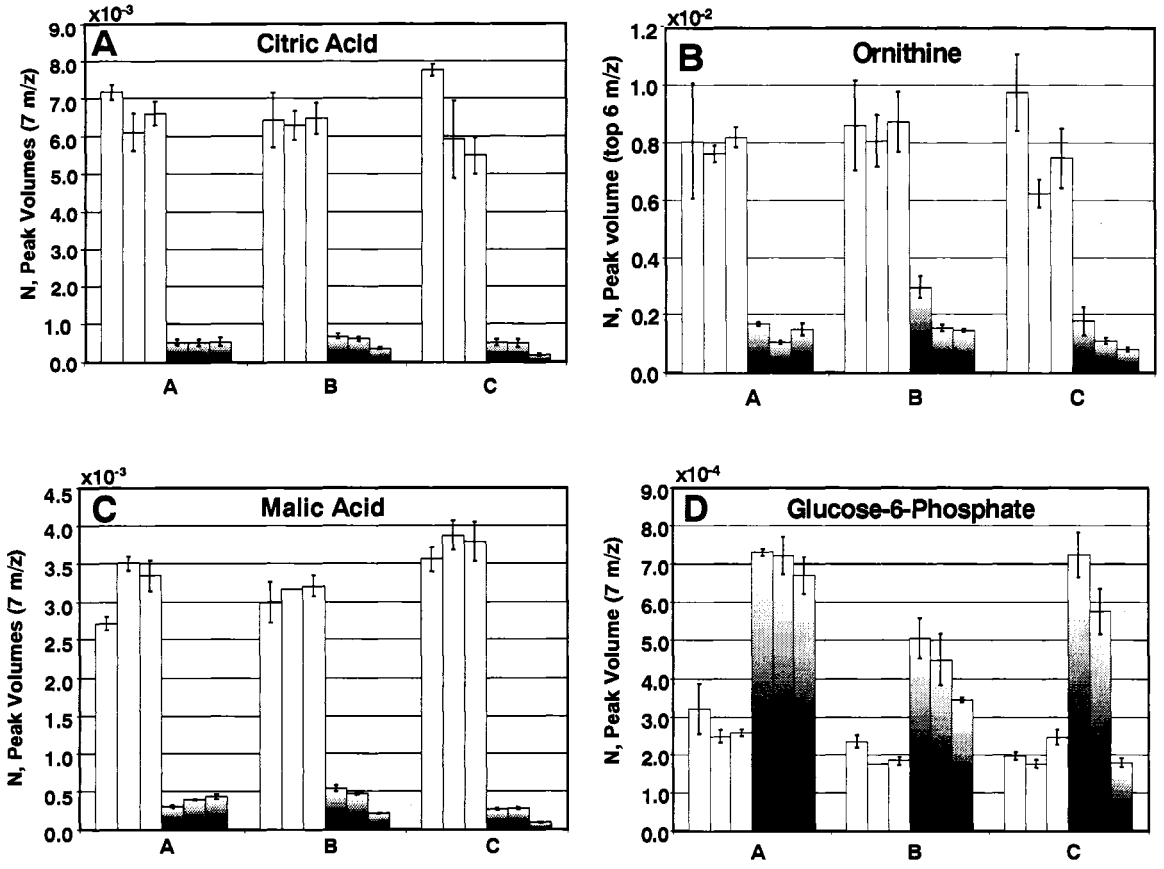


Figure 3.9 A,B: Detailed analysis of the peak volumes, N, obtained using PARAFAC Variation in extract-to-extract as well as culture-to-culture comparisons for R and DR samples is demonstrated. The DR is shown in the white bars and the R by the shaded bars. The error bars represent the injection-to-injection standard deviation. In citric acid (A), ornithine (B), malic acid (C), the R and DR peak volumes from each of the cultures fall within the same range. In glucose-6-phosphate (D) the R and DR are not as clearly separated as in (A), (B) and (C). This is in agreement with what is shown in the scores plot of m/z 387, Figure 3.5C. the R and DR peak volumes from each of the cultures fall within the same range. If no error bars are shown then only two injections were quantified.

Notes to Chapter 3

(1) Hope, J.; Sinha, A.; Prazen, B.; Synovec, R. *J. Chromatogr. A* **2005**, *1086*,185-192.

(2) Sinha, A. E.; Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1027*, 269-277.

(3) Song, S. M.; Marriott, P.; Kotsos, A.; Drummer, O. H.; Wynne, P. *Forensic Science International* **2004**, *143*, 87-101.

(4) Lu, X.; Cai, J.; Kong, H.; Wu, M.; Hua, R.; Zhao, M.; Liu, J.; Xu, G; *Anal. Chem.* **2003**, *75*, 4441-51.

(5) Focant, J.; Sjödin, A.; Turner, W. E; Patterson, Jr. D.G. *Anal. Chem.* **2004**, *76*, 6313-6320.

(6) Sinha, A.E.; Prazen, B.J.; Synovec, R.E. *Anal. Bioanal. Chem.* **2004**, *378*, 1948-1951.

(7) Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Lidstrom, M. E.; Synovec, R. E. *Talanta* **2005**, *65*, 380-388.

(8) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Jack, R. M.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1058*, 209-215.

(9) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1056*, 145-154.

(10) Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmerman, R.; Fiehn, O.; *Metabolomics* **2005**, *1*, 65-73.

(11) Shellie, R. A.; Welthagen, W.; Zrostliková, J; Spranger, J.; Ristow, M.; Fiehn, O.; Zimmermann, R. *J Chromatogr. A* **2005**, *1086*, 83-90.

(12) Jonsson, P.; Gullberg, J.; Nordström, A.; Kusano, M.; Kowalczyk, M.; Sjöström, M.; Mortiz, T. *Anal. Chem.* **2004**, *76*, 1738-1745.

(13) Jonsson, P.; Johansson, A. L.; Gullberg, J.; Trygg, J.; A, J.; Grung, B.; Marklund, S.; Sjöström, M.; Antti, H.; Mortiz, T. *Anal. Chem.* **2005**, *77*, 5635-5642.

(14) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S.J.; Oliver, S.G. *Phytochemistry* **2003**, *62*, 929-937.

(15) Sherman, F. *Methods Enzymol.* **1991**, *194*, 3-21.

(16) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Anal. Chem.* **2000**, *72*, 3573-3580.

(17) Johnson, K. J.; Wright, B. W.; Jarman, K. H.; Synovec, R. E. *J. Chromatogr. A* **2003**, *996*, 141-155.

(18) Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2000**, *52*, 1-4.

(19) A, Jiye; Trygg, J.; Gullberg, J.; Johansson, A. I.; Jonsson, P.; Antti, H.; Marklunk, S.; Moritz, T. *Anal. Chem.* **2005**. *77*, 8086-8094.

(20) Harvey, D. J.; Horning, M.G.; *J. Chromatogr.* **1973**, *76*, 51-62.

(21) Fiehn, O.; Kopka, J. Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nature Biotechnology* **2000**, *18*, 1157-1161.

(22) Voit, E. O. *J. Theor. Biol.* **2003**, *223*, 55-78.

(23) Young, E.T.; Dombek, K. M.; Tachibana, C.; Ideker, T. *J. Biol. Chem.* **2003**, *278*, 26146-26158.

(24) Wu, J.; Zhang, N.; Hayes, A.; Panoutsopoulou, K.; Oliver, S.G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3148-3153.

Chapter 4. Comprehensive Analysis of Yeast Metabolite GCxGC-TOFMS Data: Combining Discovery-Mode and Deconvolution Chemometric Software[§]

4.1 Introduction

Comprehensive two dimensional gas chromatography, GC x GC, is a very powerful analytical technique first introduced in the early 1990s.[1] This technique is based on the gas phase separation of analytes on two capillary columns with complementary stationary phases. Column 1 effluent is repeatedly transferred to column 2 via a modulation device. This modulation device is typically a cryogenic modulator,[2,3] although valve-based modulators have also been reported.[4-6] The separation through two capillary columns results in a multiplicative increase in peak capacity over one-dimensional (1D) separations and is therefore of great interest in the separations of complex mixtures.[7]

Recently, metabolite studies have been performed using comprehensive two-dimensional gas chromatography coupled to a time-of-flight mass spectrometer (GCxGC-TOFMS).[8-13] This instrumentation is a very powerful technology, with the ability to collect up to 500 spectra/s over the mass range of 5-1000 *m/z*. One of the problematic features associated with metabolomic data is the complexity and large volume of the three-dimensional (3D) data, *i.e.* column 1 retention time, column 2 retention time and *m/z*, provided by the GCxGC-TOFMS.[14-18] The two-dimensional (2D) separation space contains a large number of peaks, some of which overlap,[8,19] even though there is

---

[§] Large portions of this chapter have been previously published:
Rachel E. Mohler, Kenneth M. Dombek, Jamin C. Hoggard, Karisa M. Pierce, Elton T. Young, Robert E. Synovec, Analyst, 2007, 132, 756-767

separation improvement from 1D GC.[9] It is particularly important to analyze the data without overlooking important information, while at the same time reduce the data to rapidly find locations of interest or biological significance. Once the metabolites are found and identified, they must be accurately quantified. The post-injection analysis must be compiled into discovery-based high-throughput and user-friendly software using objective data reduction in order to save time.[20,21] The current commercially available software is not designed to facilitate this discovery-based approach. There are chemometric tools available, however, that could potentially be compiled to sort and quantify 3D data using the multivariate selectivity advantage provided by third order instruments such as the GCxGC-TOFMS. The order of an instrument is determined by the number of data dimensions per sample.

One popular chemometric tool of choice used for differentiating complex samples is Principal Component Analysis (PCA).[22] PCA is an unsupervised tool since no class information is given. The algorithm assumes that each sample is different, and clusters the samples based on similarity of the sample data profile. PCA calculates the signal variance between the samples in the data set and projects the results onto principal components (PCs). The first PC captures the largest signal variance with each decreasing PC capturing less. Generally only the first few PCs are studied because they model the chemical variance and provide the desired insight into the analysis. Each PC contains a score for each sample and a loadings value for each variable. Samples from the same class (or with similar chromatographic features) will have scores that cluster together and *vice versa*. Since it is not specified from which class each sample originates, within-class

variation can be severely detrimental to the clustering. However, PCA has the advantage of visually depicting how similar or different the samples are by the tightness of the clustering. The loadings plots identify which chromatographic features differentiated the sample clusters. PCA applied to the entire data cube, *i.e.* every mass channel, has not yet been compiled in an automated fashion, further limiting its usefulness in discovery-based metabolomics.

A supervised tool commonly used for determining the differences between classes of samples is based on calculating Fisher ratio values for a multi-class system.[12, 23, 24] The Fisher ratio is calculated based on the signal intensity differences between samples at each point in the separation space for the entire data set, then features are sorted in ascending order, thus providing a sorted list of features mathematically defined to correspond to class-discriminating power. This analysis method requires the prior knowledge of sample type, but is more robust than PCA against within class variation when trying to differentiate samples.[12] Instead of outputting one score value for each 2D chromatogram like PCA, the Fisher ratio algorithm outputs a Fisher ratio at each point in the separation space. The Fisher ratio is defined as the class-to-class variation divided by the sum of the within-class variations.[23] Thus, as a peak tends to increasingly discriminate the classes, the Fisher ratio increases for that peak. Recently, the Synovec Group applied a Fisher ratios method to 3D GCxGC-TOFMS data using a novel indexing scheme to find chromatographic locations differentiating three sample classes.[12] This new Fisher ratio method utilized the entire GCxGC-TOFMS data cube, *i.e.* column 1,

column 2, full mass spectra, to locate class-distinguishing locations in the 2D chromatographic separation space for all monitored mass channels.

Another key component of discovery-based high throughput data analysis is PARAFAC.[22, 25, 26] PARAFAC uses the 3D data to build deconvoluted peak models for the specific sections of the GCxGC separation space that can be used for multiple chromatograms. PARAFAC uses alternating least squares to remove noise and mathematically separate (or deconvolute) overlapping components to produce pure chromatographic profiles and pure mass spectra. Thus, each chromatographic peak has a unique mass spectrum, column 1 and column 2 profiles. The purified 3D peak can then be reconstructed and integrated to give a peak volume.

In chapter 3 we reported the use of GCxGC-TOFMS with PCA applied to discover differentiating metabolites of yeast cells grown under two conditions, fermenting (glucose repressed, R) and respiring (derepressed, DR).[8] In this study PCA was used on three selected mass channels (*m/z*) to locate metabolites, and resulted in the discovery of 26 class-differentiating metabolite peak locations.[8] At the time of the study, no third order chemometric classification tool for analysis of four dimensional (4D) data, *i.e.* column 1 retention time , column 2 retention time, *m/z*, and sample number, was available for comprehensive analysis of every *m/z* collected. Recently in the Synovec lab, a Fisher ratios algorithm for GCxGC-TOFMS data was developed, which is user-friendly and uses the entire raw data cube to calculate a weighted Fisher ratio at each point in the 2D separation space for each *m/z*. This Fisher ratio information provides insight into the class differentiating power of every chromatographic location for every

m/z. With this newly developed Fisher ratio method, it was decided that this technique should be applied to the DR/R yeast data in an exhaustive fashion to explore the limits of determining class-differentiating metabolites. Once the class differentiating locations are determined, PARAFAC can then be used to obtain accurate quantification and a full mass spectrum for definitive peak identification. PARAFAC outputs peak volumes that provide concentration information to measure the extent of class distinguishing power as well as aid in determining the metabolic pathway activity of the cells. A Student's *t*-test is used on the deconvoluted data to determine if the relative concentration of metabolites in the two classes is statistically different, thus indicating whether or not the concentration ratio, N_{DR}/N_R, is statistically significantly different than one. The reported mass spectral match values (MV) provide the confidence of the metabolite identification. Also, since the PCA study, an in-house metabolite library was created to confirm elution times and include metabolites not found NIST or Max Planck Institute of Molecular and Plant Physiology libraries. This is the first comprehensive GC x GC-TOFMS report of the identification of differentiating metabolites between the fermenting and respiring conditions using 4D chemometric tools.

4.2 Experimental

4.2.1 Yeast strain and growth conditions

The experimental conditions were described in Chapter 3, however a brief description follows. The yeast strain used in this study was W303-1a (*MATa ade2 can1-100 his3-11,15 leu2-13,112 trp1-1 ura3-1*). Cells were grown at 30°C in synthetic complete medium (SC) [27] containing either 5% glucose to prepare fermenting cells (R) or

3% ethanol and 0.05% glucose followed by incubation for 6 hours to prepare cells metabolizing ethanol by respiration (DR). Three R cultures (AR, BR, CR) and three DR cultures (ADR, BDR, CDR) were grown to obtain a measure of biological variability.

4.2.2 Extraction and derivatization of metabolites

As described in the initial report, the quenching of the metabolic activity and extraction of the small polar metabolites were performed in a similar fashion to that reported by Castrillo et. al.[28] A volume of each culture containing $1x10^7$ cells was quenched and collected for extraction. The metabolite extracts were dried in a SpeedVac at room temperature and stored at -80°C under argon until GCxGC-TOFMS analysis. The dried metabolite extracts were methoximated (30 μL, 20 mg/ml methoxyamine in pyridine) and trimethylsilylated (70 μL, BSTFA: TMCS, 99:1) in order to replace the active hydrogen atoms with the trimethylsilyl (TMS) functional group.[29] The methoximation took place at 30°C for 90 minutes and the silylation step at 60°C for 60 minutes. Each extraction was performed in triplicate, *e.g.* A1R, A2R, and A3R.

4.2.3 Metabolite standards

One hundred and three individual metabolite standards from potentially active metabolic pathways (Tables 4.1-4.8) were prepared at a concentration of 1.0 mM. Derivatization conditions were the same as reported for the yeast cells.

4.2.4 Instrumental parameters

An Agilent 6890N gas chromatograph equipped with an Agilent 7683 auto-injector (Agilent Technologies, Santa Clara, CA, USA) coupled with a LECO Pegasus III time-of-flight mass spectrometer with a 4D thermal modulator upgrade (LECO, St.

Joseph, MI, USA) was used to collect the 3D GCxGC-TOFMS data, which could then be stacked to create 4D data (column 1 retention, column 2 retention, *m/z*, and sample number). The GC inlet and transfer line were held constant at 280°C. One micro liter split-less injections were made onto an RTX-5MS column, *i.e.,* column 1 of the GCxGC, (20 m x 250 μm i.d. x 0.5 μm, Restek, Bellefonte, PA, USA). The collected column 1 effluent was transferred to an RTX-200MS ,*i.e.,* column 2 of the GCxGC, (2 m x 180 μm i.d. x 0.2 μm, Restek, Bellefonte, PA, USA) every 1.5 s. Column 1 was held at 60°C for 0.25 minutes and then increased at 8°/min to 280°C where it was held for 10 minutes. Column 2 was initially set at 70°C and followed the same temperature program as column 1 giving a total run time of 37.75 minutes. The modulator was kept 40° higher than the column 1. A constant flow rate of 1 ml/min was held at the head of column 1. The ion source was set to 200°C. Data were collected at a rate of 100 spectra/s after a 5-minute solvent peak delay. Mass channels 73-500 *m/z* were used for the analysis. Each extract had 4 replicate injections except A1R and A1DR, which were injected in triplicate. This resulted in 70 injections (70 GC x GC-TOFMS chromatograms) for the analysis.

4.2.5 Data Analysis

The LECO ChromaTOF software v.2.21 (LECO, St. Joseph, MI, USA) was used to collect the data and obtain a signal-to-noise ratio (S/N) at the locations of interest. The PARAFAC-deconvoluted mass spectra were submitted to three mass spectral matching libraries: the NIST main library, an in-house metabolite library, and the metabolite library

obtained from the Max Planck Institute of Molecular and Plant Physiology (http://www.mpimp-golm.mpg.de/mms-library/index-e.html).

The previously developed Fisher ratio method for GCxGC-TOFMS was used to analyze all samples at each mass channel, as illustrated in Figure 4.1. The algorithm automatically imports raw ChromaTOF data and converts LECO format chromatographic data into Matlab variables, providing unfiltered, raw data in Matlab format. Since the Fisher ratio algorithm is based on signal intensities, the raw data was baseline corrected and normalized to the total ion chromatogram, thus correcting sample and injection variations. Given class information, the 4D Fisher ratio algorithm calculates a Fisher ratio at every point in the chromatographic space for each mass channel *(m/z)* of that baseline corrected, normalized data. The Fisher ratio is defined as the class-to-class variation divided by the sum of the within class variations. The Fisher ratio for each *m/z* is then weighted by the mean signal at each point to substantially minimize any noise points that might be correlated with the different classes. In this study the weighted Fisher ratios for each peak in each *m/z* are then summed to give a 2D Sum of Fisher Ratios plot, then sorted in descending order to output a list of peak locations in the chromatographic space that potentially differentiate the sample classes. The analyst must then decide upon a threshold Fisher ratio value that mathematically separates the class-distinguishing features from the uninformative features. This greatly reduces the data for submission to PARAFAC.[12] All Fisher ratios reported in this report are weighted as specified earlier.

A PARAFAC GUI developed in-house and implemented using the PLS toolbox version 3.51 (Eigenvector Research, Manson, WA) was used to quantify the metabolites that the weighted Fisher ratios indicated had variability between R and DR sample extracts. The important chromatographic sub-regions of the peaks located using 2D Sum of Fisher Ratios were submitted to the PARAFAC GUI to obtain the deconvoluted 2D chromatographic profiles as well as the deconvoluted mass spectra. This GUI provides an automated deconvolution for all metabolites in the queue and an objective determination of the number of factors used for the deconvolution.[26]

To validate the statistical significance of the peak volumes obtained from each sample type, (R and DR) a comparison of means was made using the Student's *t*-test.[30] Prior to calculating a *t* value for each metabolite an F-test was performed to determine the similarity of the standard deviations for each metabolite from each sample type. Whether a *t*-test for equal or unequal variances was used based on the outcome of this F-test. Using the t-test to validate the discriminating peaks may seem like circular reasoning, but the Fisher ratios were applied to the entire data set to reduce the data set by finding subregions of interest before deconvolution. Only those subregions of interest were submitted to the deconvolution software, and the *t*-test was applied to the deconvolution results to validate that the pure, deconvoluted peak volumes were statistically different between the sample classes to a given confidence (95%).

4.3 Results and Discussion

4.3.1 Data reduction

The yeast metabolic extracts were shown to be sufficiently complex and ideally suited for GCxGC-TOFMS analysis. In the chapter 3 analysis, however, information from the entire cube of data was not comprehensively utilized.[8] Typically, when metabolite samples are compared using a gas chromatograph interfaced with a mass spectrometer, either *m/z* 73 or the TIC are analyzed. Mass channel 73 is frequently used because the derivatized yeast metabolite extracts have been methoximated and trimethylsilylated, and all trimethylsilylated adducts will produce a significant fragment at *m/z* 73, as seen in Figure 4.2.[9, 10, 31, 32] Although there are a large number of components at *m/z* 73, there are many other *m/z* that should be included in the data analysis. Peaks in these other mass channels are more likely to be free of significantly overlapping components than peaks at *m/z* 73.[19] It is difficult to visually compare the chromatograms from the different sample types, *e.g.*, R and DR, at *m/z* 73, and manually locating differing peaks would be difficult considering the large number of peaks present and the volume of data to be compared. The Fisher ratio method reduced the entire data cube in an automated fashion with little preprocessing to locate differentiating peak locations.[12] After processing the data as shown in Figure 4.1, the weighted Fisher ratios calculated by comparing all 70 sample chromatograms for mass channels 73-500 were summed to give a 2D Sum of Fisher Ratios plot, Figure 4.3. A comparison of the 2D Sum of Fisher Ratios plot, in Figure 4.3, to the original chromatogram, in Figure 4.2, showed that the number of peaks in the data is drastically reduced thereby greatly

decreasing the amount of data included in the comprehensive study. The streaking across the 2nd dimension in Figure 4.3 at approximately 30 minutes was due to trehalose, which at the unique *m/z*, has a S/N >10,000 in the DR but has a S/N ~450 in the R samples. Unlike PCA, the Fisher ratio method did not require deletion of chromatographic sections to locate class-to-class distinguishing peaks. Also, the Fisher ratio method is more robust to within-class variation. [12]

The differentiating peaks found in the GCxGC-TOFMS separations of R and DR samples were visualized by projecting the 2D Sum of Fisher Ratios onto the column 1 separation dimension, as shown in Figure 4.4. This projection allows the magnitude of the weighted Fisher ratios to be readily visible with the assumption that the larger 2D Sum of Fisher Ratios corresponds to analyte locations that most readily distinguish the DR and R samples. At first glance, in Figure 4.4A, the Fisher ratio method does not appear to find many differing peak locations. However, upon closer inspection, a larger number of peaks present at lower Fisher ratio sums is observed as shown in Figure 4.4B, which is a sub-region of Figure 4.4A. If the 2D Sum of Fisher Ratios plot is further amplified the Fisher "noise" level can be determined, where the noise or "LOD" level is defined as the level below which the ratio of chromatographic signals no longer differentiates the classes. In this study the noise level in the 2D Sum of Fisher Ratios plot is slightly less than $3x10^{-4}$ as shown in Figure 4.4C, which is another sub-region of Figure 4.4A. The range of the weighted Fisher ratios that truly differentiated classes lies between 1.43 (the largest 2D Sum of Fisher Ratios see Table 4.9) and $3x10^{-4}$ (the minimum 2D Sum of Fisher Ratios).

4.3.2 Metabolite standards

Prior to mining the locations found using the 2D Sum of Fisher Ratios, metabolite standards were injected into the instrument and run under the same conditions as the yeast metabolite extracts. Definitive identification of compounds can be difficult as shown in Chapter 3. To address this problem as well as to gain an understanding of the types of compounds amenable to GC analysis, 103 metabolites ranging from cofactors to pentose phosphate intermediates to amino acids were prepared and injected individually. Seventy-nine of those metabolites injected were determined amenable to GC analysis and the remaining 24 were not detected as the intact compound, Tables 4.1-4.8. It was determined that the solubility of the metabolites in pyridine varied widely leading to large differences in signal at *m/z* 73, with the sugar phosphates giving a much lower signal than the amino acids and sugar alcohols. Although the larger compounds, *e.g.* cofactors and folic acid, dissolved in pyridine, they decomposed in the injection inlet and were not detected as the intact trimethylsilylated compound. Succinyl CoA decomposed to produce a signal for succinate that is an eighth the signal produced by a succinate standard at the same concentration. The peak for ribose 5 phosphate also had two contributors, NADH and ribose 5 phosphate. The ribose 5 phosphate signal was a factor of two larger than the NADH contribution at the same concentration. Glucose 1 phosphate arose from both glucose 1 phosphate and the decomposition of UDP glucose. Other than these three cases, metabolite peaks arise from only one source and can give retention time confirmation to aid in the definitive identification of eluting compounds.

4.3.3 2D Sum of Fisher ratio thresholds

The 2D Sum of Fisher Ratios threshold was set to consecutively decreasing values to establish a noise baseline, i.e. the threshold value at which the number of reagent artifact peaks begins to outnumber the known and unknown metabolite peaks. These results are summarized in Figure 4.5. The 'x' label indicates the number of identified (or known) metabolites. The '+' label indicates the number of the known metabolites that are statistically changing between the DR and R sample types according to the *t*-test with a 95% confidence limit. A significant change was defined as a N_{DR}/N_R that is statistically different than one. The number of known metabolite peaks increased to 73 as the 2D Sum of Fisher Ratios threshold is decreased. Of these 73, there were 54 locations (summarized in Tables 4.9-4.10) that correspond to identified and significantly differentiating metabolites. Fisher analysis is generally a data reduction method used to rapidly locate class distinguishing features. However, we consecutively set the 2D Sum of Fisher Ratios lower to analyze the data in a comprehensive fashion, so even metabolites that did not differentiate classes were located as shown in Table 4.11. The Fisher ratio method found 19 additional metabolites that did not have a statistically significant N_{DR}/N_R, according to Student's *t*-test. These 19 metabolites are listed in Table 4.11 and will be described below. The total number of all metabolites (squares) either known or unknown, reagent artifacts (diamonds) and the number of unknown metabolites (triangles) are also indicated in Figure 4.5. The 81 unknown metabolite peaks along with their 2D Sum of Fisher Ratios are listed in Tables 4.12-4.13. An unknown metabolite is a peak that was absent in the reagent blank, and had a pure mass

spectrum output by PARAFAC that received a poor mass spectral match, because it was absent in all of the mass spectral libraries.

The plot shown in Figure 4.5 can be divided into roughly three regions: identified metabolites dominated, metabolite discovery, and Fisher ratio noise limited. The identified metabolite dominated region roughly covers the first three thresholds ($1x10^{-2}$, $5x10^{-3}$, $3x10^{-3}$) and in this region the identified metabolites significantly outnumber both the unknown metabolites and the reagent artifacts. The last three thresholds ($1x10^{-3}$, $5x10^{-4}$, $3x10^{-4}$) include what has been termed the metabolite-discovery section, where the number of unknown metabolites (and reagent artifacts) rapidly increase, to approach and finally outnumber the known metabolites. In the metabolite discovery region the power of the GCxGC-TOFMS data combined with the Fisher ratio method becomes evident, illuminating potentially important metabolites that were previously difficult to see. Mass spectral results from the metabolite discovery region can serve as a database for later structure elucidation of unknown metabolites. The region to the left of the last threshold point ($<3x10^{-4}$) is termed the noise-limited region, where a vast majority of the metabolites do not significantly differentiate the classes and this is where most reagent artifacts are located. A peak finding algorithm located an average of 590 peaks in the chromatograms at *m/z* 73. However, using the Fisher ratio method, at the lowest threshold analyzed, only 157 metabolite peaks were found. This shows that the Fisher ratio method significantly reduced the data prior to data mining with PARAFAC because over 400 peaks were determined to insignificantly distinguish the classes and thus did not require further in-depth PARAFAC analysis. The majority of the 400 peaks arise from

reagent artifacts and column stationary phase bleed, and the remaining peaks (an estimated 20-80) known and unknown metabolites not changing between sample types. Since each Fisher ratio is weighted by the average signal at each point the smaller intensity but potentially important peaks, i.e. lower S/N, are generally found at the lower 2D Sum of Fisher Ratio values, in Tables 4.9 and 4.10. Conversely, when a peak with a large intensity is located at a low 2D Sum of Fisher Ratios value, that peak is often not differentiating the classes, and appears in Table 4.11.

