2006

Chemometric Analysis of Multivariate Liquid Chromatography Data: Applications in Pharmacokinetics, Metabolomics, and Toxicology

Sarah Elizabeth Graham Porter
Virginia Commonwealth University

CHEMOMETRIC ANALYSIS OF MULTIVARIATE LIQUID CHROMATOGRAPHY DATA: APPLICATIONS IN PHARMACOKINETICS, METABOLOMICS, AND TOXICOLOGY

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry at Virginia Commonwealth University.

by

SARAH ELIZABETH GRAHAM PORTER
M.S., Virginia Commonwealth University, 2000
B.S., University of Virginia, 1998

Director: Sarah C. Rutan
Professor, Department of Chemistry

Virginia Commonwealth University
Richmond, Virginia
December 2006

Acknowledgement

I would of course like to thank my advisor, Prof. Sarah C. Rutan, first of all for talking me into coming to graduate school in the first place, and second for making sure I finished. I am grateful for the many opportunities that she has given me and everything she has taught me. She has consistently gone above and beyond the duty of any graduate advisor, and I am truly lucky to have had the opportunity to work with her.

I would also like to thank Prof. Pete Carr, my "academic godfather", and his graduate student, Dwight Stoll. Pete has adopted me as a surrogate graduate student and provided a great deal of feedback on my research and my writing. Dwight has provided me and the Rutan group with multitudinous LC-DAD data to analyze.

From the VCU chemistry department, I would like to acknowledge the Rutan group members, Dennis Thekkudan, Hope Bailey, and Carol Holden, for their support and company. I am truly grateful and indebted to Brad Mangrum for spending hours working with me in the mass spec center (*aka* "crazy town"). I would also like to thank the undergraduate researchers with whom I have had the pleasure of working – Richard Keithley, Marc Cantwell, and Patrick Still – they have taught me more than I could teach them. Finally I would like to thank Chrissy Winschel for all of her help and support.

On a personal note, I would like to thank all of my "parental units" – John and Dian Graham, Dave and Elaine Reubush, Robin and Anne Porter, and Pam and Brian Marth – for all of their support as I finish school for the last time.

Last but not least, I would like to dedicate this dissertation to my husband, Scott Porter. He has done more for me than he will ever know.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

List of Abbreviations

%RSD	Percent relative standard deviation
2D-LC	Two dimensional liquid chromatography
ADHD	Attention deficit hyperactivity disorder
ALS	Alternating least squares
APCI	Atmospheric pressure chemical ionization
BCA	Bureau of Criminal Apprehension
CANDECOMP	Canonical decomposition
CORCONDIA	Core consistency diagnostic
CYP2D6	Cytochrome P450 2D6
CYPs	Cytochromes P450
DAD	Diode array detector
DP	Discriminating power
EIA	Enzyme immunoassay
ESI	Electrospray ionization
F5	Pentafluorobenzene
fALS	Flexible constraint alternating least squares
FID	Flame ionization detector
FLD	Fluorescence detector
FN	False negative
FP	False positive
FWD	Fixed wavelength UV detector
GC-MS	Gas chromatography – mass spectrometry
GE	General enzyme method
GRAM	Generalized rank annihilation method
HPLC	High performance or high pressure liquid chromatography
IAA	Indole-3-acetic acid
LC	Liquid chromatography
MCR	Multivariate curve resolution
MDA	3,4-methylenedioxyamphetamine
MDMA	3,4-methylenedioxymethamphetamine

MLL	Mean list length
MM	Michaelis-Menten
MS	Mass spectrometry
MS-MS	Tandem mass spectrometry
NADPH	Nicotinamide adenine dinucleotide
NAS	Net analyte signal
NPV	Negative predictive value
orp	Mutant *orange pericarp*
PARAFAC	Parallel factor analysis
PCA	Principal component analysis
PHMA	*p*-hydroxymethamphetamine
PMA	*p*-methoxyamphetamine
PMMA	*p*-methoxymethamphetamine
PPV	Positive predictive value
RPLC	Reversed phase liquid chromatography
RSD	Residual standard deviation
SPE	Solid phase extraction
SRM	Selected reaction monitoring
SS	Steady-state
SSRI	Selective serotonin reuptake inhibitor
SVD	Singular value decomposition
TFA	Target factor analysis
TN	True negative
TP	True positive
Tri-PLS	Trilinear partial least squares
WTTFA	Window target testing factor analysis

Abstract

CHEMOMETRIC ANALYSIS OF MULTIVARIATE LIQUID CHROMATOGRAPHY DATA: APPLICATIONS IN PHARMACOKINETICS, METABOLOMICS, AND TOXICOLOGY

By Sarah Elizabeth Graham Porter

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry at Virginia Commonwealth University.

Virginia Commonwealth University, December 2006

Major Director: Sarah C. Rutan
Professor of Chemistry

In the first part of this work, LC-MS data were used to calculate the *in-vitro* intrinsic clearances (CL_{int}) for the metabolism of *p*-methoxymethamphetamine (PMMA) and fluoxetine by the CYP2D6 enzyme using a steady-state (SS) approach and a new general enzyme (GE) screening method. For PMMA, the SS experiment resulted in a CL_{int} of 2.7 ± 0.2 μL pmol $2D6^{-1}$ min^{-1} and the GE experiment resulted in a CL_{int} of 3.0 ±

0.6 μL pmol $2D6^{-1}$ min^{-1}. For fluoxetine, the SS experiment resulted in a CL_{int} of 0.33 ± 0.17 μL pmol $2D6^{-1}$ min^{-1} and the GE experiment resulted in a CL_{int} of 0.188 ± 0.013 μL pmol $2D6^{-1}$ min^{-1}. The inhibition of PMMA metabolism by fluoxetine was also demonstrated.

In the second part of the work, target factor analysis was used as part of a library search algorithm for the identification of drugs in LC-DAD chromatograms. The ability to resolve highly overlapped peaks using the spectral data afforded by the DAD is what distinguished this method from conventional library searching methods. A validation data set of 70 chromatograms was used to calculate the sensitivity (correct identification of positives) and specificity (correct identification of negatives) of the method, which were 92 % and 94 % respectively.

Finally, the last part of the work shows the development of data analysis methods for four-way data generated by two-dimensional liquid chromatography separations with DAD. Maize seedlings were analyzed, specifically focusing on indole-3-acetic acid (IAA) and related compounds. Window target testing factor analysis was used to identify the spectral groups represented by the standards in the mutant and wild-type chromatograms. Two curve resolution algorithms were applied to resolve overlapped components in the data and to demonstrate the quantitative potential of these methods. A total of 95 peaks were resolved. Of those peaks, 45 were found in both the mutant and wild-type maize, 16 peaks were unique to the mutants, 13 peaks were unique to the wild-types, and the remaining peaks were standards. Several IAA conjugates were quantified in the maize samples at levels of 0.3 - 2 μg/g plant material.

CHAPTER 1. Overview and Objectives

The overarching theme of the work presented in this dissertation is the development and application of chemometric methods to multivariate chromatographic data of varying levels of complexity. Liquid chromatography (LC) is a widely used technique in analytical chemistry for the separation of complex mixtures. LC separation techniques are used extensively in the pharmaceutical industry and in forensic laboratories, as well as for proteomic and metabolomic studies. In this work, LC separations were used to develop new methods for measuring *in-vitro* reactions between drugs and purified enzymes. Additionally, fast, high-temperature LC methods were investigated as drug screening techniques and as a second dimension for comprehensive two-dimensional LC. The detectors used in this work (and in most modern LC separations) were multichannel detectors, including diode array detectors and mass spectrometers. The data obtained when using LC with such detectors often require data analysis techniques in order to extract relevant information, and chemometric methods afford such an opportunity.

Chemometric methods have been used for many years to analyze chromatographic separations with multichannel detectors [1]. The data arising from such techniques are called *multivariate* data. As an example, a *two-way* data set obtained by running a mixture sample on an LC with a spectrometric detector will contain a matrix of intensity values as a function of retention time (chromatographic dimension) and

wavelength or m/z (spectral dimension). Running several samples of different concentrations (as in a calibration data set) adds a third dimension to the data (sample or concentration dimension) to create *three-way* data. Such multivariate data can readily be resolved into individual components using multivariate curve resolution (MCR) techniques such as parallel factor analysis (PARAFAC) and alternating least squares (ALS) [2]. These methods are particularly useful when the mixtures are complex and the chromatograms contain many overlapped peaks. This work will explore the application of MCR and other chemometric methods to liquid chromatography – diode array data (LC-DAD) and LC data with mass spectrometric detection (LC-MS). Two-way and three-way data will be discussed and finally the analysis of four-way data in the form of two-dimensional LC with DAD (2D-LC-DAD) will be investigated.

The application of MCR and kinetic fitting routines to LC-MS data is described in detail in Chapter 4 and in reference [3]. LC-MS methods were developed for the quantification of drug metabolites in order to determine the pharmacokinetic constants of drugs *in-vitro*. Several drug systems were studied, including dextromethorphan, an over the counter cough suppressant, *p*-methoxymethamphetamine (PMMA) and 3,4-methylenedioxymethamphetamine (MDMA), two "designer" amphetamine analogs, and fluoxetine, sold commercially as Prozac®. The goal of this project was to develop methods for measuring the kinetics of the *in-vitro* reaction between the drug (substrate) and a microsomal preparation of insect baculovirus expressed cytochrome P450 2D6 (CYP2D6).

A steady-state approach is often used to study the pharmacokinetics of drugs *in-vitro,* where several different concentrations of the drug are incubated in the presence of an enzyme for a specific amount of time, and the formation of product is measured at each substrate level [4]. In this work, a new experimental method using fast gradient LC-MS was designed to complement the information gained from these classical *in-vitro* steady-state incubation experiments. The new method is referred to as a general enzyme incubation method. The general enzyme method resembles general kinetic experiments, where the change in concentration of the reactants and products is measured as a function of time. Here, the changes in the concentrations of the drug (substrate) and the metabolite (product) in an enzymatic reaction were measured using a fast LC-MS method and modeled using kinetic fitting routines previously described [5].

A steady-state least squares kinetic fitting algorithm was used to fit the data obtained from the steady-state experiments [6], and a flexible kinetic modeling program previously developed [5] was used to fit the data obtained from the new general enzyme method. Both methods were used to determine the intrinsic clearance, a useful pharmacokinetic parameter, and the intrinsic clearance values that were determined for PMMA and fluoxetine were consistent between the new method and the steady-state method. The new method was also used to study metabolic drug interactions; the clearance of PMMA was shown to decrease by nearly an order of magnitude in the presence of an equimolar amount of fluoxetine.

The work described in Chapter 5 and in reference [7] details the development of a library search algorithm based on target factor analysis (TFA) that was used to analyze

LC-DAD data. The purpose of the project was to develop a very fast and robust screening method for the identification of samples containing toxicologically relevant compounds. Screening methods are used in forensic testing laboratories to quickly identify samples that are likely to contain drugs, and then longer confirmatory tests are performed on those samples for evidential purposes. The sample throughput of a forensic laboratory can be increased by quickly identifying those samples that do not contain any compounds of interest, thereby eliminating the need for a confirmatory test.

In this work, samples containing toxicologically relevant drugs and their metabolites were analyzed by a fast gradient LC-DAD system developed by Stoll *et al.* [8]. Samples both in biological matrices (blood and urine) and in "clean" matrix (mobile phase or pure solvent) were included in the study. The chromatographic method was evaluated on its discriminating power [9] and mean list length [10] (to allow comparison with previously published methods), and the library search method was evaluated on its sensitivity and selectivity. These parameters estimate the method's ability to correctly identify positive and negative samples. A positive test was based on both a retention time match and a spectral match to a library of 47 compounds. This method allowed for the resolution of highly overlapped peaks in a chromatogram, and was able to identify peaks under conditions where conventional library search algorithms fail. The combination of the fast chromatographic method and the library search algorithm allowed for very efficient screening of biological samples for drugs of abuse and their metabolites. It was shown that the screening method has the potential to nearly double the throughput of a laboratory relative to the longer confirmatory testing.

Finally, in Chapter 6, quadrilinear four-way data obtained using 2D-LC-DAD were analyzed using several different approaches. The work presented in Chapter 6 can also be found in reference [11]. The purpose of this project was to develop data analysis methods for extracting qualitative and quantitative information from higher order data. Extracts of the homozygous mutant *orange pericarp (orp)* maize seedlings and wild-type maize seedlings were analyzed by 2D-LC-DAD along with a set of standards. The standards included in the study were compounds related to the biosynthetic pathways of indole-3-acetic acid (IAA), an important plant hormone. Two data analysis methods were used: a TFA algorithm for targeted metabolic profiling based on a method published by Lohnes *et al.* [12], and two MCR algorithms for quantitative comparison between samples. The TFA algorithm allowed a chromatogram to be searched for the spectra of specific analytes (in this case, the indolic standards) and different samples could be qualitatively compared based on the presence or absence of the target analytes. The flexible constraint alternating least squares (fALS) algorithm described in reference [13] and the quadrilinear PARAFAC model [14] were applied to resolve all of the components (both known and unknown) present in the maize extracts. The unique combination of these two MCR algorithms allowed the compositions of wild-type and *orp* maize samples to be quantitatively compared based on the relative contribution of the components present. In addition several of the indolic standards were identified and quantified in the maize samples.

The results of these three projects will be discussed in detail in this dissertation in Chapters 4-6. A brief introduction of the chromatographic and chemometric techniques

used will be presented in Chapters 2 and 3. Finally, the broad scope and the possibility of the future applications of the results presented here will be discussed in Chapter 7. The chromatographic data that were studied in this work, including ultra fast high temperature LC, 2D-LC, and LC-MS, represent some of the latest advances in LC research, and the application of the chemometric methods are an important contribution to the field.

CHAPTER 2. Liquid Chromatography Separation Methods and Detectors

2.1. Liquid Chromatography

For decades, LC separations have been used by analytical chemists to separate the components present in mixtures. It was in the early 1970's that new developments in stationary phase materials allowed for the rapid rise in the laboratory use of high performance or high pressure LC (HPLC), which allowed it to become one of the most popular chromatography applications in analytical chemistry laboratories [15]. In fact, LC is typically used synonymously with HPLC in chromatography literature (and will be throughout this dissertation as well).

Although gas chromatography with mass spectrometric detection (GC-MS) remains a mainstay in many analytical laboratories, it has several disadvantages that make LC a more practical choice for many applications. Perhaps the most important distinction is that GC analytes must be both volatile and thermally stabile due to the high temperatures required for GC analysis (usually >200 ºC). Many less volatile, polar compounds (including many drugs of abuse and small molecule metabolites) can only be analyzed by GC after derivatization [16]. LC is a more universal technique: any analyte that is soluble in a suitable solvent can be analyzed by LC methods if an appropriate detector is chosen.

LC methods generally fall into two categories, normal phase and reversed phase, based on the relative polarity of the mobile phase and the stationary phase. Reversed phase LC (RPLC) is by far the more common technique, because many organic compounds are soluble in suitable solvents. RPLC uses non-polar stationary phases, most notably long chain alkyl groups (*i.e.*, octadecyl- or octyl- groups), bonded to a silica support. The mobile phases used in RPLC are polar organic solvents (acetonitrile or methanol) mixed with aqueous buffers. Gradient LC is often used to allow the separation of analytes with widely varying retention characteristics in a reasonable amount of time, much like temperature programming in GC separations [15]. The term gradient refers to a change in the composition of the mobile phase over the course of the separation.

2.2. Two Dimensional Liquid Chromatography

Two-dimensional separation methods have gained popularity in recent years, and increasingly complex samples are being analyzed using these approaches, especially in proteomics research. 2D-LC techniques are used in order to take advantage of the superior resolving power afforded by using orthogonal separation mechanisms. The significant increase in separation space is particularly suited to complex samples such as those encountered in proteomic studies, where the samples may contain hundreds of components. Multi-dimensional separations in proteomics research were recently reviewed by Issaq *et al.* [17]. Although relatively few applications to metabolomics have been published, it is clear from the apparent complexity of the metabolome that such powerful separation tools will soon be invaluable to this field as well [18].

Two-dimensional separation techniques were described in detail by Giddings in 1984 [19]. Giddings defined two-dimensional separations as those in which a sample is subjected to two separation techniques that are at right angles (or orthogonal) to one another. They can be carried out with any separation method, including GC, LC, electrophoresis, and thin-layer chromatography; however, this discussion will pertain mainly to 2D-LC. The result is that there is significantly more separation "space" than in one-dimensional separations. It was clear to Giddings even in 1984 the limitation of one-dimensional separations in evaluating the highly complex samples usually encountered in biological and environmental studies [19].

2D-LC techniques can generally be grouped into two categories: heart-cutting and comprehensive [20]. Heart-cutting involves collecting a specific fraction (either manually or automatically) of a chromatogram and subjecting it to a second separation with an orthogonal method (*e.g.*, a different RPLC column or a different mode such as ion-exchange). This method is often used to analyze a particular peak in a one-dimensional separation that is suspected to contain multiple components that were not separated on the first column. Comprehensive 2D-LC, as described by Bushey and Jorgenson [21], involves the sampling of sequential fractions of the effluent from the first dimension column and subjecting every fraction to the second dimension separation. Heart-cutting is the more popular method, due to the fact that it can be done without specialized instrumentation; however, descriptions of comprehensive 2D-LC methods are becoming more common in the literature [22-25].

While previously published comprehensive 2D-LC methods required analysis times of hours or days [25], more recent work has significantly decreased the necessary analysis times [23]. Stoll and Carr [24] have recently described a novel approach for fast comprehensive 2D-LC analysis using gradient ion exchange in the first dimension and ultra-fast, high temperature reversed phase gradient elution in the second dimension. The improved speed of the separation is achieved by the use of high temperatures (> 100 ºC), short columns, and high linear velocities in the second dimension separation. Their system was able to separate the compounds present in a tryptic digest of bovine serum albumin in less than 30 minutes. Stoll *et al.* [26] have also demonstrated the use of high temperatures to greatly speed up comprehensive 2D-LC with reversed phase gradient elution in both chromatographic dimensions. They were able to generate a peak capacity of 870 in 25 minutes using a single second dimension column. Based on the fact that each 1^{st} dimension peak appeared in at least two consecutive 2^{nd} dimension chromatograms, this system also moved closer to meeting the sampling rate requirement for comprehensive 2D-LC suggested by Murphy, Schure, and Foley [20].

The techniques developed by Stoll *et al.* [26] were applied to extracts of wild-type and *orp* mutant maize seedling tissues and a set of indolic metabolite standards. The chromatographic methods were presented as a first step in demonstrating that comprehensive 2D-LC is a practical high speed analytical methodology for small molecule metabolites. The data collected by the system described in reference [26] were analyzed in this work by several different chemometric methods. The methods discussed

in Chapter 6 of this work (and in reference [11]) show that the combination of 2D-LC-DAD and chemometrics will be an important tool in metabolomic studies.

2.3. Detectors for Liquid Chromatography

2.3.1 Absorbance Detectors

There are many choices for online detection for LC, but by far the most popular are spectroscopic detectors, including fixed wavelength UV (FWD), DAD, and fluorescence detectors (FLD) [15]. All such detectors are solute property detectors, which measure a particular chemical or physical property of a solute that is independent of the solvent (mobile phase). In particular, FWDs and DADs are useful for the analysis of many organic compounds, especially pharmaceutical compounds, since many of these compounds absorb in the UV region. Early FWDs collected absorbance data at 254 nm, the wavelength at which low-pressure mercury lamps emit most of their light. Many samples absorb at 254 nm, including proteins, nucleic acids, and many other organic compounds, so this limited wavelength was acceptable for many applications [27]. However, using only a single wavelength is a significant restriction on the utility and identification power of LC because only the retention times are useful for identification of the compounds in the mixture, and the absorbance of a particular compound may be limited at the chosen wavelength.

Simultaneous multiple wavelength detection has many applications in analytical chemistry, including in spectroscopy (*e.g.*, atomic absorption, atomic emission, and UV-visible absorption) [28]. Two different approaches were developed: multiplexing, in

which the emitted energies are rapidly scanned, and multichannel detection, in which the emitted energies are simultaneously detected by individual detector elements. Early reports of the multiplexing approach were reported by Denton *et al.* [29] and Saitoh and Suzuki [30]. The linear photodiode array detector (a multichannel detector) has found favor as an LC detector for its speed and relatively low noise levels. Advances in computer processing in the late 1970's further helped to cement the place of the DAD detector as an invaluable tool for chromatographers [28]. The modern-day DAD relies on an array of diodes (as the name implies) to detect many wavelengths of light simultaneously.

The information gained upon the addition of a DAD to an LC method could be seen early on in the example provided by George [31] where a possible impurity peak was identified in a mixture of cyanocobalamin (vitamin B12) and riboflavin. The B12 peak showed possible fronting; however it was impossible to determine from only a single wavelength analysis whether the peak was truly fronting or whether the shoulder was in fact an impurity in the mixture. Analysis with a DAD revealed that the "fronting" was an impurity with a unique spectrum and a relatively low absorbance at 254 nm.

Despite their great advantages over single channel detection, the relatively low sensitivity and selectivity of DAD are primary limitations on using these detectors. Only analytes that have significant absorption in the UV range can be analyzed, and many compounds with similar structures have nearly identical spectra, which makes the selectivity of the detector low compared to MS. A further limitation of the DAD is that the dynamic range of the detector is limited. As the absorbance values approach 3 AU

(where the transmittance is approaching 0.001) the response of the detector is no longer linear with concentration. In certain instruments, this situation can result in negative absorbance values being recorded or in chromatography peaks that are severely distorted at their apex. Quantitative analysis becomes impossible in these cases. However, despite its shortcomings, the information content of the DAD is still far greater than any single channel detector. The chemometric methods that will be discussed in this work allow quantitative and qualitative information to be obtained, even from a detector with substantially lower selectivity than a mass spectrometer.

2.3.2. Mass Spectrometry

The use of MS detectors for LC has increase dramatically over the past decade. A search on Web of Science for the terms "liquid chromatography and mass spectrometry" shows that in the past 15 years, the number of publications dealing with LC-MS has increased by nearly an order of magnitude [32]. It was unarguably the introduction of electrospray ionization (ESI) by Yamashita and Fenn in 1984 [33] that led to the improvement in the coupling of LC to MS. ESI also allowed for the ionization of large molecules like proteins, and made LC-MS an important tool in proteomics research. In addition to its many applications in proteomics, LC-MS has found favor in the analysis of many pharmaceutical compounds. Most methods require very little sample preparation beyond dilution and filtration and relatively low molecular weight compounds can be detected readily with single quadrupole MS instruments in positive or negative ion detection mode. In particular, the basic drugs discussed in Chapter 4 of this dissertation

are amenable to positive ion detection mode with the buffers typically used in low pH RPLC separations. Triple quadrupole, ion trap, or quadrupole-time of flight mass analyzers can further improve sensitivity and selectivity by allowing tandem mass spectrometry (MS-MS) and selected reaction monitoring (SRM) experiments to be carried out. MS-MS analysis is particularly attractive for analyzing isobaric compounds (analytes with the same nominal mass).

In early ESI sources, the flow rate of the analyte solution being introduced into the source was limited to 1-10 μL min^{-1} due to the mechanism of nebulization, which was solely by the use of a charged capillary. Such low flow rates are impractical for LC analyses with standard analytical columns. However, the introduction of pneumatically-assisted ESI sources allowed for higher flow rates (up to 200 μL min^{-1}), and the addition of a stream of heated drying gas (usually nitrogen) to the system finally brought the flow rates into a range where the ESI source was compatible with typical LC flow rates (500 – 1000 μL min^{-1} for newer sources) [16].

Although modern ESI sources are compatible with the higher flow rates encountered with LC analysis, sensitivity and limits of detection are still typically much better when lower flow rates are used [16]. A standard analytical LC column usually has dimensions of 4.6 mm x 150 mm with a particle size of 5 μm. Typical flow rates for a column of this size are between 500 and 1000 μL min^{-1}. Decreasing the flow rate significantly can lead to prohibitively long analysis and re-equilibration times and band broadening due to longitudinal diffusion of the analytes. Adapting LC methods to use lower flow rates, while still maintaining reasonable analysis time, can be accomplished in

several ways. Two possibilities are the use of narrower bore or shorter columns, or the use of a post column flow splitter.

A flow splitter can be used to reduce the flow rate into the ESI source without having to adjust the chromatographic method. The splitter sends the higher flow to waste and the lower flow to the ESI. For a 20:1 splitter, if an analysis uses a flow rate of 1000 $\mu L\ min^{-1}$, the flow rate into the ESI source would be 50 $\mu L\ min^{-1}$. The main drawback to this method is that flow splitters, even those sold as "low dead volume", increase extra-column volume, which can lead to significant peak broadening, decreased resolution, or otherwise poorly shaped chromatographic peaks [15].

Using shorter or narrower columns in the chromatographic method is another way to reduce flow rate [15]. Decreasing the inner diameter (ID) of the column and/or the length of the column can allow lower flow rates without significantly increasing analysis time, although using shorter columns will often compromise resolution. Chromatographic columns are sold in 2.1 mm ID and 1 mm ID sizes that can be used for LC-MS analysis, as well as the standard 4.6 mm ID column with lengths down to 30 mm. Smaller particle sizes, down to 3 μm, are also available that can improve the efficiency and resolution of the column.

Another drawback to using MS detection with LC is that there tends to be a higher amount of background noise in LC-MS data in the lower mass region ($m/z < 500$ amu) as compared to other detection methods such as DAD and FLD. This problem can usually be overcome by using SRM or by using chemometric techniques to resolve the background components from the analytes and reduce noise [34]. Another issue often

encountered using an ESI source with LC is matrix effects, which were recently reviewed by Taylor [35]. Matrix effects occur when co-eluting compounds or components of the LC mobile phase interfere with or alter the ionization efficiency of the analyte molecule at the ionization source. The effects are particularly pronounced with ESI [36]. Matrix effects were first described by Tang and Kebarle [37] who noted that the electrospray responses of organic bases were affected by changes in the concentrations of other organic bases in the solution. Mobile phase composition and sample composition can have a detrimental effect on the ionization efficiency of the analytes and can negatively affect quantification. The best ways to combat these effects are to use selective extraction procedures for biological samples [38], optimized solvent systems [39], and chromatographic systems that completely resolve analytes from each other and from matrix components. Atmospheric pressure chemical ionization sources (APCI) do not usually have such pronounced effects [38, 39].

APCI sources work by ionizing the samples in the vapor phase using a corona discharge. A reactant gas, usually methane, is charged by interaction with a 70 eV electron beam and then allowed to undergo a proton transfer reaction with the sample being introduced. The result is a "soft" ionization technique in which the m/z value of the analyte is its molecular weight plus a proton (MH^+) [40]. Because ESI ionizes all of the polar components of the mixture in the liquid phase, non-volatile compounds such as proteins, salts, and other impurities are ionized and analyzed along with the analytes of interest. APCI occurs in the gas phase, and thus only the volatile components are ionized.

MS detectors for LC continue to increase in popularity. The high sensitivity and specificity of these detectors often outweigh their disadvantages. However, DAD detectors are still the detector of choice when a robust, inexpensive method is needed. The work presented in this dissertation will showcase both MS and DAD detectors for LC and highlight the great advantages gained when using any multichannel detector for chromatographic separations.

CHAPTER 3. Chemometrics

3.1. Introduction and Notation

Chemometrics can be broadly defined as the application of mathematical and statistical methods to chemical data. As discussed in Chapter 2, the multichannel detectors often used for LC separations result in higher order data that can benefit from treatment with chemometric methods. Specifically, chemometric methods allow overlapped chromatographic peaks to be resolved that traditional single channel detectors would miss. For example, peaks with less than ideal chromatographic resolution (R_s < 1.0) can readily be quantified. The purpose of this chapter is to present some terminology and nomenclature commonly used in chemometrics and to introduce the methods that will be used to analyze the data in this dissertation.

For the purposes of this work, the standard nomenclature and notation for multi-way analysis as published by Kiers [41] will be used. Scalars are denoted by lowercase italicized variables. Vectors and two-way data arrays (matrices) will be denoted by capital bold-faced variables, $\mathbf{X}$ (I x J) where the capital italicized variables I and J represent the number of rows and columns in the matrix, respectively. For a column vector, $\mathbf{X}$, J is equal to one, and for a row vector, I is equal to one. Three-way and higher arrays will be denoted by capital, bold-faced and underlined variables, as in $\underline{\mathbf{X}}$ (I x J x K). The elements of an array are denoted by the same variable as the array, but lowercase and

italicized, with subscripts indicating the indices, so for example, x_{ijk} represents the i^{th} row, the j^{th} column, and the k^{th} slice of the three-way array $\underline{\mathbf{X}}$. The indices are always given as rows first (i), then columns (j), then slices (k, for three-way), and so on for higher order arrays. The symbol $^{\mathbf{T}}$ represents the transpose of a matrix and the symbol † represents the pseudo-inverse of a non-square matrix.

3.2. Multi-way data

Hyphenated instruments such as LC-DAD or LC-MS produce *multivariate* data. For example, an LC experiment with a multichannel spectrometric detector produces *two-way* data, which can be modeled mathematically by

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{B}^{\mathbf{T}} + \mathbf{E} \qquad (1)$$

where $\mathbf{X}$ (I x J) is the data matrix as it is collected from the instrument, $\mathbf{A}$ contains the chromatographic profiles of the individual components of the mixture (intensity information as a function of elution time), $\mathbf{B}^{\mathbf{T}}$ is the transpose of the spectral profiles of the individual components (intensity as a function of wavelength), and $\mathbf{E}$ (I x J) is an error term containing the residual variance in the data [2]. The dimensions of $\mathbf{A}$ and $\mathbf{B}$ are I x R and J x R, respectively, where A is the number of chromatographic time points (also the number of rows in the data matrix), B is the number of wavelengths or m/z values in each spectrum (also the number of columns in the data set), and R is the number of components in the mixture, also referred to as the *rank* of the data.

The goal of MCR algorithms is to estimate the chromatographic and spectral profiles, $\mathbf{A}$ and $\mathbf{B}$, and to minimize the error term, $\mathbf{E}$. This problem is often solved by a

least-squares method (or ALS), which consists of iterative calculations of **A** and **B** starting from an initial estimate of one or the other of the profiles. There are many different methods for getting an initial estimate of the profiles. Singular value decomposition (SVD) is one common starting point for ALS calculations. The following equations are then used to solve Equation 1 for either **A** or **B** iteratively:

$$\mathbf{A}_{\mathbf{pred}} = \mathbf{X} \bullet (\mathbf{B}^{\mathbf{T}})^{\dagger} \tag{2}$$

$$\mathbf{B}_{\mathbf{pred}} = \mathbf{A}^{\dagger} \bullet \mathbf{X} \tag{3}$$

where $\mathbf{A}_{\mathbf{pred}}$ and $\mathbf{B}_{\mathbf{pred}}$ are the predicted profiles, based on the initial estimate of either **A** or **B**. The algorithm iterates between these two equations, each time calculating a predicted data matrix, $\mathbf{X}_{\mathbf{pred}}$, and determining the least squares error between **X** and $\mathbf{X}_{\mathbf{pred}}$. When this error is minimized, or when a maximum number of iterations has been carried out, the resolved profiles, **A** and **B**, are obtained. Higher order data can be analyzed as discussed below.

The earliest work in multi-way analysis was developed by psychologists and dubbed "psychometrics," and the field of chemometrics followed closely on its heels. One of the earliest publications detailing a three-way analysis was by Ledyard Tucker in 1964 (later completed in 1966), in which he showed that a three-way array, $\underline{\mathbf{X}}$, could be decomposed into loading matrices **A**, **B**, and **C** and a three-way core matrix $\underline{\mathbf{G}}$ [42]. This model was later named the Tucker3 model. In 1970, a three-way model that was based on the Tucker3 model was introduced independently by Carroll and Chang [43], who called their model canonical decomposition (CANDECOMP) and Harshman, who called his model parallel factor analysis (PARAFAC) [44]. Both the CANDECOMP and the

PARAFAC models are a special instance of the Tucker3 model in which the core matrix, $\underline{\mathbf{G}}$, is a super-identity array and all modes have the same number of factors. For the purposes of this work, the term PARAFAC will be used, but the CANDECOMP and PARAFAC models are the same.

The PARAFAC model can be generalized for an *N-way* multivariate array $\underline{\mathbf{X}}$ by

$$x_{i_1 i_2 \ldots i_N} = \sum_{r=1}^{R} \left(\prod_{n=1}^{N} a_{i_n r,n} \right) + e_{i_1 i_2 \ldots i_N} \tag{4}$$

for N modes, corresponding to $\mathbf{A_1}$, $\mathbf{A_2}$., ... $\mathbf{A_n}$ where R is the rank of the data and $e_{i_1 i_2 \ldots i_N}$ are the elements of the matrix containing residual unexplained variance [41]. The three-way PARAFAC model is depicted graphically in Figure 1. The model is usually solved using an ALS algorithm as described above. In each case, the number of rows in a given loadings matrix $\mathbf{A_n}$ is equal to the number of data points in that mode, represented by *I*, *J*, *K*, *etc.*, and the number of columns in $\mathbf{A_n}$ corresponds to the number of individual chemical species (*R*) contributing to the model. A component is generally a chemical compound contributing to a chromatogram (*i.e.*, a pure chromatographic peak), but a component can also arise from the gradient backgrounds often observed in gradient LC.

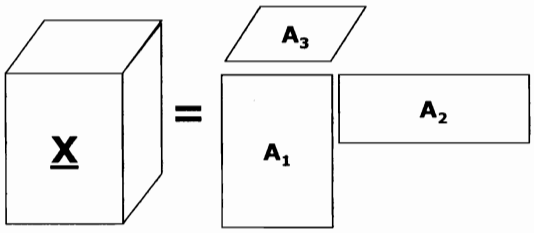

Figure 1. Graphical depiction of a three-way PARAFAC model where $\underline{\mathbf{X}}$ is the data set being analyzed, and $\mathbf{A_1}$, $\mathbf{A_2}$, and $\mathbf{A_3}$ are the loadings of the three modes.

To prevent rotational ambiguity in the resolved profiles that may result when using ALS to solve the PARAFAC model, chemically relevant constraints are usually applied. The non-negativity constraint is used when negative profiles are not meaningful; for example, negative concentrations or negative ion counts in MS data are impossible. A unimodality constraint can be applied to the chromatographic profiles, which imposes the restriction that a profile can have only one maximum. The algorithms used throughout this work allow these constraints to be applied or left out, depending on the data being analyzed. In particular, the ALS algorithm developed by Bezemer and Rutan [13] allows the constraints to be applied flexibly to any or all of the components or modes in the data, and will be referred to as 'fALS' throughout this work to distinguish it from other available ALS algorithms.

3.3. Rank determination

One of the first steps to analyzing multivariate data is to determine the number of the latent variables in the data, or *R*. There are many techniques for determining *R* [45], although none of them provide completely unambiguous results. For two-way data, principal component analysis (PCA) methods can be used. SVD [2] is a particular PCA algorithm that decomposes a two-way data matrix into three matrices,

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^{T} \tag{5}$$

where the columns of **U** (*I x J*) are abstract eigenvectors in the row space, the columns of **V** (*J x J*) are abstract eigenvectors in the column space, and **S** (*J x J*) is a diagonalized matrix containing the square roots of the eigenvalues of **X**. If **X** is an LC-DAD

chromatogram, the number of rows (I) will be equal to the number of time points in the chromatogram, and the number of columns (J) will equal the number of wavelengths in the DAD spectra. **U** represents the *abstract chromatograms* while **V** represents the *abstract spectra* when **X** is decomposed. The magnitudes of the eigenvalues (the square of the diagonal elements of **S**) are related to the importance of the associated spectrum and chromatogram as they contribute to the variance in the data. As a result, the data can be described, within experimental error, by only R factors, although the existence of noise in the data creates a number of eigenvalues that is equal to the number of columns or rows (whichever is smaller) in the data matrix.

The method of residual standard deviation (RSD) [45], which compares the RSD of the eigenvalues of the matrix, can be used to calculate R based on the noise that is known to be present in the system. The eigenvalues (λ_i) are determined by

$$\lambda_i = s_{ii}^2 \tag{6}$$

where s_{ii} are the diagonal elements of the singular value matrix, **S** [2]. The RSD is calculated as

$$RSD = \left(\frac{\sum_{r+1}^{j} \lambda_i^0}{I(J-R)} \right)^{1/2} \tag{7}$$

where I and J are the number of rows and columns in the data, respectively, R is the number of factors being considered, and λ_i^0 are the eigenvalues attributed to noise rather than to real chemical information [45]. The RSD is determined for a single factor (R = 1), and compared to the estimated experimental error in the data. If the calculated RSD is

greater than the estimated error inherent in the data, more factors are required to adequately describe the data. The calculation is repeated for two factors, and then three, and so on until all of the eigenvalues have been included (up to $R = J\text{-}1$). When the RSD is approximately equal to the estimated random error inherent in the measurement, the correct number of factors has been determined. Although this method of rank determination is not exact unless the noise level of the data is well established, it is simple to use and provides a reasonable estimate the number of components that are represented within a data set.

Another method of determining rank using SVD is to examine a plot of the singular values versus component number (called a scree plot); a break in the continuity in the scree plot is an indication of the number of components [46]. A more quantitative method that is also based on singular values is to calculate the percent variance explained by adding components to the model [46]. If all of the principal components are used, the explained variance in the data would be 100 %, by definition. A threshold can be chosen so that the number of components chosen represents, for example, 95 % of the variance in the data. The choice of the percentage of explained variance is highly dependent on the amount of noise and background present in the data.

For rank determination in three-way and higher order arrays, the array can be unfolded into two-way matrices, and the unfolded matrices can be analyzed by SVD as described above. For example, a four-way data array, $\underline{\mathbf{X}}$ ($I \times J \times K \times L$) can be unfolded as follows: $\mathbf{X}_1$ ($I \times JKL$), $\mathbf{X}_2$ ($J \times IKL$), $\mathbf{X}_3$ ($K \times IJL$) and $\mathbf{X}_4$ ($L \times IJK$) [2, 47] The subscripts on $\mathbf{X}$ indicate which mode is preserved, and the parenthetical notation

represents the dimensions of the unfolded matrix. When SVD analysis is carried out on each of these unfolded matrices, the diagonal of the singular value matrix contains the rank information for the mode that is preserved. Often in higher order data, each dimension may have a different rank. The PARAFAC model requires that the rank be the same in all of the modes, so in these cases the maximum rank of the four dimensions can be used to calculate the model. The result will usually be that some of the profiles in the lower rank modes will be highly correlated.

3.4. Multivariate Selectivity

The multivariate selectivity for a component in a mixture refers to the precision with which the PARAFAC model can resolve the component in the presence of overlapped signals in a multi-way data set. Multivariate selectivity can be used to evaluate the relative precision of quantification for specific compounds in a mixture relative to that of the pure (single component) sample. The magnitude of the multivariate selectivity for a component *r* is an indication of how orthogonal, or unique, the signal for that component is in the data set relative to all other components in the data. Multivariate selectivity (SEL_r) can be expressed as

$$SEL_r = \frac{\|NAS_r\|}{\|TS\|} \tag{8}$$

where $||NAS_r||$ is the norm of the net analyte signal matrix for component *r* (that is, the signal unique to that component), and ||TS|| is the norm of the total signal for the sample collected [48, 49]. In geometrical terms, the NAS is the sine of the angle between the

signals for component *r* projected onto the subspace defined by all other components in the sample [50].

The multivariate selectivity for component *r* in an N-way data set can also be defined as the diagonal of the matrix obtained by

$$SEL_r = \left\{ \left[\left(A_2^T A_2 \right) \bullet \left(A_3^T A_3 \right) \bullet \cdots \left(A_n^T A_n \right) \right]^{-1} \right\}_{rr}^{-\frac{1}{2}} \quad (9)$$

where $\mathbf{A_2}$, $\mathbf{A_3}$, ... $\mathbf{A_n}$ are the resolved PARAFAC profiles and the symbol • is the element-wise Hadamard product [51, 52]. The concentration profile, $\mathbf{A_1}$, is left out of the selectivity equation because the profiles are normalized by vector length to provide a maximum selectivity of one. By this definition, a selectivity of one indicates that a compound in a mixture can be quantified with the same precision as if it were in a pure sample. A selectivity of zero indicates that the analyte is completely overlapped and cannot be quantified at all. This particular definition of multivariate selectivity, derived by Olivieri [51], is useful because the resolved PARAFAC profiles of a data set can be used to calculate the selectivity directly and because the calculation is not computationally intensive. There are several other mathematical and matrix based derivations of the selectivity equations that have been published that are equivalent to Olivieri's version [49, 51-53].

3.5. Target Factor Analysis

TFA is a chemometric technique that is used to determine the presence or absence of a known reference spectrum within a set of overlapped spectra [54]. An LC-DAD

chromatogram can be considered to be a collection of spectra organized by retention time. Hence, the TFA algorithm is useful for finding a reference spectrum within such a data set. The first step of the procedure is to apply SVD and determine the rank of the data using one of the methods described above (RSD, scree plot, or percent explained variance). The first R columns of the abstract spectral matrix $\mathbf{V}$ are represented by $\overline{\mathbf{V}}$ (I x R). By selecting only the first R columns of $\mathbf{V}$, those contributions to the data that are not significant (*i.e.*, the noise) are excluded. The abstract spectra are aptly named; they do not have any chemical meaning and are merely mathematical representations of the data. In order to compare this set of spectra to a chemically meaningful reference spectrum, target transformation is carried out. The linear combination of the R abstract spectra are tested against the reference (target) spectrum to create a transformation matrix $\mathbf{T}$ (also called a rotation matrix), according to the equation

$$\mathbf{T} = \overline{\mathbf{V}}^{\dagger} \cdot \mathbf{L} \tag{10}$$

where $\mathbf{L}$ is the target spectrum and $\overline{\mathbf{V}}^{\dagger}$ is the pseudo-inverse of $\overline{\mathbf{V}}$. The transformation matrix $\mathbf{T}$ is used to rotate $\overline{\mathbf{V}}$ and transform it into a chemically meaningful spectrum, according to Equation 11,

$$\hat{\mathbf{L}} = \overline{\mathbf{V}} \cdot \mathbf{T} \tag{11}$$

where $\hat{\mathbf{L}}$ is the predicted spectrum. This transformation finds the linear combination of the first R abstract spectra that most closely represents the target spectrum.

The degree of correlation between $\mathbf{L}$ and $\hat{\mathbf{L}}$ is measured by the angle theta (θ), where

$$\theta = \cos^{-1}\rho \quad (12)$$

and ρ is the square root of the correlation coefficient between the predicted spectrum and the reference spectrum [28]. The two spectra can be represented as vectors in *J*-dimensional space (where *J* is the number of wavelengths in the spectrum) and θ is the angle between those two vectors. For two spectra that are identical ($\rho = 1.000$), the angle between them (θ) is 0° and is independent of the magnitude of the vectors. However, due to the presence of noise, it is more likely that two independently obtained spectra will have a small but non-zero value of θ. It has been suggested that a value of θ less than 10° indicates a high probability that two spectra are the same [12]; however, this variable is highly dependent on the amount of noise present in the system. Figure 2 shows an example of the results of a target test for a chromatogram against two different target spectra. In Figure 2A, the target spectrum (solid line) was a match to the predicted spectrum (dashed line), indicating that the target spectrum is present in the chromatogram that was tested. In Figure 2B, the target spectrum is not a match, which is confirmed by the higher value for θ.

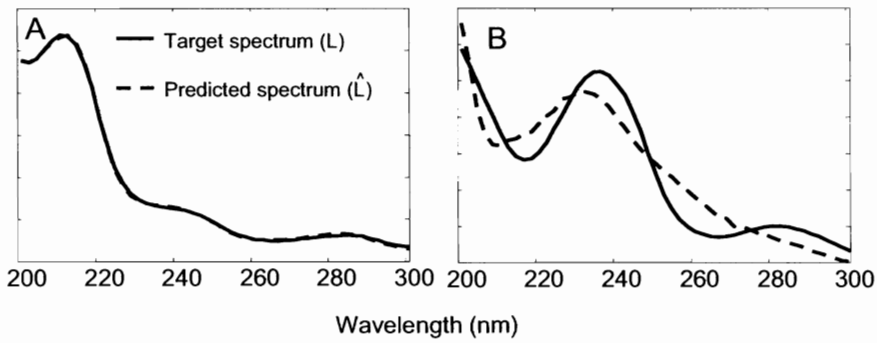

Figure 2. Comparison of a predicted spectrum (dotted line) and a target spectrum (solid line) where the target spectrum is (A) a match ($\theta = 0.8°$) and (B) not a match ($\theta = 7.7°$)

Lohnes *et al.* [12] introduced a specific application of TFA that they called window target testing factor analysis (WTTFA). The TFA algorithm described above is initially performed for the first W data points of a chromatogram (where W is a user defined window); after the analysis is completed in the first window, the window is incremented one time unit, and the analysis repeated. The window is then moved one data point at a time along the chromatographic time axis until the end of the chromatogram. The effect of applying the TFA algorithm in this sequential manner is that the location in the chromatographic separation space of a spectral match is revealed by plotting θ for each window as a function of time. While TFA alone only indicates the presence or absence of a spectrum somewhere within the chromatogram, WTTFA can specifically identify the retention time of a library match. An example of the WTTFA analysis of a chromatogram containing several unknown peaks is shown in Figure 3. In this case, the spectrum of benzoylecgonine was used as a target against the unknown chromatogram and the results show that a spectrum matching the target was located at approximately 0.8 minutes. This technique is particularly powerful in chromatograms with multiple unknown peaks where one or a few target compounds can be identified, and in chromatograms with overlapped components, where more than one target spectrum might be found in a single peak.

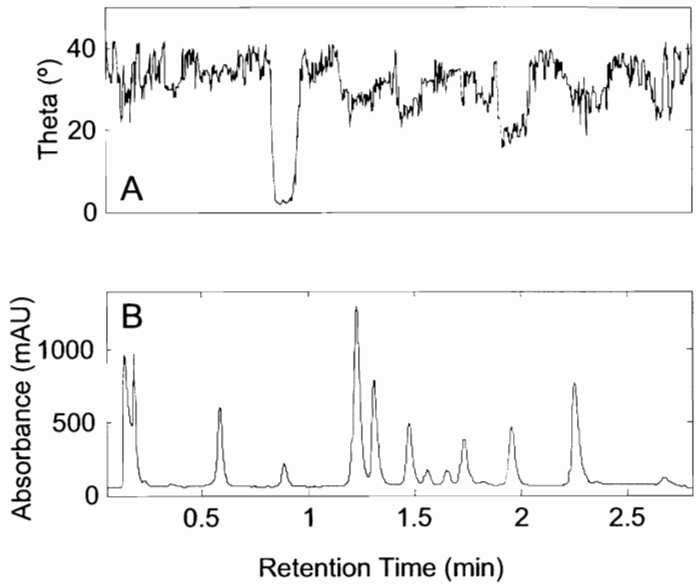

Figure 3. Results of WTTFA analysis on a chromatogram containing several unknown peaks. (A) Plot of theta vs. retention time using benzoylecgonine as the target spectrum, and (B) the chromatogram shown at 205 nm.

The methods presented in this chapter were used to analyze LC data throughout this research. The applications of these chemometric techniques to specific problems in analytical chemistry and LC separations in particular play a key role in the fields of toxicology, drug metabolism, and metabolomics.

CHAPTER 4. Liquid Chromatography – Mass Spectrometry: Development of an *In-vitro* Incubation Procedure for Screening of CYP2D6 Intrinsic Clearance Values

The goal of the work presented in this chapter was to use LC-MS and kinetic fitting algorithms as tools for the analysis of *in-vitro* drug metabolism reactions. Specifically, this project focused on the development of methods for studying in-vitro drug metabolism rather than identifying new metabolic pathways. To these ends, known drug systems, including PMMA, fluoxetine, MDMA, and dextromethorphan were used to validate the new methods developed here. This chapter is reproduced in part from reference [3], published in the Journal of Chromatography B ().

4.1. Pharmacokinetics and drugs of abuse

In-vitro pharmacokinetic studies, and specifically metabolism studies, have always been an important part of the drug discovery and development process [55]. Pharmacokinetics is the field of study that encompasses the disposition of drugs in the body, specifically absorption, distribution, metabolism, and excretion [56]. In particular, metabolism studies include the determination of the specific enzymes responsible for breaking down a drug, the kinetic parameters of enzyme interactions, and the products of the reactions [55, 57]. Understanding the metabolism of a new drug candidate is

important for predicting *in-vivo* clearance, and assessing potentially toxic or biologically active metabolites is also necessary before *in-vivo* testing of any new drug candidate can begin [55]. *In-vitro* methods can also be used to assess potential drug inhibition and drug-drug interactions, which are important because multi-drug use is common for the treatment of many diseases [58] and among illicit drug users [59].

Phenylalkylamine drugs, in particular MDMA, have received a great deal of attention in recent years due to the increasing incidence of their abuse among young people [60]. Potential interactions between these drugs and selective serotonin reuptake inhibitor drugs (SSRIs) such as fluoxetine (Prozac®) are of particular interest because these popular antidepressant drugs are often taken in combination with stimulant drugs like MDMA [59]. PMMA is a relatively new designer drug that is not as widely abused, but which has similar physiological effects as MDMA. Studies of trained rats indicate that PMMA has similar response properties as MDMA, but without the stimulant character of MDMA [61]. Both PMMA and a close structural analog, *p*-methoxyamphetamine (PMA) have been blamed for fatalities around the world. In one particular report, three case histories are cited in which the users were taking 'Ecstasy' tablets and died days later of complications believed to be from an overdose of PMMA or PMA [62].

4.2. Cytochrome P450

Cytochromes P450 (CYPs) are a group of membrane bound enzymes present in the liver that are responsible for the catalysis of numerous oxidative reactions involving

carbon, oxygen, nitrogen, and sulfur atoms in thousands of different substrates with many diverse structures [63]. CYPs have characteristics, including broad substrate specificity and broad regio- and stereoselectivity, that allow a vast number of compounds to be metabolized by a limited number of isozymes (about 100) [63]. The CYPs require a reductase as well as a cofactor such as nicotinamide adenine dinucleotide (NADPH) in the presence of oxygen in order to carry out the metabolism of many different xenobiotic compounds (any compound that is foreign to the body). The isozymes of CYPs in humans are divided into eighteen different families (denoted by an Arabic number) and forty-two subfamilies (denoted by a letter) based on the similarity of the amino acid sequences of the isozymes. Individual alleles are designated with another Arabic number following the subfamily designation [64].

The isozyme CYP2D6 is reported to be involved in the metabolism of about 12% of the most commonly prescribed pharmaceuticals [58], despite the fact that it accounts for only a small percentage (about 2%) of the CYPs found in the liver [65]. Some common CYP2D6 substrates include debrisoquine, tricyclic antidepressants, SSRIs including fluoxetine, various phenylalkylamines, and dextromethorphan, an over the counter cough suppressant [66]. CYP2D6 is a particularly interesting isozyme to study because it is known to be under-expressed in certain populations, which can lead to differences in drug metabolism between individuals [56]. It was first observed in the 1970s that a few volunteers participating in a clinical study of debrisoquine, a drug used to treat high blood pressure, suffered unexpected, adverse side effects. It was later determined that these volunteers were deficient in the enzyme required for the oxidation

of the drug, CYP2D6, and similar observations were later made for many other drugs [67].

There are several alleles of CYP2D6 with varying activity, as well as some that are not active at all, resulting in a range of diverse phenotypes [56]. Phenotypes are designated as poor or extensive metabolizers. It is reported that 5 – 10 % of Caucasians can be classified as CYP2D6 poor metabolizers. The enzyme in poor metabolizers is under-expressed or only the inactive alleles are expressed, and thus metabolism is not efficient [66, 67]. Extensive metabolizers express all active alleles. Dextromethorphan and debrisoquine, among others, are often used as probe drugs to phenotype an individual as a poor or extensive metabolizer of CYP2D6 substrates [67, 68]. Differences in drug metabolism can lead to severe toxicity or other adverse effects by altering the relationship between dose and blood concentration of the pharmacologically active drug. For example, there is a good deal of literature that describes the relationship between CYP2D6 activity and antidepressant toxicity and response [69]. Although many found no correlation, Rau *et al.* made a convincing case that CYP2D6 phenotype may have an important impact on the response and toxicity of patients to antidepressant therapy [70]. By the same logic, CYP2D6 phenotype may have an impact on the toxicity and potential for abuse of commonly abused drugs.

Drug interactions can occur when two substrates of the same enzyme are co-administered as the compounds compete for the active site in the enzyme. When drugs are ingested together, dangerous and potentially fatal overdoses can occur rapidly [65]. Understanding the mechanisms behind such interactions and the fate of the drugs in the

liver is of interest in both clinical and forensic laboratory settings. The development of rapid and robust *in-vitro* methods for characterizing CYP2D6 reactions will help to pave the way for understanding common drug interactions and help to identify or predict unknown interactions. In this work, the CYPD6 mediated reactions of PMMA and fluoxetine, shown in Figure 4, were used as model systems for developing a new *in-vitro* method to determine intrinsic clearance values.

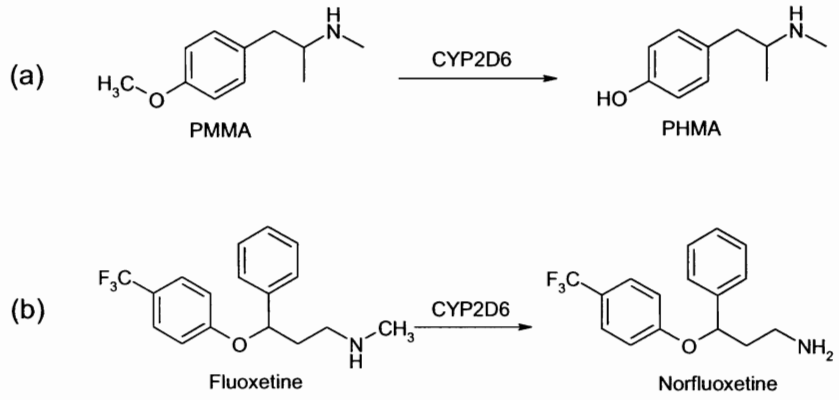

Figure 4. CYP2D6 mediated reactions discussed in Chapter 4. (a) PMMA O-demethylation to *p*-hydroxymethamphetamine (PHMA) [71] and (b) fluoxetine N-demethylation to norfluoxetine [72].

4.3. Enzyme kinetics and *in-vitro* methods

The simplest model for an enzyme-catalyzed reaction consists of a reversible binding of a substrate (S) to an enzyme (E) to form an enzyme-substrate complex (ES). This complex can break down irreversibly to form product (P). The reaction is shown in Equation 13,

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} P \tag{13}$$

where k_1, k_2, and k_3 are the micro-rate constants for each of the reactions. Because this model is not always sufficient to describe the observed experimental data, more complex models are often required. Methods for describing and fitting data to these models are useful in pharmacokinetic studies.

The most common type of experiment to determine *in-vitro* pharmacokinetic parameters is the steady-state (SS) approach [4]. The data obtained in an SS experiment is described by the Michaelis-Menten (MM) equation for the initial velocity (v_0) of the reaction as

$$\left(\frac{d[P]}{dt}\right)_{t=0} = v_0 = \frac{v_{max}[S]_0}{K_m + [S]_0} \tag{14}$$

where K_m is the Michaelis constant and v_{max} is the maximum velocity of the reaction. The key assumption in the derivation of Equation 14 is that the enzyme-substrate complex is at a steady-state and not changing appreciably (that is, d[ES]/dt = 0). The constants in the MM equation are related to the micro-rate constants by

$$K_m = \frac{k_2 + k_3}{k_1} \tag{15}$$

$$v_{max} = k_3[E]_0 \tag{16a}$$

$$v_{max} = k_3 \tag{16b}$$

where $[E]_0$ is the initial concentration of the enzyme species in the reaction mixture [73]. Equation 16b is used for v_{max} if the initial rate, v_0, is normalized to the initial enzyme concentration. The initial rate is measured at a single time point over a range of initial substrate concentrations. The selected reaction time must occur where the rate of

formation of product is still within the linear range (which is considered equivalent to t = 0). The SS plot of the initial rate of product formation (usually reported in moles of product normalized by time and enzyme concentration) is used to fit the data and obtain the constants K_m and v_{max}. This plot usually results in the familiar profile where the product formation approaches v_{max} asymptotically at high substrate levels, and K_m is equal to the substrate concentration and one-half of v_{max}.

Because the *in-vivo* substrate concentrations are usually likely to be much less than K_m, $[S]_0$ can be neglected in the denominator of Equation 14 [4], which allows the prediction of *in-vitro* intrinsic clearance as

$$CL_{int} = \frac{v_{max}}{K_m} \tag{17}$$

where CL_{int} is the intrinsic clearance. Multiplying Equation 17 by the substrate concentration, along with the appropriate scaling factors, permits the estimation of the rate of metabolism *in-vivo* from *in-vitro* parameters [4].

The MM equation successfully describes many enzyme systems, but it does not describe multiple substrate systems or atypical kinetic profiles. Because many xenobiotics do not follow conventional MM kinetics *in-vitro,* alternative models have been developed for atypical profiles using the steady-state assumption [74]. A commonly observed model in CYP2D6 systems is the two-site model, where the enzyme has two binding sites for the substrate but only one site is active and forms product [75]. This model is also referred to as a substrate inhibition model because at high substrate levels, the formation of product is inhibited by the substrate occupying the inactive site on the

enzyme. The steady-state curve for this reaction is characterized by a decrease in product formation at higher substrate levels. A Matlab® routine was developed by Sánchez-Ponce and Rutan based on Cha's method to mathematically fit any steady-state model defined by the user [6, 76]. The steady-state assumption is generally experimentally valid as long as the initial enzyme concentration is much less than the initial substrate concentration [56].

Recently published alternatives to the standard SS experiment include a direct injection LC/MS/MS technique [77], a pseudo first order kinetics method [78] and a substrate depletion approach [79]. Bhoopathy *et al.* [77] suggested that CL_{int} can be estimated using a direct injection technique with no stirring of the reaction mixture in a temperature-controlled LC autosampler tray. The first-order rate constant of elimination can be determined if the concentration of the probe drug is much less than K_m. They were able to determine CL_{int} by monitoring the depletion of substrate only. Schnell and Mendoza [78] examined the mathematical derivation of pseudo-first order (PFO) kinetics and discussed the validity of such approaches. They stated that approximating PFO conditions experimentally requires only that $[S]_0 << K_m$ and is independent of the initial enzyme concentration, contrary to previous reports [80, 81] that claimed that one of the reactants, $[E]_0$ or $[S]_0$, should be in large excess. Jones and Houston reported the application of a substrate depletion method, which has the advantage that the specific metabolic pathways of the drug do not need to be known [79].

In this work, a new general enzyme (GE) method was developed that uses the rate laws for the elementary reaction steps and an ordinary differential equation solver to find

mathematical solutions to the differential equations [5]. For the enzyme reaction mechanism shown in Equation 13, the change in product with time is expressed generally as

$$\frac{d[P]}{dt} = k_3[ES] \tag{18}$$

and analogous differential equations can be written for each of the other species (E, S, and ES) involved in the general enzyme reaction shown in Equation 13. Using the basic rules of chemical kinetics, any enzyme mechanism can be modeled by representing the rate of each step in the mechanism with a differential equation and by monitoring only one or two of the species participating in the reaction [82]. Bezemer and Rutan have previously described an approach for the fitting of kinetic data to any model [5], and more recently presented the method specifically for the fitting of enzyme kinetic data to the general kinetic model [83]. This approach uses the ordinary differential equation solver in Matlab® to find numerical solutions to the differential rate equations.

4.4. Methods and Materials

The authentic standards for PMMA, fluoxetine, and norfluoxetine were obtained from Alltech (State College, PA, USA) as unscheduled chromatographic standard solutions of 1 mg/mL of the free base in methanol. PHMA and NADPH tetrasodium salt were obtained from Sigma (St. Louis, MO, USA). Formic acid, 98 %, was obtained from Fluka (Steinheim, Germany), 6 M ammonium hydroxide was obtained from Ricca Chemical Company (Arlington, TX, USA), and acetonitrile was obtained from EMD (Gibbstown, NJ, USA). Sodium hydrogen phosphate was obtained from EM Science

(Cherry Hill, NJ, USA), and phosphoric acid (85 %) was obtained from Fisher Scientific (Pittsburgh, PA, USA). Ultrapure 18 MΩ-cm water dispensed in house was used to prepare all chromatographic eluents and buffers. For the enzyme incubation experiments, CYP2D6*1 Supersomes™, (baculovirus-insect cell expressed with coexpression of CYP450 reductase) and control Supersomes™ (from wild-type baculovirus-insect cells) and an NADPH regenerating system (solutions A and B) were all obtained from BD Biosciences (Bedford, MA, USA).

4.4.1. Incubation Experiments

Steady-State Incubations. The SS incubation method used in this work was developed based on previously published methods [68, 84, 85]. A 0.10 M phosphate buffer was prepared with Na_2HPO_4, and H_3PO_4 was used to adjust the pH to 7.4. A 10 mM solution of NADPH was prepared on the day of analysis by dissolving the appropriate amount of the tetrasodium salt in phosphate buffer. Stock solutions of PHMA, PMMA, fluoxetine, and norfluoxetine were prepared in phosphate buffer and diluted to make calibration standards as outlined in Table 1. The reaction components were added to a 1.5 mL microcentrifuge tube in the following order: enzyme, substrate(s), phosphate buffer, and then NADPH to start the reaction. The total incubation volume was brought to 500 µL with phosphate buffer. The concentration of NADPH in the final reaction mixture was 600 µM, and the concentrations of the other reactants were as shown in Table 1. The tubes were placed in a Precision metabolic shaker (Winchester, VA, USA) at 37 °C immediately after the NADPH was added. After

the time indicated in Table 1, the samples tubes were placed on ice, and 750 µL of ice cold acetonitrile was added to stop the reaction. Samples were centrifuged on a Biofuge 17R centrifuge from Baxter Scientific Products (West Chester, PA, USA) at 9,300 g (12,000 rpm) for 10 min. The supernatant was filtered through a 0.2 µm nylon filter, and placed into an autosampler vial for HPLC analysis. Unanalyzed portions were stored in the freezer.