4.3.4 PARAFAC deconvolution/quantification

At each of the locations exposed using the Fisher ratio algorithm, the Leco ChromaTOF software supplied with the instrument can search standard libraries to obtain a compound identification. Even though ChromaTOF has a 3D deconvolution program built into the data processing, it is beneficial to confirm the results using full mass spectral PARAFAC deconvolution, which is built for 3D data arrays. The PARAFAC GUI offers rapid analysis in which the number of mass channels included in the deconvolution can be specified, the three profiles obtained (column 1, column 2 and mass spectral), and the resulting fully deconvoluted mass spectrum for each component in the time window can be exported to the NIST Mass Spectral Database for a mass spectral search. If the fully-deconvoluted mass spectrum does not receive a high mass spectral match to a component already in the libraries, the NIST mass spectral database adds the unknown spectrum to a new in-house library. The retention times of the deconvoluted profiles are also compared to the retention times of the matched library profiles, to minimize incorrect assignments. The power of a PARAFAC full mass spectral

deconvolution is evident in the results for methionine and aspartic acid analysis shown in Figures 4.6-4.7. At *m/z* 73 these two metabolites have a chromatographic resolution of about 0.3 on both column 1 and column 2 separation dimensions. The two metabolite peaks must be distinguished from each other as the mass spectra change across the peak. This challenging GCxGC-TOFMS data sub-region was submitted to PARAFAC, which produced the deconvoluted column 1 and column 2 profiles. The deconvoluted profiles show that these two metabolites have similar intensities and overlap significantly on both the column 1 and column 2 dimensions, Figure 4.6. Aspartic acid and methionine were two of the 103 metabolite standards injected under the same GC conditions as the yeast extracts. The retention times obtained for aspartic acid and methionine standards were three seconds apart on the first dimension and nearly co-linear in the second, which confirms the retention times output by PARAFAC. Standards were run individually so it is expected that there would be some retention time shifting from matrix effects. Identified metabolites that were run on the GCxGC-TOFMS system as in-house library standards providing retention time validation are labeled with a "+" after the MVs in Tables 4.9-4.11.

In addition to the column 1 and column 2 deconvoluted profiles for the derivatized aspartic acid and methionine, the pure mass spectral profiles are also produced by PARAFAC, Figure 4.7. This is a challenging application for PARAFAC since both spectra are very similar, with shared mass fragments at *m/z* 45, 73, 100, 133, 147 and 218. Hence, mass fragments 117, and 232 in the methionine spectrum are residuals from aspartic acid, and therefore lower the mass spectral similarity values

(MV). The correlation of all masses in the sample spectrum to the all masses in the library spectrum is listed as the first MV given. The second MV (after the comma) is the reverse match factor, which only uses the mass fragments present in the library spectrum to obtain a correlation value. In some cases the reverse value can be considered more accurate, especially if mass fragments below a signal threshold are removed from the library spectrum. The MVs listed in Tables 4.9-4.11 are the reverse values for the full, *i.e* all *m/z*, deconvoluted spectra. Since these two components overlap extensively on column 1 and column 2 and because the two components share many of same *m/z*, it is beneficial to use only selected *m/z* for the deconvolution. The use of selected *m/z* also speeds up the analysis. For the quantification of methionine, *m/z* 45, 61, 128 and 160 were used.

Following positive identification of the metabolite peaks, concentration ratios, N_{DR}/N_R, were obtained using PARAFAC, Tables 4.9-4.11. Since it was previously determined that the average biological variation outweighed the average extraction variation only the first extract from each culture and all three biological replicates were used in the quantification.[79] The peak volumes from the injection replicates were averaged followed by the biological replicate averaging, prior to calculating N_{DR}/N_R. The standard Student's *t*-test was used to determine if the concentration ratios following PARAFAC deconvolution were statistically different than one with 95% confidence. In order to determine the degree-of-freedom (d-o-f) for each metabolite, an F-test was performed.[30] In some cases the within-class variation was sufficiently large causing the d-o-f to decrease from 4 to 2, which in turn raises the *t*-value in the Student's *t* table. The

ratio of the calculated *t*-value (from the Student's *t*-test) to the table *t*-value, t_{calc}/t_{table}, will be greater than one for a given metabolite peak that is statistically different between the DR and R classes. Using this procedure, the quantities of 28 metabolites that were present in both sample types were determined to be statistically different, Table 4.9. There were an additional 26 metabolites that were abundant in one sample type, but absent in the other sample type, Table 4.10. These results more than double the number of distinguishing metabolite peaks reported in the initial study using PCA. Nineteen additional metabolite peaks were found using the Fisher ratio method that did not differ between the R and DR sample types according to the *t*-test, Table 4.11. A few of the metabolites in this table would pass the 95% confidence limit if there were less within-class variation, *e.g.,* ornithine, pyruvate and glucose-6-phosphate (G6P 1518). For a d-o-f equal to 4, the t_{table} is 2.776 and for ornithine the t_{calc} value was 3.86, but since this metabolite did not pass the F-test, it was determined that the d-o-f equals two which corresponds to a t_{table} value of 4.303. Pyruvate and G6P 1518 are two other examples with a t_{calc} of 3.20 and 3.79, respectively. Even though these 19 metabolites in Table 4.11 did not pass the 95% *t*-test, it is still interesting to note the metabolites detected in the samples.

In addition to the known metabolites, there were also 81 unknown metabolites found using the Fisher ratio method, Tables 4.12-4.13. An unknown metabolite in this case is defined as a peak in the 2D chromatographic space that is not in the reagent blank and its mass spectrum has specific properties similar to a metabolite mass spectrum, but with a low metabolite mass spectral match value. Figure 4.8 shows an example of an

unknown metabolite located at a column 1 retention time of 1153 s and a column 2 retention time of 0.4 s. The same PARAFAC procedure described above for the known metabolites was used to obtain the full deconvoluted profiles, Figure 4.9. The highest forward match value obtained for this mass spectrum was 761 and, in the library spectrum with the highest match, the fragment at the significant *m/z* of 231 was missing. In some cases there were interferences overlapping with the unknown metabolite peaks Figure 4.8. Once the mass spectral profile was deconvoluted, it was then exported into the NIST Mass Spectral Database, and compiled into an in-house library. This library spectrum could then be used to select *m/z* for the quantification of the unknown metabolites using PARAFAC in order to calculate N_{DR}/N_R, Table 4.12. A number of these unknown derivatized metabolites have distinguishing mass fragments at higher *m/z*. Two examples are shown in Figure 4.10. The compound unk1744 has a large fragment at a *m/z* of 406 (Figure 4.10A), and in spite of a very low S/N (Table 4.13) a good mass spectrum was obtained. The unknown with the highest N_{DR}/N_R, unk1174, also had a large fragment at a higher *m/z*, Figure 4.10B. These deconvoluted mass spectra can be posted online for anyone interested to examine, interpret, and utilize. The unknown metabolites that are found to distinguish samples in future yeast studies can be potentially identified using updated mass spectral libraries, or the mass spectral fragmentation patterns.[33] Certain mass fragments are indicative of specific compound fragments. For example, a mass fragment at *m/z* 190 is indicative of a compound containing $CH(OTMS)_2$ and *m/z* 147 represents $(CH_3)_2$=O-Si$(CH_3)_3$ which stems a compound which contains at least two trimethylsilyl groups. The most class differentiating concentration

ratio for the unknown metabolites was 0.1 (essentially 10:1 in favor of R), whereas the largest for the known metabolites was 66.8. As with the known metabolites, the unknown metabolites were also submitted to the Student's two-method *t*-test, and the t_{calc}/t_{table} value is given in Table 4.12. The locations with high S/N but lower 2D Sum of Fisher Ratio values were not statistically different between the samples types and the t_{calc}/t_{table} was less than one.

4.3.5 Biological interpretations

The full biological implications of the information reported are beyond the scope of this chapter, however a few interesting results should be noted. The addition of fructose 2,6-bisphosphate to the in-house metabolite standard library identifies the unknown C6 sugar phosphate1 peak located in the prior study[79]. Fructose 2,6-bisphosphate is an allosteric regulator of the glycolytic pathway. During glycolysis fructose 2,6-bisphosphate activates phosphofructokinase, which stimulates the conversion of fructose-6-phosphate to fructose 1,6-bisphosphate. The greater abundance of fructose 2,6-bisphosphate and fructose 1,6-bisphosphate in R cells is consistent with a highly activated glycolytic carbon flux. Fructose 1,6-bisphosphate is an allosteric activator of pyruvate kinase. The reaction catalyzed by pyruvate kinase produces pyruvate near the end of the glycolytic pathway and pyruvate was also found to be of greater abundance in R cells although not to the 95% confidence level, Table 4.11. Phosphofructokinase and pyruvate kinase are key regulated enzymes of glycolysis.[34] Fructose 2,6-bisphosphate also allosterically inhibits fructose 1,6-bisphophate phosphatase which catalyzes the reverse reaction of phosphofructokinase. This prevents the opposing pathways of

glycolysis and gluconeogenesis from creating a futile cycle where ATP is hydrolyzed without creating biomolecules or fueling energy intensive processes useful to the cell. Previously studies in metabolomics and transcriptomics showed that intermediates of the TCA cycle, were more prominent in DR cells.[8, 35, 36] The previous metabolomic study reported only two TCA cycle intermediates (citric acid and malic acid) to be up regulated in respiring conditions; however, in this more comprehensive study succinate and fumarate were also found to be significantly more abundant in the DR cells. This is consistent with an active TCA cycle in derepressed cells and its use to more completely oxidize carbon sources to carbon dioxide for the production of ATP. It is also consistent with repressed cells obtaining ATP primarily through the inefficient oxidation of glucose by the glycolytic pathway. These are just a few of the connections that can be made between the metabolites discovered and their metabolic pathways.

4.4 Conclusions

Significant progress has been made on the automated chemometric software used to analyze large volumes of data collected on the GCxGC-TOFMS instrument in its entirety. The Fisher ratio method has the ability to rapidly identify class-distinguishing peaks in the 2D chromatogram using all mass channels. The amount of data reduction obtained depends on where the 2D Sum of Fisher Ratios threshold is placed. In this study, where the threshold was set down to the noise level, the data was reduced from more than 590 peaks to 157 metabolite peaks, yielding a 4-fold data reduction. The PARAFAC GUI was then used to quantify the interesting locations found by the Fisher ratio algorithm. The complete deconvoluted PARAFAC mass spectrum in conjunction

with 2D retention time data from standards confirmed the identification of important compounds thus improving the confidence in the mass spectral identifications. This procedure or a similar one will be important for biological interpretations of the data. This three-step method of (1) using the Fisher ratios algorithm to find potential sub-regions of interest, (2) submitting those sub-regions to PARAFAC, then (3) checking the pure peak volume ratios with a Students' *t*-test, is meant to be a high-throughput method for discovering and quantifying statistical differences among 4D sample data that preserves and takes advantage of the third order structure of the data set. This software uses the complete dimensionality of the data and, because it is semi-automated, is ideal for high-throughput metabolomic analyses. The following considerations should be pointed out regarding the analysis method described in this work. This technique is the application of a univariate technique to select features for a multivariate model. Thus, in some cases the combination of two or more compounds can be significant where each individual contributing compound is not. These combinations might not be detected by the Fisher ratios method. Users should be aware of sources of variation other than chemical variations that can arise from the data collection set up. Since discriminating features are defined mathematically as any feature with a Fisher ratio greater than some threshold, any source of variation that causes a signal change that happens to correlate with the sample types may be misidentified as a chemical source of variation.

Table 4.1: Cofactor standards.

KEGG number	Metabolite	Molecular Weight	GC amenable?
c00002	ATP	551.14	No
c00008	ADP	427.2	No
c00020	AMP	365.24	No
c00575	cAMP	369.2	No
c00003	NAD	663.43	No
c00006	NADP	787.37	No
c00004	NADH	709.4	No
c00010	CoA	809.57	No
c00068	thiamine diphosphate	460.77	Yes
c00248	lipoamide	205.34	No
c00016	FAD	829.51	No

Table 4.2: Glycolysis, Tricarboxylic Acid (TCA) cycle, and glyoxylate standards.
[a] not located using Fisher analysis
[#]metabolites not amenable to GC analysis
*decomposes in the injection inlet to produce a small amount of succinate.

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c00118	d-glyceraldehyde 3 phosphate	170	---[a]
c00022	pyruvate	110.04	0.4
c00024	acetyl-CoA[#]	809.57	---[a]
c01172	d-glucose 6 phosphate	282.12	0.3
c05378	d-fructose 1,6 bisphosphate	406.06	R only
c00074	phosphoenolpyruvate	208.04	---[a]
c00197	d-glycerate 3 phosphate	230.02	1.8
c00111	dihydroxyacetone phosphate	181.92	---[a]
c00665	d-fructose 2,6 bisphosphate	362.1	R only
c00149	l-malate	134.09	7.9
c00036	oxaloacetate	132.07	---[a]
c00026	2-oxoglutarate	168.08	---[a]
c00091	succinyl-CoA[#,*]	867.61	---[a]
c00158	citrate	192.12	13
c00122	fumarate	160.04	8.5
c00417	cis-aconitate	174.11	---[a]
c00042	succinate	162.05	3.0
c00048	glyoxylate	114.03	---[a]

Table 4.3: Fermentable and non-fermentable carbon source standards.
Three out of 15 metabolites were not GC amenable, mainly due to the low boiling point of the metabolites. These metabolites with low boiling points elute within the reagent region.
[a] not located using Fisher analysis
[#]metabolites not amenable to GC analysis
*does not elute in the time frame of GC method developed for yeast extracts.

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c00093	sn-glycerol 3 phosphate	324.13	2.8
c00163	propanoate[#]	96.06	---[a]
c00186	lactate	90.08	1.7
c00116	glycerol	92.05	0.1
c00033	acetate[#]	136.08	'---[a]
c00794	d-sorbitol	182.17	'---[a]
c00392	d-mannitol	182.17	0.4
c05402	melibiose	342.3	---[a]
c00492	raffinose[#,*]	594.51	---[a]
c00208	maltose	360.31	---[a]
c00089	sucrose	342.31	---[a]
c00124	galactose	180.16	---[a]
c00095	fructose	180.16	R only
c00031	glucose	180.16	R only
c00243	lactose	342.31	---[a]

Table 4.4: Oxidative and non-oxidative pentose phosphate standards.
[a] not located using Fisher analysis
[#]decomposes in the injection inlet to produce a d-glucose 1 phosphate fragment.
%detected as ribofuranose

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c00532	l-arabitol	152.15	---[a]
c00259	l-arabinose	150.13	---[a]
c00379	xylitol	152.15	R only
c00181	d-xylose	150.13	---[a]
c00103	d-glucose 1 phosphate	304.1	0.6
c00029	UDP glucose[#]	566.3	---[a]
c00199	d-ribulose 5 phospate	274.07	---[a]
c00345	6-phospho-d-gluconate	342.08	R only
c00279	d-erythrose 4 phosphate	222.07	---[a]
c00257	d-gluconate	218.14	---[a]
c01801	deoxyribose	134.13	---[a]
c00121	d-ribose	150.13	R only[%]

Table 4.5: Amino acid standards.
[a] not located using Fisher analysis

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c00062	l-arginine	210.66	1.1
c00049	l-aspartic acid	133.1	---[a]
c00025	l-glutamic acid	169.11	3.1
c00135	l-histidine	209.63	---[a]
c00407	l-isoleucine	131.17	1.7
c00123	l-leucine	131.17	5
c00047	l-lysine	182.65	3.2
c00073	l-methionine	149.21	1.6
c00079	l-phenylalanine	165.19	1
c00065	l-serine	105.09	1
c00188	l-threonine	119.12	3.3
c00078	l-tryptophan	204.23	---[a]
c00082	l-tyrosine	181.19	3.9
c00183	l-valine	117.15	1.0
c00041	l-alanine	89.05	0.9
c00491	l-cystine	240.3	---[a]
c00037	glycine	75.1	---[a]
c00148	l-proline	115.1	1.3
c00152	l-asparagine	150.1	1.9
c00064	l-glutamine	146.2	2.3

Table 4.6: Vitamin standards.
[a] not located using Fisher analysis
[#] not amenable to GC analysis

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c00106	uracil	112.1	---[a]
c00147	adenine	135.13	DR only
c00137	myo-inositol	180.16	DR only
c00568	p-aminobenzoic acid	159.12	---[a]
c00378	thiamin	346.27	---[a]
c00120	biotin	244.31	---[a]
c00314	pyridoxine	205.64	---[a]
c00864	d-pantothenic acid	238.27	---[a]
c00504	folic acid[#]	441.4	---[a]
c00178	thymine	126.11	---[a]

Table 4.7: Lipids/Fatty acid β-oxidation standards.

[a] not located using Fisher analysis

[#] not amenable to GC analysis

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c00083	malonyl-CoA[#]	853.58	---[a]
c00383	malonate	104.06	---[a]
c00249	palmitic acid	256.42	---[a]
c01530	stearic acid	284.48	1.1
c02679	lauric acid	200.18	---[a]
c06424	myristic acid	228.21	---[a]
c00114	choline[#]	139.62	---[a]
c00307	CDPcholine[#]	564.34	---[a]
c00356	3-hyroxy-3-methylglutaryl-CoA[#]	1009.67	---[a]
c00332	acetoacetyl-CoA[#]	971.6	---[a]
c00164	acetoacetate	108.02	---[a]
c00877	crotonoyl-CoA[#]	835.61	---[a]
c01771	crotonic acid[#]	86.09	---[a]

Table 4.8: Miscellaneous metabolite standards.

[a] not located using Fisher analysis

KEGG number	Metabolite	Molecular weight	N_{DR}/N_R
c03878	N-acetyl-d-glucosamine	221.09	---[a]
c01083	trehalose	342.12	66.8
---	methionine sulfone	181.21	---[a]
---	methionine sulfoxide	165.21	---[a]

Table 4.9: Statistically different metabolites found in both R and DR samples. 2D Sum of Fisher Ratios for the methoximated and trimethylsilylated metabolites in both sample types (DR and R) that were found to be statistically different between the samples to the 95% confidence interval. The mass spectral similarity value (MV) is a comparison of the fragments in the library spectrum to the sample spectrum of the full deconvoluted mass spectrum. The ChromaTOF signal-to-noise value (S/N) is the ratio from the more concentrated class, *e.g.* citrate S/N from DR sample, at the unique *m/z*. The lower Fisher Ratios often represent less concentrated, but important metabolites. #Adenosine, 5'-S-methyl-5'-thio-N-(trimethylsilyl)-2',3'-bis-O-(trimethylsilyl). + retention times from in-house library match the sample elution time. * the fragment 143 in the sample spectrum is missing, but other *m/z* look similar. ¯ a small portion of succinate could arise from the decomposition of succinyl CoA.

Metabolite	t_{R1}, min	t_{R2}, s	MV	S/N	Fisher sum	t_{calc}/t_{table}	N_{DR}/N_R
trehalose	29.7	0.68	919	>10,000	1.4350	37.97	66.8
glucose	20.4	0.42	936+	>10,000	0.8998	1.17	0.02
glycerol	11.0	0.43	945+	>10,000	0.3036	3.23	0.1
lysine	20.5	0.59	926+	>10,000	0.1495	1.39	3.2
citrate	19.3	0.65	846+	3650	0.1029	1.41	13.0
threonine	13.0	0.58	952+	9722	0.0878	2.08	3.3
malate	14.5	0.73	942+	>10,000	0.0719	3.41	7.9
5-oxoproline	15.1	1.35	949+	>10,000	0.0639	1.82	2.4
glutamic acid	16.5	0.75	936+	9103	0.0491	2.55	3.1
tyrosine	20.8	0.64	950+	4067	0.0252	4.71	3.9
homoserine	13.9	0.58	853	5826	0.0214	2.53	5.5
glutamine	18.6	0.99	935+	>10,000	0.0188	1.72	1.9
leucine	11.0	0.62	893+	9995	0.0174	2.44	5.0
succinate	11.6	0.88	939+	8402	0.0148	1.82	3.0
Adenosine, 5'-S#	30.1	1.08	892	1849	0.0119	3.42	5.3
mannitol	20.9	0.43	893+	2600	0.0082	3.32	0.4
fumarate	12.2	0.98	925+	1361	0.0058	3.93	8.5
cystathionine	23.9	0.68	895	1901	0.0054	1.20	0.3
glycolic acid	7.5	0.69	935	4834	0.0048	6.12	2.6
glycerol 3 phosphate	18.6	0.89	917+	613	0.0024	2.09	2.8
glucose 1 phosphate	18.6	0.46	898+	1731	0.0020	1.79	0.6
ethyl succinate*	10.5	1.02	872	1152	0.0018	1.98	1.6
asparginine	17.2	0.94	927+	1109	0.0016	2.40	1.9
octanoic acid	10.7	0.64	906+	314	0.0011	2.67	0.6
threonic acid	15.5	0.5	940	212	0.0006	1.84	0.3
glutamine 3TMS	14.4	0.08	885+	1103	0.0005	1.77	2.8
glucose 6 phosphate	25.5	0.65	924+	110	0.0004	1.21	0.3
pseudo uridine	25.3	0.51	819	556	0.0004	3.91	1.7

Table 4.10: Metabolites with significant concentration in only one sample type. The mass spectral similarity value (MV) is a comparison of the fragments in the library spectrum to the sample spectrum. The S/N values were obtained from the ChromaTOF software. [+]identified using the in-house library with the observed retention times matching the elution times of the standards.

Metabolite	t_{R1}, min	t_{R2}, s	MV	S/N	Fisher sum	N_{DR}/N_R
glucose	20.7	0.45	929[+]	9418	0.6443	R only
glucopyranose	21.4	0.4	914	5284	0.0825	R only
glucopyranose	20.3	0.42	895	7869	0.0546	R only
butanoic acid 2,4 TMS TMS ester	13.3	0.57	892	1090	0.0146	DR only
methyl citrate	20.5	0.68	839	552	0.0104	R only
glucose 1 phosphate	19.5	0.41	755[+]	101	0.0098	R only
myo-inositol	22.7	0.51	910[+]	595	0.0056	DR only
adenine	19.9	0.71	917[+]	595	0.0037	DR only
galactofuranose	19.7	0.38	872	473	0.0036	R only
fructose	20.1	0.39	941[+]	1055	0.0034	R only
gluconic acid	17.1	0.76	911	1832	0.0031	R only
butanoic acid 2 amino	9.2	0.62	895	1553	0.0027	DR only
glucose	20.1	0.48	923[+]	352	0.0020	R only
orotic acid	18.3	0.61	916	531	0.0019	R only
2 desoxy-pentose	18.5	0.43	822	657	0.0017	R only
2-o-glycerol-α-d-galactopyranoside	24.5	0.38	932	190	0.0009	R only
fructose 1,6 bis phosphate	29.4	1.26	881[+]	355	0.0008	R only
fructose 2,6 bis phosphate	23.9	0.77	825[+]	42	0.0007	R only
ribofuranose	16.5	0.39	835	153	0.0006	R only
xylonic acid	16.8	1.03	914	252	0.0006	R only
arabinonic acid lactone	17.3	1.29	850	168	0.0005	R only
xylitol	17.8	0.54	822[+]	146	0.0005	R only
fructose 1,6 bis phosphate	29.4	1.19	903[+]	143	0.0004	R only
6-phospho-d-gluconate	26.4	0.67	912[+]	163	0.0004	R only
2-amino-adipinic acid	17.8	0.74	930	114	0.0003	R only
n-acetyl glutamic acid	18.7	1.45	833	96	0.0003	R only

Table 4.11: False positives.
Metabolites that are not statistically different, but metabolites in the sample, nonetheless. The mass spectral similarity value (MV) is a comparison of the fragments in the library spectrum to the sample spectrum. The ChromaTOF signal-to-noise value (S/N) is the ratio from the more concentrated class, *e.g.* ornithine S/N from DR sample. *UDP glucose shares this peak location with glucose 1 phosphate. +retention times from in-house library match the sample elution time. # MV for the isotopically labeled compound.

Metabolite	t_{R1}, min	t_{R2}, s	MV	S/N	Fisher sum	t_{calc}/t_{table}	N_{DR}/N_R
ornithine	19.3	0.58	954	4185	0.1082	0.90	4.2
isoleucine	11.4	0.61	928+	>10,000	0.0136	0.88	1.7
lactate	7.3	0.63	956+	6272	0.0130	0.93	1.7
serine	12.5	0.6	907+#	5466	0.0063	0.02	1.0
alanine	8.0	0.63	885+	>10,000	0.0043	0.28	0.9
stearic acid	24.0	0.53	927+	9256	0.0032	0.22	1.1
glucose 6 phosphate	25.3	0.61	948+	462	0.0026	0.88	0.4
3-OH Propionic	8.7	0.67	906	1643	0.0023	0.71	2.0
methionine	15.0	0.74	843+	4755	0.0018	0.47	1.6
glucose 1 phosphate*	19.5	0.48	889+	638	0.0013	0.36	0.6
glycerate 3 phosphate	19.2	1.02	936+	829	0.0011	0.79	1.8
proline	11.4	0.63	920+	5218	0.0008	0.46	1.3
o-toluic acid	12.5	0.79	885	9262	0.0007	0.11	1.0
pyruvate	7.1	0.58	864+	234	0.0005	0.74	0.4
asparagine	14.7	1.08	912+	524	0.0005	0.98	1.5
CoA phosphoric acid fragment	11.1	1.21	928+	8280	0.0004	0.09	1.0
arginine	14.1	0.66	941+	3324	0.0004	0.17	1.1
valine	10.0	0.62	933+	5079	0.0004	0.05	1.0
phenylalanine	16.6	0.67	927+	3576	0.0004	0.05	1.0

Table 4.12: Statistically different unknown metabolites from R and DR samples. 2D Sum of Fisher Ratios for the derivatized metabolite unknowns discovered in both sample types. Unknowns were quantified using selected mass channels from the full PARAFAC mass spectra of peaks at the correct retention times. The S/N obtained from ChromaTOF is listed to give an idea of the intensities of each peak. Metabolites that are statically different to the 95% confidence interval have a *t*-ratio greater than 1. *could not calculate the d-o-f due to too S/N in one of the three biological class samples; the average of the other samples were used in the DR/R calculation.

Metabolite	t_{R1}, min	t_{R2}, s	S/N	Fisher sum	t_{calc}/t_{table}	N_{DR}/N_R
unk1173	19.6	0.52	5092	0.0190	2.9	3.2
unk922	15.4	0.56	858	0.0051	1.3	0.1
unk1179	19.7	0.71	3847	0.0049	0.4	6.2
unk1069	17.8	0.49	2852	0.0043	2.7	2.3
unk1215	20.3	0.51	5445	0.0024	0.9	0.4
unk907	15.1	0.79	282	0.0013	1.2	1.7
unk1215 63	20.3	0.63	57.4	0.0012	*	0.2
unk1200	20.0	0.59	1196	0.0012	1.2	0.2
unk592	9.9	0.59	157	0.0011	1.4	0.4
unk1108	18.5	0.71	238	0.0010	1.2	4.6
unk574	9.6	0.59	482	0.0009	*	0.5
unk1101	18.4	0.72	982	0.0008	0.4	7.5
unk1035	17.3	0.44	1365	0.0008	1.0	0.4
unk1657	27.6	0.69	851	0.0008	0.4	0.5
unk843	14.1	0.54	281	0.0008	0.2	1.0
unk768	12.8	0.56	502	0.0007	1.4	0.6
unk1491	24.9	0.43	214	0.0007	1.0	0.2
unk1551	25.9	0.71	819	0.0006	0.5	2.6
unk1359	22.7	0.18	1384	0.0006	0.0	1.0
unk1633	27.2	0.69	934	0.0006	3.5	7.1
unk472	7.9	0.49	293	0.0006	1.3	0.6
unk484	8.1	0.47	1200	0.0005	0.9	0.9
unk1272	21.2	1	2150	0.0005	0.9	0.7
unk579	9.7	0.53	1065	0.0005	0.7	0.8
unk1191	19.9	0.45	>10,000	0.0005	1.6	1.3
unk1147	19.1	0.53	332	0.0005	0.5	1.5
unk1029	17.2	1.26	1170	0.0005	1.5	1.5
unk1081	18.0	0.18	5323	0.0005	0.3	1.2
unk1632	27.2	0.53	62	0.0004	2.0	0.4
unk945	15.8	0.55	172	0.0004	6.5	3.0
unk556	9.3	1.05	1252	0.0003	1.4	1.3
unk1680	28.0	0.43	206	0.0003	0.8	0.3

Table 4.13: Unknown metabolites discovered in only one sample type.