Table 1. Experimental conditions for steady-state experiments.

Analyte	Calibration Range (µM)	Incubation Concentrations (µM)	Enzyme Concentration (µM)	Incubation Time (minutes)
PMMA	2 – 500	0 – 700 [a]	0.02 [a]	45
PHMA	5 – 80	0 – 250 [b]	--	--
Fluoxetine	0.2 – 10	0 – 500 [a]	0.02 [a]	120
Norfluoxetine	0.05 – 5	0 – 20 [b]	--	--

[a] Concentration added to incubation mixture
[b] Concentrations detected by LC-MS

General Enzyme Kinetics Incubations. Calibration standards for all analytes were prepared as shown in Table 2. NADPH regenerating solution A and solution B were mixed in a ratio of 25:5 and kept at 37 °C until added to the reaction vessel. The following components were mixed in an HPLC autosampler vial: enzyme, substrate (PMMA and/or fluoxetine), and phosphate buffer for a total incubation volume of 400 µL. The incubation mixtures were injected directly into the LC-MS system without further preparation. One injection was made before the NADPH mixture was added, and then an injection was made every 4 (for PMMA) or 6 (for fluoxetine) minutes after the

NADPH (30 μL of the mixed regenerating solution) was added to the mixture. The autosampler tray was held at 37 °C for the duration of the incubation. For the inhibition experiment, 5 μM of fluoxetine was added to the reaction mixture before adding the NADPH.

Table 2. Experimental conditions for the general enzyme experiments.

Analyte	Calibration Range (μM)	Incubation Concentrations (μM)	Enzyme Concentration (μM)	Incubation Time (minutes)
PMMA	0.4 – 12	5 [a]	0.02 [a]	0 – 75
PHMA	0.4 – 12	0 – 5 [b]	--	--
Fluoxetine	2 – 10	10 [a]	0.05 [a]	0 – 150
Norfluoxetine	2 – 10	0 – 4 [b]	--	--

[a] Concentration added to incubation mixture
[b] Concentrations detected by LC-MS

4.4.2. LC-MS Analysis

All chromatographic separations were carried out on a Waters Alliance 2795 LC system equipped with a heated autosampler and column compartment (Waters Corp., Milford, MA, USA). The column was thermostated to 40 °C unless otherwise noted. A guard column and an in-line filter were used for all chromatographic separations. All mobile phases were filtered through 0.45 μm membrane filters before use. The injection volume from the autosampler was 10 μL. Detection was accomplished with a Thermo LCQ XP Deca Plus ion trap mass spectrometer equipped with an ESI source (Thermo Electron Corp., Waltham, MA, USA). A divert valve was used in front of the

electrospray source to avoid contamination from buffer salts in the chromatographic dead volume.

Calibration curves for each analyte were constructed from the standards as shown in Tables 1 and 2. Standards were made in both "clean" phosphate buffer matrix, and in matrix containing 0.020 μM of insect control enzyme. The insect control standards were used for quantification of the analytes in the incubation mixtures. Levallorphan was used as an internal standard for the fluoxetine incubations (added to the mixtures), and a post-column infusion [86] of *d*-amphetamine was used as an internal standard for the PMMA incubations. However, the use of the internal standard for the quantification of the components resulted in a degradation of the precision of the calibration parameters; therefore direct calibration in conjunction with the resolved responses from the curve fitting analysis was employed for quantification.

Analysis of Steady-State Incubations of PMMA. The LC conditions for the SS incubations of PMMA (PMMA-SS method) were as follows: mobile phase A was 98 % deionized water, 2 % acetonitrile, and 0.01 % formic acid; mobile phase B was 98 % acetonitrile, 2 % water, and 0.01 % formic acid. The pH of the aqueous phase was approximately 3. The flow rate used was 250 μL min^{-1} and the column output went into the electrospray source via the divert valve on the mass spectrometer. The mobile phase gradient was from 8 % to 20 % mobile phase B from 2 to 12 minutes, then from 20 % to 30 % B from 12 – 14 minutes, then from 30 % back to 8 % B from 14 – 18 minutes. The total run time was 25 minutes, including column re-equilibration time. The chromatographic column used was a 50 x 4.6 mm Phenomenex Luna C18(2) stationary

phase, with 5 μm particles (Phenomenex, Torrance, CA, USA). The ESI settings were as follows: the spray voltage was set at 5.50 kV, and the capillary temperature was set at 275 °C. Nitrogen was used for the drying gas and the auxiliary gas. Full scan mode was used for detection of analytes.

Analysis of General Enzyme Kinetics Incubations of PMMA. The LC conditions for the GE incubations of PMMA alone and the inhibition experiment (PMMA-GE method) were as follows: mobile phase A was a 10 mM ammonium formate buffer, prepared gravimetrically with appropriate amounts of formic acid and ammonium hydroxide to achieve a pH of 3.6. Mobile phase B was 100 % acetonitrile. The flow rate used was 650 μL min^{-1}, and the effluent from the column went to the ionization source via the divert valve on the mass spectrometer. The column was thermostated to 50 °C. The gradient was from 3 % to 33 % mobile phase B from 0.6 – 0.7 minutes, stayed at 33 % B until 1.9 minutes, and then went from 33 % to 3 % B from 1.9 – 2.0 minutes. The total run time was 3 minutes, including column re-equilibration time. The chromatographic column used was a 20 x 2.1 mm Betasil C18 DASH HTS (Thermo Electron Corp. Waltham, MA, USA) with 5 μm particles. The mass spectrometer was used in selected reaction monitoring (SRM) mode to select the fragmentation products of PMMA and PHMA (m/z 150 and 135 respectively). The ESI settings were as follows: the spray voltage was set at 4.5 kV, and the capillary temperature was set at 200 °C. Nitrogen was used for the drying gas and the auxiliary gas.

Analysis of Fluoxetine Incubations. Both the SS and GE kinetics incubations of fluoxetine and CYP2D6 were analyzed as follows (FLX method): The mobile phases

were the same as for the PMMA-GE method, and the column was the same as for the PMMA-SS method. The separation was carried out under isocratic conditions with 33 % mobile phase B. The total run time was 6 minutes. The mass spectrometer was used in full scan mode for the detection of the analytes. The ESI settings were as follows: the spray voltage was set at 4.5 kV, and the capillary temperature was set at 200 °C. Nitrogen was used for the drying gas and the auxiliary gas.

4.4.3. Data Analysis

The XCalibur® software program (Thermo Electron Corp., Waltham, MA, USA) was used to determine the peak areas of the analytes from the LC-MS chromatograms. The file converter tool in XCalibur® was used to convert collected chromatograms into text files, and a Pascal program written in house for MS-DOS was used to convert the text files into a matrix format suitable for analysis in Matlab®. Kinetic analysis of all collected data was carried out in the Matlab programming environment, using Matlab®, ver. 7.0.4 (Mathworks, Natick, MA, USA). The fALS routine described in reference [13] was used to resolve the chromatographic, spectral, and concentration profiles from the LC-MS data. A least-squares fitting routine with a built in steady-state constraint [6] was used to fit the curves obtained from the SS experiments, and a general kinetic fitting function [5] was used to fit the curves obtained from the GE experiments.

4.5. Comparison of General Enzyme and Steady-State Approach

Rather than following the classical steady-state approach, wherein multiple substrate levels are monitored after a fixed incubation time [56], the new method presented in this work uses a general kinetic approach (GE method). The change in concentration of a single substrate level is measured as a function of time, and the resulting data is fit using an ordinary differential equation solver and a kinetic fitting routine previously developed [5]. The advantages to this method are that fewer raw materials are needed (including costly enzyme preparations), sample preparation time is significantly less, and the concentrations of the species can be tailored to suit the detection limits of the instrument being used.

PMMA and fluoxetine were used as the systems to serve for validating the new method. Several systems were considered over the course of this project in addition to the ones presented here, including the O-demethylation of dextromethorphan to dextrorphan, and the O-demethylation of MDMA to ±-3,4-dihydroxymethamphetamine. Several SS experiments were carried out with these systems and many simulations were run to predict the experimental results. Due to the expense of the instrumentation and the reagents involved, many replicate experiments could not be performed, and so the experiments had to be planned very carefully. Both dextromethorphan and MDMA have well characterized metabolic profiles that would have made them suitable test cases. However, the lack of an authentic standard for the MDMA metabolite and the fast kinetics of the dextromethorphan made them less than ideal candidates. The detection limits for the PMMA, fluoxetine, and their metabolites were adequate, and simulations

based on published pharmacokinetic parameters of both drugs indicated that they were well suited to test the new GE method.

A fast LC-MS method was used to quantify the substrates and product(s) present in enzyme incubation samples. The results discussed below show that consistent results could be obtained for the intrinsic clearance of fluoxetine and PMMA using both the GE and SS methods. Although the fit constants (the micro-rate constants for the GE method and the K_m and v_{max} values for the SS method) cannot be calculated with high precision by either fitting approach due to the fact that they all co-vary significantly, the intrinsic clearance can be calculated using Equations 15 – 17. Using Monte Carlo error estimation methods confirmed that the error in CL_{int} is relatively low.

The enantiomeric specificity of enzymatic reactions has received a good deal of attention in the literature. Both PMMA and fluoxetine have chiral centers; therefore, the possibility of differential *in-vitro* metabolism of the stereoisomers exists. The drugs used in this study were racemic mixtures (as are the corresponding street and marketed drugs) and no attempt at differentiating the metabolism of the steroisomers was made. However, several studies have been published discussing the possibility of chirality playing a role in the metabolism of chiral drugs [87-90]. Caldwell's thorough review article described the effect of enantiomeric discrimination in drug metabolism for several systems at both the substrate and product level [87]. There are no studies specifically on the stereoselective metabolism of PMMA; however, several sources have published values for K_m and v_{max} of the R and S isomers of MDMA. Tucker *et al.* reported a K_m of 1.72 ± 0.12 μM and 2.90 ± 0.10 for (+)-MDMA and (–)-MDMA, respectively [91]. They

also cited several older reports that the neurotoxic effects of MDMA were isomer specific and that the enantiomers of MDMA and MDA may have different behavioral effects, however they did note that the differences in the *in-vitro* metabolism parameters were small.

Based on these published reports, it would not be unexpected to find some difference in the metabolism of the enantiomers of PMMA and fluoxetine. Simulations of the kinetic results for racemic mixtures of MDMA (based on the reports of the enantioselective pharmacokinetic constants) indicated that it is unlikely that the clearance values for the isomers could be resolved in practice. The simulations of the MDMA *in-vitro* reactions showed that a racemic mixture would result in a K_m and v_{max} value that are approximately averages of the (+) and (-) isomers and within the standard error of the reported values. A detailed analysis of stereospecific metabolism is beyond the scope of the screening method described here.

4.5.1. PMMA Incubations

Carrying out *in-vitro* metabolism experiments under classical SS conditions allows for the calculation of the constants K_m and v_{max} by fitting the data to a steady-state model as described in reference [6]. The intrinsic clearance can then be estimated using the calculated parameters. Figure 5 shows the results of the SS experiment used to characterize the *in-vitro* metabolism of PMMA. The LC-MS data were resolved using the fALS algorithm described in reference [13], which allows flexible implementation of chemically relevant constraints on a component-by-component basis. Spectral selectivity

and unimodality constraints were used for the analyte components and non-negativity was applied to all of the components (analytes and background).

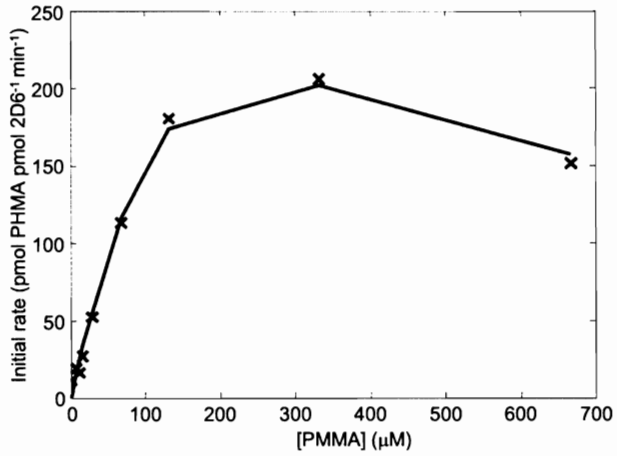

Figure 5. Results of the steady-state experiment for PMMA, fit to a substrate inhibition profile. $CL_{int} = 2.7 \pm 0.2$ μL pmol 2D6^{-1} min^{-1}, $SE_{fit} = 10.2$.

The results from the SS experiment with PMMA and CYP2D6 showed an atypical kinetic profile, with the concentration of PHMA decreasing at the highest substrate concentration. This pattern is characteristic of a substrate inhibition model [4], and forcing a fit to the MM equation resulted in an overestimation of the CL_{int} and a poor fit quality. The data were instead fit to the substrate inhibition model suggested by Tracy [4]. The calculated intrinsic clearance of 2.7 ± 0.2 μL pmol 2D6^{-1} min^{-1} was approximately 10-fold less than the intrinsic clearance determined from the data of Staack *et al.* [71]. The discrepancy is likely due to the fact that they used 5 mM Mg^{+2} in their incubations, while Mg^{+2} was not used in these experiments. Some experimental evidence will be required in order to confirm this suspicion. Adjustment of the reaction parameters (*e.g.*, concentration, ionic strength, order of addition of reactants) will be the

subject of future studies. Another factor that may have led to the significant difference in the reported kinetic parameters is that substrate concentrations greater than 400 μM were not employed [71].

The results from the GE incubation of PMMA with CYP2D6 are shown in Figure 6. These results show a general case where the depletion of substrate and the formation of product are measured as a function of time. The data were fit to the general enzyme model shown in Equation 13 to determine intrinsic clearance. The GE CL_{int} of PMMA, 3.0 ± 0.6 μL pmol $2D6^{-1}$ min^{-1}, was within experimental error of the value calculated from the SS experiment. Atypical kinetic profiles (*i.e.*, substrate inhibition) were not used for the GE experiment because the substrate concentration was low (5 μM) and substrate inhibition only occurs at high substrate concentrations. These results show that the estimation of intrinsic clearance obtained using the GE method is comparable to that obtained using the SS method.

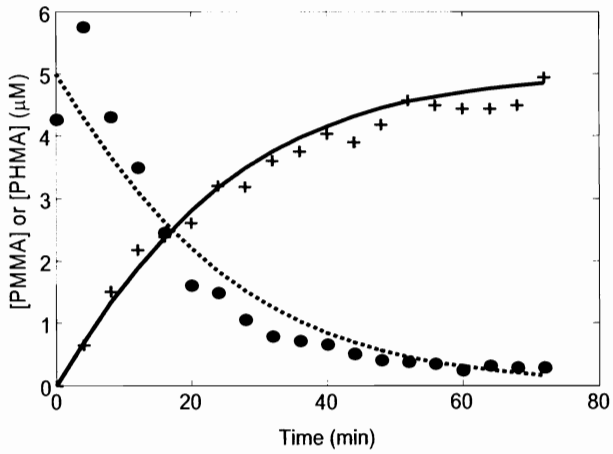

Figure 6. Results of GE experiment for PMMA. The fit for the model shown in Equation 13 is shown for the formation of product (solid line and '+') and the depletion of substrate (dotted line and '●'). CL_{int} = 3.0 ± 0.6 μL pmol $2D6^{-1}$ min^{-1}, SE_{fit} = 0.40.

4.5.2. Fluoxetine Incubations

The results from both the SS and GE experiments with fluoxetine and CYP2D6 strongly indicated the existence of atypical kinetic profiles. Atypical kinetic profiles have been previously observed for the metabolism of fluoxetine in the literature. Margolis *et al.* [92] showed that fluoxetine follows MM kinetics at relatively low concentrations, but Ring *et al.* [72] have shown that the pure R-fluoxetine enantiomer follows a substrate inhibition profile. There have also been previous reports that fluoxetine is metabolized into multiple other metabolites including hippuric acid and p-trifluoromethylphenol [93]. The m/z values for hippuric acid and p-trifluoromethylphenol are 179 and 162 amu, respectively, but these two compounds were not detected in the present experiments because the ion intensity data was only collected for m/z values ranging from 250 to 350 amu. The flexibility of the fitting algorithms used for both the SS experiment [6] and for the GE experiments [5] allowed a model to be designed that best described the experimental data and was consistent with previous reports of atypical kinetics observed in the fluoxetine and CYP2D6 system.

A plot of v_0 versus $[S]_0$ for the fluoxetine SS experiment is shown in Figure 7. This curve was consistent with the biphasic kinetic model suggested by Korzekwa, *et al.* [94] where the enzyme has two binding sites for fluoxetine. This particular biphasic profile indicates that one of the binding sites has a much higher affinity (*i.e.*, a lower K_m) than the other. Incorrectly forcing the data to fit to the MM model poorly predicted the norfluoxetine concentration at both the low and high fluoxetine concentrations. The

quality of the fit obtained and the precision of the calculated clearance was compromised by the high level of noise in the data. However, three points (indicated in grey in Figure 7) were omitted as outliers based on a plot of the known concentration of fluoxetine versus the resolved concentration profiles of fluoxetine.

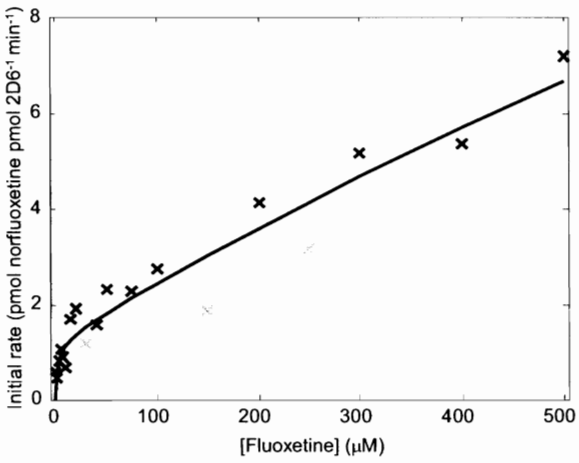

Figure 7. Results of the steady-state experiment for fluoxetine and CYP2D6. The data were fit to the model shown in Figure 9. $CL_{int} = 0.33 \pm 0.17$ μL pmol $2D6^{-1}$ min^{-1}, $SE_{fit} = 0.33$. The points indicated in grey are those that were omitted as outliers.

Using the general enzyme reaction (Equation 13) to fit this data gave a poor fit and a high standard error. The fluoxetine concentration continued to decrease after the concentration of norfluoxetine leveled off, suggesting the possibility that another product was being formed that contributed to the overall clearance of fluoxetine. Also, the rate of decrease of the fluoxetine concentration appeared to slow down toward the end of the incubation, which indicated that the enzyme was losing activity over the rather long time course of the experiment (as reported previously [95]).

A modified mechanism was developed to incorporate the atypical kinetics observed in both the SS and GE experiments. The data and the fit for the GE experiment

are shown in Figure 8. The modified mechanism is shown in Figure 9. The mechanism modeled the biphasic kinetics (9a and 9b) seen in the SS experiment, the formation of a second product (9c), and the loss of activity of the enzyme (9d) observed in the GE experiment.

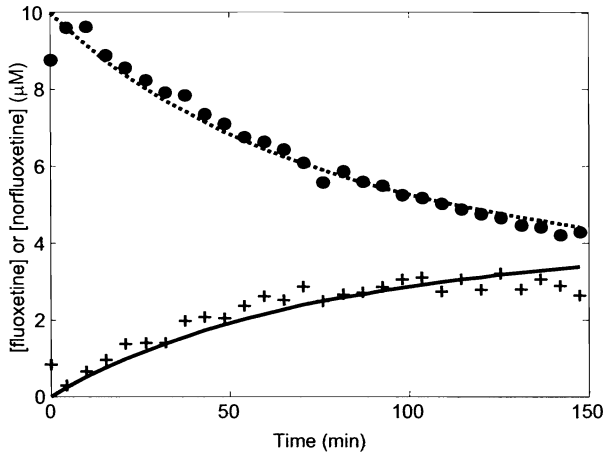

Figure 8. Results of GE experiment for fluoxetine. The fit is shown for the model in Figure 9 as the formation of product (solid line and '+') and the depletion of substrate (dotted line and '●'). $CL_{int} = 0.188 \pm 0.013$ μL pmol 2D6^{-1} min^{-1}, $SE_{fit} = 0.34$.

$$\text{(a)}\quad E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

$$\text{(b)}\quad ES \underset{k_5}{\overset{k_4}{\rightleftharpoons}} ESS \xrightarrow{k_6} E + P$$

$$\text{(c)}\quad ES \xrightarrow{k_7} E + Q$$

$$\text{(d)}\quad E \xrightarrow{k_8} DE$$

Figure 9. Modified enzyme mechanism for fluoxetine experiments. S is fluoxetine, P is norfluoxetine, Q is a second product, and DE is the deactivated enzyme. (a) General enzyme reaction; (b) second active (low affinity) site; (c) formation of second product; and (d) decay of enzyme activity.

The CL_{int} from each pathway was combined in order to obtain an overall CL_{int} [96],

$$CL_{int} = \frac{V_{max1}}{K_{m1}} + \frac{V_{max2}}{K_{m2}} + \frac{V_{max,Q}}{K_{m1}} \quad (19)$$

where v_{max1}, K_{m1}, v_{max2}, and K_{m2} are the maximum rates and the Michaelis constants of the two enzyme binding sites, and $v_{max,Q}$ is the maximum rate for the reaction of the enzyme substrate complex (ES) forming product Q. Equations 15 and 16 were used to calculate the constants in Equation 19 for each pathway. The formation of the second product Q was not included in the calculation of intrinsic clearance; for both experiments the reported clearance is with respect to the formation of norfluoxetine only, since it was the only product measured. For the GE experiment, the biphasic portion (Figure 9b) of the model shown in Figure 9 was not included in the fitting of the GE experimental data, because the low substrate concentrations employed in this experiment did not warrant it. The CL_{int} for the GE experiment with respect to the formation of norfluoxetine was 0.188 $\pm$ 0.013 μL pmol $2D6^{-1}$ min^{-1} and the CL_{int} for the SS experiment was 0.33 ± 0.17 μL pmol $2D6^{-1}$ min^{-1}.

4.5.3. Inhibition of PMMA by Fluoxetine

In the inhibition experiment, the inhibition of PMMA metabolism by fluoxetine was modeled using the GE approach. The results of a GE incubation of PMMA with CYP2D6 in the presence of 5 μM fluoxetine are shown in Figure 10. The intrinsic clearance for PMMA in this system was 0.40 ± 0.14 μL pmol $2D6^{-1}$ min^{-1}. There is

nearly an order of magnitude decrease in the clearance for PMMA in the presence of an equimolar amount of fluoxetine. Fluoxetine has previously been shown to be a potent inhibitor of CYP2D6 [97], and this relatively simple experiment demonstrates that the change in intrinsic clearance can be determined. Because fluoxetine is also a substrate of CYP2D6, this method could also be used to study the metabolic interaction between the two drugs.

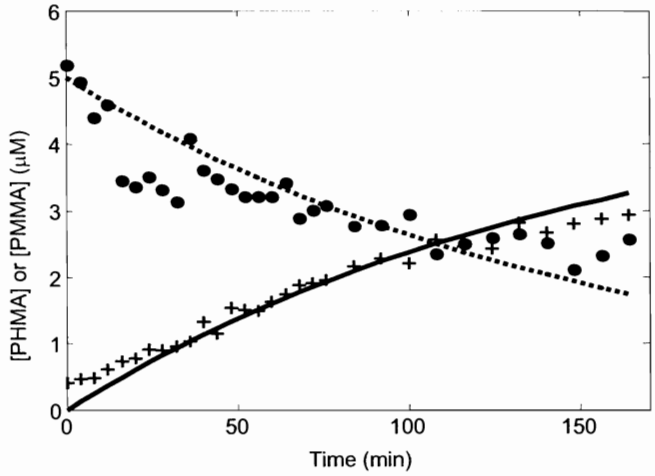

Figure 10. Inhibition of PMMA by fluoxetine. The fit is shown for the formation of PHMA (solid line and '+') and the depletion of PMMA (dotted line and '●'). $CL_{int} = 0.40 \pm 0.14$ μL pmol 2D6^{-1} min^{-1}, $SE_{fit} = 0.37$.

The results of all of the experiments presented in this chapter are summarized in Table 3. Although the GE and SS fluoxetine incubations required a modified mechanism to fit the data, it is shown here that the *overall* clearance of fluoxetine from a system with respect to norfluoxetine is consistent (within experimental error) regardless of the experimental method. The results of both of these experiments show that consistent results can be obtained using the new GE method and the SS method, even when

different models are needed to fit the experimental data. The new method proved to be a complement to traditional *in-vitro* metabolism studies.

Table 3. Summary of all results obtained in Chapter 4.

	CL_{int} – SS experiment ($\mu L\ pmol\ 2D6^{-1}\ min^{-1}$)	CL_{int} – GE experiment ($\mu L\ pmol\ 2D6^{-1}\ min^{-1}$)
PMMA	2.7 ± 0.2	3.0 ± 0.6
Fluoxetine	0.33 ± 0.17	0.188 ± 0.013
PMMA – Fluoxetine	--	0.40 ± 0.14

CHAPTER 5. LC-DAD and Target Factor Analysis as a High-throughput Screening Method for Drugs of Abuse

The work presented in this chapter represents the results of a collaboration with Prof. Peter W. Carr's group at the University of Minnesota Department of Chemistry. The goal of this project was to develop a fast screening method for drugs in biological matrices. A fast chromatographic method was used in combination with a TFA algorithm to identify samples that contained compounds of interest. The library of drugs used in this study contained 47 compounds of toxicological relevance. The work was submitted for publication in two parts to the Journal of Chromatography A. The chromatographic method development was presented in part I of the series [8], and the chemometrics and data analysis portion of the work can be found in part II of the series [7]. Parts of this chapter are reproduced in part from reference [7], published in the Journal of Chromatography A ().

5.1. Drug Screening

Forensic drug and toxicology laboratories have an on-going need for rapid, simple assays for screening biological samples suspected of containing drugs and metabolites of toxicological interest. Screening methods are used to identify samples that do not contain drugs and eliminate the expense and time required for a longer method. Current

techniques for such toxicological screening include enzyme immunoassay (EIA) [98-101] and LC-DAD [102]. EIA is usually used in screening for substances with similar properties or structures (*i.e.*, drug classes such as benzodiazepines) [103]. LC-DAD is also useful as a more specific screening method, and there are many recent publications dealing with the use of LC-MS in forensic toxicology [104-106]. LC-MS and GC-MS are typically used as confirmatory methods [98, 107]. A recent review by Maurer [108] gives an overview of chromatographic techniques commonly applied in toxicological testing.

Absorbance detectors (including DAD) are much less expensive and relatively simpler to use than MS detectors. LC-DAD is a fast and robust method for *screening* biological samples in conjunction with a library search algorithm to quickly identify those samples that require confirmatory testing. LC-DAD methods for screening in toxicology were recently reviewed by Pragst [102]. For example, Herzler *et al.* [109] showed that DAD data can be used to selectively identify abused substances in spectrochromatograms based on comparison to a library of over 2500 "toxicologically relevant" substances. Their method relied on the calculation of a 'similarity index' (related to the correlation coefficient) to determine the similarity between a spectrum in an unknown chromatogram and a library spectrum. In addition to spectral matching, a relative retention time was also used to identify the substances of interest.

5.2. Fast Gradient Liquid Chromatography

One of the main shortcomings in gradient elution RPLC has been the time required to re-equilibrate the system between runs. As a consequence, the throughput of an otherwise fast method can be significantly reduced. However, recent improvements in the methodology for ultra-fast gradient LC and significant reductions in required re-equilibration times have greatly improved the speed and efficiency by which these separations can be carried out. Schellinger *et al.* [110-112] have recently investigated the minimum essential re-equilibration time in a gradient elution to provide excellent retention time reproducibility (± 0.004 min) between runs without the requirement for full column equilibration. As a result of this work, a gradient LC-DAD method has been developed with a gradient time of less than three minutes and a re-equilibration time of less than one minute for a total cycle time of four minutes. An instrument modification employing two pumping systems further reduces the total cycle time of a single analysis to only 2.8 minutes. The development and demonstrated reproducibility of this method is discussed extensively by Stoll *et al.* [8].

In this work, the method of TFA was used as part of a library search algorithm to identify spectra contained within an LC-DAD chromatogram. The main distinctions between the method described here and previously published work are (i) the use of a high-speed, gradient elution LC method, (ii) the use of a corrected retention index rather than retention times for library matching, and (iii) the use of a factor analysis algorithm that can readily resolve overlapped chromatographic peaks. The data collected by Stoll *et al.* in reference [8] was analyzed as part of this study, in addition to drug-free blood

samples and blood samples spiked with drugs. The TFA algorithm resolves overlapped peaks in a chromatogram and is a fast and accurate method for determining which samples should go on for a confirmatory test. If no compounds of interest are detected, the analysis is complete in only a fraction of the time required for the full analysis. The results provided by LC analysis provide important information for the selection of the appropriate chromatographic and mass spectrometric conditions for subsequent confirmatory analyses.

5.3. Application of Retention Index

Shifts in retention time between runs (both systematic and random) often plague LC separations, usually due to slight changes in temperature, mobile phase composition, and column aging. Although these factors are largely controllable by column heating, gravimetric preparation of solvents and proper storage of columns, there are still uncontrollable factors that can cause a small degree of retention shifting. These shifts can become significant over time. To combat these effects, an index based on interpolation between primary and secondary standards that are run with the analytes of interest can be used for long term application of retention libraries. For a screening method that uses retention data, the method should account for day-to-day, column-to-column, and run-to-run shifts in retention. Here, the retention index as introduced by Smith [113] is used, and a modification introduced by Bogusz [114, 115] is included for further correction of shifting.

The observed retention index for compound *i*, $I_{obs,i}$, is calculated from its retention time, $t_{R,i}$, as

$$I_{obs,i} = 100n + 100\left(\frac{t_{R,i} - t_{R,n}}{t_{R,n+1} - t_{R,n}}\right) \quad (20)$$

where n is the number of carbons in the primary standard eluting immediately before compound *i*, $t_{R,n}$ is the retention time of the primary standard eluting immediately before compound *i*, and $t_{R,n+1}$ is the retention time of the primary standard eluting immediately after compound *i*. The I_{obs} values for each drug in the library were calculated and compiled into a list to represent the standard values (I^o). The difference in between-day standard deviations for retention time and retention index is shown in Figure 11. Eight different library compounds were measured over ten different days (not consecutive) and the percent relative standard deviation (% RSD) was calculated for t_R and I_{obs}. Clearly, there is significant improvement in reproducibility between days using the retention index.