Metabolite	t_{R1}, min	t_{R2}, s	S/N	Fisher sum	N_{DR}/N_R
unk955	15.9	0.59	>10,000	0.0181	R only
unk1231	20.5	0.42	6946	0.0096	R only
unk1266	21.1	0.45	806	0.0079	R only
unk1363	22.7	0.51	892	0.0056	R only
unk1501	25.0	0.51	1610	0.0048	R only
unk1182	19.7	0.5	231	0.0029	R only
unk1515	25.3	0.48	1527	0.0026	R only
unk1549	25.8	0.43	331	0.0022	R only
unk1212	20.2	0.37	857	0.0021	R only
unk1626	27.1	0.75	204	0.0021	R only
unk1588	26.5	0.44	683	0.0020	R only
unk1575	26.3	0.45	358	0.0017	R only
unk1474	24.6	0.45	452	0.0017	R only
unk1500	25.0	0.43	491	0.0013	R only
unk1648	27.5	0.53	1243	0.0013	R only
unk1195	19.9	0.41	101	0.0012	R only
unk1521	25.4	0.42	373	0.0011	R only
unk1291	21.5	0.4	96	0.0010	R only
unk1153	19.2	0.4	143	0.0008	R only
unk1651	27.5	0.54	42	0.0007	R only
unk690	11.5	0.49	225	0.0007	R only
unk1495	24.9	0.39	157	0.0007	R only
unk1744	29.1	0.61	72	0.0006	R only
unk1594	26.6	0.55	423	0.0006	R only
unk1449	24.2	0.45	47	0.0006	R only
unk1486	24.8	0.53	424	0.0006	R only
unk1612	26.9	0.73	179	0.0006	R only
unk1705	28.4	0.58	80	0.0006	R only
unk840	14.0	0.75	1557	0.0006	R only
unk1720	28.7	0.66	146	0.0006	R only
unk1614	26.9	0.89	259	0.0005	R only
unk1711	28.5	0.67	75	0.0005	R only
unk1465	24.4	0.43	96	0.0005	R only
unk1410	23.5	0.45	735	0.0005	R only
unk1473	24.6	0.86	133	0.0005	R only
unk1414	23.6	0.46	224	0.0005	R only
unk762	12.7	0.82	346	0.0005	DR only
unk1621	27.0	0.52	163	0.0005	R only
unk1723	28.7	0.64	179	0.0005	R only
unk1560	26.0	0.35	133	0.0004	R only
unk963	16.1	0.51	101	0.0004	DR only
unk1246	20.8	0.46	80	0.0004	R only
unk1603	26.7	0.5	328	0.0003	R only
unk1828	30.5	0.67	55	0.0003	R only
unk1480	24.7	0.49	36	0.0003	R only
unk1767	29.5	0.65	360	0.0003	DR only
unk1536	25.6	0.44	276	0.0003	R only
unk949	15.8	0.65	368	0.0003	R only
unk1464	24.4	0.51	74	0.0003	R only

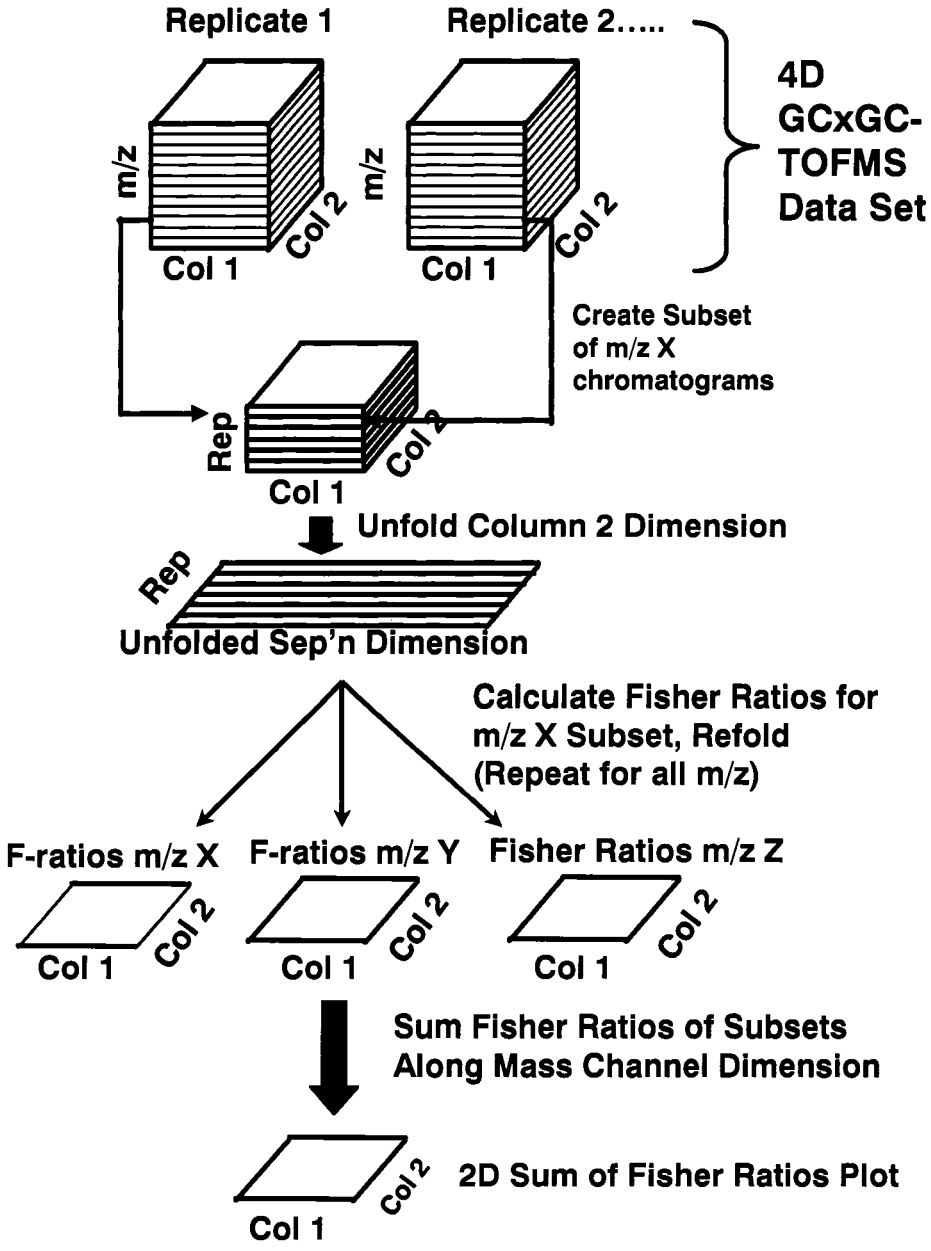


Figure 4.1: Schematic of the Fisher ratio algorithm. This algorithm calculates a Fisher ratio at every point in the 2D chromatographic space for each mass channel. The ratios from each mass channel are weighted by the mean signal at each point, and summed to produce peaks at locations that distinguish the sample types (the 2D Sum of Fisher Ratios plot), and then these peaks are sorted in descending order for subsequent PARAFAC analysis.

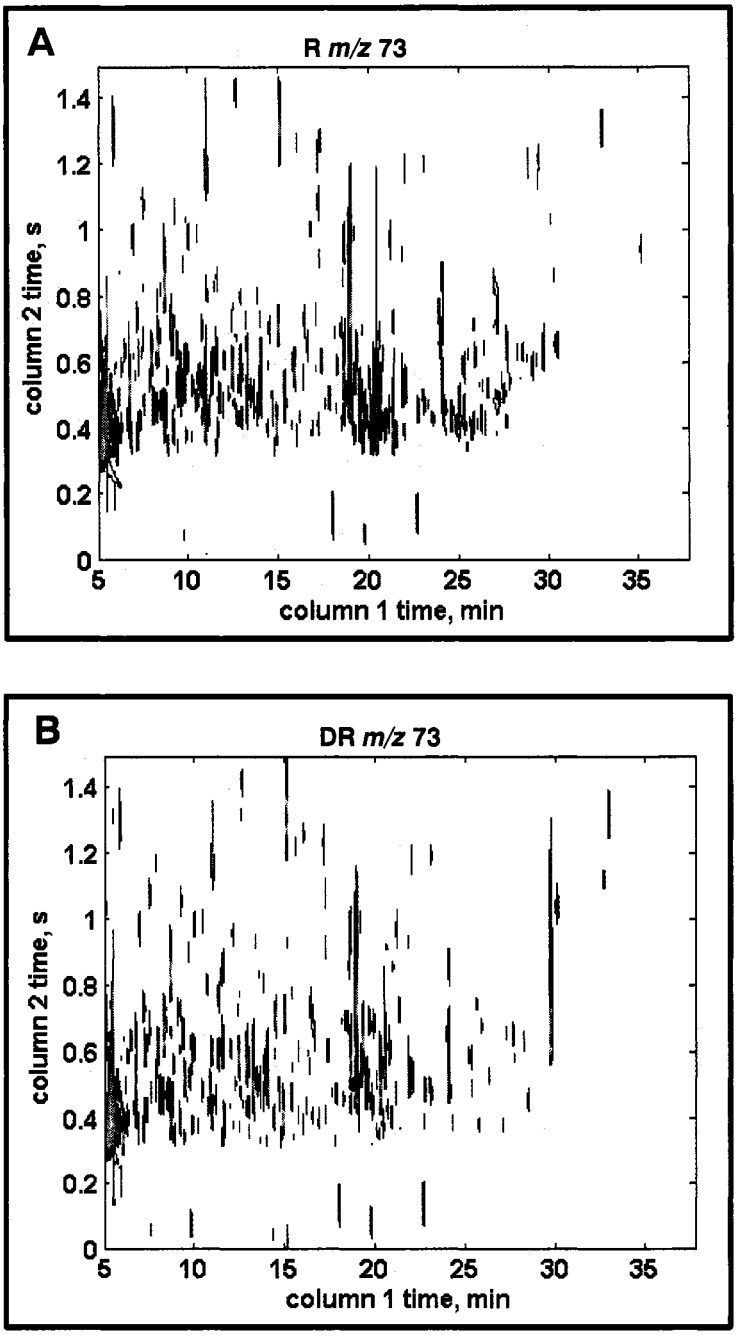


Figure 4.2: Typical GCxGC-TOFMS contour plots at a *m/z* 73. (A) fermenting yeast cells, R, and (B) respiring yeast cells, DR. Both sample types share many similar components.

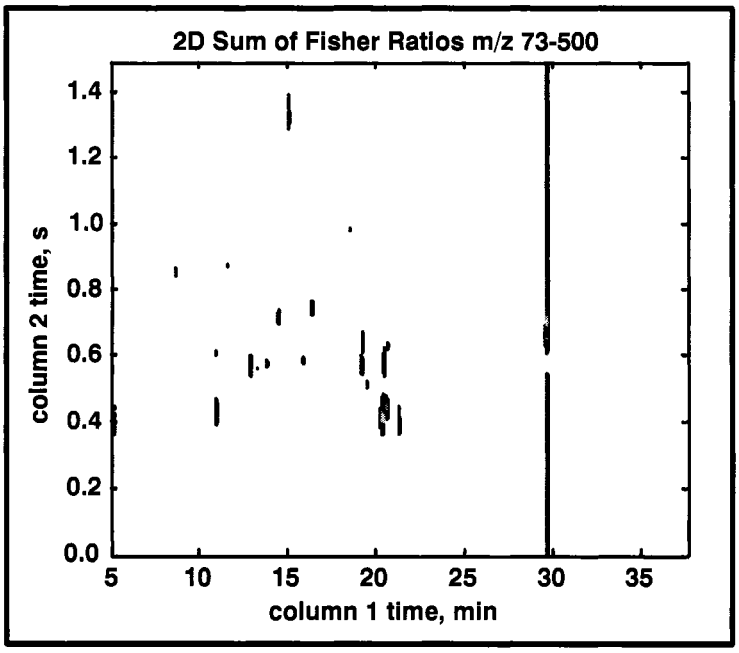


Figure 4.3: 2D Sum of Fisher Ratios plot for m/z 73-500.
Contour showing 2D Sum of Fisher Ratios above a threshold of 0.003. The peaks in the plot pinpoint the locations of chromatographic peaks that differentiate the classes in Figure 4.2(A) and (B). The streaking at approximately 30 minutes is due to trehalose, which is very concentrated in the DR samples and is very low in concentration in the R samples.

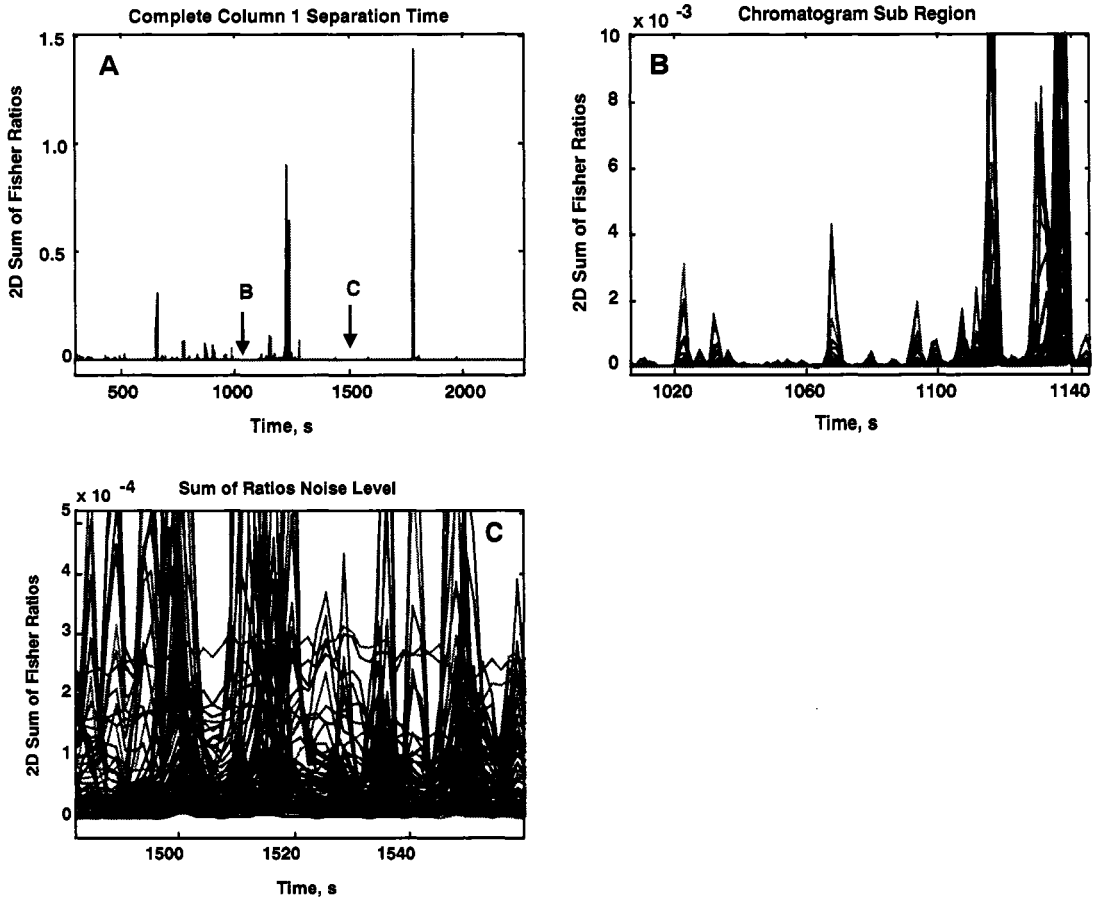


Figure 4.4: The one-dimensional projection of the 2D Sum of Fisher Ratios plot. (A) A few metabolites are found that significantly differentiate the classes, but in (B) many more peaks are at lower 2D Sum of Fisher Ratio values. (C) The noise level for the 2D Sum of Fisher Ratios is approximately $3x10^{-4}$.

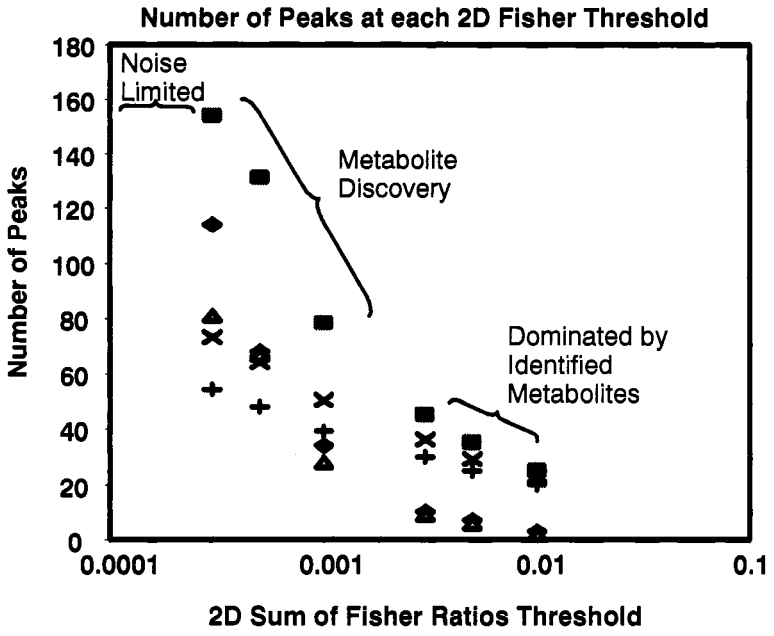


Figure 4.5: Peaks located as a function of the 2D Sum of Fisher Ratios threshold. A comparison of means by the Student's *t*-test was used to determine if the concentration means from the two samples types, DR and R, were significantly different within a 95% confidence interval following PARAFAC deconvolution. As the 2D Sum of Fisher Ratios threshold decreases, there is a substantial increase in the known metabolites (x) and known metabolites that also statistically differentiate the classes (+). The 2D Sum of Fisher Ratios noise level (LOD) as seen in Figure 4.4(C) is slightly lower than $3x10^{-4}$, and as this noise level is approached the number of unknown metabolites (triangles) and reagent artifacts (diamonds) outnumber the identified metabolites (x). The total number of unknown and known metabolites is shown by the squares.

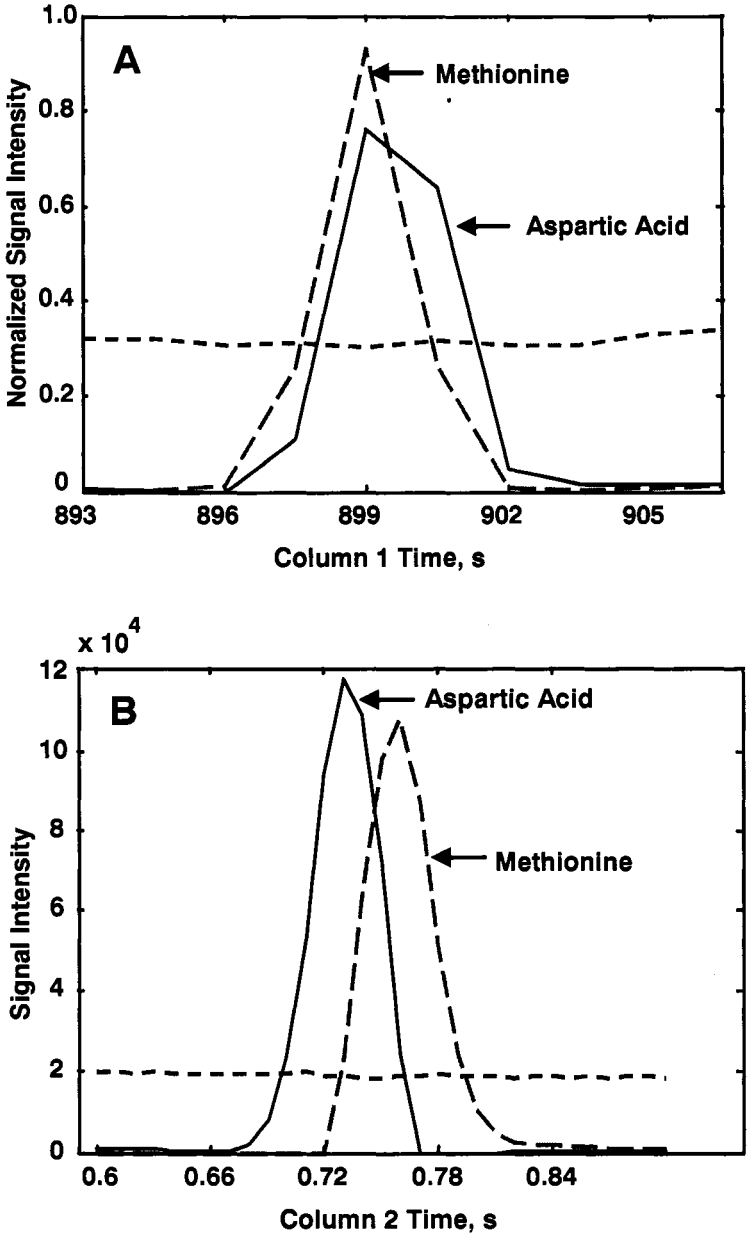


Figure 4.6: Metabolite identity confirmation using chromatographic profiles. The PARAFAC resolved profiles can be compared to the retention times of individually run standards. Pure component chromatographic profiles in (A) the column 1 separation dimension, and in (B) the column 2 separation dimension can be extracted using a full mass spectral PARAFAC deconvolution on an important sub-region of the entire 2D chromatographic space. PARAFAC is able to distinguish the two metabolites even though at *m/z* 73 the resolution is approximately 0.3.

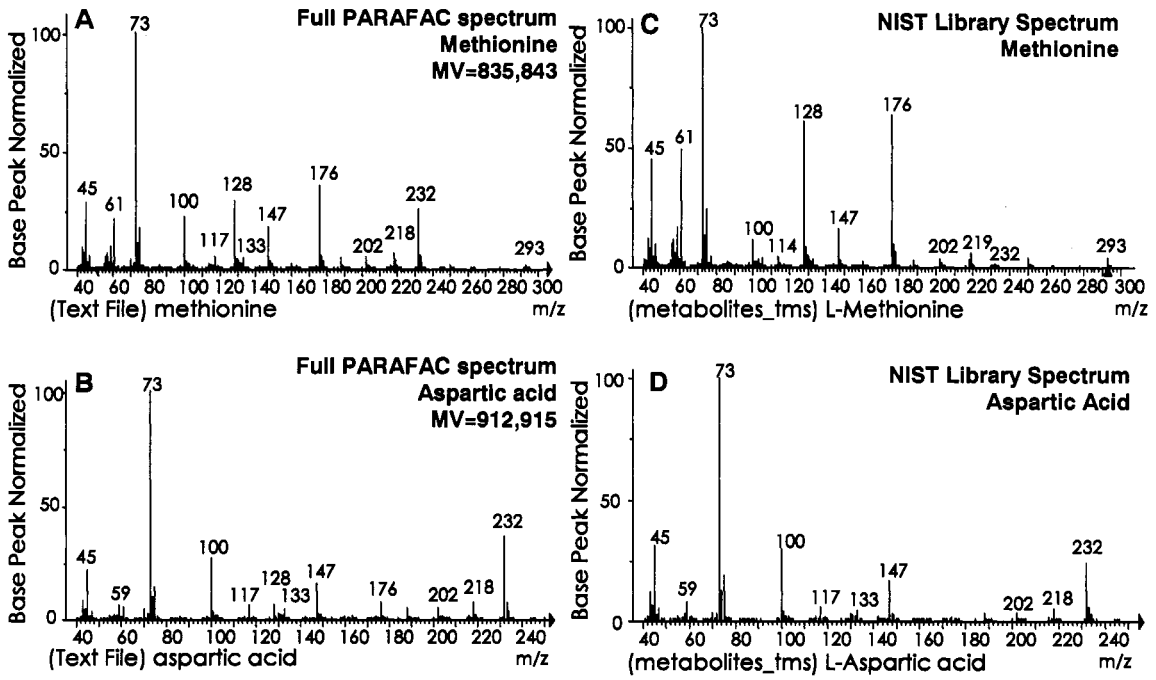


Figure 4.7: Metabolite identity confirmation using mass spectral profiles.
The full PARAFAC deconvoluted mass spectra for the two significantly overlapped metabolites from Figure 4.6: (A) methionine, and (B) aspartic acid. The NIST library mass spectra are shown for comparison where (C) is methionine, and (D) aspartic acid. The similarity of the two mass spectra gives impetus for the use of selective mass channels when quantifying overlapping components. The first MV gives the similarity of the sample spectrum to the library spectrum and the second MV gives the similarity between only the fragments present in the library spectrum with those in the sample spectrum.

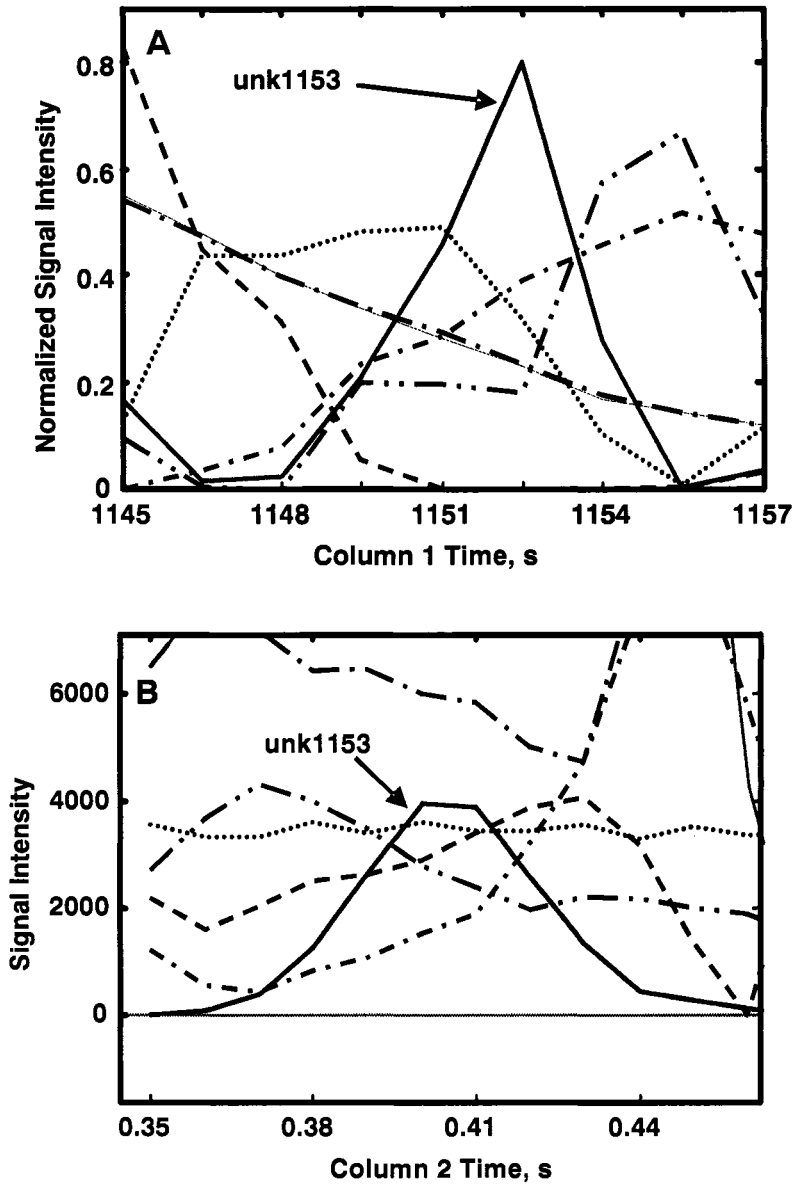


Figure 4.8: Chromatographic profiles for unknown metabolites.
An example of a metabolite found (unk1153) that was not identified, but could be added to a database. The chromatographic profiles where (A) is the column 1 profile and (B) is the column 2 profiles for each unknown can be determined and compiled into a database for later interpretation and potential identification.

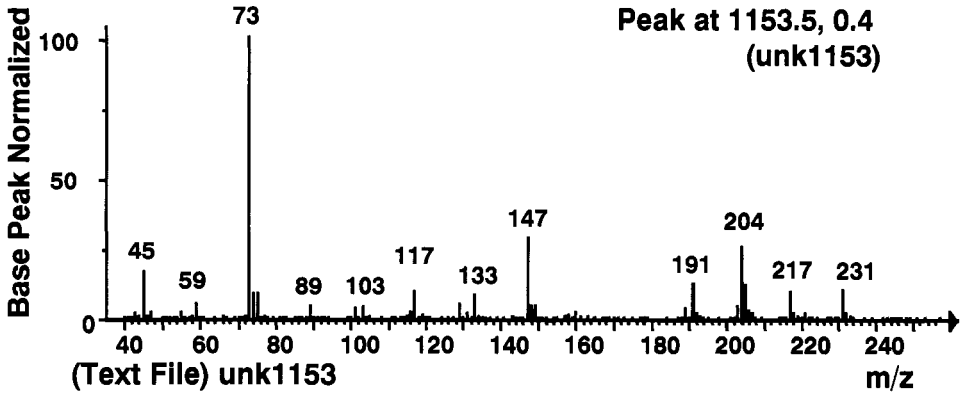


Figure 4.9: Mass spectral profiles for the unknown metabolite in Figure 4.8. The deconvoluted mass spectrum can be used as a library spectrum to quantify the unknown peaks.

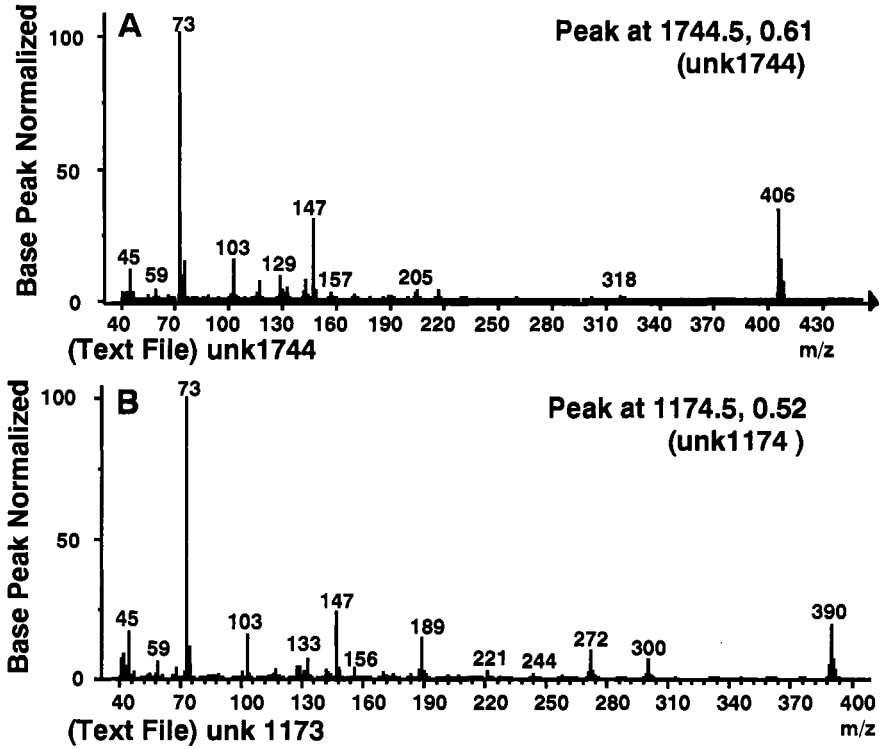


Figure 4.10: Sample mass spectra of unknown metabolites
Sample mass spectra obtained using the discovery-based software method of unknowns having strong mass fragments at higher mass channels for (A) unk1744, and (B) un1174. Many of the unknowns show this phenomenon.