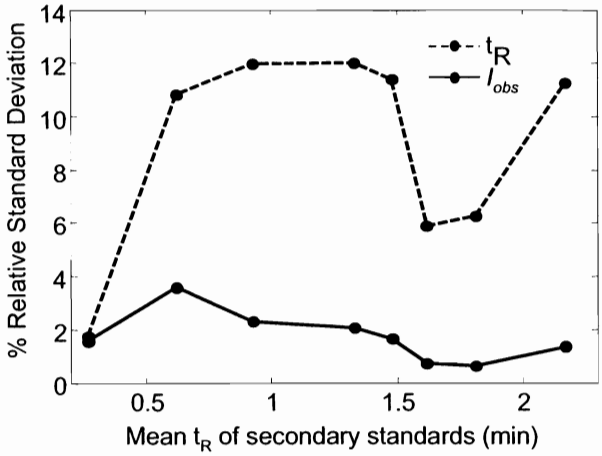

Figure 11. Percent relative standard deviation of retention time (t_R) and observed retention index (I_{obs}) of 8 secondary standard compounds injected over 10 non-consecutive days as a function of their average retention times.

Further improvement can be made by using a corrected retention index, $I_{,corr}$. This value is calculated using the retention times of secondary standards chosen from the list of compounds in the library, by Equations 21-23 [115]:

$$a = \frac{I^o_{s+1} - I^o_s}{I_{obs,s+1} - I_{obs,s}} \tag{21}$$

$$b = aI^o_{s+1} - I_{obs,s+1} \tag{22}$$

$$I_{corr,i} = aI_{obs,i} + b \tag{23}$$

where I^o_s and I^o_{s+1} are the standard retention indices of the secondary standards eluting immediately prior to and following compound i, and $I_{obs,s}$ and $I_{obs,s+1}$ are the retention indices of the secondary standards eluting immediately prior to and following compound i calculated for the day of analysis using Equation 20. The secondary standards are

library compounds chosen to encompass the retention times of all of the library compounds (early, middle and late eluters are included).

5.4. Evaluation of Screening Methods

A screening method should ideally give a minimum number of false positive and false negative results, and should quickly and accurately identify any samples that require further testing. In the case of a false positive result, the sample will be further examined by a more reliable confirmatory method (LC-MS or GC-MS), and the slower, more sophisticated assay will resolve the error. However, if the rate of false positives is too high, the benefit gained from using the fast screening method will be lost. In the case of a false negative result, the sample will not be further analyzed, so this type of error is more serious.

To quantitatively measure the effectiveness of this screening method, a 2 x 2 contingency table was used to evaluate the results for two qualitative variables, outcome and consequence [116]. The possible outcomes of a screening test are a positive test (target compounds identified) and a negative test (no target compounds identified). The consequence of a positive test is that the sample will go on for confirmatory testing, while the consequence of a negative test is that the sample will not go on for further testing. The sensitivity and specificity of the method are then calculated by

$$Sensitivity = 100\left(\frac{TP}{C_1}\right) \tag{24}$$

$$Specificity = 100\left(\frac{TN}{C_2}\right) \tag{25}$$

where TP is the number of true positives (samples correctly identified as containing library substances), TN is the number of true negatives (blank samples correctly identified as negative) and C_1 and C_2 are equal to the total number of analyzed chromatograms that contain drug peaks and the total number of blank chromatograms, respectively [116]. These two values allow the outright comparison between methods. The algorithm described in this work was developed based on two parameters, the maximum value of θ for the spectral match (see Chapter 3.5) and the maximum deviation from the standard retention index for the retention match. By varying these parameters, the method can be evaluated based on its tendency to produce false positives (sensitivity) and false negatives (specificity).

The positive predictive value (PPV) and the negative predictive value (NPV) are additional parameters that can indicate the effectiveness of a method. These values are calculated by Equations 26 and 27,

$$PPV = 100\left(\frac{TP}{TP + FP}\right) \tag{26}$$

$$NPV = 100\left(\frac{TN}{FN + TN}\right) \tag{27}$$

where FN represents the number of false negatives and FP represents the number of false positives [116]. If the analysis method indicates that a target compound is present, the PPV is the conditional probability that that compound is actually present in the sample. Conversely, the NPV is an estimate of the probability that a sample indicated as

containing no target compounds is truly a blank. In other words, PPV and NPV are estimates of the ability of the method to detect true positives and true negatives.

Two other commonly reported measures of selectivity for screening methods are the mean list length (MLL) [10] and the discriminating power (DP) [9]. The DP is a measure of the probability that two compounds in the library can be distinguished by the method. The list length (LL) for a compound *i* represents the number of compounds, *ll*, in the library that are indistinguishable from compound *i*, and the MLL is simply the mean of the LL values over all of the compounds in the library. These parameters are calculated using the following equations:

$$DP = 1 - \frac{2p}{q(q-1)} \tag{28}$$

$$MLL = \frac{\sum_{i}^{q} ll_i}{q} \tag{29}$$

where p is the number of indistinguishable substance pairs, and q is the total number of substances in the library.

5.5. Materials and Methods

All solvents were obtained from J.T. Baker (Mallinckrodt Baker, Inc., Phillipsburg, NJ), drug free blood was obtained from UTAK Laboratories Inc. (Valencia, CA), and the drug standards were obtained from Cerilliant Corporation (Round Rock, TX). The Clean Screen DAU columns used for the solid phase extraction (SPE) were

obtained from United Chemical Technologies, Inc. (Bristol, PA). Deionized water was used to prepare all aqueous solutions.

5.5.1. Preparation of blood samples

The procedure for extraction of the samples was adapted from Telepchak *et al.* and carried out by Kathryn Fuller at the Minnesota BCA [117]. Whole blood specimens from volunteers at the Minnesota Bureau of Criminal Apprehension (BCA) were collected in blood collection tubes containing 100 mg of sodium fluoride and 20 mg of potassium oxalate. Blood samples were stored at -20 °C until analyzed and warmed to room temperature and mixed on a tube rocker for at least 2 minutes before extraction. A 1.0 mL aliquot of each blood sample was added to screw cap tubes, followed by 10 μL of an internal standard solution (5 mg/mL L-erythro-methoxamine in methanol). The tubes were mixed by vortexing, 4 mL of water were added to the sample, and the samples were allowed to stand for 5 minutes. The samples were centrifuged for 10 minutes at 3000 rpm, and 2 mL of 0.1 M phosphate buffer (pH 6.0) was added to the supernatant. The pH was adjusted (if needed) to 6.0 ± 0.5 with 0.1 M mono- or dibasic sodium phosphate.

The samples were extracted using a Zymark RapidTrace automated SPE system (Zymark Corp., Hokinton, MA). The SPE columns used were 3 mL Clean Screen DAU columns with a 200 mg sorbent bed. The RapidTrace was programmed with the following parameters: the SPE cartridges were conditioned with 3 mL of methanol, followed by 3 mL of water, followed by 1 mL of 0.1 M phosphate buffer (pH 6.0), at a flow rate of 12 mL/min. After conditioning, 7.2 mL of the sample was loaded onto the

column at 1 mL/min. The RapidTrace cannula was purged with 6 mL of water followed by 6 mL of methanol at 12 mL/min, and then the SPE cartridges were rinsed with 3 mL of water followed by 1 mL of 0.1 M acetic acid at 12 mL/min. The columns were dried under nitrogen for 2 minutes, and then rinsed with 2 mL hexane at 12 mL/min. The samples were eluted with 3 mL of a 50/50 hexane/ethyl acetate eluent at 2 mL/min followed by a rinse with 3 mL of methanol at 12 mL/min (acidic fraction). The columns were again dried under nitrogen for 5 minutes, and the second fraction was eluted with 3 mL of a methylene chloride/isopropanol/ammonium hydroxide (78:20:2) solvent at 1 mL/min (basic fraction).

Following the SPE procedure, the organic solvent was evaporated from the acidic and basic fractions under nitrogen at 37 °C until approximately 300 μL of the solvent remained; 10 μL of a 1 % HCl in isopropanol (v/v) solution was added, and the samples were evaporated to dryness under nitrogen. The samples were then resuspended in 40 μL of a water / acetonitrile (95/5) mixture containing 10 mg/L of uracil and vortexed for 5 seconds. The samples were warmed with a heat gun for about 5 seconds until condensation formed around the tube about 2 cm above the residue. The samples were centrifuged at low speed (1000 rpm) for 2 minutes and the supernatant was transferred into autosampler inserts. The inserts were centrifuged at high speed (13,000 – 14,000 rpm) for 20 to 60 minutes and the dried extracts were reconstituted using 50 μL of the initial mobile phase used in the gradient elution (10/90 (v/v) acetonitrile/20 mM perchloric acid in water). Finally, the basic fraction was analyzed by the LC system described in reference [8]; the injection volume for LC analysis was 10 μL.

Blood samples from volunteers were pooled to prepare samples for matrix studies. The pooled samples were analyzed as-is or spiked with varying levels of target drugs to determine the limits of detection of the TFA method with respect to a matrix background. The spiked samples were included as part of data set B as described below.

5.5.2. Collection of Chromatograms

All chromatograms were collected on one of the fast LC-DAD systems described by Stoll *et al.* [8]. Table 4 summarizes the data sets used to develop and validate the analysis method, including the reference data set for the library. The compounds included in this study are shown in Table 5. A set of chromatograms was collected containing single component peaks of each drug in the library in order to determine the reference retention times and retention indices for each compound, which are denoted in Table 5 as t_R^o and I^o in respectively. A spectral library was created by obtaining the UV-visible spectrum for each of the drugs from 201 – 301 nm. Rather than independently collecting the spectra, they were extracted from the chromatographic peak maxima of the pure components. The chromatograms in the three data sets outlined in Table 4 were pooled together, and then separated into a training set (n=63) and a validation set (n=70). The training data were used to determine the method parameters and contained various combinations of overlapped peaks, low abundance peaks, and spectrally similar peaks in order to determine the effectiveness of the method for resolving the peaks.

Table 4. Summary of data sets evaluated in Chapter 5. All data sets included blank chromatograms.

Data Set	Compounds Included	Purpose
Reference	All	Establish reference values for spectra and retention index
A	Various mixtures	Evaluate overlapped peaks and peaks from same spectral class
B	Amitriptyline, oxycodone, zolpidem	Evaluate low intensity peaks in the presence of gradient background and blood matrix
C	Amphetamine, MDA, hydrocodone, zolpidem	Evaluate different concentration ratios of highly overlapped peaks

Table 5. Spectral class, standard retention index (I^o), and standard retention time (t_R^o) for 47 library compounds.

Class	Compound Name	I^o	t_R^o
A	2-hydroxyethylflurazepam	441.41	1.6104
B	6-Acetylmorphine	279.82	0.6094
C	7-Aminoclonazepam	218.59	0.3937
U	7-Aminoflunitrazepam	259.24	0.5369
D	Alprazolam	455.52	1.7078
U	Amitriptyline *	546.47	2.2809
E	Amphetamine	278.57	0.6050
U	Benzoylecgonine *	336.05	0.9126
U	Bromazepam	326.01	0.8480
F	Cathinone	233.18	0.4451
U	Chlordiazepoxide	394.18	1.2869
U	Clobazepam	373.54	1.1540
D	Clonazepam *	465.61	1.7775
B	Codeine	245.87	0.4898
U	Cyclobenzaprine HCl	532.10	2.1987
A	Desalkylflurazepam	439.08	1.5943
G	Diazepam *	428.86	1.5237
E	Ephedrine	246.44	0.4918
D	Estazolam	430.70	1.5364
D	Flunitrazepam	484.32	1.9067
D	Flurazepam	465.64	1.2590
H	Hydrocodone	301.32	0.6890

H	Hydromorphone	198.44	0.3262
C	Lometazepam	530.55	2.1898
C	Lorazepam	475.28	1.8443
I	MBDB	343.42	0.9601
E	Methamphetamine	302.75	0.6982
F	Methcathinone	248.42	0.4988
I	Methylenedioxyamphetamine (MDA)	289.72	0.6443
I	Methylenedioxyethylamphetamine (MDEA)	327.27	0.8561
I	Methylenedioxymethamphetamine (MDMA)	306.40	0.7217
D	Midazolam	455.30	1.7063
B	Morphine *	140.59	0.2518
U	Nitrazepam	377.05	1.1766
U	Nordiazepam	382.28	1.2103
C	Oxazepam	450.71	1.6746
H	Oxycodone *	276.53	0.5978
H	Oxymorphone	162.91	0.2805
J	Paramethoxyamphetamine (PMA)	302.03	0.6936
J	Paramethoxymethamphetamine (PMMA)	315.81	0.7823
E	Phenylpropanolamine	219.73	0.3977
G	Prazepam	549.41	2.2977
E	Pseudoephedrine	244.71	0.4857
H	Sertraline *	571.11	2.4219
C	Temazepam *	498.89	2.0073
D	Triazolam	494.14	1.9745
U	Zolpidem hemitartrate *	404.21	1.3535

* Compound used as a secondary standard

5.5.3. Data Analysis

All data analysis was performed in the Matlab® programming environment (The Mathworks, Natick, MA), version 7.0.4. A macro provided by Agilent (Agilent Technologies; Wilmington, DE) was used to convert the data collected in Chemstation (version A.10.01) into a comma separated variable (CSV) file format that could be loaded into Matlab as a variable. The SVD algorithm used was a built-in Matlab function. All other Matlab functions, including the TFA algorithm, were written in-house. The fALS

algorithm was written in-house as reported previously [13]. Analysis was carried out on a Dell® Optiplex GX280 with a Pentium 4 3.2 GHz processor with 2 GB of RAM.

5.6. Creation of retention index and spectral libraries

The library used in this study contained 47 different target compounds (*cf.* Table 5). The compounds were chosen to include frequently abused drugs, drugs of toxicological relevance, and several metabolites. Chromatograms of each target compound were collected in order to establish standard values for the retention indices of the pure components. These results constituted the *retention index library*. A matrix containing the spectra for all 47 target compounds was created that constituted the *spectral library*. These spectra were used as the target spectra as described in Chapter 3.4. A correlation matrix for the spectral library was used to determine the occurrence of similar spectra. The library contained 20 unique spectra, where highly correlated spectra ($\rho > 0.98$) were considered to be identical. Based on the results of this analysis, each target compound was either assigned to one of ten spectral classes (classes designated by 'A' – 'J' in Table 5), or was one of 10 compounds that exhibited a unique spectrum (designated by a 'U' in Table 5). Because this method discriminates based on retention indices as well as spectra, the spectral class assignment is only significant when two compounds in the same class are overlapped chromatographically. Such cases affected the calculation of the DP and MLL of the method (discussed below) and were included in both the training and validation data sets.

5.7. Discriminating Power and Mean List Length

The DP [9] and MLL [10] of the retention index library and the spectral library were calculated using Equations 28 and 29 to characterize the selectivity of the method and to compare this method to previously published methods. Compounds with a spectral correlation coefficient greater than 0.98 and/or a retention index within ±12 retention index units of one another were considered indistinguishable by the method. The value of 12 retention index units corresponds to three times the long term (13 months) standard deviation in retention index reported for the LC method in the companion study [8]. This error window was consistent with that used by Maier and Bogusz [118]. For a library of 47 compounds, the DP was 0.94 and the MLL was 3.72 based solely on spectral criteria. The DP and the MLL, when only retention index was considered, were 0.95 and 3.26 respectively. When both the retention indices *and* the spectra of the compounds were considered, the DP and MLL improved to 0.997 and 1.255, respectively, and the number of indistinguishable pairs (p) dropped from 53 to 6 (out of a total 1081 possible pairs).

A MLL value of close to one indicates that on average only one compound will be identified with a given retention index and spectral result; the MLL increases as the number of indistinguishable pairs increases, up to a maximum value of q. Conversely, DP is less than one as long as p is greater than q. DP approaches one as p approaches zero; that is, when all compounds in the library are unique. This ideal situation indicates that each library compound is only indistinguishable from itself.

The parameters DP and MLL are a reasonable indication of how well a method distinguishes compounds in a library, but they are also strongly correlated with q. Therefore, the size of the library should be considered when comparing DP and MLL between two different methods. Maier and Bogusz [118] reported a DP and MLL of 0.96 and 4.04 for a library of 56 acidic drugs when considering retention index and UV absorbance maxima. This work reports a significant improvement in DP and MLL for a similar size library by using all of the spectral information afforded by using the DAD. In comparison, Herzler *et al.* [109] reported a DP of 0.9999 and a MLL of 1.253 for a library of over 2500 compounds.

5.8. Optimization of algorithm parameters

The DP and MLL of a method are indications of how well a method discriminates compounds that are not overlapped. They are theoretical parameters based on library information including retention times (or indices) and spectra. They are not based on the analysis of real samples, which are usually mixtures of substances. To show the effectiveness of the algorithm presented here for the analysis of real chromatograms, the sensitivity and selectivity of the method were evaluated based on a large set of chromatograms. The TFA algorithm was evaluated based on the production of false positive and false negative results from the test chromatograms. Training samples were used to optimize the algorithm parameters to minimize the occurrence of errors, and a validation set was analyzed to determine the robustness of the parameters chosen.

The independent parameters that required optimization were the maximum angle requirement for θ (*i.e.*, the maximum angle between the predicted and the target spectra that can be obtained and still considered a match) and the allowed difference between I_{corr} and I^o. These parameters were evaluated using a 3^2 factorial design. The levels and the results for the training data set are summarized in Table 6. A sample was considered a true positive (TP) if any of the drug peaks in it were identified correctly; hence the consequence of a TP result is that the sample will go on for confirmatory testing. Blank samples were considered true negatives (TN) if no drug peaks were identified. In addition to the results for the chromatograms, the samples were analyzed on a *per peak* basis, where every single drug peak was accounted for. In this analysis, only true positives, false positives, and false negatives were counted. As discussed in reference [8], the demonstrated long term reproducibility of I_{corr} was about 1.0 % (relative standard deviation), or about 4 retention index units, when all factors contributing to retention shift were considered. Accordingly, retention index windows of $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ around the reference value were tested, corresponding to 4, 8, and 12 retention index units, respectively. Thresholds for θ of 5°, 7.5°, and 10° were also tested as part of the experimental design.

Table 6. Results of 3^2 factorial experiment for the training data set and for the validation data set (last row).

Level	I^o-I_{corr} Window	Theta (°)	SEN [a]	SPEC [a]	PPV [a]	NPV [a]	SEN [b]	PPV [b]
1	4	10	97%	91%	90%	97%	80%	75%
2	4	7.5	97%	91%	90%	97%	79%	80%
3	4	5	93%	97%	96%	94%	63%	88%
4	8	10	100%	85%	85%	100%	81%	62%
5	8	7.5	100%	85%	85%	100%	80%	70%
6	8	5	97%	91%	90%	97%	64%	83%
7	12	10	100%	76%	78%	100%	83%	48%
8	12	7.5	100%	76%	78%	100%	81%	59%
9	12	5	97%	85%	85%	97%	66%	75%
Validation	4	7.5	92%	94%	94%	91%	69%	77%

[a] Sensitivity, specificity, PPV and NPV calculated according to Equations 24 – 27 by chromatogram.
[b] Sensitivity (Equation 24) and PPV (Equation 26) calculated for individual target compound peaks present.

As seen in Table 6, both the angle θ and retention index window affect the results; all parameters (selectivity, specificity, PPV, and NPV) must be taken into account when choosing the appropriate threshold values for θ and I^o-I_{corr} window. For example, the results for the chromatogram analysis for the levels 4, 5, 7, and 8 all show a sensitivity of 100 %, indicating that no false negative results were obtained; however, the specificities for these levels are all below 90 %, indicating a high rate of false positive results. The last two columns of Table 6 summarize the results of the *per peak* analysis as described above. Again, levels 4, 5, 7, and 8 have the highest sensitivities; however, all of their PPV values are 70 % or lower (as low as 48 % for level 7). The consequence of a very low PPV for the *per peak* analysis is that many extraneous peaks are identified as being target compounds, which will complicate the application of confirmatory testing methods

such as LC-MS where selected ion monitoring or selected reaction monitoring is employed. Level 3 has the highest PPV on a *per peak* basis; however the sensitivity is low both on the *per peak* basis and on the chromatogram analysis. Because all of these values should ideally approach 100 %, it was logical to choose the level with the highest average over all of the factors for both the chromatogram analysis and the *per peak* analysis.

Level 2 had the highest average over all of the factors evaluated. This level did not have the highest value for any of the factors; however, a compromise was made to select the level with the best overall performance. By applying the factorial design it was determined that an unknown compound must have a value of θ less than 7.5° and a I_{corr} within 4 retention index units of the target compound in order to be considered a match. This requirement for I_{corr} corresponds to a window of approximately ± 1 %, consistent with the results reported in the companion study [8]. These parameters were applied to the validation data set, and the 2 x 2 contingency table for the validation set is shown in Table 7. The fact that the validation results were consistent with the training results confirms that the parameters applied to the method are robust and not dependent on the selection of data being analyzed. The calculated sensitivity and specificity for the validation data was 92 % and 94 % respectively. These results compare favorably to EIA screening methods, where the sensitivity tends to be high but the specificity tends to be low (meaning there is a high rate of false positive tests) [98, 100, 119].

Table 7. 2x2 contingency table for the validation data set, where TP = true positive, FP =false positive, FN = false negative, and TN = true negative.

	Drugs present	Drugs absent	Row totals
Positive test [a] (confirmatory testing)	TP = 33	FP = 2	$R_1 = 35$
Negative test [a] (no confirmatory testing)	FN = 3	TN = 32	$R_2 = 35$
Column totals	$C_1 = 36$	$C_2 = 34$	70

[a] The test results indicate whether confirmatory testing will be done (positive test) or not (negative test)

Too many false positive results (evidenced by a low PPV and sensitivity) indicate that the screening method is not useful in eliminating samples from consideration that do not need confirmatory testing. A simple calculation can indicate whether or not the screening method described here significantly increases the throughput of an analysis laboratory. The total analysis time without any screening method (T_{total}) is equal to the number of samples times the cycle time of the confirmatory method. That is, all samples are evaluated using a longer confirmatory method. Using the screening method on all samples and the confirmatory method on only those samples that test positive, the total analysis time ($T_{total,screen}$) is the total number of positive samples (TP + FP) times the cycle time of the confirmatory method plus the total number of samples time the cycle time of the screening method. When $T_{total,screen}/T_{total}$ approaches 1, the screening method is no longer advantageous in saving time. The factor that determines this crossover point

is the ratio of total positive samples (TP + FP) to the total number of samples. For the validation data set analyzed in this work, the cycle time of the long method was assumed to be 30 minutes (typical for a GC-MS analysis), the cycle time of the screening method was 4 minutes (including data analysis time), and the total number of samples was 70. The total number of positive screening tests for the validation data set was 35 (R_1 in Table 7), giving a positive rate of 50%. At this rate, where $T_{total,screen}/T_{total} = 0.63$, using the screening method has the potential to nearly double the throughput of the laboratory relative to running the full confirmatory test on every sample.

In order to carry out the data analysis of the chromatograms more efficiently, the peak integration tables obtained from the data acquisition software were used to identify the regions of the chromatograms that should be target tested. By analyzing only those regions of the chromatogram where a peak has been integrated, there is no need for a peak integration algorithm to be included. The retention times found in the data collection software can be converted to I_{corr} and compared to the standard retention indices of the target analytes. A disadvantage to this method, particularly for Hewlett Packard/Agilent LC users utilizing Chemstation, is that a single wavelength chromatogram must be used to do the integration. In other software, such as Waters Millenium 32®, maximum absorbance plots can be used, possibly allowing lower detection limits for compounds that have their λ_{max} at different wavelengths. Another potential problem occurs when peaks are overlapped, causing the observed retention time to be shifted from the expected retention time.

The application of the WTTFA algorithm to this data was also explored [12]. This algorithm is as effective as the TFA algorithm in identifying drug peaks and resolving overlapping peaks; however, it took much longer to compute the results. The analysis of the training data set (63 samples) chromatograms took hours rather than minutes. It was more efficient to analyze sections of the chromatogram determined by the retention times in Chemstation as discussed above.

5.9. Sample Results

Several different data sets were analyzed to evaluate the efficacy of the method (summarized in the Materials and Methods section of this chapter and Table 4). Data set A contained chromatograms with various mixtures of overlapping compounds, including drug/metabolite pairs, pairs with the same or similar spectra, and pairs with different spectra. This data set also contained individual chromatograms of the drugs of interest. Because the presence of overlapping peaks in a chromatogram may shift the apparent retention time (and hence I_{corr}) that is reported, chromatograms with highly overlapped peaks were included to assess the effect of this phenomenon on the results. Analyzing these sets of partially and highly overlapped target compounds can determine (a) if the TFA algorithm is successful at resolving the spectra of the overlapped peaks, and (b) what the effect on the allowable retention index window would be.

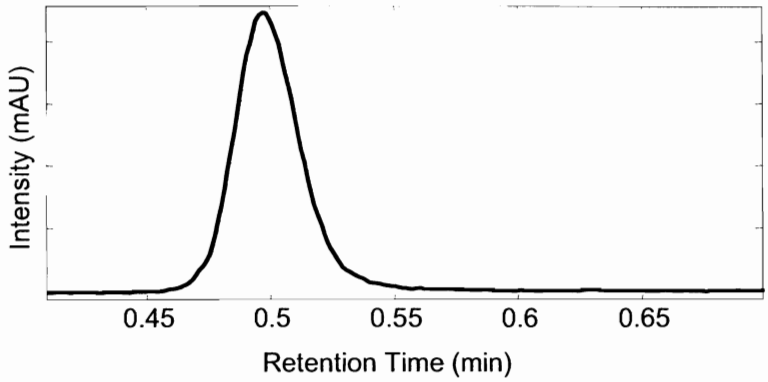

Figure 12. Example chromatogram from data set A at 210 nm illustrating overlapping peaks. This chromatogram contains ephedrine ($I^o = 246.4$), codeine ($I^o = 245.9$) and methcathinone ($I^o = 248.4$). The calculated I_{corr} for this peak is 244.6.

An example of one of the chromatograms from data set A is shown in Figure 12. The single wavelength chromatogram in Figure 12 illustrates the issue of overlapping or indistinguishable retention indices for drugs with different spectra. This peak consists of contributions from three target compounds, which is not evident from the single wavelength chromatogram. The sample contained ephedrine ($I^o = 246.4$), codeine ($I^o = 245.9$) and methcathinone ($I^o = 248.4$). The calculated I_{corr} for this peak was 244.6. Two issues are illustrated here: the observed spectrum, and the calculated retention index. The observed spectrum at the apex of this peak is a linear combination of the three spectra, which the TFA algorithm can resolve and individually identify. Without this step, traditional library searching algorithms (such as those available in Chemstation) are not able to identify any of the target compounds in this peak due to its distorted (relative to the library) spectral profile. The presence of overlapping peaks within a chromatogram will also degrade the quality of a match to standard retention index values (I^o). Codeine and methcathinone were easily identified in this chromatogram. Although in this case the

retention indices were all within 4 *I* units of I^o, the spectrum of ephedrine was not identifiable even using the TFA algorithm. The spectrum of ephedrine (spectral class E, *cf.* Table 5) does not have a maximum at any selective wavelengths, so the distortion due to the overlap with the other compounds is too much even for the TFA algorithm to resolve. Using a larger threshold for θ will allow the identification of all three drugs; however, using too high of a threshold creates too many false positive results (*cf.* Table 6).

Data set B was designed to determine if the TFA algorithm could detect low intensity peaks in a chromatogram, particularly in the presence of a changing baseline typically seen in gradient chromatography and in the presence of overlapping peaks from a blood matrix. Oxycodone, zolpidem, and amitriptyline were chosen as test compounds for this data set as these three drugs have retention times at the beginning, middle, and end of the gradient. A range of concentrations was tested from 0.2 μg/mL to 20 μg/mL in both a "clean" matrix (mobile phase buffer) and a blood matrix (spiked in certified drug-free blood). Chromatograms of these drugs in both matrices at 0.2 μg/mL (the lowest level tested) are shown in Figure 13. In the clean matrix sample (Figure 13B), the amitriptyline (peak 3) is severely overlapped with a gradient background peak that is most likely due to a mobile phase impurity. Particularly with low abundance components, such an overlap can distort the apparent spectrum of the peak relative to the library spectrum. This distortion is illustrated in Figure 13C, where the library spectrum of amitriptyline (dashed line) is overlaid with the actual spectrum from the apex of peak 3 (solid line) in Figure 13B. Using the TFA algorithm resolves the component from the

background and allows identification of the peak where traditional library search algorithms would fail. The spectrum of the background (and any other interfering peaks) is resolved by SVD and therefore the TFA algorithm can readily distinguish drug peaks from background without any need for background subtraction.

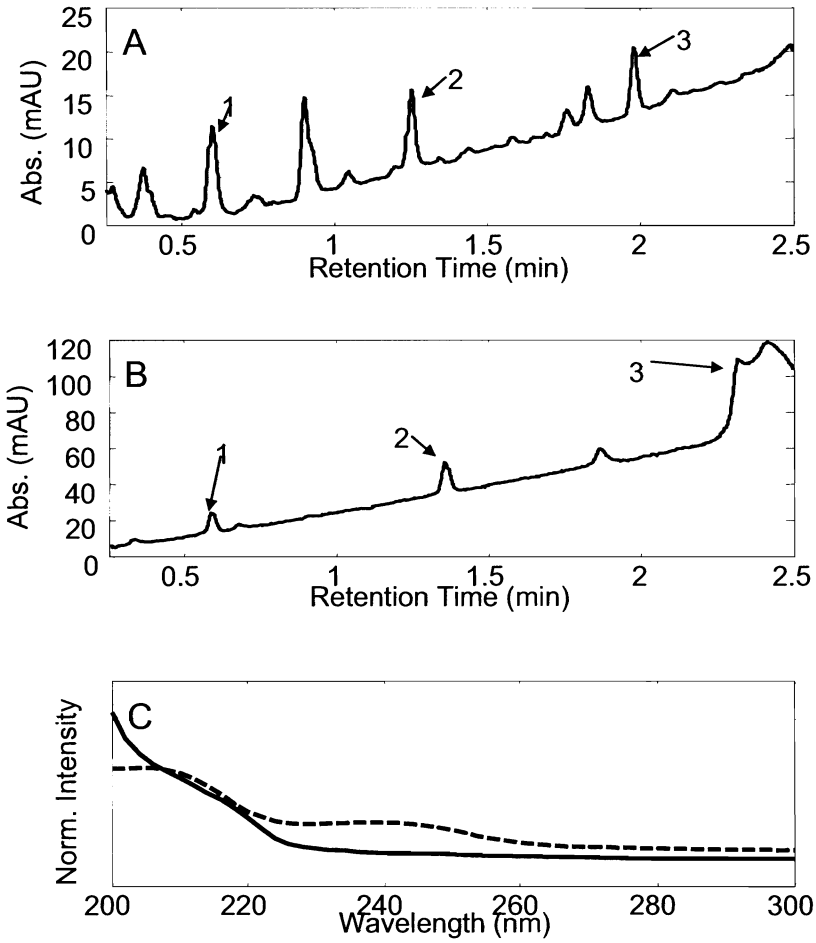

Figure 13. Oxycodone (peak 1), zolpidem (peak 2), and amitriptyline (peak 3) in (A) blood matrix, and (B) clean matrix at 0.2 μg/mL. (C) Comparison of the library spectrum (dashed line) of amitriptyline and the spectrum from the apex of peak 3 from the chromatogram in B (solid line).

A comparison of Figures 13A and 13B also shows the significant shift in retention time that can occur over long time scales and on different instruments. The

chromatogram in Figure 13A was collected on the two-pump system described in the companion study, and the chromatogram in Figure 13B was collected on the one-pump system [8]. The difference in retention times between the amitriptyline peaks (peak 3 in both figures) is clear simply from visual inspection; however, the application of the retention index method gave consistent I_{corr} values for both peaks.

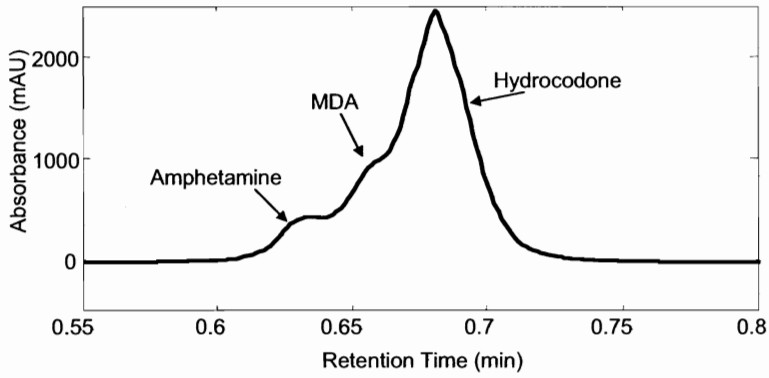

Figure 14. Single wavelength chromatogram (210 nm) selected from data set C containing amphetamine, MDA, and hydrocodone.

The third data set that was evaluated (data set C) included chromatograms of mixtures of amphetamine, 3,4-methylenedioxyamphetamine (MDA), hydrocodone, and zolpidem. Amphetamine, MDA, and hydrocodone are highly overlapped peaks, as is evident by their retention times in Table 5. A single wavelength chromatogram (210 nm) is shown in Figure 14. In this case only one peak was integrated from Chemstation, and thus the retention indices of the two earlier peaks were significantly different from the library values. If a large enough error window was allowed, all three compounds could be identified but under the final chosen conditions (± 4 retention index units) only MDA and hydrocodone were identified in most of the chromatograms in this data set. The

sample was flagged for confirmatory testing, however, so this result was still considered a true positive. Different concentration ratios of these three drugs were included in the training data to determine what the effect on the reported retention index would be, and what the allowed deviation between I_{corr} and I^o should be in order to identify the drugs. While an error window for I_{corr} could be applied that would identify all of the drugs in these mixtures, such a large window adversely affected the other results and resulted in too many false positives.

In addition to the drug mixtures discussed, all three data sets contained "blank" chromatograms. These chromatograms included blanks of the buffer used in the mobile phase, certified drug-free blood and urine blanks, blood samples obtained from volunteers at the Minnesota BCA, and samples that contained compounds not included in the library. The blank samples were used in determining the rate of false positive results obtained by our method.

5.10. Preliminary study of the application of fALS algorithm to selected data

As discussed, and illustrated in Figures 12 and 14, regions of high peak overlap can cause problems in identifying target compounds based on spectra or retention index, even when the demonstrated reproducibility of the method is very good (1 % relative standard deviation). The application of the fALS algorithm described in reference [13] can alleviate some of these issues by resolving the retention profiles and the spectral profiles and allowing the calculation of a more accurate I_{corr} for the individual components. Also the resolved spectral profiles often show a better match to the target

spectrum than simply the linear combination of the SVD abstract spectra used in the TFA algorithm.

The chromatogram shown in Figure 12 has three highly overlapped target compounds present: ephedrine, codeine, and methcathinone. Using the parameters established by the method ($\theta < 7.5°$ and I_{corr} within 4 units of I^o), only codeine and methcathinone can be identified. The resolved ephedrine spectrum has a θ of about 10° relative to the library spectrum of ephedrine; this difference is due to the high degree of overlap and incomplete resolution of the spectra. Using a higher θ threshold allows all three target compounds to be identified, but two other target compounds are falsely identified, including cathinone (same spectrum as methcathinone), and pseudoephedrine (same spectrum as ephedrine). Using the three spectra present (spectra B, E, and F, *cf.* Table 5) plus a background spectrum as initial estimates and a non-negativity constraint for the fALS algorithm, the three compounds present can readily be resolved. The results of the fALS analysis are shown in Figure 15. The fALS algorithm was only applied to the local window where the peak was located, and the first spectrum in this local window was used as the background spectrum for the initial estimate. By using the maxima of the three resolved components, new I_{corr} values can be calculated for ephedrine (244.94), codeine (243.78) and methcathinone (247.63), all of which are within 4 retention index units of the I^o value and the θ values between the library spectra and the resolved spectra are all under 7.5°.

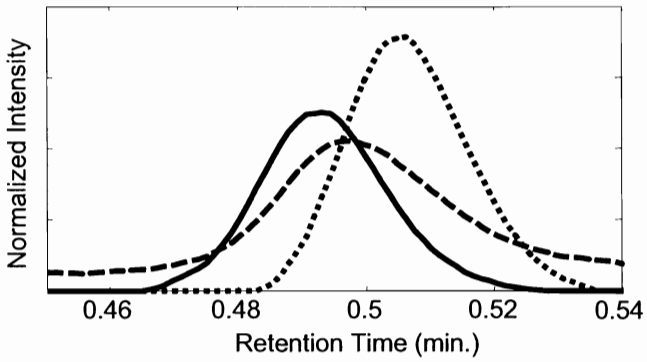

Figure 15. Resolved chromatographic profiles of three components present in Figure 12. Solid – codeine; dashed – ephedrine; dotted – methcathinone.

Although the application of the curve resolution technique would potentially improve the overall success of the screening method, the limitation lies in the difficulty of automating the algorithm to analyze a large number of samples at once. The fALS algorithm would need to be applied to each section of the chromatogram where a target drug was detected, and it would be necessary for the analyst to individually examine the results obtained from the curve resolution algorithm *for each sample* to determine the appropriate set of constraints and initial estimates that best fit the data. The added time and analyst input required to carry out the fALS analysis would negate the potential of the method for high-throughput screening. For this reason, it is not yet feasible to include the fALS analysis as part of the fast screening method.

5.11. Multivariate Selectivity Applied to LC-DAD and Orthogonal LC-DAD

One of the advantages to applying the fast LC method described in this chapter is that the possibility exists for a second, orthogonal separation to occur in a very short amount of time. Combining the retention data from two columns with the spectral data

should significantly improve the sensitivity and specificity of the method by increasing the number of compounds that can be distinguished. Orthogonal column chromatography involves collecting two independent separations of a mixture on columns with different chromatographic selectivity in order to increase the number of compounds that can be separated [120, 121]. When using a very fast chromatography method, overlapped peaks are likely to occur as more components are added to a mixture. Particularly with complex samples (such as biological samples), there are often likely to be matrix peaks or other interfering compounds that may overlap with an analyte of interest. The use of a second column can alleviate these problems by providing a second separation mechanism where compounds that were not separated on the first column can be separated on the second column.

Cantwell *et al.* [122] carried out simulations using the retention data for the 47 compounds in this study, which were obtained using a C18 column with a perchloric acid mobile phase [8], and an additional set of retention data collected on a pentafluorobenzene (F5) column and a phosphate buffer mobile phase. The purpose of these simulations was to use the multivariate selectivity metric (discussed in Chapter 3.4) to determine whether a second separation or the use of the DAD detector will improve the precision of the method. The data from the two columns were stacked to form a three-way data set with bimodal chromatographic profiles in the second mode and spectral profiles in the third mode (the first mode is sample or concentration). The data format is shown in Figure 16. The retention time data for the two columns were used to

simulate chromatographic profiles with Gaussian peaks having 4σ peak widths of 0.175 min. The selectivity for each component was calculated using Equation 9 [122].

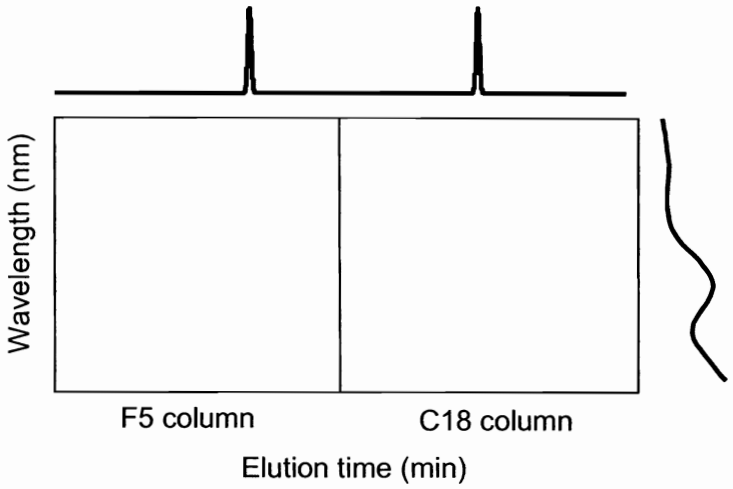

Figure 16. Graphical representation of the stacked data structure for the orthogonal LC simulations.

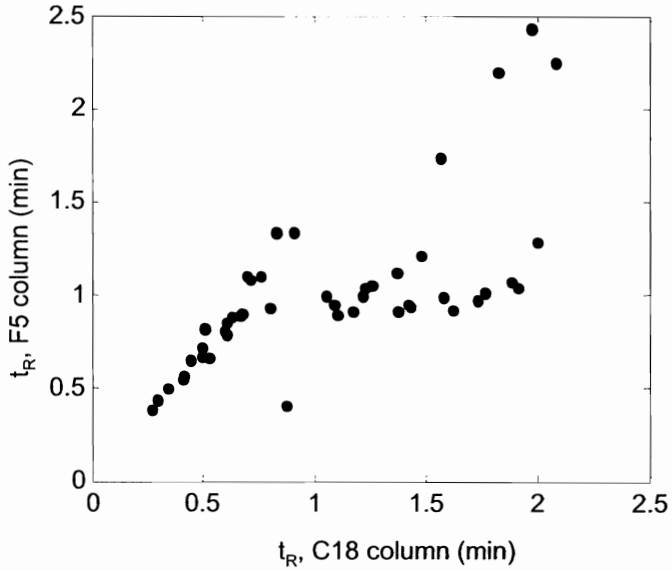

Figure 17. Plot of retention times on column 1 (C18) vs. column 2 (F5) for the orthogonal LC simulations.

Table 8. Selectivity results for orthogonal LC simulations.

Method	Average Selectivity – C18 column	Average Selectivity – F5 column
LC – single wavelength	0.59	0.29
LC – DAD	0.75	0.56
Orthogonal LC – single wavelength	0.69	
Orthogonal LC – DAD	0.78	

A plot of the retention times on the C18 column vs. the retention times on the F5 column is shown in Figure 17. This figure shows that the two separations used here were not perfectly orthogonal. Because both separations were reversed phase gradient separations, it is nearly impossible to achieve complete orthogonality. Regardless, there was still an increase in the average selectivity of the components, and for some specific components the increase in selectivity was significant. Table 8 summarizes the selectivity results obtained for this system. Specifically, upon going from a single wavelength, single column LC separation to the orthogonal LC-single wavelength separation, there is a significant increase in the average selectivity for both columns. The addition of the second chromatographic separation resulted in a marked increase in the number of peaks that could be quantified with precision approaching that possible in a single component sample. These results also show the difference in the selectivity between the two columns used for these experiments. For the single wavelength, single column separation, the average selectivity of the C18 column is twice that of the F5 column. The results indicate that using the C18 column with DAD detection will result

in a more selective method than using the orthogonal separation without the DAD; however the orthogonal separation with DAD detection still results in the highest overall selectivity.

Although the average selectivity for the 47 compounds is only increased by a small amount (see Table 8), there are several cases where the individual selectivity of a component is improved dramatically by the addition of the second separation. A list of compounds whose individual selectivities are improved by more than 50 % is shown in Table 9. This improvement indicates that adding the extra data dimension (in this case, the second chromatographic separation) more than doubles the precision with which a component can be quantified in a mixture. For example, the selectivities of 6-acetylmorphine and amphetamine are significantly improved upon the addition of the second chromatographic separation. The individual chromatographic and spectral profiles for these two compounds are shown in Figure 18. Because the spectra of the two drugs are similar and the separation on the C18 column is minimal, the F5 separation is necessary to gain an improvement in selectivity.

Table 9. List of compounds whose selectivity increased by more than 50% upon the addition of the orthogonal separation.

Drug name	C18-LC-DAD Selectivity	Orthogonal - LC – DAD Selectivity
6-Acetylmorphine	0.27	0.58
Amphetamine	0.38	0.82
Codeine	0.43	0.75
Diazepam	0.48	0.80
Oxycodone	0.35	0.54

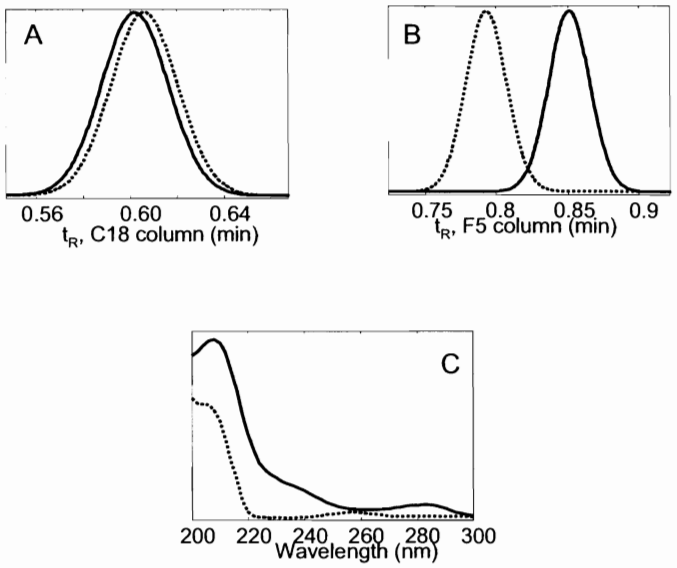

Figure 18. Resolved profiles for 6-acetylmorphine (solid line) and amphetamine (dotted line). (A) C18 chromatographic profiles; (B) F5 chromatographic profiles; (C) spectral profiles.

To show the advantage of adding a spectral dimension to a chromatographic separation, the selectivity of an orthogonal LC - single wavelength separation can be compared with that of an orthogonal LC – DAD separation. Table 8 shows that the average selectivity of all 47 components increased from 0.69 to 0.78, and Table 10 lists those specific compounds whose selectivities increased by more than 50 %. Figure 19 shows the chromatographic and spectral profiles for PMA and hydrocodone, respectively. These two compounds are poorly separated on both columns; it is only the addition of the spectral dimension that allows them to be resolved using PARAFAC (or any other curve resolution method).

Table 10. List of compounds whose selectivity increased by more than 50% upon the addition of DAD detection.

Drug name	Orthogonal-LC-single wavelength Selectivity	Orthogonal - LC – DAD Selectivity
7-Aminoclonazepam	0.45	0.72
Hydrocodone	0.35	0.61
Methamphetamine	0.41	0.71
Phenylpropanolamine	0.45	0.72
p-methoxyamphetamine (PMA)	0.30	0.62

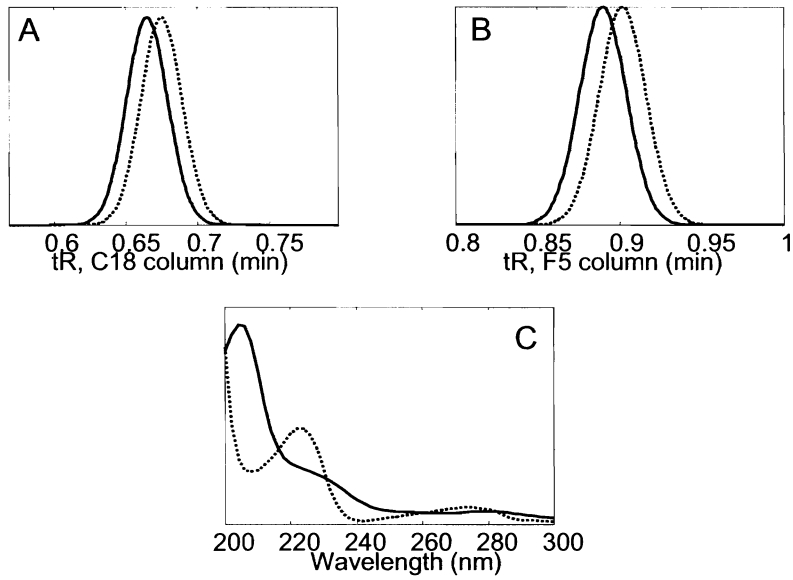

Figure 19. Resolved profiles for hydrocodone (solid line) and PMA (dashed line). (A) C18 chromatographic profiles; (B) F5 chromatographic profiles; (C) spectral profiles.

As encouraging as these results are, some cases cannot be resolved by chemometric approaches. Ephedrine and psuedoephedrine represent such a case. These compounds are stereoisomers, and as such cannot be chromatographically resolved by RPLC. Because the signals in all three modes for these two components are identical, the selectivity is low in all cases, and curve resolution algorithms will be unable to resolve the pure component profiles.

The results presented in this chapter represent an important contribution to the field of forensic drug testing. High speed chromatographic methods are becoming more and more common and data analysis methods will be required in order to resolve overlapped components in complex mixtures. The results of the selectivity studies also provide direction for the future application of orthogonal LC to the screening methods developed here.

CHAPTER 6. Two-dimensional Liquid Chromatography Diode Array Data: Applications in Metabolomics

The work presented in this chapter is also the result of collaboration with Prof. Carr. The focus of this chapter is the development and application of chemometric methods to analyze the very large data sets obtained using the 2D-LC-DAD system described by Stoll *et al.* [26]. The use of a multichannel detector provides significantly more information than what can be obtained from traditional analysis (*i.e.*, single wavelength peak integration) of two-dimensional chromatograms. In this work, the detection and analysis of compounds related to the biosynthetic pathways of IAA in maize were investigated. This chapter is reproduced in part from reference [11], published in Analytical Chemistry (Copyright © 2006 American Chemical Society. All rights reserved).

6.1. Introduction and Literature Review of Two Dimensional Chromatography Data Analysis Methods

The huge multivariate data sets obtained from two-dimensional chromatographic separations offer both a real challenge and a significant opportunity for chemical analysis, and require specialized data analysis methods to optimally extract the information within. For example, one issue that often arises in proteomic and

metabolomic studies is the detection and quantification of low abundance components in the presence of a few large dominating peaks. In addition, the samples can contain hundreds or even thousands of components [123-126]. The demonstrated peak capacity of fast 2D-LC separations is on the order of 500-2000 peaks in 30-60 minutes [26], so there is a high probability that the number of components in a complex sample will approach or exceed the peak capacity demonstrated by the method, resulting in many co-eluting components.

Using chemometric methods significantly increases the amount of information that can be obtained from 2D-LC chromatograms. They are useful for resolving overlapped chromatographic peaks, dealing with uncontrolled shifts in retention time, and taking full advantage of multichannel detectors such as MS and DAD. Because it is the more mature technology, much of the current literature in chemometrics and two-dimensional separations deals with 2D-GC separations with both single channel detectors (flame ionization (FID)) and multichannel detectors (MS). However, the data structures for 2D-LC and 2D-GC are essentially the same and so many of the same techniques are applicable to both types of separations, especially for single channel detectors and mass spectrometric detectors. The data structure for 2D-LC data is shown in Figure 20.

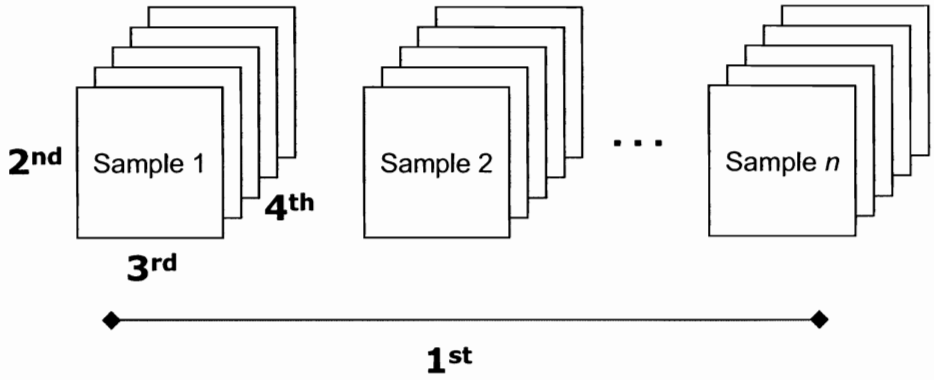

Figure 20. Data structure for the four-way 2DLC data presented in Chapter 6. The 1st mode is concentration or sample number, the 2nd mode is retention time on the first column, the 3rd mode is retention time on the second column, and the 4th mode is the spectral information (wavelength).

6.1.1. Multivariate Curve Resolution

MCR methods are a major area in which chemometrics can be used to great advantage in two-dimensional separations [127-130]. One important feature of MCR methods is that they do not require any previous knowledge of the chromatographic or spectral profiles other than an estimation of the number of components, making them very useful for analyzing samples where the composition of the samples is unknown. These methods are also useful for resolving peaks with less than ideal chromatographic resolution. The PARAFAC algorithm is described in detail in Chapter 3.2. Both PARAFAC and GRAM are eigenvalue based methods, based on the model shown in Figure 1, but GRAM is non-iterative and tends to be faster to compute. GRAM has limited applications because only two samples can be analyzed at a time; that is, the dimensions of a three-way data set $\underline{\mathbf{X}}$ are limited to $I \times J \times 2$ [2]. GRAM is generally applied where there is one known (a standard chromatogram) and one unknown in the

sample dimension. PARAFAC more generally used for three- and higher order arrays, and is usually solved with an alternating least squares algorithm. The iterative nature of the algorithm means that the results are superior to those obtained by non-iterative methods such as GRAM. The GRAM algorithm does not have well defined convergence criteria and any amount of noise in the data can complicate the calculation of the profiles [2].

GRAM has been used extensively to analyze 2D-GC separations with either MS or FID detection. Synovec's group has published several studies using GRAM and 2D-GC separations for the analysis of fuels [131-133]. GRAM was used for the quantitative analysis of methyl *tert*-butyl ether in white gasoline by 2D-GC-MS [133]. 2D-GC-FID was used to analyze aromatic isomers in jet fuel [132] and to quantify ethylbenzene and m-xylene in modified white gasoline [131]. Fraga and Corley demonstrated the use of GRAM for the quantitative analysis of overlapped peaks in 2D-LC [127], using single wavelength UV detection to analyze an aqueous test mixture of p-chlorobenzoic acid, benzoic acid, uracil, maleic acid, and phenyl phosphoric acid.

PARAFAC was used to analyze three-way 2D-GC-MS arrays of environmental data containing fuel components, pesticides, and natural products [130]. The PARAFAC model was used to resolve overlapped components and identify specific spectral signals in the data. The performance of the PARAFAC model was compared to that of conventional integration by van Mispelar *et al.* for the quantification of essential oils in perfumes by 2D-GC-FID [134]. They found that while conventional integration of the chromatograms resulted in higher precision and accuracy, the MCR models were much

faster. An application of 2D-GC-MS to metabolism studies was described by Mohler *et al.* who analyzed the metabolites of fermenting and respiring yeast cells [135]. They found that overlapped peaks could be resolved by PARAFAC, reconstructed by taking the outer product of the first and second dimension chromatographic profiles, and integrated to obtain quantitative information.

6.1.2. Multilinearity & Retention Time Alignment

All of the MCR methods discussed here require the data to be very precisely aligned in terms of both chromatographic time scales and spectral signatures. Multilinearity (*i.e.*, bi- tri- or quadrilinearity for two-way, three-way, and four-way data respectively) in multivariate data requires that instrument response of a pure component in all domains is unique, consistent, and independent of the presence of other species [136]. By this definition, all of the quantitative information is contained in one mode or domain. Clearly, in order for a pure component profile to be consistent between samples, a high degree of retention time precision is necessary. Uncontrolled shifts in retention time that often plague chromatographic separations make analysis by many MCR methods difficult. Forcing multilinearity on a data set that has retention time shifting in the chromatographic domain can lead to distorted peak shapes in the chromatographic profiles and poor precision in quantification [134, 137]. Run-to-run retention time shifts have been a major impediment to the implementation of chemometric methods for the analysis of two-dimensional chromatograms [138].

Retention time alignment algorithms (also called warping algorithms) for chromatography data abound in the literature [139-144]. However, the retention time shift in two-dimensional separations can occur in both chromatographic dimensions, creating the need for new methods to align the data. Synovec's group has published several studies of warping algorithms for comprehensive two-dimensional separations [142, 145-147]. Fraga *et al.* [145] presented a stepwise alignment function that was based on a previous report of a one-dimensional alignment function [137]. The one-dimensional function was extended to apply to peak shifts on two time axes, with the shifting that occurs on column 1 treated independently of the shifting that occurs on column 2. More recently, Johnson *et al.* [146] have reported the application of an objective retention time alignment algorithm based on windowed rank minimization alignment. They used a trilinear partial least squares algorithm (tri-PLS) to quantitatively determine the percent volume of napthalenes in jet fuel. They reported a significant improvement in the correlation between the quantitative results obtained using a standard reference method and the tri-PLS method after the alignment algorithm was applied. Pierce *et al.* [142, 147] have also introduced an alignment algorithm that analyzes small, user-defined windows of the data one at a time and shifts the data in a scalar fashion. The retention time precision was improved significantly by the application of the alignment algorithm, leading to subsequent improvements in the quantitative results.

A limitation on the warping algorithms that have been published recently is that they are only applied to a single channel (a single wavelength or mass channel) at a time. When all of the spectral information of the detector is used (as in reference [11]), an

alignment algorithm that simultaneously aligns all detector channels is required in order to avoid loss of information. Such algorithms have not surfaced in the literature as of yet, but there will obviously be a great need for them as the higher order data arrays afforded by two-dimensional separations become more common.

6.1.3. Image Analysis

Another method for analyzing two-dimensional chromatograms is to treat the single channel data as an image and use image analysis tools to obtain information. Hollingsworth *et al.* [148] have recently compared several different methods for visually comparing 2D-GC data sets as images, using retention time data and the intensity information from the detector. In this work they used gray scale difference images to determine relative concentration differences between samples. Colorized difference images also show relative concentration differences between samples while also maintaining the context of the original concentration in the sample. However, these methods are only useful for qualitative comparison of samples rather than absolute quantification. Reichenbach *et al.* [149] used image analysis methods for background removal in 2D-GC-FID chromatograms. The algorithm makes use of chromatographic 'dead bands' at the beginning and end of each second dimension separation as well as the statistical properties of the noise typically found in GC-FID data.

6.1.4. Partial Least Squares

The tri-PLS algorithm was introduced by Bro to build regression models for multi-way data sets with independent and dependent variables [150]. Prazen *et al.* [151] applied tri-PLS to three-way 2D-GC-FID data for the quantitative prediction of the aromatic and naphthene components in naphtha (a petroleum distillate). They demonstrated that fast separations with less than ideal chromatographic resolution can still be useful when chemometric techniques are used to resolve overlapped data. They compared the quantitative prediction of aromatic and naphthene components in naphtha samples to the values obtained using a standard one-dimensional GC separation. For both results, there was good agreement between the standard method and the 2D method. Johnson *et al.* [146] have also showed the application of tri-PLS to the quantitative determination of naphthalenes in jet fuel by 2D-GC-FID. Using a retention time alignment algorithm (*vide supra*) and jet fuel samples with known naphthalene concentrations, they were able to obtain good agreement between experimental results and the standard concentrations.

6.1.5. Other Data Analysis Methods

van Mispelaar *et al.* [152] have discussed several methods for analyzing two-dimensional chromatograms obtained from 2D-GC-MS instruments. They classified three different methods: target-compound analysis and group-type analysis, in which prior knowledge of the sample is required, and fingerprinting, which is an unsupervised technique. Target-compound analysis, as the name implies, indicates converting

retention times or indices and spectra into peak identities and using the peak area information (that is, the detector response) to determine the quantities of specific target analytes. The goal of group-type analysis is to obtain quantitative information on groups of analytes; *e.g.*, a specific chemical class of metabolites from a metabolomic sample. The last approach, which relies heavily on multivariate analysis, correlates the "fingerprint" (chromatogram) of an unknown to a standard to determine which components differ between the samples. This method does not require previous knowledge of the components. It is particularly appealing in systems biology, where biomarkers of interest may be hidden in a chromatogram amongst many unknown peaks and correlations between "diseased" and "healthy" populations need to be found. Pattern recognition tools can be used to develop correlations between the chromatograms and identify biomarkers.

Synovec's group [129, 153] has presented a method they call "DotMap" for the analysis of 2D-GC-MS data, which utilizes all of the spectral information available by using a multichannel detector. The DotMap algorithm, much like the TFA algorithm described in this work, uses a target analyte spectrum and searches the spectra within a two-dimensional chromatogram for signals with similar spectral signatures. It works by calculating the matrix dot product between a target mass spectral signal and the mass spectral signal at each point in the chromatogram. The dot product is similar to the correlation coefficient and the angle θ, as discussed in Chapter 3.5, where the analyte spectrum is projected in the direction of the target spectrum. However, the magnitude of the dot product is related to the similarity of the two spectra such that a larger dot product

indicates similar spectra. The location in two-dimensional space where there is a peak that matches the analyte spectrum can be found by making a contour plot of the dot-product matrix.

Two different analysis methods for dealing with 2D-LC-DAD data were investigated in this work. These methods addressed the need for identifying and quantifying targeted metabolites in metabolic profiling and comprehensive comparative studies in metabolomics [154]. For metabolic profiling studies, an algorithm is needed that can identify target metabolites in an unknown sample based on both spectral characteristics and two-dimensional retention times. The WTTFA algorithm described by Lohnes *et al.* [12] was adapted and used with 2D-LC-DAD data to identify the spectra of selected metabolites in a two-dimensional chromatogram. Towards the second goal of comprehensive comparative studies in metabolomics, quantitative comparisons must be made between samples where the identity of the components may be unknown. Curve resolution methods such as PARAFAC are well suited to such analyses [138]. Rank determination and chemically relevant constraints are important considerations when MCR methods are applied. In this work, a quadrilinear PARAFAC algorithm [14] was applied to rank deficient systems to resolve the overlapped components in the system. Further refinement of the resolved profiles was achieved by applying an alternating least squares algorithm with flexible constraints (fALS) [13].