Notes to Chapter 4

(1) Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227-231.

(2) Kinghorn, R. M.; Marriott, P. J. *J. High Resol. Chromatogr.* **1998**, *21*, 620-622.

(3) Beens, J.; Adahchour, M.; Vreuls, R. J. J.; Altena, K. v.; Brinkman, U. A. Th. *J. Chromatogr. A* **2001**, *919*, 127-132.

(4) Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **1998**, *70*, 2796-2804.

(5) Seeley, J. V.; Kramp, F.; Hicks, C. J. *Anal. Chem.* **2000**, *72*, 4346-4352.

(6) Mohler, R. E.; Prazen, B. J.; Synovec, R. E. *Anal. Chim. Acta* **2006**, *555*, 68-74.

(7) Giddings, J. C. *Unified Separation Science*; Wiley-Interscience: New York, 1991.

(8) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 2700-2709.

(9) Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O. *Metabolomics* **2005**, *1*, 65-73.

(10) Shellie, R. A.; Welthagen, W.; Zrostlikova, J.; Spranger, J.; Ristow, M.; Fiehn, O.; Zimmermann, R. *J. Chromatogr. A* **2005**, *1086*, 83-90.

(11) Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Lidstrom, M. E.; Synovec, R. E. *Talanta* **2005**, *65*, 380-388.

(12) Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagel, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 5068-5075.

(13) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Jack, R. M.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1058*, 209-215.

(14) Jonsson, P.; Johansson, A. I.; Gullberg, J.; Trygg, J.; A, J.; Grung, B.; Marklund, S.; Sjostrom, M.; Antti, H.; Moritz, T. *Anal. Chem.* **2005**, *77*, 5635-5642.

(15) Hollywood, K.; Brison, D. R.; Goodacre, R. *Proteomics* **2006**, *6*, 4716-4723.

(16) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnology* **2004**, *22*, 245-252.

(17) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends Biotechnology* **1998**, *16*, 373-378.

(18) Wang, Q.-z.; Wu, C.-y.; Chen, T.; Chen, X.; Zhao, X.-m. *Appl. Microbiol. Biotechnol.* **2006**, *70*, 151-161.

(19) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1056*, 145-154.

(20) Kell, D. B.; Oliver, S. G. *BioEssays* **2004**, *26*, 99-105.

(21) Tikunov, Y.; Lommen, A.; Vos, C. H. R. d.; Verhoeven, H. A.; Bino, R. J.; Hall, R. D.; Bovy, A. G. *Plant Physiology* **2005**, *139*, 1125-1137.

(22) Sihna, A. E.; Prazen, B. J.; Synovec, R. E. *Anal. Bioanal. Chem.* **2004**, *378*, 1948.

(23) Johnson, K. J.; Synovec, R. E. *Chemom. Intell. Lab. Syst.* **2001**, *60*, 225-237.

(24) R. O. Duda; Hart, P. E. *Pattern Classifications and Scene Analysis*; Wiley: New York, 1973.

(25) Wise, B. M.; Gallangher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, R. S. In *PLS_Toolbox 3.5 for use with MATLAB*, 2005 pp 123-127.

(26) Hoggard, J. C.; Synovec, R. E. *Anal. Chem.* **2007**, *79*, 1611-1619.

(27) Sherman, F. *Methods Enzymology* **1991**, *194*, 3-21.

(28) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J.; Oliver, S. G. *Phytochemistry* **2003**, *62*, 929-937.

(29) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Anal. Chem.* **2000**, *72*, 3573-3580.

(30) Harris, D. C. *Quantitative Chemical Analysis*, 6th. edition ed., 2003.

(31) Jonsson, P.; Gullberg, J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjostrom, M.; Moritz, T. *Anal. Chem.* **2004**, *76*, 1738-1745.

(32) Zörb, C.; Langenkamper, G.; Betsche, T.; Niehaus, K.; Barsch, A. *J. Agric. Food Chem.* **2006**, *54*, 8301-8306.

(33) McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*; University Science Books, 1993.

(34) Aoki-Kinoshita, K. F. *J. Pestic. Sci.* **2006**, *31*, 296-299.

(35) Young, E. T.; Dombek, K. M.; Tachibana, C.; Ideker, T. *J. Biol. Chem.* **2003**, *278*, 26146-26158.

(36) Wu, J.; Zhang, N.; Hayes, A.; Panoutsopoulou, K.; Oliver, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 3148-3153.

Chapter 5. Identification and Evaluation of Cycling Yeast Metabolites in GCxGC-TOFMS Data**

5.1 Introduction

Comprehensive two-dimensional gas chromatography (GC x GC) is a very powerful technique based on the separation of gas phase analytes on two capillary columns, *i.e.* column 1 and 2, with complementary interactions.[1] A modulation device collects the column 1 effluent and then rapidly injects narrow pulses onto column 2.[2-5] The peak capacity of this two-dimensional (2D) separation system is multiplicative and results in a large increase in the number of compounds the system, over one-dimensional (1D) GC, has the ability to separate.[6] The coupling of the GCxGC instrument to a univariate detector such a flame ionization detector (FID) yields informative structured 2D chromatograms, with compounds in the same class along diagonal lines.[7-9] Coupling the GC x GC to a multivariate detector, such as a mass spectrometer, yields an increase in information over a GC x GC-FID instrument, thus making the GC x GC-MS more ideal to study complex samples. GC x GC-MS instrumentation provides a third dimension that allows the analyst to distinguish compounds. In addition to the class information output by the 2D GC x GC-FID chromatograms, the third order GC x GC-MS instrumentation can generate a full mass spectrum for each analyte peak thus providing insight for the identification of compounds.

** Large portions of this chapter have been submitted for publication:
Rachel E. Mohler, Benjamin P. Tu, Kenneth, M. Dombek, Jamin C. Hoggard, Elton T. Young, Robert E. Synovec. *Special Issue J. Chromatogr. A on Recent Trends and Developments in GC,* 2007

Recently, the advent of the time-of-flight mass spectrometer (TOFMS) led to the broadening use of the third order GC x GC-MS instrumentation[10,11] (column 1, column2, and full mass spectrum) and has been applied to the study of metabolomic samples.[12-18] These metabolomic samples are sufficiently complex for analysis on a third order instrument with an estimated 200,000 different metabolites in the plant kingdom[19,20] and around 600 metabolites in yeast.[21,22] A limiting factor in the analysis of metabolomic data is the high dimensionality and complexity of the data.[23-26] In order to make biological conclusions concerning the data, some form of data reduction, *i.e.*, chemometrics, must be applied to the data to rapidly locate and sort the metabolites in order of biological importance. Metabolites are found in a wide range of concentrations, with the most concentrated metabolites not always being the most important.[27] Following identification of important metabolites, the metabolites must be accurately quantified. In order to save on the overall analysis time, the above mentioned data analysis steps must be compiled into high-throughput discovery-based software that is also user friendly.[28,29] Indeed, there is an on-going need to develop improved chemometric tools that can readily be applied to third order GC x GC-TOFMS metabolomics data in a discovery-based manner.

The first step in data analysis is to locate peaks in the 2D separation space that are changing between sample types. One method to objectively locate peaks is the use of Principal Component Analysis (PCA).[12,30] PCA is an unsupervised method that does not require class specification; hence it does not take advantage of the experimental design knowledge, so is also subject to problems due to within class variation. Each sample is

compared with all other samples to determine where the greatest variation in signal intensity lies. Those regions of the chromatograms with the greatest variation between samples are captured on the first principal component (PC1), with each decreasing principal component describing less variance. Every sample is assigned a score value on each PC and each variable (chromatographic peak) is assigned a loading value. The samples with similar scores show less variation in signal intensities. A second chemometric data reduction tool recently reported for use with GC x GC-TOFMS metabolomics data is based on Fisher ratios.[14,31] The Fisher ratio method, unlike PCA, is a supervised method making it a more ideal method for analysis of hypothesis driven research, where the origin of samples is known. Since the Fisher ratio is the average class to class variation divided by the within class variation, multiple samples from each sample type must be collected and the class of each sample specified. Some benefits to using the Fisher ratio method over PCA are that the Fisher ratio method has been automated to analyze the entire three-dimensional (3D) data output by the third order instrumentation and the Fisher method considers each point in the 2D separation space independently. One major drawback to both of these methods is that as the number of sample classes increases, the ability to find metabolites that are distinguished by class becomes more difficult and can sometimes fail.

Following data reduction, locations in the 2D separation space that contain class distinguishing metabolites must be "mined" to positively identify and quantify the metabolites. Although the 2D separation space provides a large peak capacity, some peaks might overlap (co-elute) in one of the separation dimensions. To address this issue,

a Parallel Factors Analysis (PARAFAC) graphical user interface (GUI) was developed to mathematically resolve 3D data by comparison to a target analyte specified by the user.[32,33] Prior to target analysis using PARAFAC, tentative identification of the analyte must be specified. The outputs of this GUI are pure column 1, column 2 and mass spectral profiles, which can then be transformed into a 3D peak to obtain peak volume information. Automation of this GUI to analyze multiple analytes has greatly reduced the overall analysis time of the complete data sets.

Within the last few years, two groups have reported yeast cells exhibiting periodicity in genome-wide transcription, which coordinates strongly with the dissolved oxygen in the cell.[34,35] The McKnight group reported an ultradian cycle about four to five hours in length,[34] while Klevecz and coworkers reported cycles lasting around 40 minutes in length. [35] A illustration of the oscillating molecular oxygen as observed by the McKnight group is shown in Figure 5.1 with the locations of collected time intevals marked with dots. Twelve time intervals were collected over each 5 hour cycle with each interval separated by 25 minutes. We hypothesized that the metabolites would follow a periodic pattern similar to that observed in the transcriptome data with small incremental changes in concentration throughout the cycle, but with a wide variation in metabolite concentrations.

Reported in this chapter is a method for analyzing three-way GC x GC-TOFMS data of 24 sample aliquots (*i.e.* time intervals) from yeast cells grown under nutrient limited conditions. Due to the large number of time intervals collected and the similar concentrations of metabolites in adjacent time intervals, PCA and Fisher analysis were

not ideal data reduction methods for this study. Since it was hypothesized that the samples would show periodic behavior, a method was developed that measures the amplitude (depth of modulation) of the periodic pattern based on the strongest and weakest signal intensities for each metabolite and has been termed the Sratio method. Because it was the depth of modulation that was calculated, no frequency constraints were placed on the data, thus allowing metabolites with higher or lower frequencies than that of the dissolved oxygen shown in Figure 5.1 to be located. Also the depth of modulation is independent of signal intensity. Even though the depth of modulation was calculated for each metabolite over all *m/z*, thus utilizing all mass channel information, only the three most selective *m/z* for each metabolite were ultimately used to obtain the locations of those metabolites in the 2D separation space showing periodic patterns. Following this data reduction to find the locations of cycling metabolites and tentative identification using ChromaTOF, the PARAFAC GUI was applied to determine a more accurate depth of modulation based on the deconvoluted peak volumes (Vratio). Metabolites that showed similar patterns were then determined by submitting the Vratio to PCA to better understand the cycling patterns observed. The Sratio method is significantly different from the Fourier Transform based method Murray and co-workers reported earlier this year when analyzing samples using GC-MS instrumentation.[36] Their method was set to locate metabolites cycling at the same frequency as the dissolved oxygen and did not take the discovery-based approach we report.

5.2 Theory

The yeast metabolites used in this study were expected to express a periodic cycling of concentration levels during the cell growth, similar to the pattern seen in the dissolved oxygen signal, Figure 5.1. Regular samplings from cell growth at many time intervals were submitted to GCxGC-TOFMS analysis producing a three-way array for each time interval, where the dimensions were retention time on column 1, retention time on column 2 and *m/z*. Since 24 intervals were collected for this study, the use of PCA and the Fisher Ratio Method failed to locate the differentiating locations in the 2D separation space. The data processing method developed for this study uses the ratio of chromatographic signals at each mass channel (*m/z*) between samples to find the locations in the 3D data that contain cycling metabolites and is therefore named the Sratio method, which is described below.

The Sratio method, illustrated in Figure 5.2, determined the magnitude of changes in metabolite signal intensities during cycling using the most selective *m/z* in an automated and objective fashion. As a function of individual *m/z,* each sample is removed from the original three-way sample array and compiled into a new single *m/z* three-way array, where the three dimensions are retention time on column 1, retention time on column 2, and sample number. Each sample 2D matrix in the new three-way array for a given *m/z* is then unfolded for 1D alignment along column 2. The narrowness of peaks on column 2 (50-100 ms width at base (w_b)) resulted in the retention shifting on column 2 giving more detrimental results than shifting on column 1. A majority of the shifting on column 1 was approximately a half a modulation period and resulted from out

of phase modulation, whereas the shifting on column 2 ranged from 0-30 ms. It was predicted using a Gaussian model that for a peak shifted 20 ms with a w_b of 100 ms a 37% decrease in signal intensity would be observed and for the same shifting of a 70 ms wide peak a 48% decrease in signal intensity would arise. Peaks shifted by a half a modulation period were predicted to only have a 12% decrease in signal intensity. Thus if signal intensity ratios were calculated at a specific point in the 2D separation space, retention misalignment would cause inflated ratios. The aligned chromatograms were then refolded and augmented back into the new three-way array. The new three-way array was summed along the sample dimension and in this summed 2D matrix local peak maxima were found. The locations of the peak maxima are retained in each sample, and the remaining chromatographic locations zeroed. The replicates of each growth time interval are then averaged. Since data was collected over two dissolved oxygen cycles, the algorithm finds the samples with the two greatest chromatographic signals at each retention time, *i.e.* at each peak maxima, in the chromatographic space and calculates an average, Smax. The samples with the two lowest signals at each peak maxima are also found and the average calculated, Smin. An illustration of the samples averaged for Smax and Smin are depicted by the "x" in the metabolite trace of Figure 5.1. If the Smax or Smin is lower than the user specified noise threshold, the value at that point is replaced with the noise threshold value. The Sratio is then calculated, where the Sratio is defined as the ratio of the maximum and minimum raw chromatographic signals at each 2D peak maxima location:

$$S_{ratio,X} = \frac{S_{\max,X}}{S_{\min,X}}, \text{ at } m/z \text{ X} \qquad (1)$$

thus returning the average signal amplitude, *i.e.* depth of modulation, of the repeating pattern for each peak maxima in the 2D separation space for each *m/z*. Chromatographic locations with larger Sratios have a greater change in signal during cell growth and thus pinpoint the 2D chromatographic locations to study in greater detail. These steps are then repeated for all *m/z* to obtain a Sratio "mass" spectrum for each peak maxima in the chromatographic space. The largest Sratios are averaged to give an overall Sratio at each retention time point:

$$\overline{Sratio} = \frac{\sum_{i=1}^{n} (S_{ratio,i})}{n} \qquad (2)$$

This average of at least three Sratios disposes of noise, some reagent artifacts and peaks arising from peak maxima misaligned in one or two *m/z*. The average of three or more Sratios also minimizes the bias from overlapping interferents at the less selective *m/z*. An RSD (standard deviation of n largest Sratios/average of n largest Sratios) threshold can also be specified for the $\overline{Sratio}$ to remove artifacts with more than three Sratio values, but in which the Sratios vary wildly in intensity. Peaks arising from metabolites tend to have less variation in the Sratio intensities. Specific parameters used in this study will be described in the Data Analysis section.

5.3 Experimental

5.3.1 Cell growth, extraction and derivatization

Yeast manipulations and continuous culture experiments using the diploid yeast strain CEN.PK were performed as previously described.[34] The quenching of metabolic activity and extraction of intracellular metabolites was performed using a method adapted

from Castrillo et al.[37] Following growth to high density (OD~8-9), 1 ml of the continuous chemostat culture was rapidly quenched (4ml 10mM tricine, pH 7.4, 60% methanol) at -40°, for each time interval. Cells were spun at 1,000 g for three minutes at -10°, washed with 1 ml of the quenching buffer, and resuspended in 1 ml of extraction buffer (0.5 mM Tricine, pH 7.4, 75% ethanol) at 80° for three minutes followed by incubation at 4° for five minutes. Samples were spun at 20,000 g to remove cell debris and stored at -80° until analysis. For a pilot study 12 samples, each containing 1 x 10^7 cells were collected every 25 minutes for 5 hours resulting in 12 time intervals. For the comprehensive study, 24 samples were collected over a 10 hour period. Metabolites were derivatized following the protocol described previously using methoximation (20 mg/ml methoxyamine in pyridine) and trimethylsilylation (BSTFA:TMCS, 99:1).[12]

5.3.2 Instrumental parameters

Three-way GCxGC-TOFMS data was collected on a LECO Pegasus III system (LECO, St. Joseph, MI, USA). One μL split less injections of the collected metabolite extracts were made onto a 20 m x 250 μm i.d. x 0.5 μm d_f RTX-5MS column 1 (Restek, Bellefonte, PA, USA). Every 1.5 seconds column 1 effluent was transferred onto a 2 m x 180 μm i.d. x 0.2 μm d_f RTX-200MS column 2 (Restek, Bellefonte, PA, USA). Column 1 was held at 60°C for 0.25 minutes then ramped at 8°/min to 280°C were it was held for 10 minutes. Column 2 was kept 10° higher than column 1 and the modulator 40° higher than column 1. The inlet and transfer line were kept constant at 280°C and the ion source held at 200°C. The head of column 1 had a constant flow of 1 ml/min. An acquisition

rate of 100 spectra/s was collected after a five minute solvent delay. Each sample was injected in triplicate, resulting in the collection of 72 GCxGC-TOFMS chromatograms.

5.3.3 Data Analysis

In this study 1D peak match retention alignment along column 2 was performed.[38] All samples were aligned to sample 1 injection 1. The thresholds for implementing the Sratio method are as follows: the noise threshold was defined as five times the noise at approximately 1462.5 seconds on column 1 and 0.45 seconds on column 2 and the averaged Sratio threshold was set to three with a RSD threshold of 0.4. Peaks at the locations output by the Sratio method were initially identified using the LECO ChromaTOF-GC software v. 3.22 (LECO, St. Joseph, MI, USA). ChromaTOF was also used to calculate the signal to noise (S/N) values at *m/z* 73. An in-house library, the NIST main library and the library obtained from the Max Planck Institute of Molecular and Plant Physiology (http://www.mpimp-golm.mpg.de/mms-library/index-e.html) were searched against. A full PARAFAC deconvolution was used to confirm the identification of each metabolite peak.[32,39] A PARAFAC GUI was utilized in the quantification of metabolite peaks discovered by the Sratio method. This GUI has the potential to analyze multiple analytes in a sample at once. Outputs from this GUI are pure chromatographic and mass spectral profiles as well as peak volumes. Principal component analysis (PCA) was performed on the peak volume data to obtain objective cycling patterns in the data using PLS toolbox version 3.51 (Eigenvector Research, Manson, WA).

5.4 Results and Discussion

5.4.1 Pilot Study

To quickly determine if the metabolome followed a similar pattern to the dissolved oxygen and transcriptome, 15 metabolites that appeared to be changing between time intervals were located by eye and quantified. A majority of these 15 metabolites exhibited a five hour period with a greater than 2-fold increase in peak volume between the maximum and minimum time intervals. The volume profiles for six metabolites (2hydroxy-isovaleric acid, glucose-6-phosphate, myo-inositol, isoleucine, unknown 897, aspartic acid) are shown in Figures 5.3 A-F. In addition to the fact that the metabolome does appear to be cycling, it is interesting to note that the peak volume profiles between metabolites do not have maximum volumes at the same time intervals. The metabolites in Figures 5.3 A and C are approximately 180° out of phase with each other. Indeed, these results provided impetus for a more comprehensive study in which samples were collected over two dissolved oxygen cycles.

5.4.2 Retention shifting

Prior to calculating Sratios for the metabolites over the 24 time intervals, the average shifting for 10 metabolites on both column 1 and column 2 was calculated to validate the retention time alignment along column 2, Figure 5.4. These metabolites were chosen to spread throughout the entire chromatographic run, with the first metabolite eluting at 10.3 min and the tenth metabolite eluting at 27.8 min on column 1, Figure 5.4A. The numbers beside the squares in Figure 5.4A describe the number of time intervals at that shift value. For a majority of the metabolites, the shifting was less than

the modulation period of 1.5 seconds, which translate to peaks that are out of phase, *i.e.* the modulation device collects four column 1 slices and other times only 3 column 1 slices. As stated in the Theory section, this out of phase modulation should only result in a signal decrease of around 12% and this shifting of the chromatographic signal due to out of phase peaks can not be corrected with the currently available alignment algorithms. The shifting on the column 2 was much more severe ranging by ± 30 ms, Figure 5.4B. The numbers in Figure 5.4B correspond to the elution on column 1 in ascending order. This shifting as stated in the Theory section has detrimental impact on the column 2 signals since the w_b for column 2 peaks range from 50-100 ms. It is also interesting to note that unlike column 1 shifting where a majority of time intervals showed similar shifting very few time intervals had similar shifting on column 2.

5.4.3 Data reduction

All derivatized metabolites containing a trimethylsilyl (TMS) group will result in a peak at *m/z* of 73 due to the loss of the TMS group upon ionization, Figure 5.5A. In each of the 72 yeast extract chromatograms collected there are more than 500 peaks present at this *m/z,* some of which are overlapping, so a given peak might contain more than one metabolite. The 72 chromatograms describe 24 time intervals in two yeast metabolic cycles, Figure 5.1. The 24 samples were injected in random order and in triplicate to minimize injection order bias as well as injection variation. Data was not normalized prior to Sratio calculations. The percent relative standard deviation (%RSD) in the PARAFAC peak volumes for injection replicates for 16 selected metabolites did not decrease following normalization to the summed TIC (overloaded reagent peaks

removed prior to summation) or the average of the two isotopically labeled internal standards (alanine and serine), Table 5.1.

Although Sratios were calculated for all *m/z*, three selected *m/z* (73, 128 and 217) were chosen to illustrate the data reduction. Following Sratio analysis of *m/z* 73 the data was reduced from 72 complex *m/z* 73 chromatograms down to one 2D Sratio plot and the number of peaks was significantly lowered to approximately 85 at a Sratio threshold of 3, Figure 5.5B. Sratio analysis of the more selective mass channels 128 and 217 yielded even fewer peak locations (37 and 23, respectively), thus signifying less peak overlap and more accurate Sratios, Figure 5.6A, B. The boxed regions labeled 1-4 in Figure 5.5B and 5.6 A, B indicate the locations of four metabolites empirically determined to represent four types of cycling patterns. ChromaTOF identified the four metabolites as methyl citrate, myo-inositol, glucose 6 phosphate and cystathionine, respectively. These four patterns are supported by PCA results as will be described later. In Figure 5.5B all four metabolites are present, but in Figure 5.6A and B only one of the four metabolites is present.

At each point in the 2D separation space, a Sratio "mass" spectrum covering all *m/z* analyzed can be obtained to identify the most significant *m/z* for locating cycling metabolites, *i.e.* most selective *m/z* with good S/N. The Sratio spectra for the four metabolites in the boxed regions of Figure 5.5B are shown in Figure 5.7A-D. Methyl citric gave the largest Sratio, Figure 5.7A. This metabolite at *m/z* 73 (the *m/z* of the most intense fragment) has an average signal to noise (S/N) of 1761 in the two samples with the largest concentration of this particular metabolite. In spite of this relatively low S/N

(especially at *m/z* other than 73 and the samples with lower concentration), methyl citric resulted in three Sratios that have a RSD less than 0.4. The average Sratio calculated from the three largest Sratios (40, 32 and 23) was 32 and the standard deviation of these Sratios was 8. The RSD was calculated by dividing the standard deviation by the average, giving an RSD of 0.3. The RSD threshold was set to remove reagent peaks that had significant signal intensity in one or two *m/z* but a signal intensity near the limit of detection in most other *m/z*. The lower Sratios for methyl citric are due to the Smin noise threshold. Recall from the Theory section that if Smin is below the specified noise threshold, Smin is automatically set to the noise threshold, thus resulting in a larger value to divide into Smax. Myo-inositol is the second metabolite chosen for the exhaustive study, Figure 5.7B. At first glance it might seem disturbing that there is such a large number of *m/z* with similar Sratios, when mass spectra seem to have far fewer fragments. This can be explained by the fact that the Sratio algorithm is not biased by signal intensity, as long as the peak is above the specified noise threshold. Indeed, every *m/z* in the Sratio spectrum is in the library mass spectrum for myo-inositol. This metabolite in contrast to methyl citric has a much larger S/N and no overlapping interferents, which is evidenced by the many *m/z* that have similar Sratio values. The glucose-6-phosphate Sratio spectrum in Figure 5.7C shows an equally complex pattern to that of myo-inositol, with a majority of the mass fragments giving similar Sratio values. Cystathionine, like methyl citric, does not have quite as many *m/z* with Sratios, Figure 5.7D. At *m/z* 73 (the *m/z* of the most intense fragment) the average S/N for cystathionine in the samples with

the two largest signals is only 207, however there remain at least three selective *m/z* with similar Sratios that can be averaged.

5.4.4 PARAFAC quantification

Following location of cycling peaks using the Sratio method and tentative identification using ChromaTOF software, the PARAFAC GUI was used to positively identify and accurately quantify metabolites. The resulting volume profiles over all time intervals for the four metabolites studied in depth after PARAFAC quantification are given in Figure 5.8A-D. These four metabolites all show different patterns over the five hour dissolved oxygen cycle, and are reproduced in the second cycle. The sharp spiking discovered for methyl citric, Figure 5.8A, was also present in the one cycle data collected but is not shown for brevity. Myo-inositol and cystathionine showed a much broader cycling than methyl citric, Figure 5.8B and D respectively. The only difference between the myo-inositol and cystathionine frequency was the phase shift; the two metabolites are approximately 180° out of phase. Glucose-6-phosphate is unique in that there appear to be two cycles within one dissolved oxygen cycle, Figure 5.8C. Glucose-6-phosphate and methyl citrate might not have been located if the Sratio method put frequency constraints on the data. Each of these metabolite profiles were confirmed by the data collected over one cycle.

All identifiable as well as unknown (*i.e.* unidentified in mass spectral libraries) metabolites with an Sratio greater than three were studied. At this Sratio threshold, 44 identified unique metabolites and 41 unknown metabolites were located, as summarized in Tables 5.2 and 5.3. The metabolites in the tables are listed in order of descending

PARAFAC volume ratio (Vratio). The Vratio calculation is analogous to the Sratio calculated in equation 1, where the average of the PARAFAC volumes at the two time intervals with largest volumes is divided by the average of the PARAFAC volumes at the two time intervals with the smallest volumes. Although the cycling metabolites did give the highest Sratios, a Vratio can be calculated after quantification of each peak using PARAFAC to obtain more accurate cycling amplitude. For a majority of the identified metabolites the Vratio is less than the Sratio, with an average S/V ratio of 1.6. This phenomenon arises from two causes. The first is slight peak misalignment between samples due to out of phase modulation. Since the Sratio method only retains the peak maxima found after the summation of all samples at each 2D separation time, any specified point that is less than the peak maxima will result in artificially slightly high results. The second cause for inflated Sratios is interferents. Depending on the resolution of overlapping peaks the raw signal could be more affected at some m/z.

In addition to the Sratios and Vratios, mass spectral match values (MV) were also calculated to give an indication of the quality of the identification and are reported in Table 5.2. The MVs were obtained after a complete mass range PARAFAC deconvolution of the two most intense samples. These two values were averaged to give the final MV listed in the table. A perfect match is 1000. More than 50% of the metabolites have a MV above 900 and a vast majority also have retention confirmation. The retention confirmation was obtained from an in-house library of standards run under the same instrumental conditions as the samples. To obtain the S/N listed in Table 5.2, the values ChromaTOF software calculated from the two most concentrated time

intervals for each metabolite were averaged. The method developed is not biased towards peaks of greater S/N and three metabolites in the top ten Vratios have a S/N less than 500. Ribulose 5 phosphate had one of the lowest S/N values at 147 and trehalose one of the largest with an S/N value greater than 10,000.

5.4.5 PCA classification

Following location and quantification of differentiating metabolites, objective clustering using PCA was applied to determine which of the metabolites cycled in a similar manner, shown in Figure 5.9. Prior to data input into the PCA GUI, each time interval was divided by the largest peak volume for each metabolite to give a maximum signal of 1. This forced all metabolites to the same intensity range (0-1), *i.e.* level playing field, while preserving the depth of modulation for each metabolite. At first glance, the points on the scores plot appear to be a continuum, but closer inspection reveals that information about the metabolites can be gleaned from both PC1 and PC2. The PC values as well as the grouping for both unknown and known metabolites are reported in Tables 5.3 and 5.4, respectively. The first PC captures intensity variations from both the depth of modulation and cycling frequency over the 24 time intervals, Figure 5.10A, with only the knowns shown for clarity. The phase information is captured on PC2, Figure 5.10B. If one assumes time interval 5 (center of maximum dO2 signal) is zero and then calculates a phase relative to that time interval, it is readily apparent that G2 metabolites are more in phase with the dissolved oxygen than the G1 metabolites, which are approximately 180 ° out of phase. The phase shift described on PC2 can also

been seen in Figures 5.8B, D. Cystathionine, Figure 5.8D, and myo-inositol, Figure 5.8B, have approximately the same PC1 score, but are on opposite ends of PC2.

5.4.6 Biological interpretations

The cycling patterns obtained for the metabolites agree with the gene expression data reported in literature, where it is stated that both DNA replication and cell division are compartmentalized as a function of the Yeast Metabolic Cycle (YMC).[34,35] Klevecz and co-workers reported that transcripts involved in sulfur metabolism were up regulated in the respiratory phase (time intervals 9-12, 21-24).[35] Cystathionine and homoserine, which are part of sulfur metabolism, are both more abundant in time intervals 8-12 and 20-24, Figure 5.8D and Table 5.4. Also, Tu and co-workers reported that the small metabolites ethanol and acetate, which are indicators of glycolytic metabolism, are more abundant in time intervals 1-2, 10-14 and 22-24.[34] Glucose-6-phosphate, 6-phospho-d-gluconate, and pyruvate are all up regulated during these time intervals. It is also interesting to note that metabolites in the urea cycle follow the same temporal pattern. Ornithine and arginine both experience maximum intracellular concentration around the same time intervals Table 5.4. Further discussion is beyond the scope of this dissertation; however the study of the YMC is of great importance since it is believed that higher eukaryotes experience similar patterns.

5.5 Conclusions

A data processing method termed the Sratio method was developed for 3D GC x GC-TOFMS data and proven to distinguish chromatographic locations showing changes between a large number of sample classes, in this application cycling

metabolites. The major strength of this reported algorithm is that only the most selective *m/z* were retained for each 2D chromatographic location in order to find sample (*i.e.* time interval) differentiating locations. Two additional strengths of the Sratio algorithm are the user's ability to tune the algorithm to reduce solvent and reagent artifacts as well as obtain one Sratio per chromatographic peak. Issues addressed for proper implementation of this algorithm include the dependence on accurate retention time alignment between samples and the need for a sufficient sampling of the first dimension slices to reconstruct a column 1 peak profile. Following the Sratio method by an accurate quantification tool, in this case PARAFAC, allows an improved determination into the differentiating ability of the located chromatographic sub region, and an objective cluster analysis (PCA) can be used to determine the similarity of samples. These combined chemometric tools allow the user to gain an accurate picture of the system under investigation. This method also has potential for the analysis of data exhibiting similar behavior in more than one sample type but large differences between to others. The bulk of the biological implications can be found in another report.[40]

Table 5.1: Peak volume percent relative standard deviation comparison.

The %RSD for 16 metabolite peak volumes was calculated for the raw, TIC normalized, and normalized to the average of the isotopically labeled serine and alanine.

Metabolite	Raw %RSD	TIC %RSD	13C average %RSD
2OH-Isovaleric acid	7.76	7.80	6.60
malic acid	6.08	6.50	4.76
methionine	6.39	7.24	6.67
myo-inositol	14.17	12.72	12.31
6 phospho-d-gluconate	15.62	14.28	14.15
glucose 6 phosphate	7.36	7.28	6.42
trehalose	6.90	6.91	5.65
pyruvate	7.79	8.32	5.29
glycine 2TMS	5.12	7.50	6.47
valine	5.57	7.00	4.68
glucose	5.07	6.97	5.83
lactic acid	8.55	8.34	3.52
leucine	8.95	8.34	9.43
isoleucine	8.84	8.07	8.85
proline	5.15	6.76	6.77
succinic acid	7.77	8.09	6.85

Table5.2: Metabolites located with a Sratio greater than three.
[+] retention time confirmed by standards.

Metabolite	t_{R1}, min	t_{R2}, s	MV	S/N	Sratio	Vratio
methyl citrate	19.8	0.64	912	1761	31.5	36.2
fructose[+]	20.3	0.36	916	2814	29.8	33.2
sucrose[+]	28.5	0.52	890	3378	26.7	17.2
6-phospho-d-gluconate[+]	27.9	0.7	838	3325	17.4	13.1
glycerol 3P[+]	18.8	0.97	892	448	9.6	10.1
glucopyranose	21.6	0.37	803	380	6.2	9.8
glycerol[+]	11.3	0.38	947	>10000	12.7	7.4
2OH-IsoValeric Acid	9.4	0.59	932	3074	10.1	6.7
cystathionine	24.4	0.69	854	207	4.8	4.9
myo-inositol[+]	23.0	0.51	923	4914	7.4	4.8
lysine[+]	21.0	0.54	931	409	6.6	4.2
serine[+]	12.8	0.58	889	6404	5.2	4.1
norvaline	10.7	0.59	854	116	3.4	4.1
aspartic acid[+]	15.4	0.75	957	>10000	5.2	4.0
lactic acid[+]	7.5	0.6	945	6775	5.1	3.5
ornithine	19.6	0.57	936	644	3.8	3.3
arginine[+]	16.8	0.56	946	893	4.4	3.2
succinic acid[+]	12.0	0.89	941	928	7.0	3.2
homoserine	14.2	0.56	887	398	3.9	3.0
2-methyl malic acid	14.6	0.69	862	301	3.8	3.0
threonic acid	15.7	0.49	899	333	3.4	2.9
2-amino adipinic acid	18.2	0.75	837	183	3.3	2.9
glycine[+]	11.9	0.59	913	268	4.6	2.8
glucose[+]	20.9	0.44	880	413	3.9	2.8
glycerate 3P[+]	19.4	1.16	899	2451	5.8	2.7
glucose 6P[+]	25.5	0.67	945	4584	4.9	2.6
pyruvate[+]	7.3	0.56	918	822	4.4	2.6
trehalose[+]	30.0	0.75	909	>10000	7.3	2.5
fumaric acid[+]	12.7	0.98	935	308	4.3	2.4
3OH butanoic acid	9.3	0.62	914	551	3.3	2.3
ribulose 5P[+]	23.3	0.7	826	147	3.3	2.1
proline[+]	11.8	0.61	954	3566	4.3	2.1
leucine[+]	11.3	0.6	937	1746	3.6	2.0
alanine[+]	8.3	0.61	964	>10000	3.1	2.0
isoleucine[+]	11.7	0.59	944	2934	3.7	2.0
NAD fragment[+]	22.2	1.04	878	244	3.2	1.9
thiamin diphosphate[+]	17.4	1.29	919	1332	3.9	1.9
d-ribose 5P[+]	23.1	0.71	901	336	3.8	1.9
valine[+]	10.3	0.61	954	9054	3.7	1.8
threonine[+]	13.2	0.56	957	6843	3.4	1.7
phenylalanine[+]	17.0	0.67	933	650	3.2	1.7
UDP glucose & G1P[+]	18.9	0.43	951	3108	3.8	1.6
mannitol[+]	21.1	0.41	926	2224	3.4	1.6
pseudo uridine	25.7	0.51	783	460	3.3	1.5

Table 5.3: Locations of the unknown metabolites.

The group number corresponds to the clustering described in Figure 5.9

Metabolite	t_{R1}, min	t_{R2}, s	Vratio	PC1 score	PC2 score	Group
unk 897	15.0	1.02	18.0	2.15	-1.66	G1
unk 562	9.4	0.39	10.8	1.71	0.30	G4
unk 1288	21.5	0.5	8.6	2.06	0.71	G2
unk 561	9.4	0.41	6.3	2.42	0.18	G4
unk 915	15.3	0.35	6.0	3.74	-0.01	G4
unk 1002	16.7	0.86	6.0	2.81	0.99	G2
unk 832	13.9	0.99	5.7	2.82	1.25	G2
unk 972	16.2	0.59	3.8	3.05	-0.80	G1
unk 1053	17.6	0.4	3.7	2.93	-0.19	G4
unk 711	11.9	0.5	3.6	2.65	-0.09	G4
unk 1207	20.1	0.71	3.6	2.82	-0.63	G1
unk 984	16.4	0.43	3.4	2.11	-0.07	G4
unk 844	14.1	0.44	3.2	2.49	-0.08	G4
unk 970	16.2	0.75	3.2	3.16	-0.88	G1
unk 1227	20.5	0.99	3.2	2.31	0.12	G4
unk 783	13.1	0.54	3.1	3.28	0.90	G2
unk 685	11.4	0.37	3.0	3.47	0.14	G4
unk 634	10.6	0.59	2.9	3.04	0.92	G2
unk 1084	18.1	1.31	2.9	3.72	-0.25	G4
unk 1257	21.0	0.72	2.8	2.96	-0.75	G1
unk 432	7.2	1.22	2.7	3.26	0.18	G4
unk 681	11.4	1.34	2.6	3.49	0.03	G4
unk 567	9.5	0.73	2.5	2.53	-0.02	G4
unk 934	15.6	0.98	2.4	3.67	0.56	G2
unk 1598	26.6	0.67	2.4	3.12	-0.16	G4
unk 1189	19.8	0.51	2.4	3.23	-0.13	G4
unk 1287	21.5	0.68	2.4	3.26	0.37	G4
unk 951	15.9	0.76	2.3	3.41	0.27	G4
unk 982	16.4	0.66	2.1	3.54	0.74	G2
unk 846	14.1	0.53	1.9	3.74	0.51	G2
unk 1046	17.4	1.42	1.8	3.78	-0.12	G4
unk 882	14.7	0.79	1.8	3.46	-0.14	G4
unk 798	13.3	0.59	1.8	3.84	0.06	G4
unk 1990	33.2	1.37	1.7	3.76	-0.11	G4
unk 978	16.3	0.36	1.7	3.64	-0.25	G4
unk 1042	17.4	0.84	1.6	3.79	0.02	G4
unk 910	15.2	1.08	1.6	4.00	0.35	G4
unk 1203	20.1	0.45	1.6	4.11	0.09	G4
unk 1134	18.9	0.57	1.5	4.09	-0.06	G4
unk 1155	19.3	0.54	1.5	4.09	0.02	G4
unk 792	13.2	0.4	1.3	4.27	0.05	G4

Table 5.4: Locations of identified metabolites in the PCA scores space.
The group number corresponds to the clustering described in Figure 5.9 and the phase, ϕ, was calculated relative to the dissolved oxygen.

Metabolite	PC1 score	PC2 score	Group	Vratio	ϕ
methyl citrate	0.69	-0.28	G3	36.2	----
fructose	0.65	-0.59	G3	33.2	----
sucrose	1.31	0.46	G3	17.2	----
6-pshospho-d-gluconate	1.66	-0.22	G3	13.1	----
glycerol-3P	1.13	-0.39	G3	10.1	----
glucopyranose	1.51	-0.04	G3	9.8	----
glycerol	1.43	-0.63	G3	7.4	----
2OH-IsoValeric	2.44	1.47	G2	6.7	0
cystathionine	2.78	-1.01	G1	4.9	150
myo-inositol	2.59	0.63	G2	4.8	60
lysine	2.80	-1.07	G1	4.2	120
serine	2.50	-0.82	G1	4.1	120
norvaline	2.96	1.21	G2	4.1	30
aspartic acid	2.99	1.13	G2	4.0	300 (-60)
lactic acid	2.41	-0.23	G4	3.5	----
ornithine	3.15	-0.88	G1	3.3	210
arginine	2.74	-0.62	G1	3.2	210
succinate	3.16	-0.62	G1	3.2	150
homoserine	3.25	-0.76	G1	3.0	150
2-methyl malic acid	2.87	-0.55	G1	3.0	240
threonic acid	3.27	0.90	G2	2.9	30
2-amino-adipinic acid	3.10	-0.79	G1	2.9	150
glycine	2.95	-0.30	G4	2.8	----
glucose	2.62	-0.10	G4	2.8	----
glycerate 3P	3.16	0.24	G4	2.7	----
glucose 6P	3.21	-0.12	G4	2.6	----
pyruvate	3.29	-0.74	G1	2.6	120
trehalose	3.09	0.44	G2	2.5	60
fumaric acid	2.94	0.14	G4	2.4	----
3OH butanoic acid	3.45	0.70	G2	2.3	0
ribulose 5P	3.11	-0.31	G4	2.1	----
proline	3.42	-0.28	G4	2.1	----
leucine	3.55	-0.38	G1	2.0	120
alanine	3.71	-0.04	G4	2.0	----
isoleucine	3.80	0.56	G2	2.0	60
NAD fragment	3.56	-0.13	G4	1.9	----
thiamin diphosphate	3.55	-0.04	G4	1.9	----
ribose 5P	3.47	-0.19	G4	1.9	----
valine	3.91	0.06	G4	1.8	----
threonine	3.56	-0.02	G4	1.7	----
phenylalanine	3.99	0.27	G4	1.7	----
UDP glucose & G1P	4.07	-0.01	G4	1.6	----
mannitol	4.12	-0.02	G4	1.6	----
pseudo uridine	4.02	-0.04	G4	1.5	----

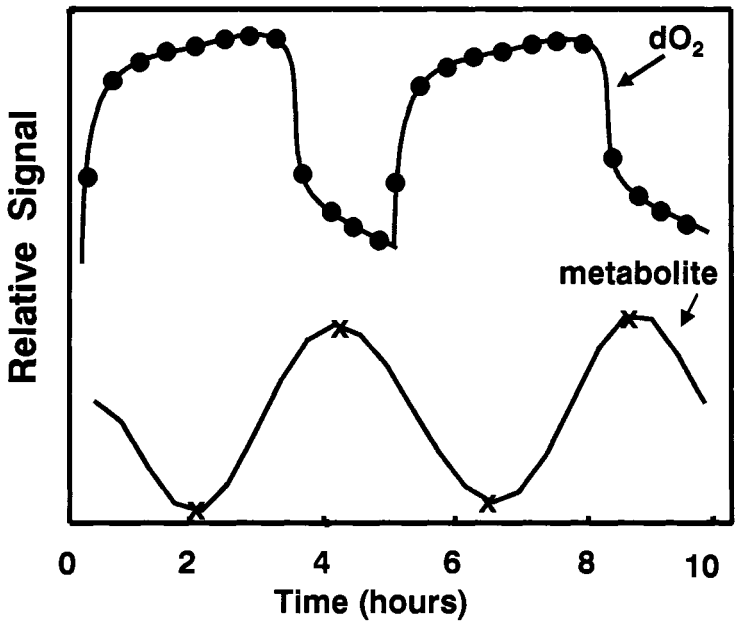


Figure 5.1: Illustration of dissolved oxygen content and a predicted metabolite trace. Yeast cells grown under continuous nutrient limited conditions exhibit cyclic patterns with each cycle lasting five hours. Samples were collected every 23-25 minutes as shown by the dots on the dissolved oxygen trace. The locations of the time intervals averaged for Smax and Smin are shown by the "x" on the predicted metabolite trace.

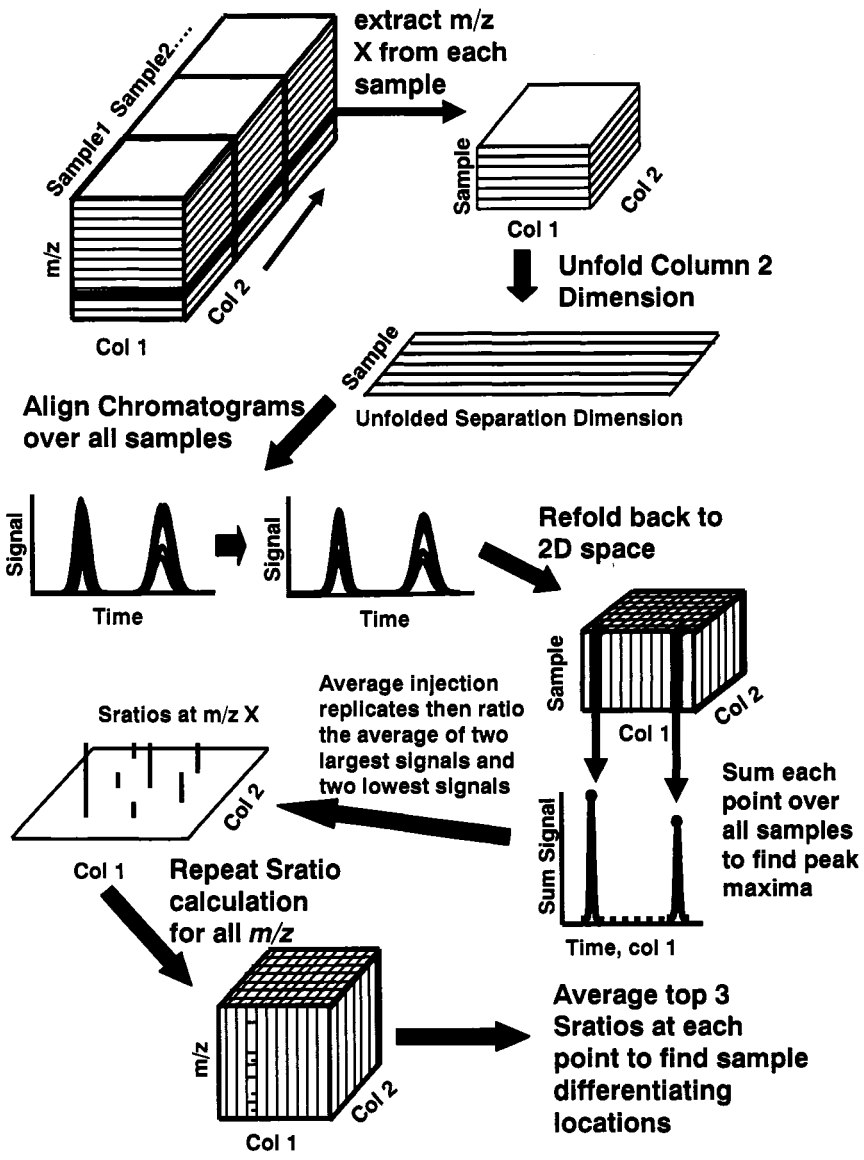


Figure 5.2: A stepwise schematic explaining the Sratio method.
One mass channel, *m/z*, is removed from each sample and compiled into a one data cube. Each sample is unfolded to remove the second dimension and aligned to the first sample. The samples are then refolded back into the data cube and each point in the 2D separation space summed over all samples to pinpoint peak maxima. The injection replicates are then averaged. The peak heights from the two samples with the largest signals at each point in the chromatographic space are averaged, Smax. The two lowest signals over all samples at each point were also averaged, Smin. A ratio of Smax to Smin was then calculated at each *m/z* to give insight into the amplitude of the cycles for each peak. This process is then repeated in an automated fashion over all *m/z* specified to obtain a Sratio spectrum at each point in the 2D chromatographic space. The three largest Sratios are averaged to find differentiating chromatographic locations.

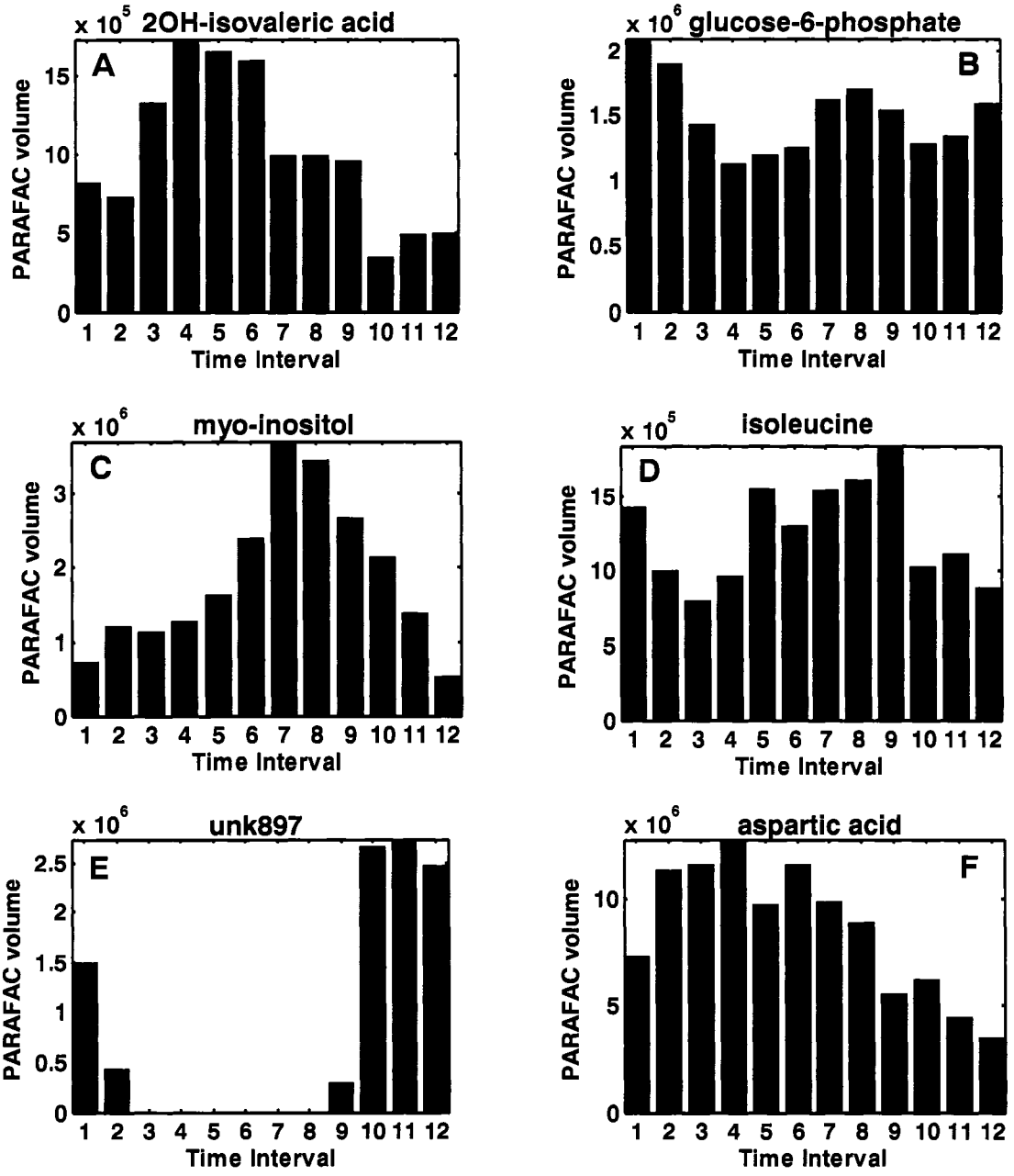


Figure 5.3: One cycle metabolite volume profiles.
Peak volume profiles for: (A) 2-hydroxy isovaleric acid, (B) glucose-6-phosphate, (C) myo-inositol, (D) isoleucine, (E) an unknown at 897 seconds on column 1, and (F) aspartic acid.

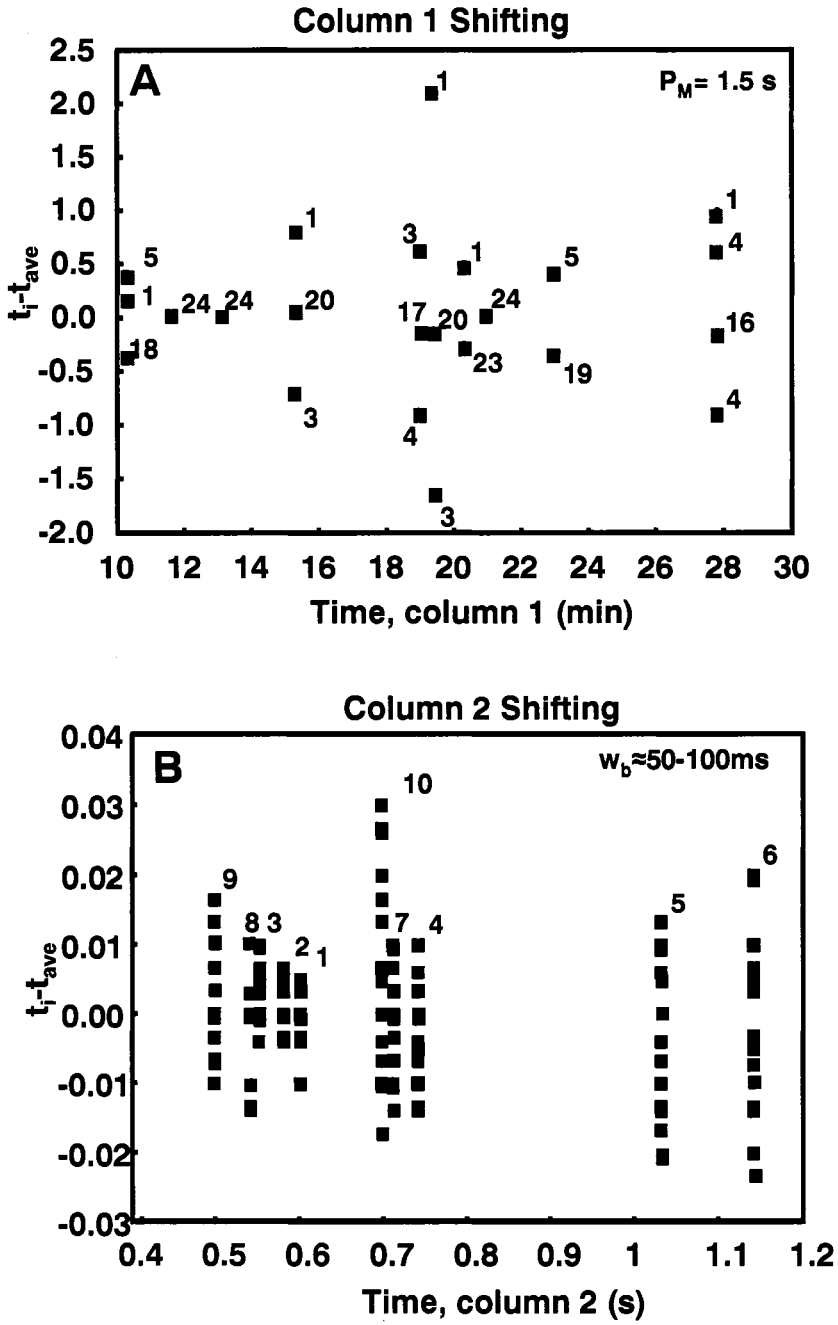


Figure5.4: Retention shifting in raw data for 24 samples on column 1 and 2.
(A) Average retention shifting on column 1 in minutes for ten selected metabolites. The number beside each square signifies the number of samples with that shift value. (B) Average retention shifting on column 2 for those ten selected metabolites in (A). The numbers correspond to the elution order on column 1.

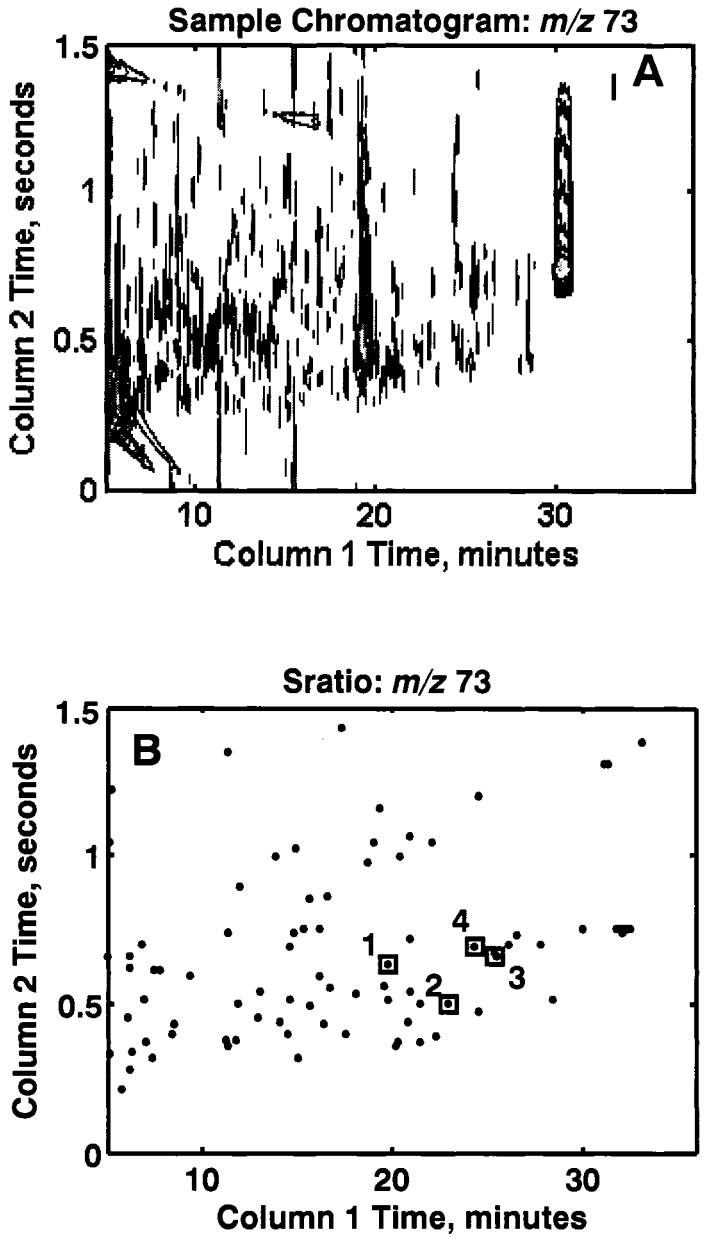


Figure 5.5: Complexity of the yeast extracts at *m/z* 73.
(A) Contour plot of the raw data at *m/z* 73. (B) The results at *m/z* 73 after the data has been analyzed by Sratio method. The locations of the metabolites for the validation studies are enclosed in the boxed regions labeled 1-4, which were identified as methyl citrate, myo-inositol, glucose 6 phosphate and cystathionine.