6.2. Rank Deficiency in Four-Way Data

One challenge in modeling multi-way data is rank deficiency. As a simplified example, an LC-DAD chromatogram with two separated peaks should have a rank of two in the chromatographic mode. However, when those two peaks have the same or very similar DAD spectra, the rank in the spectral dimension will only be one. Analyzing this data by a multi-way model such as PARAFAC, and using a fit diagnostic such as the core consistency diagnostic (CORCONDIA), will indicate that a one-component model suffices to fit the data [155]. However, if the maximum rank of the data is known, a PARAFAC model can still be calculated, and another measure of fit quality (such as sum of squares) can be used instead of the CORCONDIA. In the case of the data analyzed here, the spectra of the 26 indolic standards indicate that these data will be rank deficient in the spectral dimension; many indolic metabolites have similar spectral patterns. Similarly, highly overlapped chromatographic peaks can cause rank deficiency in either of the chromatographic modes. The maximum rank of the data was determined as discussed in Chapter 3.3.

6.3. Materials and Methods

6.3.1. Collection of 2D-LC-DAD Chromatograms

2D-LC-DAD chromatograms resulting from injection of solute-free mobile phase, a mixture containing 26 indolic standards, two *orp* mutant seedling samples, and two wild-type maize seedling samples were obtained. The details of the chromatographic system and the sample preparation for the maize seedlings are reported in detail

elsewhere [26]. The 26 standard compounds included in this study are shown in Figure 21 and include IAA (compound 13) as well as other related indolic compounds. The compounds are numbered in order of their first dimension retention time. The two-dimensional chromatogram of the indole standards, labeled with compound numbers, is shown in Figure 22 as a contour plot at 220 nm.

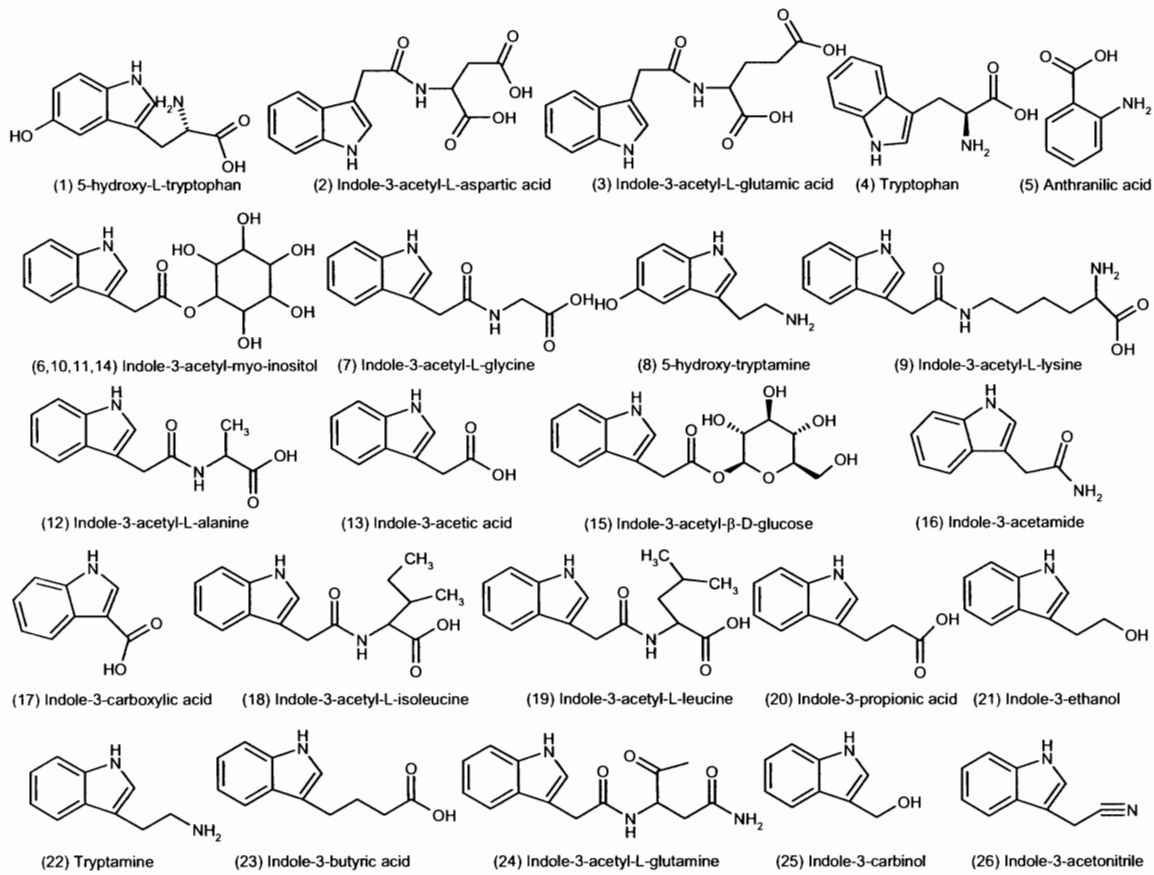

Figure 21. Structures and identification numbers of the 26 indolic metabolites discussed in Chapter 6.

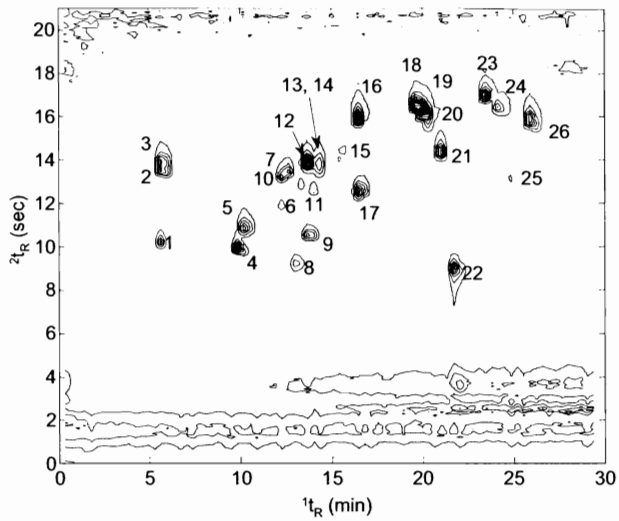

Figure 22. Contour plot of the 2D-LC chromatogram of the 26 indole standards (220 nm). Numbers refer to compounds shown in Figure 21.

6.3.2. Data Analysis

All data analysis was carried out within the Matlab® programming environment (Mathworks, Natick, MA) on a Dell Optiplex GX280 computer with a 3 GHz processor and 2 GB of RAM. Chromatograms collected in HP Chemstation (rev. A.10.01, Agilent Technologies, Palo Alto, CA) were converted into text files with a macro provided by Agilent. The WTTFA algorithm was written in house in the Matlab® programming environment based on the approach outlined by Lohnes *et al.* [12]. The fALS algorithm was written in house as described previously [13]. The PARAFAC algorithm used was part of the N-way toolbox for Matlab® developed by Andersson and Bro, which is available for free download on the internet [14]. All other functions used were built-in Matlab® functions.

6.4. Qualitative Metabolite Profiling with WTTFA

The WTTFA method introduced by Lohnes *et al.* [12] was used to determine if any of the peaks in the mutant and wild-type chromatograms were spectrally similar to the standards. The method was adapted to determine the retention time of the putative matches in two-dimensional chromatographic space. A detailed description of the WTTFA algorithm can be found in section 3.5 of this dissertation. In the analysis of the two-dimensional chromatograms in this work, 26 indolic standards (Figure 21) were used as target compounds for the qualitative analysis of *orp* mutant and wild type maize samples. The spectra of these 26 standards were placed into a single matrix, and a correlation matrix was generated using the built-in correlation coefficient function in Matlab®. Spectra that were correlated by 98.5% ($\theta < 10^{o}$) or more were considered to be identical. Using this criterion, six unique spectra, representative of all the indolic metabolite standards, were used to construct a library; these six spectra are shown in Figure 23. The compound numbers (from Figure 21) represented by each spectral factor are shown in the caption to Figure 23. The six spectra corresponded to six structural classes: IAA and its conjugates (class A), 5-hydroxy-indoles (class B), indole-3-propionic and indole-3-butyric acid (class C), and three spectra that were unique to single compounds, anthranilic acid, indole-3-acetamide, and indole-3-ethanol (classes D, E, and F, respectively).

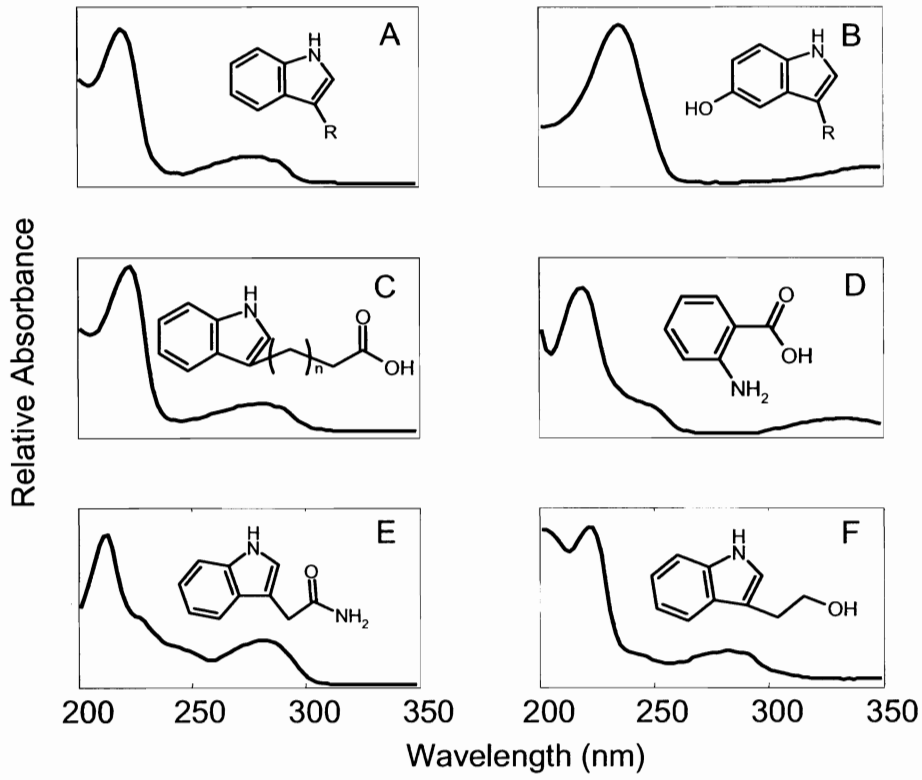

Figure 23. Six unique spectra of the indolic standards in Figure 21. The structure in each figure shows the chemical class represented by each spectrum. (A) compounds 2-4, 6-7, 9-15, 17-19, 22, 24-26; (B) compounds 1 and 8; (C) compounds 20 and 23; (D) compound 5; (E) compound 16; (F) compound 21. The key to the compound numbers is given in Figure 21.

The parameters in the WTTFA algorithm that required optimization were the window size (W), the rank, *i.e.*, the number of components expected within a window (R), and the threshold value for θ. First, the window size was determined based on the peak width of a typical peak in the indole standard chromatogram. Although changing the window size did not have a significant effect on the qualitative results obtained, there was an advantage in using a smaller window, as the computing time was significantly reduced. However, if the window size was too small (*i.e.*, much smaller than the width of a peak), many "noise" peaks were detected. The baseline width (4σ) of a typical second

dimension peak in these data was approximately 800 milliseconds, so a range of window sizes from 0.4 seconds to 1.6 seconds was tested. It was determined that the ideal window size should be slightly bigger than the 4σ width of the peak, therefore a window size (W) of 1 second was ultimately chosen. This choice resulted in the most accurate identification of the two-dimensional retention times for the known standards and the fewest number of extraneous peaks being identified. These results were not in agreement with Lohnes' *et al.* recommendations for the window size (they suggested a window width of 2σ); however they did note that the size of the window had little effect on the qualitative results that they obtained [12].

The maximum number of components (R) expected within a window was set at six, the largest number of components that could reasonably be expected to be resolved within the 1 second window. This choice was based on the complexity of the samples, and the relationship between the window size and the typical peak width. It is quite possible given the sample complexity that overlapped peaks could be present within a 1 second window; however, it was not likely that more than six overlapping compounds within a 1 second window would be observed. Although a rank of six components was somewhat of an overestimation, this parameter had little effect on the results. Finally, the threshold value for θ was set at 5° for these data, although the appropriate threshold may depend on the S/N of the chromatograms of interest and should be considered as a variable related to the data collected. The threshold for θ and the window size are closely related parameters: θ should remain below the threshold value for a time at least equal to the window size in order to be considered a true positive match. In addition to the

approaches described, a modification to this approach using a Gaussian window rather than a boxcar type window was evaluated, but no improvement in the qualitative results was achieved [156].

The algorithm was applied to each unfolded (one-dimensional) chromatogram in the data set to identify the spectral matches, and then the retention times in the first and second chromatographic dimension were determined. The retention times were determined by finding the peak apex within the regions of the first dimension chromatogram wherein θ was below the threshold for at least a period of time equal to W. The peak regions were determined by creating a "boxcar chromatogram" with zeros in the data matrix where θ was above the threshold and ones where θ was below the threshold. Subsequently, the "boxcar chromatogram" was reshaped into a two-dimensional chromatogram. A contour plot of the "boxcar chromatogram" shows the locations of the peaks in two-dimensional space that are highly correlated with one of the six spectra in the library (Figure 23). In this manner, spectral matches can be correlated with the first and second dimension retention times of the standards.

The results of the WTTFA analysis are shown in Figure 24 for one of each of the wild-type (24A) and mutant maize (24B) samples. The dots in Figure 24A and 24B represent the retention times of the indolic metabolite standards in two-dimensional separation space (based on the chromatogram shown in Figure 22). This figure gives a qualitative picture of the spectral signatures of the compounds present in the mutant and wild-type maize samples. The colored contours surround regions where θ is below the threshold of 5° for that spectral class (where the colors correspond to the spectral classes

shown in Figure 23). In those cases where a contour surrounds the location of one of the 26 standards, there is an indication that the peak observed in the unknown chromatogram is indeed a positive match for the standard known to elute in that particular region of the 2D separation space.

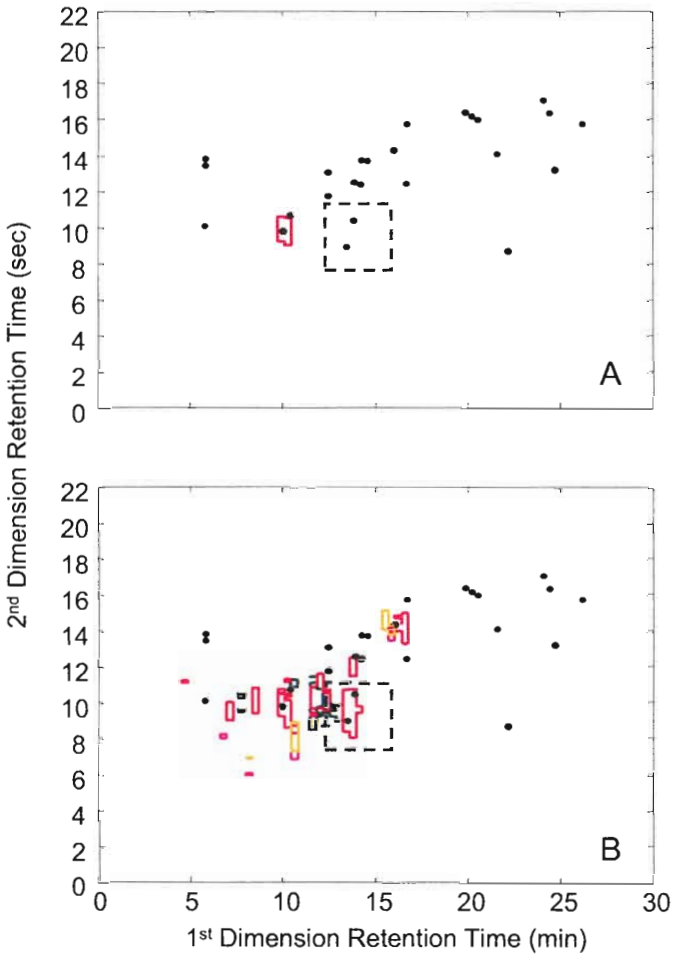

Figure 24. Comparison of WTTFA results for wild-type and mutant corn seedling extracts. Black dots represent the retention times of the 26 indolic standards. (A) Wild-type maize; only spectral component A is detected (red). (B) Homozygous *orp* maize; spectral components A (red), B (not detected), C (yellow), D (green), E (not detected), and F (purple). The boxes denoted by dashed lines in A and B represent the region chosen to illustrate the PARAFAC results in subsequent figures.

The results indicate that there is only one detectable peak in the wild-type maize samples that is consistent with any of the standards analyzed. The red contour in Figure 24A indicates that a component consistent with tryptophan is present in this sample. Based on a similar analysis, the mutant sample contains peaks with spectra that are consistent with tryptophan as well as several other components, including anthranilic acid (5), 5-hydroxytryptamine (8), indole-3-acetyl-L-lysine (9), indole-3-acetyl-myo-inositol (10), and indole-3-acetyl-β-D-glucose (15). It is also interesting to note that there are several contours in the mutant chromatogram that have spectroscopic patterns that resemble one of the six indole classes but that do not match the retention times of any of the 26 standards. These areas indicate as-yet unknown indoles that may play a significant role in the tryptophan-independent biosynthesis of IAA. For example, there are several instances in the mutant chromatogram of spectral factor D (green contours in Figure 24B), which is unique to anthranilic acid compared to the other 25 metabolites studied in this work. This result indicates the presence of other compounds that are structurally related to anthranilic acid that may be of biological interest.

6.5. Quantitative Metabolomic Studies with PARAFAC and fALS

The PARAFAC model has been used extensively for modeling multivariate chromatographic data [157-159]. PARAFAC was employed in this work for the analysis of four-way data, where the general N-way model is shown in Equation 4. Equation 4 represents a multilinear model, and the data analyzed in this work were considered to be quadrilinear. Lin *et al.* [136] defined bilinearity for second-order data, and these

principles are also true for trilinear (third-order) or quadrilinear data (fourth-order). For a data set to be quadrilinear, the instrument response due to a pure component in all four domains should be unique, consistent, and independent of the presence of other species. In other words, the first and second dimension chromatograms and the spectra for each pure component must be identical in every sample, differing only in magnitude, and the relative concentration information is contained entirely in the first mode. This constraint obviously requires a high-degree of retention time reproducibility between runs in order for the retention profiles to be consistent in the various samples. The demonstrated reproducibility of the first and second dimension retention times in these data was sufficient for the data to be considered quadrilinear [26].

While the WTTFA algorithm provides important qualitative information about the samples, quantitative information is often desired as well. LC-DAD is particularly well suited for quantitative analysis due to the high level of peak area and retention time reproducibility that can be achieved on a high quality system. The application of multivariate curve resolution methods (PARAFAC and fALS) are typically for the purpose of extracting pure component profiles and resolving overlapped peaks. These methods also provide an excellent opportunity to resolve background components from chemical components and thus allow the identification of low abundance metabolites in the mutant and wild-type maize seedlings. The PARAFAC algorithm from the N-way toolbox [14] and the fALS algorithm [13] were used together to compare the different samples within the data set and to obtain information about the relative concentrations of the components in each sample.

In principle, the four-way data set could be analyzed by a single application of the quadrilinear PARAFAC algorithm; however, in practice it is not feasible to analyze the entire data in a single analysis. Rank determination via a scree plot of the entire data set indicated the presence of at least 80 – 100 components (which includes background signals as well as chemical components); such an analysis is not computationally realistic using a desktop computer; further, the results would be very difficult to interpret. Instead, the data were broken up into more easily handled sections, which were analyzed individually and the results pooled once analysis was complete. The sections were selected based on visual estimates of the relative complexity of the chromatogram.

The data sections are shown in Figure 25, which shows an overlay of the contour plots from one mutant sample, one wild type sample, and the standard sample at 220 nm. These single wavelength chromatograms illustrate the presence of many overlapped peaks within a sample, as well as some peaks that are common among the samples. The spectral information obtained from the DAD is absolutely necessary to resolve and interpret the relevance of the peaks in these regions. The regions of the chromatograms that are devoid of peaks were not analyzed, and additionally, the dead volume peaks in the second dimension chromatograms were excluded. The rank in each mode was determined both by the scree plot method and by the 90% explained variance method, as discussed in section 3.3. In six out of the nine sections analyzed, the rank determined by visual examination of the scree plots was in close agreement with the 90% variance rank. The scree plot rank was used as a starting point for the analysis only in those cases where

there was a large difference between the two rank determination methods, otherwise the 90% variance rank was used.

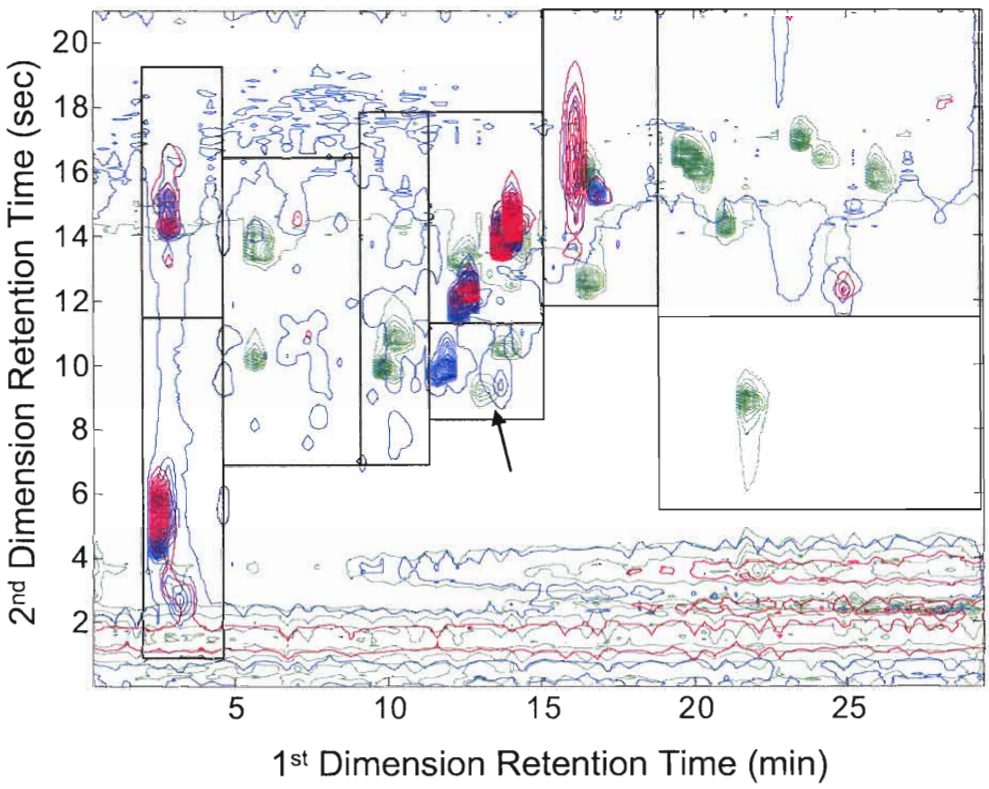

Figure 25. Overlay of contour plots from one mutant sample (blue), one wild type sample (red) and the indole standards (green) at 220 nm. Black boxes represent the sections of data that were analyzed, and the section indicated with an arrow is discussed in detail in the text.

The following discussion pertains to the section of the data indicated by the arrow in Figure 25, but the same data analysis procedure was followed for each section. The 90% variance rank of this section of the data was ten components; this rank was used as a starting point for the modeling. A ten-component model was fit using the PARAFAC function from the N-way toolbox [14]. The initial results revealed that the two standard

components present in this section, 4-hydroxytryptamine (compound 8) and indole-3-acetyl-L-lysine (compound 9), were not resolved. Instead, a single component with a bimodal chromatographic profile was resolved with a spectrum that appeared to be a linear combination of the individual spectra of the standards. Application of a unimodality constraint in the PARAFAC algorithm was not successful at resolving the two standards in the ten-component model, and a much higher number of components were required before the standards were resolved. The problem with applying a model with too many components is that the model begins to fit noise, and it becomes difficult to obtain accurate quantitative information.

The fALS algorithm, previously presented by Bezemer and Rutan [13], allows flexibility in the implementation of chemically relevant constraints such as unimodality. It proved to be very useful for solving the issues with the PARAFAC algorithm discussed above. With this method any constraint can be selectively applied to specific components and/or specific modes. The unimodality constraint is applied to only those components that represent chromatographic peaks, while the background components can have multimodal profiles. The results from the ten-component PARAFAC model (without imposing unimodality) were used to initiate the fALS algorithm (first applied without unimodality). The components giving rise to peak-shaped responses were identified, and the algorithm was run again, this time applying the unimodality constraint only to those components that appeared to be chromatographic peaks. The results from the previous fit were used to initiate the iterations. When the results from the second pass indicated that one of the unconstrained components tended to exhibit a peak-like (as opposed to a

gradient-like background) response, that component was constrained to unimodality in the subsequent fit.

A total of two background components generally sufficed to describe the background gradients in each section. In cases where there appeared to be more than two background components contributing to the model, the rank of the model was decreased, and the data were refit using the procedure described above. In cases where there were peaks evident in the raw data that were not being resolved, the rank of the model was increased. After the formal mathematical analysis was conducted the resolved factors (*i.e.*, chromatograms in both dimensions) were visually inspected to ensure that no artifacts had been introduced. The fALS function [13] does not allow true quadrilinearity to be applied; therefore the data were augmented in the spectral dimension to form a three-way data array. To complete the quadrilinear data analysis, the results from the fALS algorithm were used to initiate a final application of the four-way PARAFAC algorithm, using only the non-negativity constraint. This successive application of the two different curve resolution algorithms capitalized on the application of flexible constraints afforded by fALS, and the application of quadrilinearity afforded by PARAFAC.

The data section highlighted in Figure 25 was ultimately fit using a nine-component model, and the resolved profiles are shown in Figure 26. The two background components have been removed for clarity. The blue and red traces in Figure 26 represent the first two components resolved in this section of the data. These compounds elute at $^{1}t_{R}$ = 11.6 min; $^{2}t_{R}$ = 9.6 sec and $^{1}t_{R}$ = 11.9 min; $^{2}t_{R}$ = 9.8 sec

respectively; that is they very nearly co-elute in both chromatographic dimensions, and they have nearly identical spectra. They are present at relatively high concentration in both mutant samples, but are not present in the standard or wild type samples. The relative amount of the two components differs significantly between the two mutant samples (Figure 26D), but this is probably due to the high degree of overlap of the components in both chromatographic directions. It should be noted that the sum of these two components to the two mutant samples is reasonably consistent.

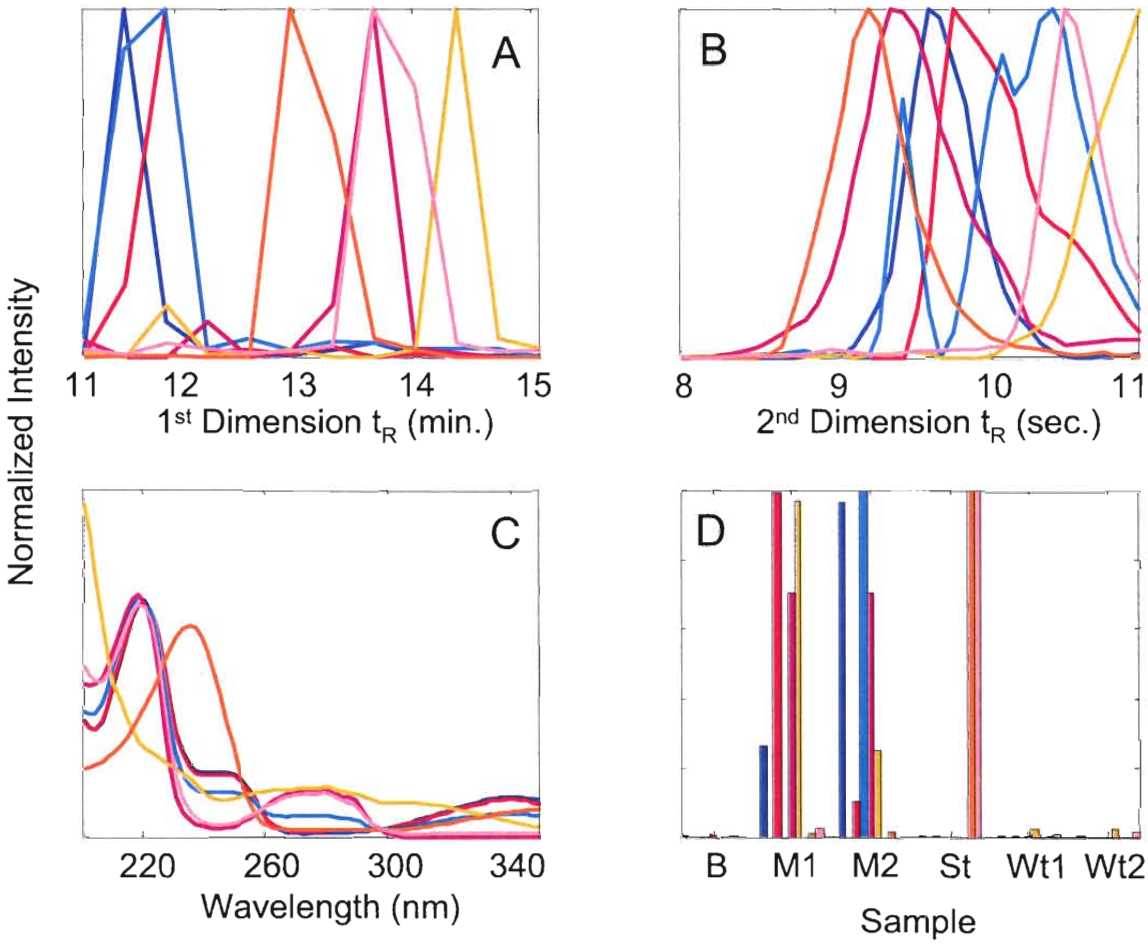

Figure 26. Resolved PARAFAC profiles of selected section (see Figure 25) for a nine-component model. (A) First dimension retention profiles; (B) second dimension retention profiles; (C) spectral profiles; (D) concentration profiles.

The two standard components in the selected section are shown by the orange and pink traces in Figure 26. These compounds were present in the standard chromatogram, and correspond to 5-hydroxytryptamine (compound 8, orange trace, ^1t_R = 13.0 min; ^2t_R = 9.2 sec) and indole-3-acetyl-L-lysine (compound 9, pink trace, ^1t_R = 13.7 min; ^2t_R = 10.5 sec), respectively. Indole-3-acetyl-L-lysine and 5-hydroxytryptamine were found at very low levels (S/N<10) in the mutant type samples, but were not detected in the wild type samples. The component represented by the magenta trace, eluting at ^1t_R = 13.7 min; ^2t_R = 9.4 sec, is well-separated chromatographically from the indole-3-acetyl-L-lysine peak in the second dimension only, and exhibits a very similar spectral response. Based on its spectrum, this component is likely an indole conjugate.