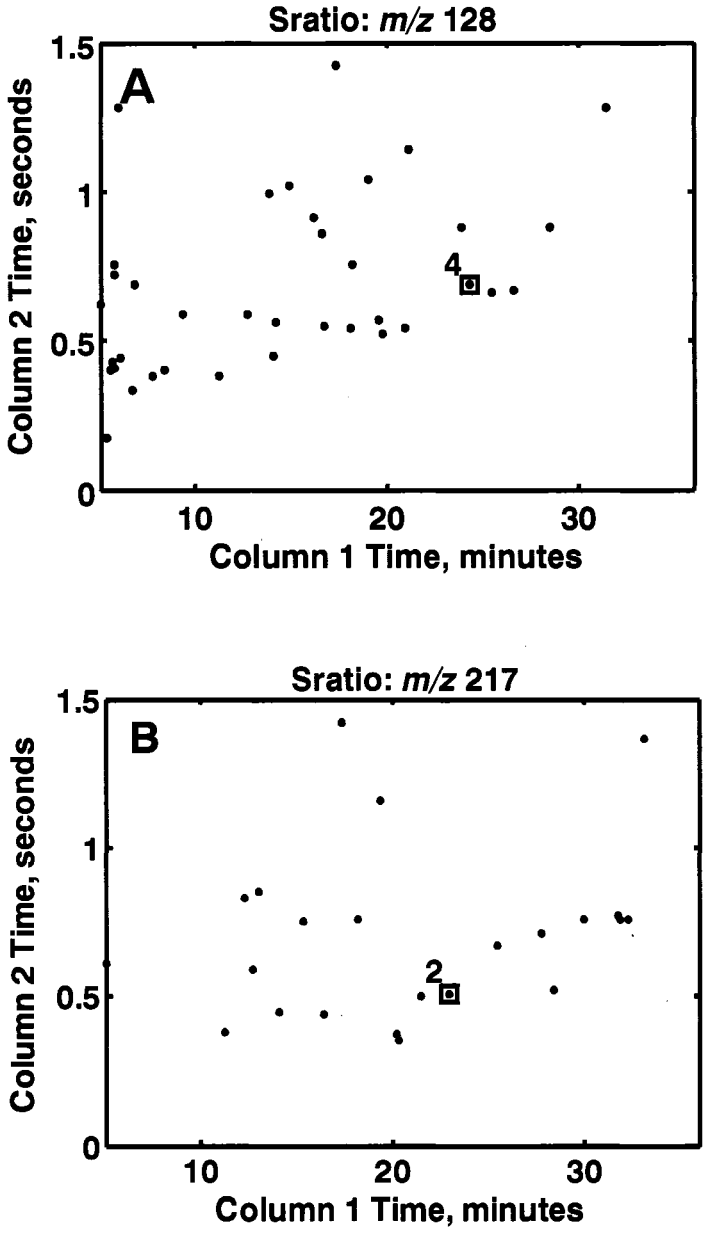
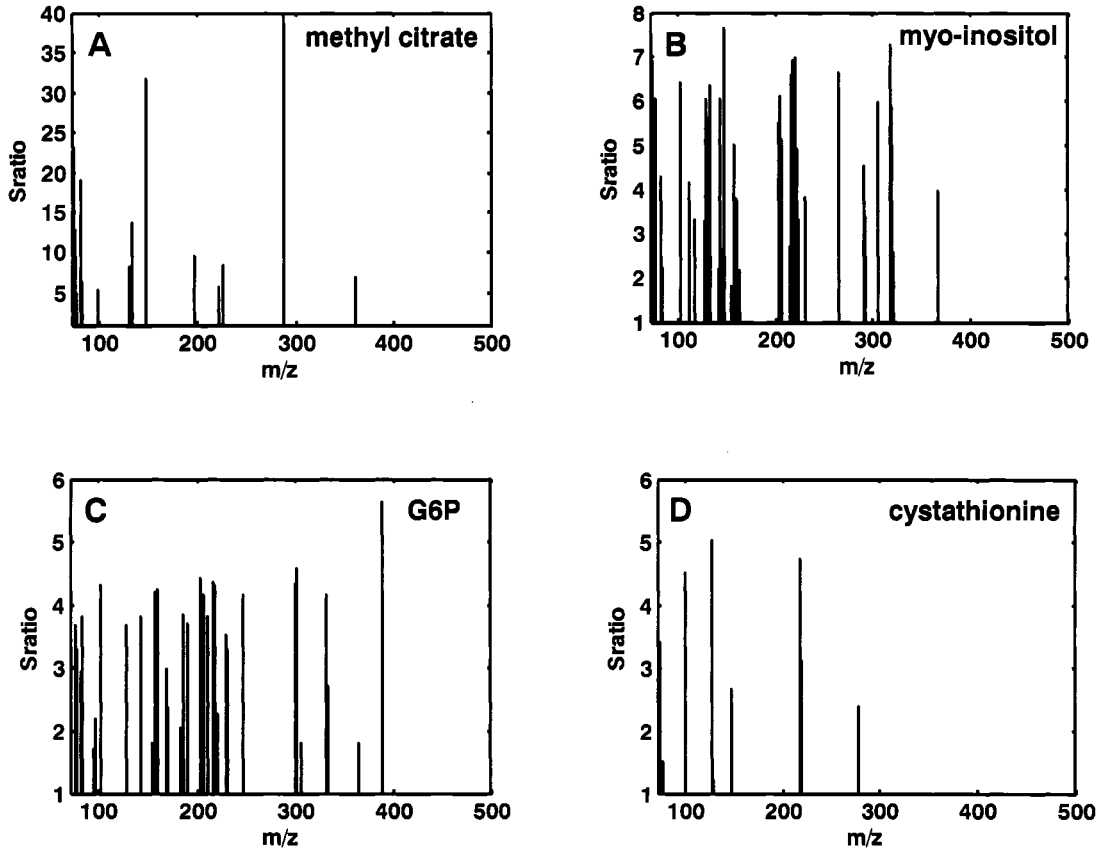


Figure 5.6: Two-dimensional Sratio plots at selective *m/z*.
(A) 2D Sratio plots at *m/z* 128 and (B) *m/z* 217. These *m/z* yield more selective information than the global *m/z* 73. Only one in four of the metabolites for the validation study are include in these *m/z*. For identification of these metabolites see the text.



5.7: The Sratio "mass" spectra for the four metabolites studied in depth.
In order of decreasing Sratios: (A) methyl citrate, (B) myo-inositol, (C) glucose-6-phosphate and (D) cystathionine.

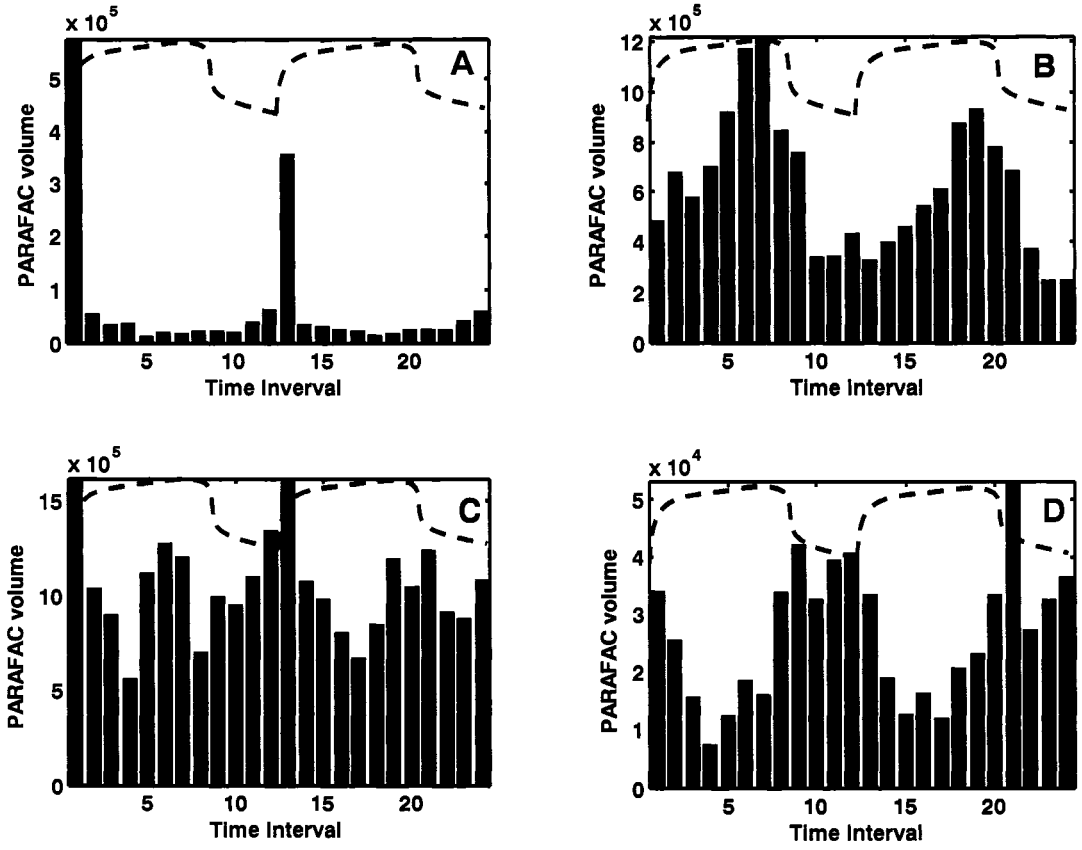


Figure 5.8: Four cycling patterns observed in the data.
(A) sharp spikes (methyl citrate), (B) unimodal oscillations maximizing around time interval 7 (myo-inositol), (C) bimodal oscillations (glucose-6-phosphate), and (D) unimodal oscillation directly out of phase with (B) (cystathionine). These patterns were determined following identification and PARAFAC quantification. The dashed line represents the dissolved oxygen profile.

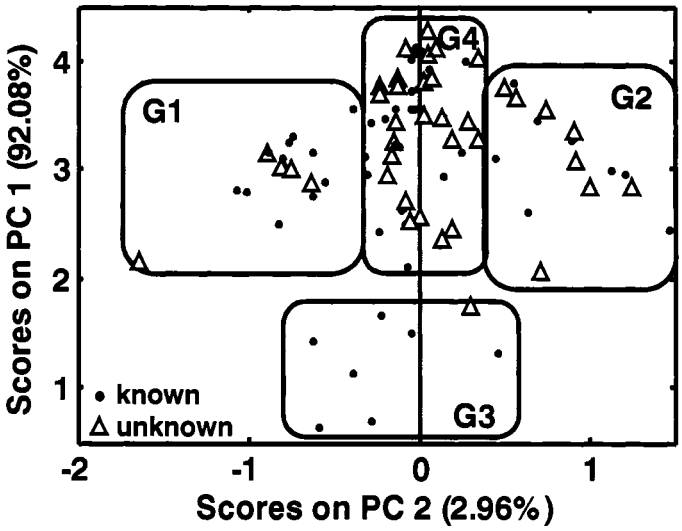


Figure 5.9: Objective classifying of the metabolites using PCA.
Metabolites were determined to fall into roughly four groups, G1-4. G1 and 2 are unimodal but out of phase, G3 metabolites show spiking traits and metabolites in G4 are either multimodal or not cycling (*i.e.* low depth of modulation).

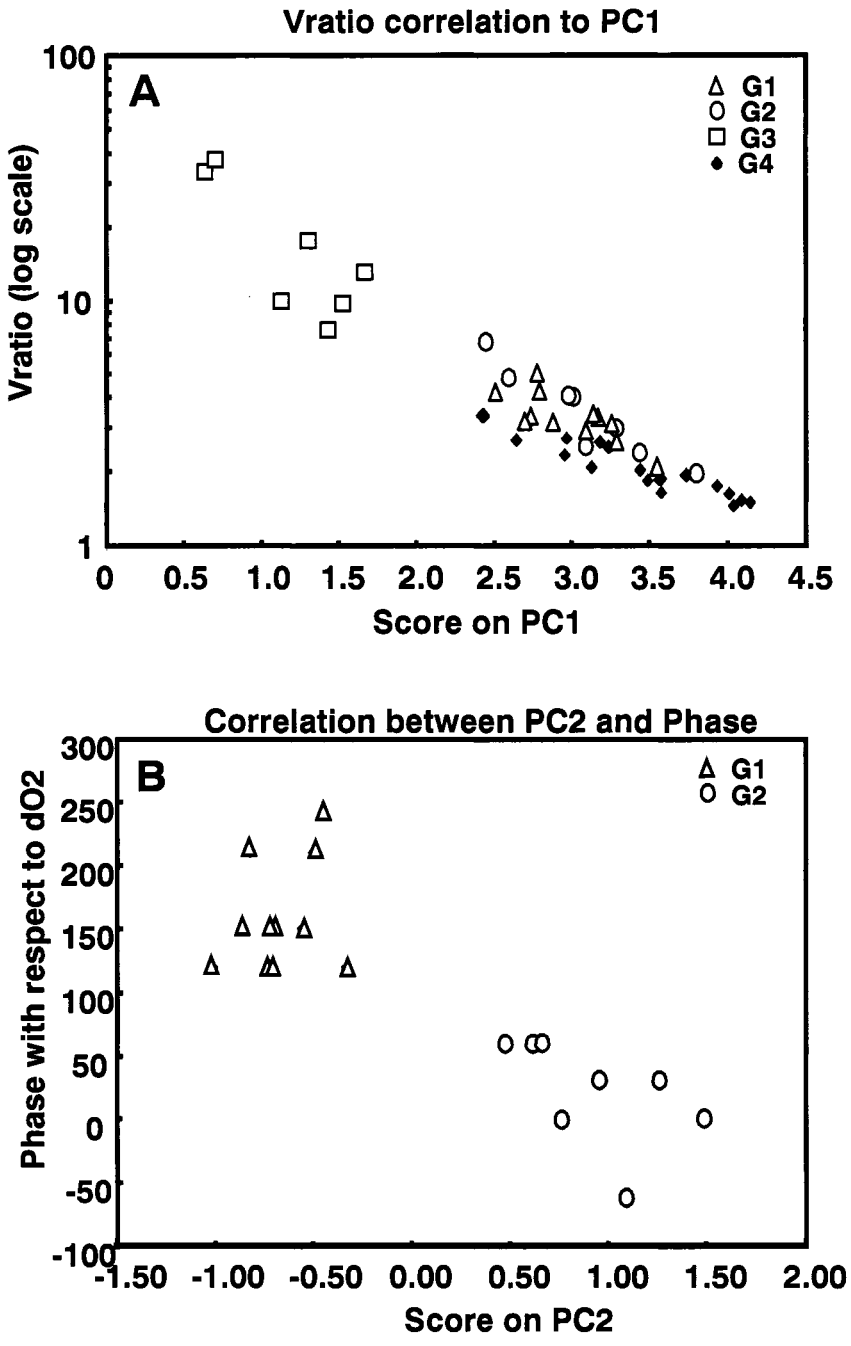


Figure 5.10: Correlation between the PCA scores and metabolites.
(A) PC1 describes the Vratios and (B) PC2 separates based on phase information. The phase was calculated relative to time interval 5 of the dissolved oxygen cycle.

Notes to Chapter 5

(1) Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227-231.

(2) Kinghorn, R. M.; Marriott, P. J. *J. High Resol. Chromatogr.* **1998**, *21*, 620-622.

(3) Beens, J.; Adahchour, M.; Vreuls, R. J. J.; Altena, K. v.; Brinkman, U. A. Th. *J. Chromatogr. A* **2001**, *919*, 127-132.

(4) Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **1998**, *70*, 2796-2804.

(5) Seeley, J. V.; Kramp, F.; Hicks, C. J. *Anal. Chem.* **2000**, *72*, 4346-4352.

(6) Giddings, J. C. *Unified Separation Science*; Wiley-Interscience: New York, 1991.

(7) Focant, J.-F.; Sjodin, A.; Jr., D. G. P. *J. Chromatogr. A* **2003**, *1019*, 143-156.

(8) Johnson, K. J.; Prazen, B. J.; Olund, R. K.; Synovec, R. E. *J. Sep. Sci.* **2002**, *25*, 297-303.

(9) Beens, J.; Brinkman, U. A. Th. *Analyst* **2005**, *130*, 123-127.

(10) Dimandja, J.-M. D.; Grainger, J.; Jr., D. G. P.; Turner, W. E.; Needham, L. L. *J. Exposure Analysis and Environmental Epidemiology* **2000**, *10*, 761-768.

(11) Deursen, M. v.; Beens, J.; Reijenga, J.; Lipman, P.; Cramers, C.; Blomberg, J. *J. High Resol. Chromatogr.* **2000**, *23*, 507-510.

(12) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 2700-2709.

(13) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1056*, 145-154.

(14) Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagel, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 5068-5075.

(15) Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O. *Metabolomics* **2005**, *1*, 65-73.

(16) Shellie, R. A.; Welthagen, W.; Zrostlikova, J.; Spranger, J.; Ristow, M.; Fiehn, O.; Zimmermann, R. *J. Chromatogr. A* **2005**, *1086*, 83-90.

(17) Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Lidstrom, M. E.; Synovec, R. E. *Talanta* **2005**, *65*, 380-388.

(18) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Jack, R. M.; Synovec, R. E. *J. Chromatogr. A* **2004**, *1058*, 209-215.

(19) Pichersky, E.; Gang, D. R. *Trends Plant Sci.* **2000**, *5*, 439-445.

(20) Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W. *Trends Plant Sci.* **2004**, *9*, 418-425.

(21) Förster, J.; Famili, I.; Fu, P.; Palsson, B. O.; Nielsen, J. *Genome Research* **2003**, *13*, 244-253.

(22) Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. *Analyst* **2005**, *130*, 606-625.

(23) Jonsson, P.; Johansson, A. I.; Gullberg, J.; Trygg, J.; A, J.; Grung, B.; Marklund, S.; Sjostrom, M.; Antti, H.; Moritz, T. *Anal. Chem.* **2005**, *77*, 5635-5642.

(24) Hollywood, K.; Brison, D. R.; Goodacre, R. *Proteomics* **2006**, *6*, 4716-4723.

(25) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245-252.

(26) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends Biotechnol.* **1998**, *16*, 373-378.

(27) Berg, R. A. v. d.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; Werf, M. J. v. d. *BMC Genomics* **2006**, *7*, 142-157.

(28) Kell, D. B.; Oliver, S. G. *BioEssays* **2004**, *26*, 99-105.

(29) Tikunov, Y.; Lommen, A.; Vos, C. H. R. d.; Verhoeven, H. A.; Bino, R. J.; Hall, R. D.; Bovy, A. G. *Plant Physiology* **2005**, *139*, 1125-1137.

(30) Sihna, A. E.; Prazen, B. J.; Synovec, R. E. *Anal. Bioanal. Chem.* **2004**, *378*, 1948.

(31) Johnson, K. J.; Synovec, R. E. *Chemom. Intell. Lab. Sys.* **2001**, *60*, 225-237.

(32) Hoggard, J. C.; Synovec, R. E. *Anal. Chem.* **2007**, *79*, 1611-1619.

(33) Wise, B. M.; Gallangher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, R. S. In *PLS_Toolbox 3.5 for use with MATLAB*, 2005, pp 123-127.

(34) Tu, B. P.; Kudlicki, A.; Rowicka, M.; McKnight, S. L. *Science* **2005**, *310*, 1152-1158.

(35) Klevecz, R. R.; Bolen, J.; Forrest, G.; Murray, D. B. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1200-1205.

(36) Murray, D. B.; Beckmann, M.; Kitano, H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2241-2246.

(37) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J.; Oliver, S. G. *Phytochemistry* **2003**, *62*, 929-937.

(38) Johnson, K. J.; Wright, B. W.; Jarman, K. H.; Synovec, R. E. *J. Chromatogr. A* **2003**, *996*, 141-155.

(39) Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Sys.* **2000**, *52*, 1-4.

(40) Tu, B. P.; Mohler, R. E; Liu, J. C.; Dombek, K. M.; Young, E. T.; Synovec, R. E.; McKnight, S. L. submitted *Proc. Natl. Acad. Sci. U. S. A.*, **2007**

Chapter 6. Research Conclusions and Future Directions

6.1 Concluding Remarks

The hypothesis that GCxGC-TOFMS instrumentation was applicable to metabolomic analysis proved correct and a more complete understanding of complex biological phenomenon was obtained. A method was successfully developed for the discovery-based analysis of yeast metabolites using GCxGC-TOFMS instrumentation to study two very different sets of yeast samples ($^{(1)}$ two sample types and $^{(2)}$ 24 sample types). The use of this third order instrument was determined suitable for yeast metabolomic studies following the proof-of-principle study on yeast cells believed to be exhibiting large differences based on previously reported gene expression data. Data analysis is the key ingredient for metabolomic analysis and data reduction tools (i.e. chemometrics) are essential for the discovery-based analysis of complex yeast metabolomic samples.

Ideally, the analyst should use all data dimensions and determine the most selective *m/z* to interpret the data not only for data reduction but also for accurate quantification. Unfortunately, very few chemometric tools exist that can take advantage of third order data in its entirety. Chemometric techniques presented in this dissertation progressed from the use of commercially available software on a subset of the data cube, Chapter 3, to in-house written chemometric software since no third order chemometric software was commercially available, Chapters 4 and 5. Initially PCA was applied to selected *m/z*, however, it was determined that a large portion of the important information included in the data cube was ignored. To address this issue a recently developed 4D

Fisher ratio method adapted from a 2D algorithm was applied to GCxGC-TOFMS data as a data reduction technique. The Fisher method worked well for metabolomic data with only two classes and large class-to-class differences, however due to the class correlation of the noise a signal weight was placed on the data, which like PCA, created a signal intensity bias. To remove signal intensity bias a 4D Sratio method was developed. This Sratio method like the Fisher method analyzes the entire cube of data collected on the GCxGC-TOFMS in an automated fashion. Although the Sratio method analyzes all *m/z* only the most selective *m/z* are retained in order to reduce signal contribution from interferents and noise. The beauty of this method is the robustness against signal bias and the ability to locate differentiating peaks between a large number of sample classes. The data reduction tools coupled with peak identification using commercially available software and PARAFAC quantification have greatly aided in the generation of accurate biological conclusions drawn from the metabolomic samples analyzed on the GCxGC-TOFMS.

There are many reports in literature stating that there is no universal instrumentation that has the ability to detect every class of compound present in the metabolome. This is certainly still the case. The limits as to the types of compounds amenable to GCxGC-TOFMS were discussed in Chapter 4. The current derivatization methods do not enable the analysis of very large metabolites, *e.g.* cofactors, using GCxGC-TOFMS. Also, the very small volatile compounds, *e.g.* acetate and ethanol, elute with the reagent artifacts and can not be detected in the same sample as the other slightly larger metabolites, *e.g* . amino acids and sugars. To obtain a more complete

description of the metabolome (metabolomics) complementary analytical techniques must be applied.

6.2 Future Directions

The development of chemometric software for two-dimensional chromatographic systems has made significant progress in the last couple of years; however there is still large room for improvement. One significant improvement would involve the development of a fast and robust 2D retention time alignment that can handle peak tailing in both separation dimension one and two. The Sratio method reported is highly dependant on alignment of chromatographic peaks. In the study reported in Chapter 5, all the data was collected on the same set of columns with the significant retention shifting along only column 2, which is readily corrected by a 1D alignment. A second area of chemometrics requiring ongoing development is the exporting of data from the instrument software into analysis software. Work presented in Chapters 3-5 utilized an export function developed in-house, however, any data preprocessing performed in the instrumentation software, *e.g.* baseline correction, is lost in the transfer.

There are many future metabolomic applications that can be studied using the GCxGC-TOFMS in concert with the chemometric techniques discussed in this dissertation, especially the Sratio method. Unlike PCA and the Fisher ratio method, the Sratio method averages the within class samples, thus reducing the dependence of the algorithm on extraction and injection variation. For illustrative purposes, ideal sample sets for Sratio analysis using four classes are described in Figure 6.1. Even though Example 3 shows the largest difference between samples, Fisher analysis would more

likely output Example 2 as being a more differentiating location since each class has a distinctively different value. Any data, *e.g.* metabolomic data, that is collected on a third order instrument can be submitted to this type of analysis to determine the locations of the largest differences in the data. The methods developed in this research were geared towards the analysis of yeast cells using a GC x GC-TOFMS however with few modifications the methods reported can be applied to other organisms such as rats and mice. In the future, it would also be beneficial to determine exact concentrations of metabolites in the samples, rather than the relative concentrations reported in this dissertation.

One other area of potential future work lies in the area of cellular processes in yeast. The understanding of cellular processes is extremely important due to the undesired mutations that can be formed and result in diseases such as cancer. Another interesting mutation study would involve the analysis of metabolites generated when specific genes or protein kinases are removed from the yeast cells. It would be beneficial to study this phenomenon by collecting both gene expression and metabolite data and determine how the different "-omics" interact.

Significant improvement is also needed in the mass spectral libraries in order to minimize the number of unknown compounds. There are many metabolites that are not included in the mass spectral library databases. This problem is due to a couple of reasons. A few important reasons for the libraries lacking so many compounds are that not all metabolites are stable and commercially available. In order to add metabolites not commercially available to library, organic chemists would have to synthesize these

compounds in a research laboratory. One other reason is that researchers have not yet been able to run all metabolites as standards on the different types of mass spectrometers. The different types of mass spectrometers will yield different fragmentation patterns, which could result in misleading mass spectral match values. Significant progress has been made, but there is still a long ways to go.

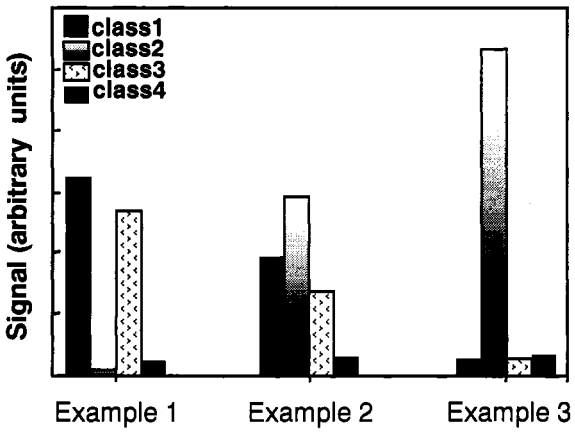


Figure 6.1: Illustration of sample sets for potential future work.
Examples 1-3 are metabolite signal patterns that can be located in GCxGC-TOFMS data using the data reduction tools discussed in Chapters 4 and 5.

Bibliography

A, J.; Trygg, J.; Gullberg, J.; Johansson, A. I.; Jonsson, P.; Antti, H.; Marklund, S. L.; Moritz, T. "Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome," *Anal. Chem.*, **2005**, 77, 8086-8094.

Adahchour, M.; van Stee, L. L. P.; Beens, J.; Vreuls, R. J. J.; Batenburg, M. A.; Brinkman, U. A. Th. "Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection for the trace analysis of flavour compounds in food," *J. Chromatogr. A*, **2003**, 1019, 157-172.

Adams, M. A.; Chen, Z.; Landman, P.; Colmer, T. D. "Simultaneous Determination by Capillary Gas Chromatography of Organic Acids, Sugars, and Sugar Alcohols in Plant Tissue Extracts as Their Trimethylsilyl Derivatives," *Anal. Biochem.*, **1999**, 266, 77-84.

Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. "High-throughput classification of yeast mutants for functional genomics using metabolic footprinting," *Nature Biotehnol.*, **2003**, 21, 692-696.

Aoki-Kinoshita, K. F. "Overview of KEGG applications to omics-related research," *J. Pestic. Sci.*, **2006**, 31, 296-299.

Bajad, S.; Shulaev, V. "Highly-parallel metabolomics approaches using LC-MS^2 for pharmaceutical and environmental analysis," *Trends Anal. Chem.*, **2007**, 26, 625-636.

Batta, A. K.; Salen, G.; Rapole, K. R.; Batta, M.; Batta, P.; Alberts, D.; Earnest, D. "Highly simplified method for gas-liquid chromatographic quantitation of bile acids and sterols in human stool," *J. Lipid Res.*, **1999**, 40, 1148-1154.

Batta, A. K.; Salen, G.; Batta, P.; Tint, G. S.; Alberts, D.; Earnest, D. "Simultaneous quantitation of fatty acids, sterols and bile acids in human stool by capillary gas-liquid chromatography," *J. Chromatogr. B*, **2002**, 775, 153-161.

Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics a Practical Guide*. New York: John Wiley & Sons, Inc. **1998**.

Beens, J.; Boelens, H.; Tijssen, R.; Blomberg, J. "Quantitative Aspects of Comprehensive Two-Dimensional Gas Chromatography (GC x GC)," *J. High Resol. Chromatogr.*, **1998**, 21, 47-54.

Beens, J.; Blomberg, J.; Schoenmakers, P. J. "Proper tuning of comprehensive two-dimensional gas chromatography (GC x GC) to optimize the separation of complex oil fractions," *J. High Resol. Chromatogr.*, **2000**, 23, 182-188.

Beens, J.; Adahchour, M.; Vreuls, R. J. J.; van Altena, K.; Brinkman, U. A. Th. "Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, **2001**, 919, 127-132.