These results are consistent with the WTTFA analysis. It can be seen in Figure 24 (in the section indicated by the dashed box) that a red contour surrounding the two standards was found in the mutant, but not in the wild-type maize. This is a strong indication that there is a peak in the mutant maize sample consistent with the spectra and retention time of 5-hydroxytryptamine, as confirmed by the PARAFAC-ALS analysis.

Figure 27 shows a demonstration of the ability of the curve resolution methods to minimize noise and remove gradient background signals. Panel A shows a surface plot of the 2D chromatogram at 220 nm of one of the wild type samples. There is a large amount of background present in this section of the chromatogram, and while there appears to be a peak present, it is has a very low S/N and is barely detectable. Panel B shows the ALS model for these data after the resolved profiles were reconstructed according to Equation 4; clearly, a great deal of noise has been removed by the fitting

procedure. Finally, the two background components that were resolved from the data were omitted before carrying out the reconstruction, as shown in panel C. The background signal is virtually eliminated, and two peaks are now clearly evident in the surface plot. These peaks correspond to the blue and yellow components in Figure 26.

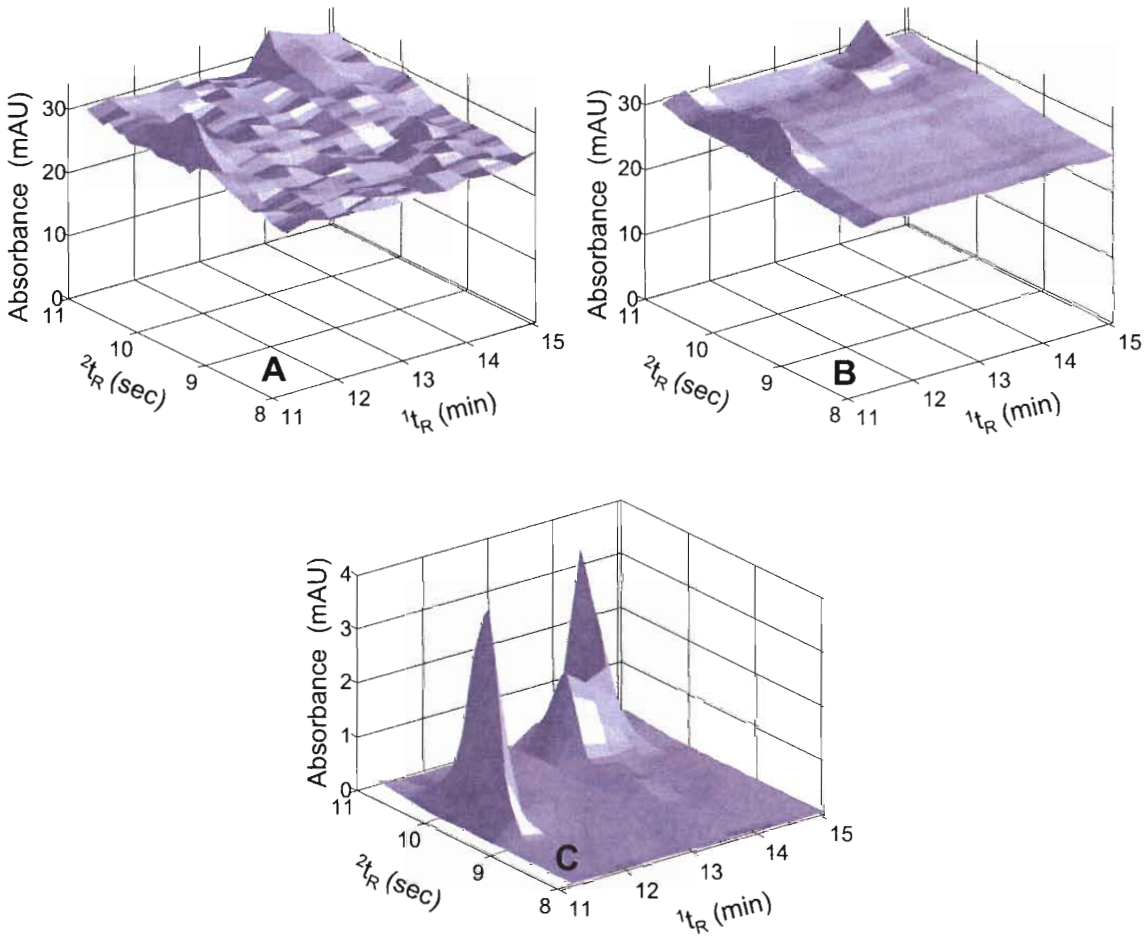

Figure 27. Comparison of chromatograms at 220 nm of the selected section of the data for one of the wild-type samples. (A) Raw data; (B) reconstructed data from PARAFAC profiles; and (C) reconstructed data from PARAFAC profiles with background components omitted.

Each section of the data outlined in Figure 25 was analyzed by the same procedure described above. The pooled results of the analysis of the entire four-way data set are shown in Figure 28. The dots in Figure 28 show the locations in two-dimensional chromatographic space of every peak that was resolved by the PARAFAC and fALS algorithms. The colors correspond to the sample type (mutant, wild-type, or standard) in which the peaks were found with a S/N of at least 10. Table 11 summarizes these results. A total of 95 components were resolved, not including the background components. The only standard that was identified in both the mutant and wild type maize samples was tryptophan (indicated by the black dot in Figure 28) – this result is consistent with the qualitative WTTFA results. In addition, 5-hydroxy-L-tryptophan and tryptamine were identified in the mutant samples. The other standards that were identified in the wild-type and mutant maize (magenta and yellow dots) were conjugates of IAA.

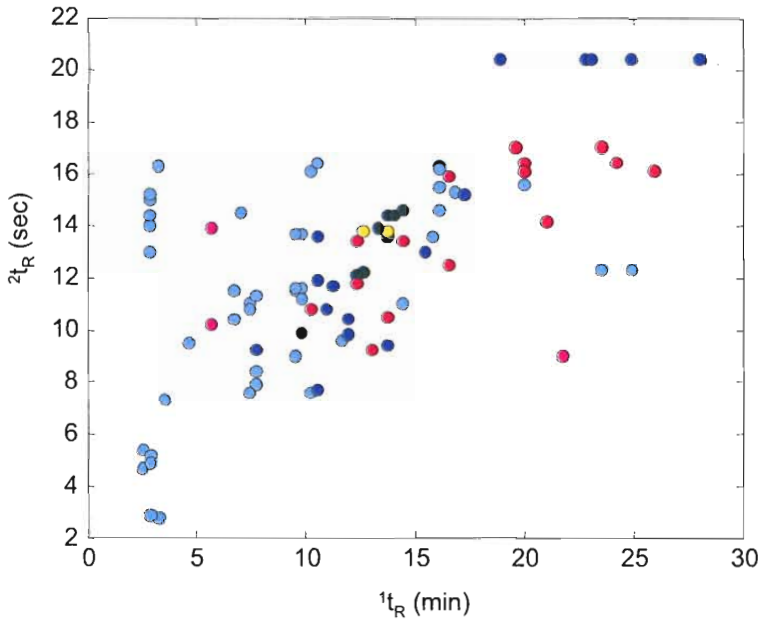

Figure 28. All peaks resolved by the PARAFAC-fALS algorithms. Blue – mutant only; green – wild type only; red – standard only; cyan – mutant and wild type; magenta – mutant and standard; yellow – wild type and standard; black – mutant, wild type, and standard. The only black dot represents tryptophan.

Table 11. Summary of the peaks resolved in each sample at or above a S/N of 10. Retention times are shown in Figure 28.

	Mutant only	WT only	Std. only	Mutant & WT	Mutant & Std.	WT & Std.	Mutant, WT, & Std.
No. of peaks resolved	16	13	15	45	3	2	1
Identity of standards (compound number) [a]	--	--	5-6, 8-9, 11, 13/14, 16-21, 23-24, 26[b]	--	1, 2/3[c], 22	7/10[c], 12	4

[a] For compound numbers, see Figure 21.
[b] Standards 15 and 25 were not detected.
[c] Standards 2 & 3, 7 & 10, and 13 & 14 were unresolved pairs.

The quantitative results for 5-hydroxy-L-tryptophan, tryptophan, indole-3-acetyl-L-alanine and tryptamine are summarized in Table 12. The levels for tryptophan in the maize seedlings are consistent with previously reported levels of tryptophan in Lemna (duckweed), which has been used as a model system similar to maize [160]. Quantitative results for the IAA conjugates of glutamine and aspartic acid (compounds 2 & 3) and indole-3-acetyl-L-glycine and –myoinositol (compounds 7 & 10) could not be obtained because these species were unresolved by the PARAFAC algorithm (*cf.* Table 11).

Table 12. Quantitative results of PARAFAC analysis for selected standards. Concentrations are reported in μg standard per gram plant material.

	Mutant 1	Mutant 2	WT1	WT2
5-hydroxy-L-tryptophan	ND*	0.6	ND	ND
Tryptophan	1.4	1.1	1.8	1.6
Indole-3-acetyl-L-alanine	ND	ND	0.8	0.3
Tryptamine	1.9	ND	ND	ND

*ND – not detected

6.6. Biological Relevance of Results

IAA was not detected in either the wild type or the mutant maize samples. IAA eluted close to a retention time corresponding to a very large peak that saturated the detector and resulted in negative absorbance values at the apex of the peak. The second dimension chromatogram in this region had to be omitted from the analysis so these peaks could not be resolved. However, several of the other compounds that were

identified in the mutant and wild type samples are of biological relevance. It is known that anthranilate is a precursor to IAA in the biosynthetic pathway of tryptophan in maize and other plants [161]; it is therefore not surprising that anthranilic acid and related conjugates might be detected in the maize seedlings. In addition, previous studies have shown that *orp* seedlings contain a significant amount of indole-3-acetic acid as an amide conjugate [162], however its identity was unknown. Tentative identification by 2D-LC is a major step forward in sorting out the changing indolic composition in these plants.

Indole-3-acetyl-L-lysine was previously isolated from bacteria, where it appears to be involved in plant gall formation [163], but it has not previously been found in plants. However, it is a compound of unique interest in that plants also contain a class of proteins with an apparent post-translational indoleacyl modification at specific lysine residues on the proteins [164]. Hydroxytryptamines are relatively unusual in maize, although several routes for indole degradation or indole alkaloid production involve ring oxidations [165]. It is important to point out that these results could not be obtained without the extra information provided by the DAD spectra of the metabolites. Identification based on retention times alone would not be feasible given the large number of peaks detected, many of which are low abundance peaks overlapped with larger peaks.

6.7. Multivariate Selectivity in Four-Way Data

The multivariate selectivity metric can be used to define the useful information contained within a two-dimensional chromatogram. For example, Sinha *et al.* have

shown a relationship between two-dimensional chromatographic resolution (2D-R_s) and multivariate selectivity [166]. They found that above a 2D-R_s of 0.75, the multivariate selectivity approaches one, which indicates that the components can be quantified with the same precision in a mixture as in a pure sample. They also found that by using chemometric methods to mathematically resolve overlapped peaks (those with 2D-R_s of less than 0.75), peaks with a 2D-R_s down to 0.30 could be quantified.

In this work, the retention times of the 95 components and the spectra resolved by the PARAFAC algorithm were used to construct a theoretical chromatogram of peaks with peak widths of 0.7 minutes and 800 msec on the first and second dimension columns, respectively. Table 13 shows the effect of adding dimensionality to the data by comparing the average multivariate selectivity of the two independent chromatographic dimensions with and without DAD, and the 2D-LC with and without DAD. As shown in Table 13, the improvement in the selectivity appears to scale approximately with the improvement in peak capacity and allows the increase in information provided by DAD detection to be quantified. Ultimately, combining two dimensions of chromatography with DAD allows an average selectivity of 0.84 for each component, an increase of 8% relative to single wavelength detection.

Table 13. Summary of selectivity calculated with various data dimensions.

Data Dimensions	Average selectivity per component [a]	Peak Capacity [b]
Column 1, single wavelength	0.15	50
Column 1 + DAD	0.36	n.a.
Column 2, single wavelength	0.05	17.4
Column 2 + DAD	0.25	n.a.
2D-LC, single wavelength	0.78	870
2D-LC-DAD	0.84	n.a.

[a] Selectivity calculated according to Equation 9 for a total of 95 components.
[b] Peak capacity as reported in reference [26].
n.a. Peak capacity not defined.

It should be noted that the figures reported in Table 13 are the *average* selectivities over 95 components. If a particular component is well resolved chromatographically (in one or both dimensions), there will be little improvement in selectivity upon the addition of DAD. Conversely, a component that is highly overlapped chromatographically may show a dramatic improvement in selectivity when the spectral dimension is added. For example, the individual selectivity of the resolved component corresponding to standard 12 (indole-3-acetyl-L-alanine) shows a selectivity of only 0.16 in two dimensions of chromatography due to the fact that it is highly overlapped with several peaks in the standard, mutant and wild-type maize samples. Upon the addition of the DAD spectrum, the selectivity of this component is improved by over 200%.

This analysis indicates that an improvement in the precision of quantification can be achieved by using multi-channel detectors in conjunction with MCR, without the

corresponding increase in analysis time usually required to improve chromatographic peak capacity. The magnitude of the increase in selectivity will be directly related to the information content of the detector; that is, a more selective detector such as MS will have a correspondingly greater increase in selectivity than a DAD. In addition, the gain in selectivity upon the addition of multi-channel detection will be more dramatic as the number of components and the sample complexity increase.

A further discussion of selectivity in multi-way methods, specifically as it applies to 2D-LC and 2D-LC-DAD, can be found in reference [122] by Cantwell *et al.* In this work, comprehensive 2D-LC chromatograms were simulated both with randomly distributed peaks (effectively using the whole separation space) and with correlated peaks (to more closely approximate the actual 2D-LC data observed for the maize samples). The correlated peaks were distributed randomly along the first dimension time axis. The second dimension retention times were simulated based on a logarithmic function of the first dimension retention times to provide a shape for the distribution that approximated that seen in Figure 28. The values were allowed to randomly deviate from the logarithmic function by the addition of normally distributed random numbers then scaling the resulting retention times to the second dimension time axis. A boundary of 2σ for the retention times at both the start and end of the time axis was imposed on both dimensions in order to force all peaks to be fully within the separation space. The peak widths in the first and second dimension and the correlated peak function were designed to approximate the metabolomics data that are discussed in this chapter.

The purpose of these simulations was to evaluate the effect of the second dimension separation time on the selectivity of a 2D-LC method. A total of 500 peaks were simulated in order to provide adequate coverage of the separation space and to provide a moderate level of interferences. The second dimension time-axis was varied from 2 to 44 sec while the first dimension time axis was fixed at 30 min. Two-dimensional chromatograms with no spectral dimension (single wavelength) constituted the three-way simulations. The simulations were repeated with a spectral dimension (four-way simulations) that contained very similar spectra in order to model the effects of an additional, relatively non-selective dimension of data. The spectra for the 500 components were randomly sampled from the 47 drug spectra described in Chapter 5.

The main advantage of comprehensive 2D-LC is the increased separation space; however, the entire separation space will never be used in any method where the two columns share some similarity in the separation mechanism. Complete orthogonality is all but impossible when the same mode of separation is used in both dimensions. The 2D-LC system described by Stoll *et al.* [26] uses reversed phase gradient elution in both dimensions. The first dimension column was less retentive on average than the second dimension column. The analysis of the four-way data described in this chapter showed that the vast majority of the components in the mixtures eluted in the upper-left corner of the separation space, and only about 70% of the total possible separation space was used (*cf.* Figure 28). However, this limitation does not negate the potential of 2D-LC to separate complex samples with many components, and the practical peak capacity of the 2D separation is still far greater than for a 1D separation. Stoll *et al.* reported a peak

capacity production of 35 peaks/min for the system described in reference [26], which is typically more than the peak capacity production for 1D systems.

For these simulations, the second dimension time axis was varied from 2 to 44 seconds to show the effect of the second dimension separation time on the average selectivity for the components in the chromatogram. The length of the second dimension time axis directly affects the sampling rate in the first dimension, and hence the number of points that can be obtained across a peak. As a result, the selectivity in the first dimension is directly related to the second dimension sampling time. An optimal time for the second dimension separation should exist that balances both the selectivity that can be attained in the second dimension and the resulting selectivity in the first dimension. In other words, increasing the second dimension separation time will necessarily increase the selectivity in that mode; however the trade-off is a loss in selectivity in the first dimension separation due to the decrease in sampling rate.

Figure 29 is a plot of the average selectivity as a function of the second dimension separation time for the correlated chromatograms. The circles represent the selectivity for the correlated peaks with the spectra included (four-way case) and the crosses are the selectivity for the correlated peaks without the spectra included (three-way case). The selectivity for both cases increases very rapidly between the two to ten second range and then levels off. While continuing to increase the second dimension separation time will improve the selectivity in that dimension, the sampling rate of the first dimension essentially goes to one, which severely diminishes the selectivity in the first dimension. The average selectivity of the correlated simulations is only about 10 % lower than that

of the random simulation, and the shapes of the curves are essentially the same. Table 14 shows a comparison of the average selectivity of the correlated retention times and the random retention times for the 18 sec second dimension separation time. Figure 29 and Table 14 both show that the addition of the spectral dimension does in fact improve the average selectivity, even for a detector with relatively low selectivity.

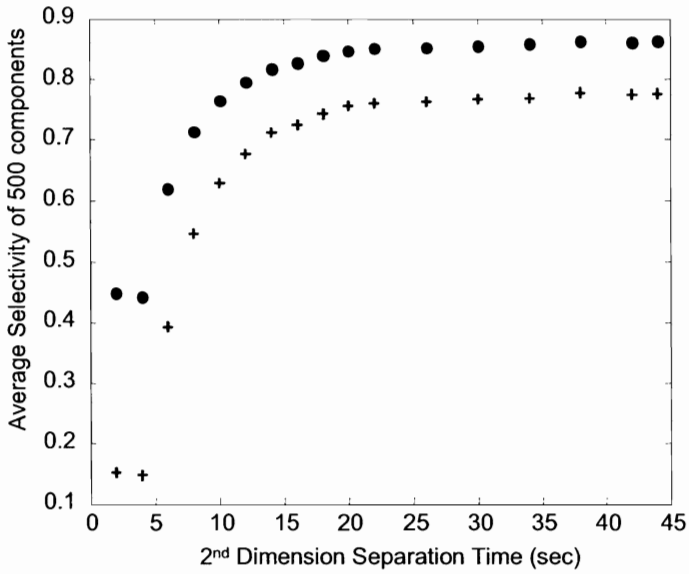

Figure 29. Comparison of the average selectivity of 500 components in the correlated comprehensive 2D-LC chromatograms for the four-way case ('●') and the three-way case ('+') as a function of the second dimension separation time.

Table 14. Comparison of the average selectivity for the 18 second 2nd dimension separation for the correlated and random retention distributions.

Data set	Average Selectivity
Four-way correlated	0.92
Three-way correlated (no spectra)	0.87
Four-way random	0.85
Three-way random (no spectra)	0.76

Murphy, Schure and Foley extensively discussed the effect of sampling rate on the peak width in the first dimension of comprehensive 2D-LC separations [20]. The

effective peak width is significantly broadened compared to the ideal Gaussian when the sampling rate is low; it is this peak broadening that effectively causes the selectivity in that dimension to decrease and limits the improvement in overall selectivity when the second dimension sampling time is increased. These results do indicate that there is an optimal second dimension separation time where the selectivity is no longer increasing and has reached a maximum. The point at which the average selectivity is 98% of the maximum value occurs at a second dimension separation time of 20 seconds for both the three-way and four-way correlated cases. The time for the second dimension separation in the work reported by Stoll *et al.* was 18 seconds plus a 3 second re-equilibration time for the second dimension column [26]. This correlation between the simulation and the actual chromatographic method upon which the simulation is loosely based is confirmation of utility of the modeling program. Figure 30 shows a theoretical first dimension chromatographic profile where the second dimension separation time is 20 seconds. This figure shows that there are at least 2 – 3 points across a first dimension peak, which comes close to meeting the sampling rate requirements suggested by Murphy, Schure and Foley [20].

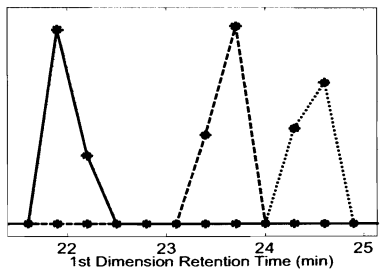

Figure 30. Example of the first dimension chromatographic profiles where the second dimension separation time is 20 seconds, showing that there are 2 – 3 points across each peak.

The results of the work discussed in this chapter will help to guide future work in the area of chemometrics and two-dimensional chromatography. Although the techniques discussed here were specifically applied to DAD data, they should be applicable to MS detection or any other multichannel detector with a two-dimensional separation. These methods take advantage of the full four-way data structure of 2D-LC-DAD, and it is anticipated that further development of the methods will enhance metabolomics studies.

CHAPTER 7. Conclusions and Future Work

The goals of all of the projects presented in this dissertation are ultimately the same: to develop and apply chemometric methods to liquid chromatography data of varying levels of complexity. Beginning with the data obtained in an *in-vitro* drug metabolism experiment on an LC-MS instrument, and ending with the highly complex metabolomic data obtained from two-dimensional chromatography with a multi-channel detector, unique combinations of existing chemometric methods have been used to obtain relevant information from these data.

Two kinetic fitting routines were used to determine the *in-vitro* pharmacokinetic parameters of several drug systems [3]. PMMA and fluoxetine were used as models to validate a new method of screening intrinsic clearance values from *in-vitro* kinetic data. Steady-state experiments served as test cases to compare the "traditional" method for determining intrinsic clearances to the new method, and a steady-state kinetics curve fitting algorithm was used to fit the data [6]. Both PMMA and fluoxetine showed evidence of atypical kinetic profiles.

The results of the GE experiments showed that a simple kinetics method can be used to predict intrinsic clearance with similar precision to conventional SS experiments. By simply monitoring the formation of product or the depletion of substrate as a function of time, the micro-rate constants of any kinetic model can be calculated and used to

estimate the intrinsic clearance. The PMMA incubation fit the general enzyme model at the low substrate level studied, but the fluoxetine required a more complex model to adequately fit the data and to determine intrinsic clearance. The clearances for both the PMMA and fluoxetine were in agreement between the two different methods, despite the fact that the concentration levels and incubation conditions were vastly different between the two methods. The GE method was modified to study the interaction between PMMA and fluoxetine. Adding an equimolar amount of fluoxetine to the mixture caused the clearance of PMMA to be reduced by nearly an order of magnitude. The significance of this work is the presentation of new methodology for studying *in-vitro* pharmacokinetics.

Future studies in this area should focus on characterizing other drug – drug interactions using similar experiments and using multiple inhibitor concentrations to calculate inhibition constants. The extension of this technique to new drug entities should be possible as long as the mass spectrometer used for detection has adequate sensitivity in the full scan mode to detect substrate and products without prior knowledge of their structures or masses. Time-of-flight mass spectrometry would be ideal for this purpose. In this work, an ion trap spectrometer was used, which necessitated the use of SRM detection in some cases to obtain adequate detection sensitivity (which required knowledge of the masses of the analytes).

The next chapter detailed the analysis of two-way LC-DAD data using a TFA algorithm derived for library searching of spectra within a chromatogram. Fast LC-DAD chromatography and the TFA algorithm were used to design a screening method for drugs in biological matrices [7]. The application of a corrected retention index

calculation for drug identification allowed the comparison of the peaks in a chromatogram to a retention library, and compensated for day-to-day and column-to-column variability in retention times over a period of 13 months. The screening method was able to identify those samples that required confirmatory testing with a low rate of false positive and false negative results. A training data set was used to optimize the method parameters, and a validation data set was used to calculate a sensitivity of 92% and a specificity of 94%. By requiring both spectral and retention index matches, the method effectively identified samples that contain drug peaks, and more importantly, could specify which drugs are present. The ability to more specifically identify the target compounds present in a sample is a key improvement in this method relative to traditional immunoassays that are used for drug screening. Knowledge of the potential analytes will make the application of a confirmatory method such as LC-MS more efficient by allowing the analyst to tailor the detector parameters for the compounds found in the screening phase.

Future work in this area could involve refining the parameters of the screening method to be more universally applicable (as opposed to specifically designed for the data being analyzed) and the application of MCR to improve the sensitivity and specificity of the method by further resolving overlapped peaks. Preliminary studies indicate that the fALS algorithm described in reference [13] would be useful for this purpose; however, automating the process will be a challenge. Orthogonal column LC-DAD could also be explored for its potential in improving the results obtained with the screening method. As discussed in Chapter 5, simulations have shown that the use of

orthogonal LC-DAD can improve the multivariate selectivity of a method. With the very fast LC methods that were developed in reference [8], an orthogonal separation would likely improve the method significantly, especially as more compounds are added to the library.

Finally, four-way 2D-LC-DAD data were analyzed using three chemometric methods, WTTFA, PARAFAC, and fALS [11]. The quadrilinear data generated by analyzing several related samples using 2D-LC-DAD is complex at best, unmanageable at worst. WTTFA provided a qualitative analysis of the data set by comparing a set of known metabolite spectra to the spectra resolved in the unknown samples (the wild type and mutant maize samples). The use of the PARAFAC model allowed for the quantitative comparison of all of the resolved components present in the data. The unique combination of a flexibly constrained algorithm (fALS) and quadrilinear data analysis (PARAFAC) are what distinguished this analysis from previously published work. The ability to selectively implement the unimodality constraint within the fALS algorithm was found to be particularly useful in these studies. Coupling these two algorithms allowed for the resolution of overlapped peaks and background components from complex mixtures of plant metabolites. These techniques also greatly enhanced the S/N of the instrument response for low abundance peaks.

Of the 26 indolic metabolite standards included in this study, several of the standards were identified at significant levels (S/N > 10) in both the mutant and wild type maize samples, including tryptophan, which was identified in both. The results of the WTTFA analysis indicated that a few additional peaks observed only in the mutant

chromatograms were 'indole-like', based on their spectroscopic characteristics, and the PARAFAC results confirmed this observation. These peaks represent important potential targets for future biological studies. Given the huge number of potential compounds in the typical plant metabolome (estimated at up to 200,000 [154]) it is not surprising that only a small fraction of the unknown components could be identified using a limited number of standard compounds in the spectral library. The relative concentration information that can be obtained using multivariate methods also provided valuable information about the differences in the levels of metabolites between different sample types. For example, based on this preliminary work with the *orp* mutant, there are at least 19 compounds that differentially accumulate in the *orp* mutant relative to wild type maize.

One important result of the work presented here is the possibility for analyzing four-way data using the quadrilinearity constraint imposed by PARAFAC while allowing flexibility in the other chemically relevant constraints such as unimodality and non-negativity. While the work here used two separate algorithms to complete the analysis, the combination of the two methods into a more efficient program should be the focus of future work. In addition, quantitative studies of organisms generally require multiple replicate samples in order to draw any statistical conclusions regarding the chemical make-up of the organism [154]. Standard addition experiments will improve the quantitative results obtained, and retention time alignment algorithms will be useful for data that deviate from multilinearity.

The results of all of these studies provide important direction for future work in the area of chemometrics as applied to chromatographic data. Mass spectrometric detectors are becoming the mainstream in many analysis laboratories due to their high sensitivity and analyte selectivity. The development of a fast LC-MS method and a general screening method, as described in Chapter 4, for the analysis of *in-vitro* drug metabolism is an important contribution to the field of illicit drug metabolism. As chromatographic separations become faster and more efficient, samples of increasing complexity are being analyzed. High throughput methods such as those described in Chapter 5 will help to increase the efficiency and accuracy of testing labs. Finally, although 2D-LC is still a relatively young technology, its popularity is growing rapidly. It is clear that the multivariate data obtained from such instrumentation will require data analysis techniques beyond simple peak integration. Methods such as the ones described in Chapter 6 will advance the analysis of such data to allow quantitative and qualitative analysis of complex proteomic and metabolomic samples by 2D-LC with multichannel detection. The chromatographic methods discussed in this work represent the state-of-the-art in liquid chromatographic separations, and the application of chemometric methods will maximize the value of these instrumental methods.

Literature Cited

Literature Cited

[1] W. Lindberg, J. Ohman, S. Wold, "Multivariate Resolution of Overlapped Peaks in Liquid-Chromatography Using Diode-Array Detection" Anal. Chem. 58 (1986) 299.

[2] A. Smilde, R. Bro, P. Geladi, Multi-way Analysis with Applications in the Chemical Sciences. John Wiley & Sons, Ltd., Hoboken, NJ, 2004.

[3] S. E. G. Porter, R. B. Keithley, S. C. Rutan, "Development of an *In-Vitro* Incubation Procedure for Screening CYP2D6 Intrinsic Clearance" J. Chromatogr. B doi:10.1016/j.jchromb.2006.11.006 (2006).

[4] T. S. Tracy, "Atypical Enzyme Kinetics: Their Effect on *in vitro-in vivo* Pharmacokinetic Predictions and Drug Interactions" Curr. Drug Metabol. 4 (2003) 341.

[5] E. Bezemer, S. C. Rutan, "Multivariate Curve Resolution with Non-Linear Fitting of Kinetic Profiles" Chemom. Intell. Lab. Syst. 59 (2001) 19.

[6] R. Sánchez-Ponce, S. C. Rutan, "Steady State Kinetic Model Constraint for Multivariate Curve Resolution - Alternating Least Squares Analysis" Chemom. Intell. Lab. Syst. 77 (2005) 50.

[7] S. E. G. Porter, D. R. Stoll, C. Paek, S. C. Rutan, P. W. Carr, "Fast Gradient Elution Reversed-Phase Liquid Chromatography with Diode-Array Detection as a High Throughput Screening Method for Drugs of Abuse: II. Data Analysis" J. Chromatogr. A 1137 (2006) 163.

[8] D. R. Stoll, C. Paek, P. W. Carr, "Fast Gradient Elution Reversed-Phase High Performance Liquid Chromatography with Diode-Array Detection as a High Throughput Screening Method for Drugs of Abuse: I. Chromatographic Conditions" J. Chromatogr. A 1137 (2006) 153.

[9] A. C. Moffat, K. W. Smalldon, C. Brown, "Optimum Use of Paper, Thin-Layer and Gas-Liquid Chromatography for the Identification of Basic Drugs I. Determination of Effectiveness for a Series of Chromatographic Systems." J. Chromatogr. 90 (1974) 1.

[10] P. G. Schepers, J. P. Franke, R. A. de Zeeuw, "System Evaluation and Substance Identification in Systematic Toxicological Analysis by the Mean List Length Approach" J. Anal Toxicol. 7 (1983) 272.