Bergstroem, S. K.; Dahlin, A. P.; Ramstroem, M.; Andersson, M.; Markides, K. E.; Bergquist, J. "A simplified multidimensional approach for analysis of complex biological samples: on-line LC-CE-MS," *Anal. Chem.*, **2006**, 131, 791-798.

Berlanga, T. M.; Peinado, R.; Millán, C.; Mauricio, J. C.; Ortega, J. M. "Influence of Blending on the Content of Different Compounds in the Biological Aging of Sherry Dry Wines," *J. Agric. Food Chem.*, **2004**, 52, 2577-2581.

Bertsch, W. "Two-Dimensional Gas Chromatography. Concepts, Instrumentation, and Applications – Part 1: Fundamentals, Conventional Two-Dimensional Gas Chromatography, Selected Applications," *J. High Resol. Chromatogr.*, **1999**, 22, 647-665.

Bertsch, W. "Two-Dimensional Gas Chromatography. Concepts, Instrumentation, and Applications – Part 2: Comprehensive Two-Dimensional Gas Chromatography," *J. High Resol. Chromatogr.*, **2000**, 23, 167-181.

Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W. "Potential of metabolomics as a functional genomics tool," *Trends Plant Sci.*, **2004**, 9, 418-425.

Blomberg, J.; Schoenmakers, P. J.; Beens, J.; Tijssen, R. "Comprehensive two-dimensional gas chromatography (GCxGC) and its applicability to the characterization of complex (petrochemical) mixtures," *J. High Resol. Chromatogr.*, **1997**, 20, 539-544.

Blumberg, L. M. "Comprehensive two-dimensional gas chromatography: metrics, potentials, limits," *J. Chromatogr. A*, **2003**, 985, 29-38.

Bro, R.; Andersson, C. A.; Kiers, H. A. L. "PARAFAC2-part II. Modeling chromatographic data with retention time shifts," *Chemometrics*, **1999**, 13, 295-309.

Bueno Jr., P. A.; Seeley, J. V. "Flow-switching device for comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, **2004**, 1027, 3-10.

Bull, I. D.; Betancourt, P. P.; Evershed, R. P. "An Organic Geochemical Investigation of the Practice of Manuring at a Minoan Site on Pseira Island, Crete," *Geoarcheology: An International Journal*, **2001**, 16, 223-242.

Bull, I. D.; Lockheart, M. J.; Elhmmali, M. M.; Roberts, D. J.; Evershed, R. P. "The origin of faeces by means of biomarker detection," *Environ. International*, **2002**, 27, 647-654.

Bushey, M. M.; Jorgenson, J. W. "Automated instrumentation for comprehensive two-dimensional high-performance liquid chromatography of proteins," *Anal. Chem.*, **1990**, 62, 161-167.

Bushey, M. M.; Jorgenson, J. W. "Automated instrumentation for comprehensive tow-dimensional high-performance liquid chromatography/capillary zone electrophoresis," *Anal. Chem.*, **1990**, 62, 978-984.

Brauer, M. J.; Saldanha, A. J.; Dolinski, K.; Botstein, D. "Homeostatic Adjustment and Metabolic Remodeling in Glucose-limited Yeast Cultures," *Mol. Biol. Cell*, **2005**, 16, 2503-2517.

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

Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. "Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis," *Anal. Chem.*, **1998**, 70, 2796-2804.

Buchholz, A.; Takors, R.; Wandrey, C. "Quantification of Intracellular Metabolites in Escherichia coli K12 Using Liquid Chromatographic-Electrospray Ionization Tandem Mass Spectrometric Techniques," *Anal. Biochem.*, **2001**, 295, 129-137.

Campion, A.; Kambhampati, P. "Surface-enhanced Raman scattering," *Chem. Soc. Rev.*, **1998**, 27, 241-250.

Carlson, M. "Regulation of glucose utilization in yeast," *Curr. Opin. Genet. Dev.* **1998**, 8, 560-564.

Castle, A. L.; Fiehn, O.; Kaddurah-Daouk, R.; Lindon, G. C. "Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results," *Briefings Bioinform.*, **2006**, 7, 159-165.

Cavagnino, D.; Magni, P.; Zilioli, G.; Trestianu, S. "Comprehensive two-dimensional gas chromatography using large sample volume injection for the determination of

polynuclear aromatic hydrocarbons in complex matrices," *J. Chromatogr. A*, **2003**, 1019, 211-220.

Chan, M. P. L.; Mohd, M. A. "Analysis of endosulfan and its metabolites in rat plasma and selected tissue samples by gas chromatography-mass spectrometry," *Environ. Toxicol.*, **2005**, 20, 45-52.

Chapman Jr., G. W.; Horvat, R. J. "Determination of Nonvolatile Acids and Sugars from Fruits and Sweet Potato Extracts by Capillary GLC and GLC/MS," *J. Agric. Food Chem.*, **1989**, 37, 947-950.

Chen, S.; Lee, M. L. "Automated instrumentation for comprehensive isotachophoresis-capillary zone electrophoresis," *Anal. Chem.*, **2000**, 72, 816-820.

Colquhoun, I. J.; Gall, G. L.; Elliott, K. A.; Mellon, F. A.; Michael, A. J. "Shall I compare thee to a GM potato?" *Trends Gen.*, **2006**, 22, 525-528.

Culha, M.; Stokes, D.; Allain, L. R.; Vo-Dinh, T. "Surface-Enhanced Raman Scattering Substrate Based on a Self-Assembled Monolayer for use in Gene Diagnostics," *Anal. Chem.*, **2003**, 75, 6196-6201.

Dallüge, J.; Vreuls, R. J. J.; Beens, J.; Brinkman, U. A. Th. "Optimization and characterization of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection (GC x GC-TOF MS)," *J. Sep. Sci.*, **2002**, 25, 201-214.

Dallüge, J.; van Rijn, M.; Beens, J.; Vreuls, R. J. J.; Beens, J.; Brinkman, U. A. Th. "Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection applied to the determination of pesticides in food extracts," *J. Chromatogr. A*, **2002**, 965, 207-217.

Dallüge, J.; van Stee, L. L. P.; Xu, X.; Williams, J.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. Th. "Unravelling the composition of very complex samples by comprehensive gas chromatography coupled to time-of-flight mass spectrometry Cigarette smoke," *J. Chromatogr. A*, **2002**, 974, 169-184.

Damian, D.; Orešič, M.; Verheij, E.; Meulman, J.; Friedman, J.; Adourian, A.; Morel, N.; Smilde, A.; van der Greef, J. "Applications of a new subspace clustering algorithm (COSA) in medical systems biology," *Metabolomics*, **2007**, 3, 69-77.

Danielsson, R.; Baeckstroem, D.; Ullsten, S. "Rapid multivariate analysis of LC/GC/CE data (single or multiple channel detection) without prior peak alignment," *Chemom. Intell. Lab. Sys.*, **2006**, 84, 33-39.

Drozd, J. *chemical derivatization in gas chromatography*. Amsterdam: Elsevier Scientific Publishing Company, **1981**.

de Koning, S.; Janssen, H.-G.; Brinkman, U. A. Th. "Group-type characterization of mineral oil samples by two-dimensional comprehensive normal-phase liquid chromatography-gas chromatography with time-of-flight mass spectrometric detection," *J. Chromatogr. A*, **2004**, 1058, 217-221.

de Koning, S.; Janssen, H.-G.; van Deursen, M.; Brinkman, U. A. Th. "Automated on-line comprehensive two-dimensional LC x GC and LC x GC-TOFMS: Instrument design and application to edible oil and fat analysis," *J. Sep. Sci.*, **2004**, 27, 397-409.

Denkert, C.; Budczies, J.; Kind, T.; Weichert, W.; Tablack, P.; Sehouli, J.; Niesporek, S.; Könsgen, D.; Dietel, M.; Fiehn, O. "Mass Spectrometry-Based Metabolic Profiling Reveals Different Metabolite Patterns in Invasive Ovarian Carcinomas and Ovarian Borderline Tumers," *Cancer Res.*, **2006**, 66, 10795-10804.

DeGraff, B. A.; Hennip, M.; Jones, J. M.; Salter, C.; Schaertel, S. A. "An Inexpensive Laser Raman Spectrometer Based on CCD Detection," *Chem. Educator*, **2002**, 7, 15-18.

Di, X.; Shellie, R. A.; Marriott, P. J.; Huie, C. W. "Application of headspace solid-phase, microextraction (HS-SPME) and comprehensive two-dimensional gas chromatography (GCxGC) for the chemical profiling of volatile oils in complex herbal mixtures," *J. Sep. Sci.*, **2004**, 27, 451-458.

Dimandja, J.-M. D.; Grainger, J.; Jr. D. G. P.; Turner, W. E.; Needham, L. L. "Measurement for assessing environmental exposures to children using serum and urine: state-of-the-art,"*J. Exposure Analysis and Environmental Epidemiology*, **2000**, 10, 761-768.

Dimandja, J.-M. D.; Clouden, G. C.; Colón, I.; Focant, J.-F.; Cabey, W. V.; Parry, R. C. "Standardized test mixture for the characterization of comprehensive two-dimensional gas chromatography columns: the Phillips mix," *J. Chromatogr. A*, **2003**, 1019, 261-272.

Dorman, F.; Wittrig, R. "Introducing Stop-Flow GC, for High-Speed/High-Resolution GC Analysis," Restek notes, **2004**.

Doyle III, F. J.; Stelling J. "Systems interface biology," *J. R. Soc. Interface*, **2006**, 3, 603-616.

Dugo, P.; Skerikova, V.; Kumm, T.; Trozzi, A.; Jandera, P.; Mondello, L. "Elucidation of Carotenoid Patterns in Citrus Products by Means of Comprehensive Normal-Phase x Reversed-Phase Liquid Chromatography," *Anal. Chem.*, **2006**, 78, 7743-7750.

Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. "Measuring the metabolome: current analytical technologies," *Analyst*, **2005**, 130, 606-625.

Elbein, A. D.; Pan, Y. T.; Pastuszak, I.; Carroll, D. "New insights on trehalose: a multifunctional molecule," *Glycobiology*, **2003**, 13, 17R-27R.

Erni, F.; Frei, R. W. "Two-dimensional column liquid chromatographic technique for resolution of complex mixture," *J. Chromatogr.*, **1978**, 149, 561-569.

Evans, C. R.; Jorgenson, J. W. "Multidimensional LC-LC and LC-CE for high-resolution separations of biological molecules," *Anal. Bioanal. Chem.*, **2004**, 378, 1952-1961.

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

Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. " Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Qaudrupole Mass Spectrometry," *Anal. Chem.*, **2000**, 72, 3573-3580.

Fiehn, O. "Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks," *Comp. Funct. Genom.*, **2001**, 2, 155-168.

Fiehn, O.; Spranger, J. "Can metabolomics be used for assessing nutritive-dependent human diseases?" *Phytochem. Rev.*, **2002**, 1, 223-230.

Fiehn, O. "Metabolic networks of Cucurbita maxima phloem," *Phytochemistry*, **2003**, 62, 875-886.

Focant, J.-F.; Sjödin, A.; Patterson Jr., D. G. "Qualitative evaluation of thermal desorption-programmable temperature vaporization-comprehensive two-dimensional gas chromatography—time-of-flight mass spectrometry for the analysis of selected halogenated contaminants," *J. Chromatogr. A*, **2003**, 1019, 143-156.

Focant, J.-F.; Sjödin, A.; Turner, W. E.; Patterson Jr., D. G. "Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC x

GC-TOFMS) for drug screening and confirmation," *Anal. Chem.*, **2004**, 76, 6313-6320.

Förster, J.; Famili, I.; Fu, P.; Palsson, B. Ø.; Nielsen, J. "Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network," *Genome Res.*, **2003**, 13, 244-253.

Fraga, C. G.; Prazen, B. J.; Synovec, R. E. "Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the High-Speed Quantitative Analysis of Aromatic Isomers in a Jet Fuel Using the Standard Addition Method and an Objective Retention Time Alignment Algorithm," *Anal. Chem.*, **2000**, 72, 4154-4162.

Fraga, C. G.; Prazen, B. J.; Synovec, R. E. "Enhancing the Limit of Detection for Comprehensive Two-Dimensional Gas Chromatography (GCxGC) using Bilinear Chemometric Analysis," *J. High Resol. Chromatogr.*, **2000**, 23, 215-224.

Fraga, C. G.; Bruckner, C. A.; Synovec, R. E. "Increasing the Number of Analyzable Peaks in Comprehensive Two-Dimensional Separations through Chemometrics," *Anal. Chem.*, **2001**, 73-675-683.

Fraga, C. G.; Corley, C. A. "The chemometric resolution and quantification of overlapped peaks form comprehensive two-dimensional liquid chromatography," *J. Chromatogr. A*, **2005**, 40-49.

Frysinger, G. S.; Gaines, R. B.; Xu, L.; Reddy, C. M. "Resolving the Unresolved Complex Mixture in Petroleum-Contaminated Sediments," *Environ. Sci. Technol.*, **2003**, 37, 1653-1662.

Funcke, W.; von Sonntag, C. "Syn and anti forms of some monosaccharide O-methyl oximes: a 13C-n.m.r and g.l.c. study," *Carbohydr. Res.*, **1979**, 69, 247-251.

Gancedo, C.; Flores, C-L. "The importance of a functional trehalose biosynthetic pathway for the life of yeasts and fungi," *FEMS Yeast Research*, **2004**, 4, 351-359.

Garrell, R. L.; Shaw, K. D.; Krimm, S. "Surface Enhanced Raman Spectroscopy of Halide Ions on Colloidal Silver: Morphology and Coverage Dependence," *Sur. Sci.*, **1983**, 124, 613-624.

Garrell, R. L. "Surface-Enhanced Raman Spectroscopy," *Anal. Chem.*, **1989**, 61, 401A-411A.

Giddings, J. C. "Two-dimensional separations: concepts and promise," *Anal. Chem.*, **1984**, 6, 1258A-1270A.

Giddings, J. C. *Unified Separation Science*. New York: John Wiley & Sons, Inc. **1991**.

Giddings, J. C. "Sample dimensionality: a predictor of order-disorder in component peak distribution in multidimensional separation," *J. Chromatogr. A*, **1995**, 703, 3-15.

Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. "Metabolomics by numbers: acquiring and understanding global metabolite data," *Trends Biotechnol.*, **2004**, 22, 245-252.

Goodacre, R. "Metabolomics-the way forward," *Metabolomics*, **2005**, 1, 1-2.

Górecki, T.; Harynuk, J.; Panić, O. "The evolution of comprehensive two-dimensional gas chromatography (GCxGC)," *J. Sep. Sci.*, **2004**, 27, 359-379.

Griffin, J. L.; Shockcor, J. P. "Metabolic Profiles of Cancer Cells," *Nature Rev.*, **2004**, 4, 551-561.

Griffin, J. L. "Understanding mouse models of disease through metabolomics," *Curr. Opin. Chem. Biol.*, **2006**, 10, 309-315.

Gross, G. M.; Prazen, B. J.; Grate, J. W.; Synovec, R. E. "High-Speed Gas Chromatography Using Synchronized Dual-Valve Injection," *Anal. Chem.* **2004**, 76, 3517-3524.

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

Guttman, A.; Varoglu, M.; Khandurina, J. "Multidimensional separations in the pharamaceutical arena," *Drug Discovery Today*, **2004**, 9, 136-144.

Hao, C.; Headley, J. V.; Peru, K. M.; Frank, R.; Yang, P.; Solomon, K. R. "Characterization and pattern recognition of oil-sand naphthenic acids using comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry," *J. Chromatogr. A*, **2005**, 1067, 277-284.

Hall, R.; Beale, M.; Fiehn, O.; Hardy, N.; Sumner, L.; Bino, R. "Plant Metabolomics: The Missing Link in Functional Genomics Strategies," *The Plant Cell*, **2002**, 14, 1437-1440.

Harrigan, G. G.; LaPlante, R. H.; Cosma, G. N.; Cockerell, G.; Goodacre, R.; Maddox, J. F.; Luyendyk, J. P.; Ganey, P. E.; Roth, R. A. "Application of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity," *Toxicol. Lett.*, **2004**, 146, 197-205.

Harris, D. C. *Quantitative Chemical Analysis*, 6th. edition ed. New York: W. H. Freeman and Company. **2003**.

Harvey, D. J.; Horning, M. G. "Derivatives for the characterization of alkyl and aminoalkylphosphonates by gas chromatography and gas chromatography mass spectrometry," *J. Chromatogr.*, **1973**, 76, 51-62.

Harwood, M. M.; Christains, E. S.; Abul Fazal, Md.; Dovichi, N. J. "Single-cell protein analysis of a single mouse embryo by two-dimensional capillary electrophoresis," *J. Chromatogr. A*, **2006**, 1130, 190-194.

Harynuk, J.; Górecki, T. "Comprehensive two-dimensional gas chromatography in stop-flow mode," *J. Sep. Sci.* **2004**, 27, 431-441.

Haswell, S. J. *Practical Guide to Chemometrics*. New York: Marcel Dekker, Inc. **1992**.

Haubitz, M.; Wittke, S.; Weissinger, E. M.; Walden, M.; Rupprecht, H. D.; Floege, J.; Haller, H.; Mischak, H. "Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy," *Kidney Int.*, **2005**, 67, 2313-2320.

Hinshaw, J. V. "Measuring Column Flow and Velocity," *LC·GC Europe*, **2003**, 16, 82, 84-85.

Hirata, Y.; Hashiguchi, T.; Kawata, E. "Development of comprehensive two-dimensional packed column supercritical fluid chromatography," *J. Sep. Sci.*, **2003**, 26, 531-535.

Hoggard, J. C.; Synovec, R. E. "Parallel factor analysis (PARAFAC) of target analytes in GC x GC-TOFMS data: Automated selection of a model with an appropriate number of factors," *Anal. Chem.*, **2007**, 79, 1611-1619.

Hollywood, K.; Brison, D. R.; Goodacre, R. "Metabolomics: Current technologies and future trends," *Proteomics*, **2006**, 4716-4723.

Holm, T. "Aspects of the mechanism of the flame ionization detector," *J. Chromatogr. A*, **1999**, 842, 221-227.

Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Lidstrom, M. E.; Synovec, R. E. "Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection: analysis of amino acid and organic acid trimethylsilyl derivatives, with application to the analysis of metabolites in rye grass samples," *Talanta*, **2005**, 65, 380-388.

Hope, J.; Sinha, A.; Prazen, B.; Synovec, R. "Evaluation of the DotMap algorithm for locating analytes of interest based on mass spectral similarity in data collected using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry," *J. Chromatogr. A*, **2005**, 1086, 185-192.

Horning, E. C.; Horning, M. G. "Human metabolic profiles obtained by GC [gas chromatography] and GC/MS [gas chromatography/mass spectrometry]," *J. Chromatogr. Sci.*, **1971**, 9, 129-140.

Howe, T. J.; Mahieu, G.; Marichal, P.; Tabruyn, T.; Vugts, P. "Data reduction and representation in drug discovery," *Drug Discovery Today*, **2007**, 12, 45-53.

Hu, L.; Chen, X.; Kong, L.; Su, X.; Ye, M.; Zou, H. "Improved performance of comprehensive two-dimensional HPLC separation of traditional Chinese medicines by using a silica monolithic column and normalization of peak heights," *J. Chromatogr. A*, **2005**, 1092, 191-198.

Hu, S.; Michels, D. A.; Fazal, Md. A.; Ratisoontorn, C.; Cunningham, M. L.; Dovichi, N. J. "Capillary Sieving Electrophoresis/Micellar Electrokinetic Capillary Chromatography for Two-Dimensional Protein Fingerprinting of Single Mammalian Cells," *Anal. Chem.*, **2004**, 76, 4044-4049.

Jain, N.; Thatte, J.; Braceale, T.; Ley, K.; O'Connell, M.; Lee, J. K. "Local-pooled-error test for identifying differentially expressed genes with small number of replicated microarrays," *Bioinformatics*, **2003**, 19, 1945-1951.

Jang, S.; Kim, S.; Shin, S.; Joo, S.-W. "Stucture of 4-biphenylthiolate on Au nanoparticle surfaces studied by UV-Vis absorption spectroscopy, transmission electron microscopy and surface-enhanced Raman scattering," *Surf. Interface Anal.*, **2004**, 36, 43-48.

Janssen, H.-G.; Boers, W.; Steenbergen, H.; Horsten, R.; Flöter, E. "Comprehensive two-dimensional liquid chromatography x gas chromatography: evaluation of the applicability for the analysis of edible oils and fats," *J. Chromatogr. A*, **2003**, 1000, 385-400.

Johnson, K. J.; Prazen, B. J.; Olund, R. K.; Synovec, R. E. "GC x GC temperature programming requirements to produce bilinear data for chemometric analysis." *J. Sep. Sci.*, **2002**, 25, 297-303.

Johnson, K. J.; Synovec, R. E. "Pattern recognition of jet fuels: comprehensive GC x GC with ANOVA-based feature selection and principal component analysis," *Chemom. Intell. Lab. Syst.*, **2002**, 60, 225-237.

Johnson, K. J.; Wright, B. W.; Jarman, K. H.; Synovec, R. E. " High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analyis," *J. Chromatogr. A*, **2003**, 996, 141-155.

Johnson, K. J.; Prazen, B. J.; Young, D. C.; Synovec, R. E. "Quantification of naphthalenes in jet fuel with GC x GC/Tri-PLS and windowed rank minimization retention time alignment," *J. Sep. Sci.*, **2004**, 27, 410-416.

Jonsson, P.; Gullberg, J.; Nordström, A.; Kusano, M.; Kowalczyk, M.; Sjöström, M.; Moritz, T. "A Strategy for Identifying Differences in Large Series of Metabolomic Samples Analyzed by GC/MS," *Anal. Chem.* **2004**, 76, 1738-1745.

Jonsson, P.; Bruce, S. J.; Moritz, T.; Trygg, J.; Sjöström, M.; Plumb, R.; Granger, J.; Maibaum, E.; Nicholson, J. K.; Holmes, E.; Antti, H. "Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets," *Analyst*, **2005**, 130, 701-707.

Jonsson, P.; Johansson, A. I.; Gullberg, J.; Trygg, J.; A, J.; Grung, B.; Marklund, S.; Sjöström, M.; Antti, H.; Mortiz, T. "High-Throughput Data Analysis for Detecting and Identifying Differences between Samples in GC/MS-Based Metabolomic Analysis," *Anal. Chem.*, **2005**, 77, 5635-5642.

Jover, E.; Adahchour, M.; Bayona, J. M.; Vreuls, R. J. J.; Brinkman, U. A. Th. "Characterization of lipids in complex samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry," *J. Chromatogr. A*, **2005**, 1086, 2-11.

Kallio, M.; Hyötyläinen, T.; Jussila, M.; Hartonen, K.; Palonen, S.; Shimmo, M.; Riekkola, M.-L. "Semi-rotating cryogenic modulator for comprehensive two-dimensional gas chromatography," *Anal. Bioanal. Chem.*, **2003**, 375, 725-731.

Kallio, M.; Hyötyläinen, T.; Lehtonen, M.; Jussila, M.; Hartonen, K.; Shimmo, M.; Riekkola, M.-L. "Comprehensive two-dimensional gas chromatography in the analysis of urban aerosols," *J. Chromatogr. A*, **2003**, 1019, 251-260.

Kell, D. B.; Oliver, S. G. "Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era," *BioEssays*, **2003**, 26, 99-105.

Kinghorn, R. M.; Marriott, P. J. "Comprehensive Two-Dimensional Gas Chromatography using a Modulating Cryogenic Trap," *J. High Resol. Chromatogr.*, **1998**, 21, 620-622.

Klevecz, R. R.; Bolen, J.; Forrest, G.; Murray, D. B. "A genomewide oscillation in transcription gates DNA replication and cell cycle," *Proc. Natl. Acad. Sci. U.S.A.*, **2004**, 101, 1200-1205.

Kowalski, B. R. "Chemometrics. Views and propositions." *J. Chem. Inf. Comp. Sci.*, **1975**, 15, 201-203.

Kneipp, K.; Wang, Y.; Kneipp, H.; Itzkan, I.; Dasari, R. R.; Feld, M. S. "Population Pumping of Excited Vibrational States by Spontaneous Surfaced-Enhanced Raman Scattering," *Phys. Rev. Lett.*, **1996**, 76, 2444-2447.

Kraly, J. R.; Jones, M. R.; Gomez, D. G.; Dickerson, J. A.; Harwood, M. M.; Eggertson, M.; Paulson, T. G.; Sanchez, C. A.; Odze, R.; Feng, Z.; Reid, B. J.; Dovichi, N. J. "Reproducible Two-Dimensional Capillary Electrophoresis Analysis of Barrett's Esophagus Tissues," *Anal. Chem.*, **2006**, 78, 5977-5986.

Kramer, R. *Chemometric Techniques for Quantitative Analysis*. New York: Marcel Dekker. **1998**.

Krebs, M. D.; Kang, J.-M.; Cohen, S. J.; Lozow, J. B.; Tingley, R. D.; Davis, C. E. "Two-dimensional alignment of differential mobility spectrometer data," *Sensors and Actuators B*, **2006**, 119, 475-482.

Kresnowati, M.T.A.P.; van Winden, W. A.; Almering, M. J. H.; ten Pierick, A.; Ras, C.; Knijnenburg, T. A.; Daran-Lapujade, P.; Pronk, J. T.; Heijnen, J. J.; Daran, J. M. "When transcriptome meets metabolome: fast cellular response of yeast to sudden relief of glucose limitation," *Mol. Sys. Biol.*, **2006**, 2, 49-65.

Kundu, S.; Mandal, M.; Ghosh, S. K.; Pal, T. "Photochemical deposition of SERS active silver nanoparticles on silica gel and their application as catalysts fro the reduction of aromatic nitro compounds," *J. Colloid Interface Sci.*, **2004**, 272, 134-144.

Lay Jr. J. O.; Borgmann, S.; Liyanage, R.; Wilkins, C. L. "Problems with the "omics"," *Trends Anal. Chem.*, **2006**, 25, 1046-1056.

Lee, A. L.; Bartle, K. D.; Lewis, A. C. "A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography," *Anal. Chem.*, **2001**, 73, 1330-1335.

Liu, Z.; Phillips, J. B. "Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface," *J. Chromatogr. Sci.*, **1991**, 29, 227-231.

Liu, Z.; Ostrovsky, I.; Farnsworth, P. B.; Lee, M. L. "Instrumentation for comprehensive two-dimensional capillary supercritical fluid-gas chromatography," *Chromatographia*, **1993**, 35, 567-573.

Liu, Z.; Sirlmanne, S. R.; Patterson Jr., D. G.; Needham, L. L. "Comprehensive Two-Dimensional Gas Chromatography for the Fast Separation and Determination of Pesticides Extracted from Human Serum," *Anal. Chem.*, **1994**, 66, 3086-3092.

Lu, P.; Rangan, A.; Chan, S. Y.; Appling, D. R.; Hoffman, D. W.; Marcotte, E. M. "Global metabolic changes following loss of a feedback loop reveal dynamic steady states of the yeast metabolome," *Metabolic Eng.*, **2007**, 9, 8-20.

Lu, X.; Cai, J.; Kong, H.; Wu, M.; Hua, R.; Zhao, M.; Liu, J.; Xu, G. "Analysis of Cigarette Smoke Condensates by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry I Acidic Fraction," *Anal. Chem.*, **2003**, 75, 4441-4451.

Mahoney, N.; Molyneux, R. J.; Smith, L. R.; Schoch, T. K.; Rolshausen, P. E.; Gubler, W. D. "Dying-Arm Disease in Grapevines: Diagnosis of Infection with Eutypa lata by Metabolite Analysis," *J. Agric. Food Chem.*, **2005**, 53, 8148-8155.

Marriott, P. J.; Kinghorn, R. M.; Ong, R.; Morrison, P.; Haglund, P.; Harju, M. "Comparison of Thermal Sweeper and Cyrogenic Modulator Technology for Comprehensive Gas Chromatography," *J. High Resol. Chromatogr.*, **2000**, 23, 253-258.

McLafferty, F. W.; Tureček, F. *Interpretation of Mass Spectra*, 4th ed. California: University Science Books. **1993**.

Michaels, A. M.; Nirmal, M.; Brus, L. E. "Surface Enhanced Raman Spectroscopy of Individual Rhodamine 6G Molecules on Large Ag Nanocrystals," *J. Am. Chem. Soc.*, **1999**, 121, 9932-9939.

Michels, D. A.; Hu, S.; Dambrowitz, K. A.; Eggertson, M. J.; Lauterbach, K.; Dovichi, N. J. "Capillary sieving electrophoresis-micellar electrokinetic chromatography fully automated two-dimensional capillary electrophoresis analysis of

Deinococcus radiodurans protein homogenate," *Electrophoresis*, **2004**, 25, 3098-3105.

Micyus, N. J.; Seeley, S. K.; Seeley, J. V. "Method for reducing the ambiguity of comprehensive two-dimensional chromatography retention times," *J. Chromatogr. A*, **2005**, 1086, 171-174.

Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. "Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells," *Anal. Chem.*, **2006**, 78, 2700-2709.

Mohler, R. E.; Prazen, B. J.; Synovec, R. E. "Total-Transfer, valve-based comprehensive two-dimensional gas chromatography," *Anal. Chim. Acta*, **2006**, 555, 68-74.

Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Pierce, K. M.; Young, E. T.; Synovec, R. E. "Comprehensive analysis of yeast metabolite GC x GC-TOFMS data: combining discovery-mode and deconvolution chemometric software," *Analyst*, **2007**, 132, 756-767.

Mohler, R. E.; Tu, B. P.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. "Identification and Evaluation of Cycling Yeast Metabolites in GC x GC-TOFMS Data," submitted *J. Chromatogr. A*, **2007**.

Moldoveanu, S. C.; David, V. *sample preparation in chromatography*. Amsterdam: Elsevier. **2002**.

Mondello, L.; Casilli, A.; Tranchida, P. Q.; Dugo, P.; Dugo, G. "Detailed analysis and group-type separation of natural fats and oils using comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, **2003**, 1019, 187-196.

Moody, R. L.; Vo-Dinh, T.; Fletcher, W.H. "Investigation of Experimental Parameters for Surface-Enhanced Raman Scattering (SERS) Using Silver-Coated Microsphere Substrates," *Appl. Spectrosc.*, **1987**, 41, 966-970.

Morales-Muňoz, S.; Vreuls, R. J. J.; Luque de Castro, M. D. "Dynamic ultrasound-assisted extraction of environmental pollutants from marine sediments for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection," *J. Chromatogr. A*, **2005**, 1086, 122-127.

Mourougou-Candoni, N.; Naud, C.; Thibaudau, F. "Adsorption of Thiolated Oligonucleotides on Gold Surfaces: An Atomic Force Microscopy Study," *Langmuir*, **2003**, 19, 682-686.

Mulvaney, S. P.; Keating, C. D. "Raman Spectroscopy," *Anal. Chem.*, **2000**, 72, 145R-157R.

Murphy, R.; Schure, M.; Foley, J. "One-and Two-Dimensional Chromatographic Analysis of Alcohol Ethoxylates," *Anal. Chem.*, **1998**, 70, 1585-1594.

Murray, D. B.; Beckmann, M.; Kitano, H. "Regulation of yeast oscillatory dynamics," *Proc. Natl. Acad. Sci. U.S.A.*, **2007**, 104, 2241-2246.

Nicholson, J. K.; Lindon, J. C.; Holmes, E. " "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, **1999**, 29, 1181-1189.

Nie, S.; Emory, S. R. "Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering," *Science*, **1997**, 275, 1102-1106.

Nielsen, J.; Oliver, S. "The next wave in metabolome analysis," *Trends Biotechnol.*, **2005**, 23, 544-546.

Nikiforova, V. J.; Kopka, J.; Tolstikov, V.; Fiehn, O.; Hopkins, L.; Hawkesford, M. J.; Hesse, H.; Hoefgen, R. "Systems Rebalancing of Metabolism in Response to Sulfur Deprivation, as Revealed by Metabolome Analysis of Arabidopsis Plants," *Plant Physiol.*, **2005**, 138, 304-318.

Norrod, K. L.; Sudnik, L. M.; Rousell, D.; Rowlen, K. L. "Quantitative Comparison of Five SERS Substrates: Sensitivity and Limit of Detection," *Appl. Spectrosc.*, **1997**, 51, 994-1001.

O'Hagan, S.; Dunn, W. B.; Knowles, J. D.; Broadhurst, D.; Williams, R.; Ashworth, J. J.; Cameron, M.; Kell, D. B. "Closed-Loop, Multiobjective Optimization of Two-Dimensional Gas Chromatography/Mass Spectrometry for Serum Metabolomics," *Anal. Chem.*, **2007**, 79, 464-476.

Oliver S. "Guilt-by-association goes global," *Nature*, **2000**, 403, 601-603.

Oliver, S. G.; Winson, M. K. "Systematic functional analysis of the yeast genome," *Trends Biotechnol.*, **1998**, 16, 373-378.

Ong, R.; Shellie, R.; Marriott, P. "Observation of non-linear chromatographic peaks in comprehensive two-dimensional gas chromatography," *J. Sep. Sci.*, **2001**, 24, 367-377.

Ong, R. C. Y.; Marriott, P. J. "A Review of Basic Concepts in Comprehensive Two-Dimensional Gas Chromatography," *J. Chromatogr. Sci.*, **2002**, 40, 276-291.

Panic, O.; Górecki, T. "Comprehensive two-dimensional gas chromatography (GC x GC) in environmental analysis and monitoring," *Anal. Bioanal. Chem.*, **2006**, 386, 1013-1023.

Patil, K. R.; Nielsen, J. "Uncovering transcriptional regulation of metabolism by using metabolic network topology," *Proc. Natl. Acad. Sci. U. S. A.*, **2005**, 102, 2685-2689.

Peters, S.; Vivó-Truyols, G.; Marriott, P. J.; Schoenmakers, P. J. "Development of a resolution metric for comprehensive two-dimensional chromatography," *J. Chromatogr. A*, **2007**, 1146, 232-241.

Phillips, J. B.; Ledford, E. B. "Thermal modulation: a chemical instrumentation component of potential value in improving portability," *Field Anal. Chem. Tech*, **1996**, 1, 23-29.

Phillips, J. B.; Beens, J. "Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions," *J. Chromatogr. A*, **1999**, 856, 331-347.

Pichersky, E.; Gang, D. R. "Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective," *Trends Plant Sci.*, **2000**, 5, 439-445.

Pierce, K. M.; Wood, L. F.; Wright, B. W.; Synovec, R. E. "A Comprehensive Two-Dimensional Retention Time Alignment Algorithm to Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data," *Anal. Chem.*, **2005**, 77, 7735-7743.

Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. "Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts," *Anal. Chem.*, **2006**, 78, 5068-5075.

Pisano, J. J.; Bronzert, T. J. "Analysis of Amino Acid Phenylthiohydantoins by Gas Chromatography," *J. Biol. Chem.*, **1969**, 244, 5597-5607.

Pohjanen, E.; Thysell, E.; Lindberg, J.; Schuppe-Koistinen, I.; Moritz, T.; Jonsson, P.; Antti, H. "Statistical multivariate metabolite profiling for aiding biomarker pattern detection and mechanistic interpretations in GC/MS based metabolomics," *Metabolomics*, **2006**, 2, 257-267.

Pope, S. A. S.; Clayton, P. T.; Muller, D. P. R. "A New Method for the Analysis of Urinary Vitamin E Metabolites and the Tentative Identification of a Novel Group of Compounds," *Arch. Biochem. Biophys.*, **2000**, 381, 8-15.

Porter, S. E. G.; Stoll, D. R.; Rutan, S. C.; Carr, P. W.; Cohen, J. D. "Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics," *Anal. Chem.*, **2006**, 78, 5559-5569.

Prazen, B. J.; Johnson, K. J.; Weber, A.; Synovec, R. E. "Two-Dimensional Gas Chromatography and Trilinear Partial Least Squares for the Quantitative Analysis of Aromatic and Naphthene Content in Naphtha," *Anal. Chem.*, **2001**, 73, 5677-5682.

Quigley, W. W. C.; Fraga, C. G.; Synovec, R. E. "Comprehensive LC x GC for enhanced headspace analysis," *J. Microcolumn Sep.*, **2000**, 12, 160-166.

Raamsdonk, L. M.; Diderich, J. A.; Kuiper, A.; van Gaalen, M.; Kruckberg, A. L.; Berden, J. A., van Dam, K. "Co-consumption of sugars or ethanol and glucose in a Saccharomyces cerevisiae strain deleted in the HXK2 gene," *Yeast*, **2001**, 18, 1023-1033.

Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; van Dam, K.; Oliver, S. G. "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations," *Nature Biotechnol.*, **2001**, 19, 45-50.

Ramos, L. S.; Burger, J. E.; Kowalski, B. R. "Third-Order Chromatography: Multivariate Analysis of Data from Parallel-Column Chromatography with Multichannel Detection," *Anal. Chem.*, **1985**, 57, 2620-2625.

Reichenbach, S. E.; Ni, M.; Kottapalli, V.; Visvanathan, A.; Ledford Jr., E. B.; Oostdijk, J.; Trap, H. C. "Chemical Warfare Agent Detection in Complex Environments with Comprehensive Two-Dimensional Gas Chromatography,"*Proc. of SPIE*, **2003**, 5085, 28-30.

Reinke, H.; Gatfield, D. "Genome-wide oscillation of transcription in yeast," *Trends Biochem. Sci.*, **2006**, 31, 189-191.

Robards, K.; Haddad, P. R.; Jackson, P. E. *Principles and Practice of Modern Chromatographic Methods*. San Diego: Academic Press. **1994**.

Ronen, M.; Botstein, D. "Transciptional response of steady-state yeast cultures to transient perturbations in carbon source," *Proc. Natl. Acad. Sci. U. S. A.*, **2006**, 103, 389-394.

Rorabacher, D. B. "Statistical Treatment for Rejection of Deviant Values: Critical Values of Dixon's "Q" Parameter and Related Subrange Ratios at the 95% Confidence Level," *Anal. Chem.*, **1991**, 63, 139-146.

Sahota, R. S.; Morgan, S. L. "Recognition of Chemical Markers in Chromatographic Data by an Individual Feature Reliability Approach to Classification," *Anal. Chem.*, **1992**, 64, 2383-2392.

Sanz, M. L.; Sanz, J.; Martínez-Castro, I. "Disaccharides by Gas Chromatography-Mass Spectrometry," *Chromatographia*, **2002**, 56, 617-622.

Sanz, M. L.; Diez-Barrio, M. T.; Sanz, J.; Martínez-Castro, I. "GC Behavior of Disaccharide Trimethylsilyl Oximes," *J. Chromatogr. Sci.*, **2003**, 41, 205-208.

Saragenti, S. R.; Vichnewski, W. "Sonication and Liquid Chromatography as a Rapid Technique for Extraction and Fractionation of Plant Material," *Phytochem. Anal.*, **2000**, 11, 69-73.

Seeley, J. V.; Kramp, F.; Hicks, C. J. "Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation," *Anal. Chem.*, **2000**, 72, 4346-4352.

Seeley, J. V. " Theoretical study of incomplete sampling of the first dimension in comprehensive two-dimensional chromatography," *J. Chromatogr. A*, **2002**, 962, 21-27.

Semin, D. J.; Rowlen, K. L. "Influence of Vapor Deposition Parameters on SERS Active Ag Film Morphology and Optical Properties," *Anal. Chem.*, **1994**, 66, 4324-4331.

Semin, D. J.; Lo, A.; Roark, S. E.; Skodje, R. T.; Rowlen, K. L. " Time-dependent morphology changes in thin silver films on mica: A scaling analysis of atomic force microscopy results," *J. Chem. Phys.*, **1996**, 105, 5542-5551.

Shellie, R. Mondello, L.; Marriott, P.; Dugo, G. "Characterisation of lavender essential oils by using gas chromatography—mass spectrometry with correlation of linear retention indices and comparison with comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, **2002**, 970, 225-234.

Shellie, R.; Marriott, P.; Leus, M.; Dufour, J.-P.; Mondello, L.; Dugo, G.; Sun, K.; Winniford, B.; Griffith, J.; Luong, J. "Retention time reproducibility in

comprehensive two-dimensional gas chromatography using cryogenic modulation II. An interlaboratory study," *J. Chromatogr. A*, **2003**, 1019, 273-278.

Shellie, R.; Marriott, P.; Morrison, P. "Comprehensive Two-Dimensional Gas Chromatography with Flame Ionization and Time-of-Flight Mass Spectrometry Detection: Qualitative and Quantitative Analysis of West Australian Sandalwood Oil," *J. Chromatogr. Sci.*, **2004**, 42, 417-422.

Shellie, R.; Marriott, P.; Morrison, P.; Mondello, L. "Effects of pressure drop on absolute retention matching in comprehensive two-dimensional gas chromatography," *J. Sep. Sci.*, **2004**, 27, 504-512.

Shellie, R. A. "Comprehensive Two-Dimensional Gas Chromatograph--Mass Spectrometry and its Use in High-Resolution Metabolomics," *Aust. J. Chem.*, **2005**, 58, 619.

Shellie, R. A.; Welthagen, W.; Zrostliková, J.; Spranger, J.; Ristow, M.; Fiehn, O.; Zimmermann, R. "Statistical methods for comparing comprehensive two-dimensional gas chromatography—time-of-flight mass spectrometry results: Metabolomic analysis of mouse tissue extracts," *J. Chromatogr. A*, **2005**, 1086, 83-90.

Simpson, I. A.; van Bergen, P. F.; Perret, V.; Elhmmali, M. M.; Roberts, D. J.; Evershed, R. P. "Lipid biomarkers of manuring practice in relict anthropogenic soils," *Holocene*, **1999**, 9, 223-229.

Sinha, A. E.; Johnson, K. J.; Prazen, B. J.; Lucas, S. V.; Fraga, C. G.; Synovec, R. E. "Comprehensive two-dimensional gas chromatography of volatile and semi-volatile components using a diaphragm valve-based instrument," *J. Chromatogr. A*, **2003**,983, 195-204.

Sinha, A. E.; Prazen, B. J.; Fraga, C. G.; Synovec, R. E. "Valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection: instrumentation and figures-of-merit," *J. Chromatogr. A*, **2003**, 1019, 79-87.

Sinha, A. E.; Fraga, C. G.; Prazen, B. J.; Synovec, R. E. "Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography--time-of-flight mass spectrometry data," *J. Chromatogr. A* **2004**, 1027, 269-277.

Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. "Multivariate selectivity as a metric for evaluating comprehensive two-dimensional gas chromatography--time-of-flight mass spectrometry subjected to chemometric peak deconvolution," *J. Chromatogr. A*, **2004**, 1056, 145-154.

Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Jack, R. M.; Synovec, R. E. "Algorithm for locating analytes of interest based on mass spectral similarity in GC x GC--TOF-MS data: analysis of metabolites in human infant urine," *J. Chromatogr. A*, **2004**, 1058, 209-215.

Sinha, A. E.; Prazen, B. J.; Synovec, R. E. " Trends in chemometric analysis of comprehensive two-dimensional separations," *Anal. Bioanal. Chem.*, **2004**, 378, 1948-1951.

Skoog D. A.; Holler, F. J.; Nieman, T. A. *Principles of Instrumental Analysis*. 5th ed. Philadelphia: Harcourt Brace College Publishers. **1998**.

Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. "Simultaneous Determination of Anionic Intermediates for Bacillus subtilis Metabolic Pathways by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry," *Anal. Chem.*, **2002**, 74, 2233-2239.

Song, S. M.; Marriott, P.; Kotsos, A.; Drummer, O. H.; Wynne, P. "Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC x GC-TOFMS) for drug screening and confirmation," *Forensic Science International*, **2004**, 143, 87-101.

Stevenson, C. L.; Vo-Dinh, T. *Modern Techniques in Raman Spectroscopy*. New York: Wiley. **1996**.

Strahl, B. D.; Allis, C. D. "The language of covalent histone modifications," *Nature*, **2000**, 403, 41-45.

Styczynski, M. P.; Moxley, J. F.; Tong, L. V.; Walther, J. L.; Jensen, K. L.; Stephanopoulos, G. N. "Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics and Biomarker Discovery," *Anal. Chem.*, **2007**, 79, 966-973.

Sun, K; Winniford, W.; Griffith, J.; Colura, K.; Green, S; Pursch, M.; Luong, J. "Comprehensive Two-Dimensional Gas Chromatography for Fast Screening of Wash Oils," *J. Chromatogr. Sci.*, **2003**, 41, 506-518.

Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Orešič, M. "Normalization method for metabolomics data using optimal selection of multiple internal standards," *BMC Bioinformatics*, **2007**, 8, 93.

Tanaka, K.; West-Dull, A.; Hine, D. G.; Lynn, T. B. "Gas-Chromatographic Method of Analysis for Urinary Organic Acids. I. Retention Indices of 155 Metabolically Important Compounds," *Clin. Chem.*, **1980**, 26, 1839-1846.

Tanaka, K.; West-Dull, A.; Hine, D. G.; Lynn, T. B.; Lowe, T. "Gas-Chromatographic Method of Analysis for Urinary Organic Acids. II. Description of the Procedure, and Its Application to Diagnosis of Patients with Organic Acidurias," *Clin. Chem.*, **1980**, 26, 1847-1853.

Theobald, U.; Mailinger, W.; Reuss, M.; Rizzi, M. "In Vivo Analysis of Glucose-Induced Fast Changes in Yeast Adenine Nucleotide Pool Applying a Rapid Sampling Technique," *Anal. Biochem.*, **1993**, 214, 31-37.

Theodorescu, D.; Fliser, D.; Wittke, S.; Mischak, H.; Krebs, R.; Walden, M.; Ross, M.; Eltze, E.; Bettendorf, O.; Wulfing, C.; Semjonow, A. "Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine," *Electrophoresis*, **2005**, 23, 2797-2808.

Thierry, A.; Maillard, M-B,; Yvon, M. "Conversion of l-Leucine to Isovaleric Acid by Propionibacterium freudenreichii TL 34 and ITGP23," *Appl. Environ. Microbiol.*, **2002**, 68, 608-615.

Tikunov, Y.; Lommen, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Bino, R. J.; Hall, R. D.; Bovy, A. G. "A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles," *Plant Physiology*, **2005**, 139, 1125-1137.

Tobin, M. C. *Laser Raman Spectroscopy*. New York: Wiley. **1971**.

Tranchida, P. Q.; Dugo, P.; Dugo, G.; Mondello, L. "Comprehensive two-dimensional chromatography in food analysis," *J. Chromatogr. A*, **2004**, 1054, 3-16.

Tu, B. P.; Tikunov, Y.; Lommen, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Bino, R. J.; Hall, R. D.; Bovy, A. G. *Plant Physiology*, **2005**, 139, 1125-1137.

Tu, B. P.; Kudlicki, A.; Rowicka, M.; McKnight, S. L. "Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes," *Science*, **2005**, 310, 1152-1158.

Tu, B. P.; Mohler, R. E.; Liu, J. C.; Dombek, K. M.; Young, E. T.; Synovec, R. E.; McKnight, S. L. "Cyclic Changes in Metabolic State During the Life of a Yeast Cell," submitted *Proc. Natl. Acad. Sci. U. S. A.*, **2007**

Tweeddale, H.; Notley-McRobb, L.; Ferenci, T. "Effect of slow metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis," *J. Bacteriology*, **1998**, 180, 5109-5116.

Vaidyanathan, S.; Dell, D. B.; Goodacre, R. "Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification," *J. Am. Soc. Mass Spectrom.*, **2002**, 13, 118-128.

Vaidyanathan, S.; Jones, D.; Broadhurst D. I.; Ellis, J.; Jenkins, T.; Dunn, W. B.; Hayes, A.; Burton, N.; Oliver, S. G.; Kell, D. B.; Goodacre, R.; "A laser desorption ionization mass spectrometry approach for high throughput metabolomics," *Metabolomics*, **2005**, 1, 243-250.

van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. "Centering, scaling, and transformations: improving the biological information content of metabolomics data," *BMC Genomics*, **2006**, 7, 142-157.

van der Horst, A.; Schoenmakers, P. J. "Comprehensive two-dimensional liquid chromatography of polymers," *J. Chromatogr. A*, **2003**, 1000, 693-709.

van Deursen, M.; Beens, J.; Reijenga, J.; Lipman, P.; Cramers, C.; Blomberg, J. "Group-type identification of oil samples using comprehensive two-dimensional gas chromatography coupled to a time-of-flight mass spectrometer (GC x GC-TOF)," *J. High Resol. Chromatogr.* **2000**, 23, 507-510.

van Dijken, J. P.; Bauer, J.; Brambilla, L.; Duboc, P.; Francois, J. M.; Gancedo, C.; Giuseppin, M. L. F.; Heijnen, J. J.; Hoare, M.; Lange, H. C.; Madden, E. A.; Niederberger, P.; Nielsen, J.; Parrou, J. L.; Petit, T.; Porro, D.; Reuss, M.; van Riel, N.; Rizzi, M.; Steensma, H. Y.; Verrips, C. T.; Vindeløv, J.; Pronk, J. T. " An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains," *Enzyme Microb. Technol.*, **2000**, 26, 706-714.

van Duyne, R. P.; Hulteen, J. C.; Treichel, D. A. "Atomic force microscopy and surface-enhanced Raman spectroscopy. I. Ag island films and Ag film over polymer nanosphere surfaces supported on glass," *J. Chem. Phys.*, **1993**, 99, 2101-2115.

van Gyseghem, E.; Dejaegher, B.; Put, R.; Forlay-Frick, P.; Elkihel, A.; Daszykowski, M.; Heberger, K.; Massart, D. L.; Heyden, Y. V. "Stationary phases in the screening of drug/impurity profiles and in their separation method development: Identification of columns with different and similar selectivities," *J. Pharm. Biomed. Anal.*, **2006**, 41, 141-151.

van Mispelaar, V. G.; Tas, A. C.; Smilde, A. K.; Schoenmakers, P. J.; van Asten, A. C. "Quantitative analysis of target components by comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, **2003**, 1019, 15-29.

Veal, E. A.; Ross, S. J.; Malakasi, P.; Peacock, E.; Morgan, B. A. "Ybp1 Is Required for the Hydrogen Peroxide-induced Oxidation of the Yap1 Transcription Factor," *J. Bio. Chem.*, **2003**, 278, 30895-30904.

Veriotti, T.; Sacks, R. "High-Speed GC and GC/MS with Series-Coupled Column Ensemble Using Stop-Flow Operation," *Anal. Chem.*, **2001**, 73, 3045-3050.

Veriotti, T.; Sacks, R. "High-Speed GC and GC/Time-of-Flight MS of Lemon and Lime Oil Samples," *Anal. Chem.*, **2001**, 73, 4395-4402.

Veriotti, T. "Analysis of Metabolomics Samples Using Comprehensive Two-Dimensional Gas Chromatography (GCxGC) with TOFMS Detection," *LCGC North America*, **2007**, (Suppl.), 13.

Villas-Bôas, S. G.; Højer-Pedersen, J.; Äkesson, M.; Smedsgaard, J.; Nielsen, J. "Global metabolite analysis of yeast: evaluation of sample preparation methods," *Yeast*, **2005**, 22, 1155-1169.

Villas-Bôas, S. G.; Rasmussen, S.; Lane, G. A. "Metabolomics or metabolite profiles?" *Trends Biotechnol.*, **2005**, 23, 385-386.

Vo-Dinh, T.; Allain, L. R.; Stokes, D. L. "Cancer gene detection using surface-enhanced Raman scattering (SERS)," *J. Raman Spectrosc.*, **2002**, 33, 511-516.

Vogh, J. W. "Isolation and Analysis of Carbonyl Compounds as Oximes," *Anal. Chem.*, **1971**, 43, 1618-1623.

Voit, E. O. "Biochemical and genomic regulation of the trehalose cycle in yeast: review of observations and canonical model analysis," *J. Theor. Biol.*, **2003**, 223, 55-78.

Vorst, O.; de Vos, C. H. R.; Lommen, A.; Staps, R. V.; Visser, R. G. F.; Bino, R. J.; Hall, R. D. "A non-directed approach to the differential analysis of multiple LC—MS-derived metabolic profiles," *Metabolomics*, **2005**, 1, 169-180.

Wang, M.; Lamer, R-J. A. N.; Korthout, H. A. A. J.; van Nesselrooij, J. H. J.; Witkamp, R. F.; van der Heijden, R.; Voshol, P. J.; Havekes, L. M.; Verpoorte, R.; van der Greef, J. "Metabolomics in the Context of Systems Biology: Bridging Traditional Chinese Medicine and Molecular Pharmacology," *Phytother. Res.* **2005**, 19, 173-182.

Wang, F. C.-Y.; Robbins, W. K.; Di Sanzo, F. P.; McElroy, R. C. "Speciation of Sulfur-Containing Compounds in Diesel by Comprehensive Two-Dimensional Gas Chromatography," *J. Chromatogr. Sci.*, **2003**, 519-523.

Wang, Q-z.; Wu, C-y.; Chen, T.; Chen, X.; Zhao, X-m. "Integrating metabolomics into systems biology framework to exploit metabolic complexity: strategies and applications in microorganisms," *Appl. Microbiol. Biotechnol.*, **2006**, 70, 151-161.

Weckwerth, W.; Fiehn, O. "Can we discover novel pathways using metabolomic analysis?" *Curr. Opin. Biotechnol.*, **2002**, 13, 156-160.

Weckwerth, W. "Metabolomics in Systems Biology," *Annu. Rev. Plant Biol.* **2003**, 54, 669-689.

Weckwerth, W.; Loureiro, M. E.; Wenzel, K.; Fiehn, O. "Differential metabolic networks unravel the effects of silent plant phenotypes," *Proc. Natl. Acad. Sci. U. S. A.*, **2004**, 101, 7809-7814.

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

Welthagen, W.; Schnelle-Kreis, J.; Zimmerman, R. "Search criteria and rules for comprehensive two-dimensional gas chromatography—time-of-flight mass spectrometry analysis of airborne particulate matter," *J. Chromatogr. A*, **2003**, 1019, 233-249.

Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O. "Comprehensive two-dimensional gas chromatography—time-of-flight mass spectrometry (GC x GC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice," *Metabolomics*, **2005**, 1, 65-73.

Weng, L.; Dai, H.; Zhan, Y.; He, Y.; Stepaniants, S. B.; Bassett, D. E. "Rosetta error model for gene expression analysis," *Bioinformatics*, **2006**, 22, 1111-1121.

Werner, E. "The Future and Limits of Systems Biology," *Sci. STKE*, **2005**, 278, 16-18.

Western, R. J.; Marriott, P. J. "Methods for generating second dimension retention index data in comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, **2003**, 1019, 3-14.

Williams, N. "Yeast Genome Sequence Ferments New Research," *Science*, **1996**, 272, 481.

Williams, R.; Lenz, E. M.; Wilson, A. J.; Granger, J.; Wilson, I. D.; Major, H.; Stumpf, C.; Plumb, R. "A multi-analytical platform approach to the metabonomic analysis of plasma from normal and zucker (fa/fa) obese rats," *Mol. BioSyst.*, **2006**, 2, 174-183.

Wise, B. M.; Gallangher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, R. S. *PLS_Toolbox 3.5 for use with Matlab*, Washington: Eigenvector Research, Inc. **2005**.

Wittke, S.; Mischak, H.; Walden, M.; Kolch, W.; Rädler, T.; Wiedemann, K. "Discovery of biomakers in human urine and cerebrospinal fluid by capillary electrophoresis coupled to mass spectrometry: Towards new diagnostic and therapeutic approaches," *Electrophoresis*, **2005**, 23, 1476-1487.

Wittrig, R. E.; Dorman, F. L.; English, C. M.; Sacks, R. D. " High-speed analysis of residual solvents by flow-modulation gas chromatography," *J. Chromatogr. A*, **2004**, 1027, 75-82.

Wolfender, J.-L.; Ndjoko, K.; Hostettmann, K. "Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectrometry: a powerful combination for the on-line structural investigation of plant metabolites," *J. Chromatogr. A*, **2003**, 1000, 437-455.

Woodward, L. A. *Raman Spectroscopy* v. 1. New York: Plenum Press. **1967**.

Xie, L.; Marriott, P. J.; Adams, M. "Chemometric analysis of comprehensive two-dimensional gas chromatography data using cryogenic modulation," *Anal. Chim. Acta*, **2003**, 500, 211-222.

Yau, W. W. "Characterizing Skewed Chromatographic Band Broadening," *Anal. Chem.*, **1977**, 49, 395-398.

Young, M. W. "An ultradian clock shapes genome expression in yeast," *Proc. Natl. Acad. Sci. U.S.A.*, **2004**, 101, 1118-1119.

Young, E. T.; Dombek, K. M.; Tachibana, C.; Ideker, T. "Multiple Pathways Are Co-regulated by the Protein Kinase Snf1 and the Transcription Factors Adr1 and Cat8," *J. Biol. Chem.*, **2003**, 278, 26146-26158.

Zini, C. A.; de Assis, T. F.; Ledford Jr., E. B.; Dariva, C.; Fachel, J.; Christensen, E.; Pawliszyn, J. "Correlations between Pulp Properties of Eucalyptus Clones and

Leaf Volatiles Using Automated Solid-Phase Microextraction," *J. Agric. Food Chem.*, **2003**, 51, 7848-7853.

Zörb, C.; Langenkämper, G.; Betsche, T.; Niehaus, K.; Barsch, A. "Metabolite Profiling of Wheat Grains (Triticum aestivum L.) from Organic and Conventional Agriculture," *J. Agric. Food Chem.*, **2006**, 54, 8301-8306.

Zrostlíková, J.; Hajšlová, J.; Čajka, T. "Evaluation of two-dimensional gas chromatography –time-of-flight mass spectrometry for the determination of multiple pesticide residues in fruit," *J. Chromatogr. A*, **2003**, 1019, 173-186.

http://web.lemoyne.edu/~giunta/pasteur.html

http://www.genome.jp/kegg/

http://www.mpimp-golm.mpg.de/mms-library/index-e.html

http://www.yeastgenome.org/community/W303.html

Vita

Rachel Mohler was born to a young military couple on September 20, 1980 at Camp Lejeune, North Carolina. She inherited her father's love for science and obtained a Bachelor of Science in chemistry at the University of California, Riverside in 2002. She immediately proceeded to graduate school at the University of Washington and obtained a Master of Science in August 2004. After spending nine years in higher education, she was awarded a Doctor of Philosophy in Analytical Chemistry in 2007.