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

[12] M. T. Lohnes, R. D. Guy, P. D. Wentzell, "Window Target-Testing Factor Analysis: Theory and Application to the Chromatographic Analysis of Complex Mixtures with Multiwavelength Fluorescence Detection" Anal. Chim. Acta 389 (1999) 95.

[13] E. Bezemer, S. C. Rutan, "Three-Way Alternating Least Squares Using Three-Dimensional Tensors in MATLAB®" Chemom. Intell. Lab. Syst. 60 (2002) 239.

[14] C. A. Andersson, R. Bro, "The N-Way Toolbox for MATLAB" Chemom. Intell. Lab. Syst. 52 (2000) 1.

[15] C. F. Poole, The Essence of Chromatography. Elsevier, New York, NY, 2003.

[16] P. Marquet, "Progress of Liquid Chromatography-Mass Spectrometry in Clinical and Forensic Toxicology" Ther. Drug Monit. 24 (2002) 255.

[17] H. J. Issaq, K. C. Chan, G. M. Janini, T. P. Conrads, T. D. Veenstra, "Multidimensional Separation of Peptides for Effective Proteomic Analysis" J. Chromatogr. B 817 (2005) 35.

[18] K. M. Oksman-Caldentey, D. Inze, M. Oresic, "Connecting Genes to Metabolites by a Systems Biology Approach" Proc. Nat. Acad. Sci. 101 (2004) 9949.

[19] J. C. Giddings, "Two-Dimensional Separations: Concept and Premise" Anal. Chem. 56 (1984) 1258A.

[20] R. E. Murphy, M. R. Schure, J. P. Foley, "Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography" Anal. Chem. 70 (1998) 1585.

[21] M. M. Bushey, J. W. Jorgenson, "Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid-Chromatography of Proteins" Anal. Chem. 62 (1990) 161.

[22] M. J. Gray, G. R. Dennis, P. J. Slonecker, R. A. Shalliker, "Comprehensive Two-Dimensional Separations of Complex Mixtures Using Reversed-Phase Liquid Chromatography" J. Chromatogr. A 1041 (2004) 101.

[23] T. Ikegami, T. Hara, H. Kimura, H. Kobayashi, K. Hosoya, K. Cabrera, N. Tanaka, "Two-Dimensional Reversed-Phase Liquid Chromatography Using Two

Monolithic Silica C18 Columns and Different Mobile Phase Modifiers in the Two Dimensions" J. Chromatogr. A 1106 (2006) 112.

[24] D. R. Stoll, P. W. Carr, "Fast, Comprehensive Two-Dimensional HPLC Separation of Tryptic Peptides Based on High-Temperature HPLC" J. Am. Chem. Soc. 127 (2005) 5034.

[25] G. J. Opiteck, S. M. Ramirez, J. W. Jorgenson, I. I. I. Moseley, "Comprehensive Two-Dimensional High-Performance Liquid Chromatography for the Isolation of Overexpressed Proteins and Proteome Mapping" Anal. Biochem. 258 (1998) 349.

[26] D. R. Stoll, J. D. Cohen, P. W. Carr, "Fast, Comprehensive Online Two-Dimensional High Performance Liquid Chromatography Through the Use of High Temperature Ultra-Fast Gradient Elution Reversed-Phase Liquid Chromatography" J. Chromatogr. A 1122 (2006) 123.

[27] M. S. Lesney, in J. F. Ryan (Editor), Chromatography: Creating a Central Science. American Chemical Society, Washington, D.C., 2006.

[28] L. Huber, S. A. George, Diode Array Detection in HPLC. M. Dekker, New York, NY, 1993.

[29] M. S. Denton, T. P. Deangelis, A. M. Yacynych, W. R. Heineman, T. W. Gilbert, "Oscillating Mirror Rapid Scanning Ultraviolet-Visible Spectrometer as A Detector for Liquid-Chromatography" Anal. Chem. 48 (1976) 20.

[30] K. Saitoh, N. Suzuki, "Multiwavelength Detection for Liquid-Chromatography with a Repeat-Scanning Ultraviolet-Visible Spectrophotometer" Anal. Chem. 51 (1979) 1683.

[31] S. George, "Three-Dimensional Data Presentation in HPLC" LC/GC 1 (1983) 158.

[32] Web of Science, The Thompson Corporation, 2006.

[33] M. Yamashita, J. B. Fenn, "Electrospray Ion Source. Another Variation on the Free-Jet Theme." J. Phys. Chem. 88 (1984) 4451.

[34] A. Felinger, M. Kare, "Wavelet Analysis of the Baseline Noise in HPLC" Chemom. Intell. Lab. Syst. 72 (2004) 232.

[35] P. J. Taylor, "Matrix Effects: the Achilles Heel of Quantitative High-Performance Liquid Chromatography-Electrospray-Tandem Mass Spectrometry" Clin. Biochem. 38 (2005) 328.

[36] C. G. Enke, "A Predictive Model for Matrix and Analyte Effects in Electrospray Ionization of Singly-Charged Ionic Analytes" Anal. Chem. 69 (1997) 4885.

[37] L. Tang, P. Kebarle, "Dependence of Ion Intensity in Electrospray Mass Spectrometry on the Concentration of the Analytes in the Electrosprayed Solution" Anal. Chem. 65 (1993) 3654.

[38] R. Dams, M. A. Huestis, W. Lambert, C. M. Murphy, "Matrix Effect in Bio-Analysis of Illicit Drugs with LC-MS/MS: Influence of Ionization Type, Sample Preparation, and Biofluid" J. Am. Soc. Mass Spectr. 14 (2003) 1290.

[39] R. Dams, T. Benijts, W. Günther, W. Lambert, A. De Leenheer, "Influence of the Eluent Composition on the Ionization Efficiency for Morphine of Pneumatically Assisted Electrospray, Atmospheric-Pressure Chemical Ionization and Sonic Spray" Rap. Comm. Mass Spectr. 16 (2002) 1072.

[40] K. A. Rubinson, J. F. Rubinson, Contemporary Instrumental Analysis. Prentice Hall, Upper Saddle River, NJ, 2000.

[41] H. A. L. Kiers, "Towards a Standardized Notation and Terminology in Multiway Analysis" J. Chemom. 14 (2000) 105.

[42] L. R. Tucker, "Some Mathematical Notes on 3-Mode Factor Analysis" Psychometrika 31 (1966) 279.

[43] J. D. Carroll, J. J. Chang, "Analysis of Individual Differences in Multidimensional Scaling Via an N-Way Generalization of Eckart-Young Decomposition" Psychometrika 35 (1970) 283-&.

[44] R. A. Harshman, "Foundations of the PARAFAC Procedure: Models and Conditions for an 'Explanatory' Multi-Modal Factor Analysis" UCLA Working Papers in Phonetics 16 (1970) 1.

[45] E. R. Malinowski, Factor Analysis in Chemistry. John Wiley & Sons,Inc., New York, 1991.

[46] M. Otto, Chemometrics. Wiley-VCH, New York, NY, 1999.

[47] R. A. Harshman, "An Index Formalism that Generalizes the Capabilities of Matrix Notation and Algebra to N-Way Arrays" J. Chemom. 15 (2001) 689.

[48] K. S. Booksh, Z. H. Lin, Z. Y. Wang, B. R. Kowalski, "Extension of Trilinear Decomposition Method with an Application to the Flow Probe Sensor" Anal. Chem. 66 (1994) 2561.

[49] A. Lorber, "Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations" Anal. Chem. 58 (1986) 1167.

[50] A. Lorber, K. Faber, B. R. Kowalski, "Net Analyte Signal Calculation in Multivariate Calibration" Anal. Chem. 69 (1997) 1620.

[51] A. C. Olivieri, "Computing Sensitivity and Selectivity in Parallel Factor Analysis and Related Multiway Techniques: The Need for Further Developments in Net Analyte Signal Theory" Anal. Chem. 77 (2005) 4936.

[52] N. J. Messick, J. H. Kalivas, P. M. Lang, "Selectivity and Related Measures for N^{th} Order Data" Anal. Chem. 68 (1996) 1572.

[53] G. Bergmann, B. Vonoepen, P. Zinn, "Improvement in the Definitions of Sensitivity and Selectivity" Anal. Chem. 59 (1987) 2522.

[54] M. McCue, E. R. Malinowski, "Target Factor-Analysis of the Ultraviolet-Spectra of Unresolved Liquid-Chromatographic Fractions" Appl. Spectr. 37 (1983) 463.

[55] J. H. Lin, A. Y. Lu, "Role of Pharmacokinetics and Metabolism in Drug Discovery and Development" Pharmacol. Rev. 49 (1997) 403.

[56] L. Shargel, A. Yu, Applied Biopharmaceutics and Pharmacokinetics. Appleton & Lange, 1999.

[57] R. Kostiainen, T. Kotiaho, T. Kuuranne, S. Auriola, "Liquid Chromatography/Atmospheric Pressure Ionization-Mass Spectrometry in Drug Metabolism Studies" J. Mass Spectr. 38 (2003) 357.

[58] L. C. Wienkers, T. G. Heath, "Predicting *in vivo* Drug Interactions From *in vitro* Drug Discovery Data" Nature Rev. Drug Disc. 4 (2005) 825.

[59] S. T. Walters, B. D. Foy, R. J. Castro, "The Agony of Ecstasy: Responding to Growing MDMA Use among College Students" J. Am. Coll. Health 51 (2002) 139.

[60] A. R. Green, "MDMA: Fact and Fallacy, and the Need to Increase Knowledge in Both the Scientific and Popular Press" Psychopharmacology 173 (2004) 231.

[61] R. A. Glennon, R. Young, M. Dukat, Y. Cheng, "Initial Characterization of PMMA as a Discriminative Stimulus" Pharmacol. Biochem. Behav. 57 (1997) 151.

[62] S. S. Johansen, A. C. Hansen, I. B. Müller, J. B. Lundemose, M. B. Franzmann, "Three Fatal Cases of PMA and PMMA Poisoning in Denmark" J. Anal Toxicol. 27 (2003) 253.

[63] L. A. Howard, E. M. Sellers, R. F. Tyndale, "The Role of Pharmacogenetically-Variable Cytochrome P450 Enzymes in Drug Abuse and Dependence" Pharmacogenomics 3 (2002) 185.

[64] D. W. Nebert, D. W. Russell, "Clinical Importance of the Cytochromes P450" The Lancet 360 (2002) 1155.

[65] P. J. Jannetto, S. H. Wong, S. B. Gock, E. Laleli-Sahin, B. C. Schur, J. M. Jentzen, "Pharmacogenomics as Molecular Autopsy for Postmortem Forensic Toxicology: Genotyping Cytochrome P450 2D6 for Oxycodone Cases" J. Anal Toxicol. 26 (2002) 438.

[66] Y. Ramamoorthy, R. F. Tyndale, E. M. Sellers, "Cytochrome P450 2D6.1 and Cytochrome P450 2D6.10 Differ in Catalytic Activity for Multiple Substrates" Pharmacogenetics 11 (2001) 477.

[67] U. M. Zanger, S. Raimundo, M. Eichelbaum, "Cytochrome P450 2D6: Overview and Update on Pharmacology, Genetics, Biochemistry" Naunyn-Schmiedeberg's Arch. Pharmacol. 369 (2004) 23.

[68] B. Rege, C. March, M. A. Sarkar, "Development of a Rapid and Sensitive High-Performance Liquid Chromatographic Method to Determine CYP2D6 Phenotype in Human Liver Microsomes" Biomedical Chromatography 16 (2002) 31.

[69] P. J. Wedlund, J. de Leon, "Cytochrome P450 2D6 and Antidepressant Toxicity and Response: What Is the Evidence?" Clin. Pharmacol. Ther. 75 (2004) 373.

[70] T. Rau, G. Wohileben, H. Wuttke, N. Thuerauf, J. Lunkenheimer, M. Lanczik, T. Eschenhagen, "CYP2D6 Genotype: Impact on Adverse Effects and Nonresponse During Treatment with Antidepressants - a Pilot Study" Clin. Pharmacol. Ther. 75 (2004) 386.

[71] R. F. Staack, D. S. Theobald, L. S. Paul, D. Springer, T. Kraemer, H. H. Maurer, "Identification of Human Cytochrome P450 2D6 as Major Enzyme Involved in the O-Demethylation of the Designer Drug p-Methoxymethamphetamine" Drug Metab. Dispos. 32 (2004) 379.

[72] B. J. Ring, J. A. Eckstein, J. S. Gillespie, S. N. Binkley, M. Vandenbranden, S. A. Wrighton, "Identification of the Human Cytochromes P450 Responsible for *in vitro* Formation of R- and S-Norfluoxetine" J. Pharmacol. Exp. Ther. 297 (2001) 1044.

[73] A. L. Lehninger, D. L. Nelson, M. M. Cox, Principles of Biochemistry. Worth Publishers, New York, NY, 1993.

[74] J. M. Hutzler, T. S. Tracy, "Atypical Kinetic Profiles in Drug Metabolism Reactions" Drug Metab. Dispos. 30 (2001) 355.

[75] Y. Lin, P. Lu, C. Tang, Q. Mei, G. Sandig, A. D. Rodrigues, T. H. Rushmore, M. Shou, "Substrate Inhibition Kinetics for Cytochrome P450-Catalyzed Reactions" Drug Metab. Dispos. 29 (2001) 368.

[76] S. Cha, "A Simple Method for Derivation of Rate Equations for Enzyme-Catalyzed Reactions under the Rapid Equilibrium Assumption or Combined Assumption of Equilibrium and Steady-State" J. Biol. Chem. 243 (1968) 820.

[77] S. Bhoopathy, B. Xin, S. E. Unger, H. T. Karnes, "A Novel Incubation Direct Injection LC/MS/MS Technique for *in vitro* Drug Metabolism Screening Studies Involving the CYP2D6 and the CYP3A4 Isozymes" J. Pharm. Biomed. Anal. 37 (2005) 739.

[78] S. Schnell, C. Mendoza, "The Condition for Pseudo-First-Order Kinetics in Enzymatic Reactions is Independent of the Initial Enzyme Concentration" Biophys. Chem. 107 (2004) 165.

[79] H. M. Jones, J. B. Houston, "Substrate Depletion Approach for Determining *in vitro* Metabolic Clearance: Time Dependencies in Hepatocyte and Microsomal Incubations" Drug Metab. Dispos. 32 (2004) 973.

[80] J. F. Corbett, "Pseudo First-Order Kinetics" J.Chem. Educ. 49 (1972) 663-&.

[81] H. P. Kasserra, K. J. Laidler, "Transient-Phase Studies of a Trypsin-Catalyzed Reaction" Can. J. Chem. 48 (1970) 1793-&.

[82] P. Atkins, J. de Paula, Physical Chemistry. W. H. Freeman and Company, New York, 2002, p. 862.

[83] E. Bezemer, S. Rutan, "Evaluation of Synthetic Liquid Chromatography-Diode Array Detection-Mass Spectrometry Data for the Determination of Enzyme Kinetics" Anal. Chim. Acta 490 (2003) 17.

[84] F. P. Guengerich, G. P. Miller, I. H. Hanna, H. Sato, M. V. Martin, "Oxidation of Methyoxyphenethylamines by Cytochrome P450 2D6" J. Biol. Chem. 277 (2002) 33711.

[85] F.P.Guengerich, personal communication, 2004.

[86] B. K. Choi, A. I. Gusev, D. M. Hercules, "Postcolumn Introduction of an Internal Standard for Quantitative LC-MS Analysis" Anal. Chem. 71 (1999) 4107.

[87] J. Caldwell, "Stereochemical Determinants of the Nature and Consequences of Drug-Metabolism" J. Chromatogr. A 694 (1995) 39.

[88] M. Lanz, R. Brenneisen, W. Thormann, "Enantioselective Determination of 3,4-Methylenedioxymethamphetamine and Two of Its Metabolites in Human Urine by Cyclodextrin-Modified Capillary Zone Electrophoresis" Electrophoresis 18 (1997) 1035.

[89] F. Sadeghipour, J. L. Veuthey, "Enantiomeric Separation of Four Methylenedioxylated Amphetamines on Beta-Cyclodextrin Chiral Stationary Phases" Chromatographia 47 (1998) 285.

[90] E. Szökö, T. Tábi, T. Borbás, B. Dalmadi, K. Tihanyi, K. Magyar, "Assessment of the N-Oxidation of Deprenyl, and Amphetamine Enantiomers Methamphetamine by Chiral Capillary Electrophoresis: An in Vitro Metabolism Study" Electrophoresis 25 (2004) 2866.

[91] G. T. Tucker, M. S. Lennard, S. W. Ellis, H. F. Woods, A. K. Cho, L. Y. Lin, A. Hiratsuka, D. A. Schmitz, T. Y. Y. Chu, "The Demethylenation of Methylenedioxymethamphetamine ("Ecstasy") by Debrisoquine Hydroxylase (CYP2D6)" Biochemical Pharmacology 47 (1994) 1151.

[92] J. M. Margolis, J. P. O'Donnell, D. C. Mankowski, S. Ekins, R. S. Obach, "(R)-, (S)-, and Racemic Fluoxetine N-Demethylation by Human Cytochrome P450 Enzymes" Drug Metab. Dispos. 28 (2000) 1187.

[93] J. Kim, K. W. Riggs, D. W. Rurak, "Stereoselective Pharmacokinetics of Fluoxetine and Norfluoxetine Enantiomers in Pregnant Sheep" Drug Metab. Dispos. 32 (2004) 212.

[94] K. R. Korzekwa, N. Krishnamachary, M. Shou, A. Ogai, R. A. Parise, A. E. Rettie, F. Gonzalez, T. S. Tracy, "Evaluation of Atypical Cytochrome P450 Kinetics with Two-Substrate Models: Evidence That Multiple Substrates Can Simultaneously Bind to Cytochrome P450 Active Sites" Biochemistry 37 (1998) 4137.

[95] G. J. Schaaf, E. M. de Groene, R. F. Maas, J. N. M. Commandeur, J. Fink-Gremmels, "Characterization of Biotransformation Enzyme Activities in Primary Rat Proximal Tubular Cells" Chemico-Biol. Int. 134 (2001) 167.

[96] L. E. Witherow, J. B. Houston, "Sigmoidal Kinetics of CYP3A Substrates: An Approach for Scaling Dextromethorphan Metabolism in Hepatic Microsomes and Isolated Hepatocytes to Predict *in vivo* Clearance in Rat" J. Pharmacol. Exp. Ther. 290 (1999) 58.

[97] S. V. Otton, D. F. Wu, R. Joffe, S. Cheung, E. M. Sellers, "Inhibition by Fluoxetine of Cytochrome-P450 2D6 Activity" Clin. Pharmacol. Ther. 53 (1993) 401.

[98] G. Cooper, L. Wilson, C. Reid, D. Baldwin, C. Hand, V. Spiehler, "Comparison of GC-MS and EIA Results for the Analysis of Methadone in Oral Fluid" J. For. Sci. 50 (2005) 928.

[99] S. George, "Position of Immunological Techniques in Screening in Clinical Toxicology" Clin. Chem. Lab. Med. 42 (2004) 1288.

[100] Y. Hino, I. Qjanpera, I. Rasanen, E. Vuori, "Performance of Immunoassays in Screening for Opiates, Cannabinoids and Amphetamines in Post-Mortem Blood" For. Sci. Int. 131 (2003) 148.

[101] D. K. Molina, V. J. Dimaio, "The Reliability of Immunoassay for Determining the Presence of Opiates in the Forensic Setting" Am. J. For. Med. Pathol. 26 (2005) 303.

[102] F. Pragst, M. Herzler, T. Erxleben, "Sytematic Toxicological Analysis by High-Performace Liquid Chromatography with Diode Array Detection (HPLC-DAD)" Clin. Chem. Lab. Med. 42 (2004) 1325.

[103] T. Stimpfl, W. Vycudilik, "Automatic Screening in Postmortem Toxicology" For. Sci. Int. 142 (2004) 115.

[104] D. Thieme, H. Sachs, "Improved Screening Capabilities in Forensic Toxicology by Application of Liquid Chromatography-Tandem Mass Spectrometry" Anal. Chim. Acta 492 (2003) 171.

[105] C. Kratzsch, O. Tenberken, F. T. Peters, A. A. Weber, T. Kraemer, H. H. Maurer, "Screening, Library-Assisted Identification and Validated Quantification of 23 Benzodiazepines, Flumazenil, Zaleplone, Zolpidem and Zopiclone in Plasma by Liquid Chromatography/Mass Spectrometry with Atmospheric Pressure Chemical Ionization" J. Mass Spectr. 39 (2004) 856.

[106] K. Kudo, H. Tsuchihashi, N. Ikeda, "Meeting Challenges in Forensic Toxicology in Japan by Liquid Chromatography/Mass Spectrometry" Anal. Chim. Acta 492 (2003) 83.

[107] M. Laloup, G. Tilman, V. Maes, G. De Boeck, P. Wallemacq, J. Ramaekers, N. Samyn, "Validation of an ELISA-Based Screening Assay for the Detection of Amphetamine, MDMA and MDA in Blood and Oral Fluid" For. Sci. Int. 153 (2005) 29.

[108] H. H. Maurer, "Position of Chromatographic Techniques in Screening for Detection of Drugs or Poisons in Clinical and Forensic Toxicology and/or Doping Control" Clin. Chem. Lab. Med. 42 (2004) 1310.

[109] M. Herzler, S. Herre, F. Pragst, "Selectivity of Substance Identification by HPLC-DAD in Toxicological Analysis Using a UV Spectra Library of 2682 Compounds" J. Anal Toxicol. 27 (2003) 233.

[110] A. P. Schellinger, P. W. Carr, "Isocratic and Gradient Elution Chromatography: A Comparison in Terms of Speed, Retention Reproducibility and Quantitation" J. Chromatogr. A 1109 (2006) 253.

[111] A. P. Schellinger, D. R. Stoll, P. W. Carr, "High Speed Gradient Elution Reversed-Phase Liquid Chromatography" J. Chromatogr. A 1064 (2005) 143.

[112] A. P. Schellinger, D. R. Stoll, P. W. Carr, "High Speed Gradient Elution RPLC of Bases in Buffered Eluents Part I: Retention Repeatability and Column Reequilibration" J. Chromatogr. A submitted (2006).

[113] R. M. Smith, Retention Indices in Reversed Phase HPLC. New York, NY, 1987.

[114] M. Bogusz, "Correction of Retention Index Values in High-Performance Liquid-Chromatography as a Tool for Comparison of Results Obtained with Different Octadecyl Silica Phases" J. Chromatogr. 387 (1987) 404.

[115] M. J. Bogusz, in R. M. Smith (Editor), Retention and Selectivity in Liquid Chromatography. Elsevier, New York, 1995, p. 171.

[116] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A. Elsevier, New York, 1997.

[117] M. J. Telepchak, T. F. August, G. Chaney, Forensic and Clinical Applications of Solid Phase Extraction. Humana Press, Totowa, NJ, 2004.

[118] R. D. Maier, M. Bogusz, "Identification Power of a Standardized HPLC-DAD System for Systematic Toxicological Analysis" J. Anal Toxicol. 19 (1995) 79.

[119] A. J. Barnes, I. Kim, R. Schepers, E. T. Moolchan, L. Wilson, G. Cooper, C. Reid, C. Hand, M. A. Huestis, "Sensitivity, Specificity, and Efficiency in Detecting Opiates in Oral Fluid with the Cozart (R) Opiate Microplate EIA and GC-MS Following Controlled Codeine Administration" J. Anal Toxicol. 27 (2003) 402.

[120] J. Pellett, P. Lukulay, Y. Mao, W. Bowen, R. Reed, M. Ma, R. C. Munger, J. W. Dolan, L. Wrisley, K. Medwid, N. P. Toltl, C. C. Chan, M. Skibic, K. Biswas, K. A. Wells, L. R. Snyder, ""Orthogonal" Separations for Reversed-Phase Liquid Chromatography" J. Chromatogr. A 1101 (2006) 122.

[121] E. Van Gyseghem, M. Jimidar, R. Sneyers, D. Redlich, E. Verhoeven, D. L. Massart, Y. Vander Heyden, "Orthogonality and Similarity Within Silica-Based Reversed-Phased Chromatographic Systems" J. Chromatogr. A 1074 (2005) 117.

[122] M. T. Cantwell, S. E. G. Porter, S. C. Rutan, "Evaluation of the Multivariate Selectivity of Liquid Chromatography Methods" Anal. Chim. Acta manuscript in preparation (2006).

[123] J. Dai, C. H. Shieh, Q. H. Sheng, H. Zhou, R. Zeng, "Proteomic Analysis with Integrated Multiple Dimensional Liquid Chromatography/Mass Spectrometry Based on Elution of Ion Exchange Column Using pH Steps" Anal. Chem. 77 (2005) 5793.

[124] S. Komatsu, X. Zang, N. Tanaka, "Comparison of Two Proteomics Techniques Used to Identify Proteins Regulated by Gibberellin in Rice" J. Proteome Res. 5 (2006) 270.

[125] H. Saito, Y. Oda, T. Sato, J. Kuromitsu, Y. Ishihama, "Multiplexed Two-Dimensional Liquid Chromatography for MALDI and Nanoelectrospray Ionization Mass Spectrometry in Proteomics" J. Proteome Res. 5 (2006) 1803.

[126] V. Tschappat, E. Varesio, L. Signor, G. Hopfgartner, "The Application of 2-D Dual Nanoscale Liquid Chromatography and Triple Quadrupole-Linear Ion Trap System for the Identification of Proteins" J. Sep. Sci. 28 (2005) 1704.

[127] C. G. Fraga, C. A. Corley, "The Chemometric Resolution and Quantification of Overlapped Peaks Form [sic] Comprehensive Two-Dimensional Liquid Chromatography" J. Chromatogr. A 1096 (2005) 40.

[128] G. M. Gross, B. J. Prazen, R. E. Synovec, "Parallel Column Liquid Chromatography with a Single Multi-Wavelength Absorbance Detector for Enhanced Selectivity Using Chemometric Analysis" Anal. Chim. Acta 490 (2003) 197.

[129] A. E. Sinha, J. L. Hope, B. J. Prazen, E. J. Nilsson, R. M. Jack, R. E. Synovec, "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 1058 (2004) 209.

[130] A. E. Sinha, C. G. Fraga, B. J. Prazen, R. E. Synovec, "Trilinear Chemometric Analysis of Two-Dimensional Comprehensive Gas Chromatography-Time-of-Flight Mass Spectrometry Data" J. Chromatogr. A 1027 (2004) 269.

[131] C. A. Bruckner, B. J. Prazen, R. E. Synovec, "Comprehensive Two Dimensional High-Speed Gas Chromatography with Chemometric Analysis" Anal. Chem. 70 (1998) 2796.

[132] C. G. Fraga, B. J. Prazen, R. E. Synovec, "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. 72 (2000) 4154.

[133] B. J. Prazen, C. A. Bruckner, R. E. Synovec, B. R. Kowalski, "Second-Order Chemometric Standardization for High-Speed Hyphenated Gas Chromatography: Analysis of GC/MS and Comprehensive GC x GC Data" J. Microcol. Sep. 11 (1999) 97.

[134] V. G. van Mispelaar, A. C. Tas, A. K. Smilde, P. J. Schoenmakers, A. C. van Asten, "Quantitative Analysis of Target Components by Comprehensive Two-Dimensional Gas Chromatography" J. Chromatogr. A 1019 (2003) 15.

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

[136] Z. H. Lin, K. S. Booksh, L. W. Burgess, B. R. Kowalski, "2^{nd}-Order Fiber Optic Heavy-Metal Sensor Employing 2^{nd}-Order Tensorial Calibration" Anal. Chem. 66 (1994) 2552.

[137] B. J. Prazen, R. E. Synovec, B. R. Kowalski, "Standardization of Second-Order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis" Anal. Chem. 70 (1998) 218.

[138] R. E. Synovec, B. J. Prazen, K. Johnson, C. G. Fraga, in P. R. Brown, E. Grushka (Editors), Advances in Chromatography. Marcel Dekker, Inc., New York, 2003, p. 1.

[139] D. Bylund, R. Danielsson, G. Malmquist, K. E. Markides, "Chromatographic Alignment by Warping and Dynamic Programming as a Pre-Processing Tool for PARAFAC Modeling of Liquid Chromatography-Mass Spectrometry Data" J. Chromatogr. A 961 (2002) 237.

[140] P. C. Eilers, "Parametric Time Warping" Anal. Chem. 76 (2004) 411.

[141] K. J. Johnson, B. W. Wright, K. H. Jarman, R. E. Synovec, "High-Speed Peak Matching Algorithm for Retention Time Alignment of Gas Chromatographic Data for Chemometric Analysis" J. Chromatogr. A 996 (2003) 141.

[142] K. M. Pierce, L. F. Wood, B. W. Wright, R. E. Synovec, "A Comprehensive Two-Dimensional Retention Time Alignment Algorithm To Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data" Anal. Chem. 77 (2005) 7735.

[143] V. Pravdova, B. Walczak, D. L. Massart, "A Comparison of Two Algorithms for Warping of Analytical Signals" Anal. Chim. Acta 456 (2002) 77.

[144] B. Walczak, W. Wu, "Fuzzy Warping of Chromatograms" Chemom. Intell. Lab. Syst. 77 (2005) 173.

[145] C. G. Fraga, B. J. Prazen, R. E. Synovec, "Objective Data Alignment and Chemometric Analysis of Comprehensive Two-Dimensional Separations with Run-to-Run Peak Shifting on Both Dimensions" Anal. Chem. 73 (2001) 5833.

[146] K. J. Johnson, B. J. Prazen, D. C. Young, R. E. Synovec, "Quantification of Naphthalenes in Jet Fuel with GC x GC/Tri-PLS and Windowed Rank Minimization Retention Time Alignment" J. Sep. Sci. 27 (2004) 410.

[147] K. M. Pierce, J. L. Hope, K. J. Johnson, B. W. Wright, R. E. Synovec, "Classification of Gasoline Data Obtained by Gas Chromatography Using a Piecewise Alignment Algorithm Combined with Feature Selection and Principal Component Analysis" J. Chromatogr. A 1096 (2005) 101.

[148] B. V. Hollingsworth, S. E. Reichenbach, Q. Tao, A. Visvanathan, "Comparative Visualization for Comprehensive Two-Dimensional Gas Chromatography" J. Chromatogr. A 1105 (2006) 51.

[149] S. E. Reichenbach, M. T. Ni, D. M. Zhang, E. B. Ledford, "Image Background Removal in Comprehensive Two-Dimensional Gas Chromatography" J. Chromatogr. A 985 (2003) 47.

[150] R. Bro, "Multiway Calibration. Multilinear PLS" J. Chemom. 10 (1996) 47.

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

[152] V. G. van Mispelaar, H. G. Janssen, A. C. Tas, P. J. Schoenmakers, "Novel System for Classifying Chromatographic Applications, Exemplified by

Comprehensive Two-Dimensional Gas Chromatography and Multivariate Analysis" J. Chromatogr. A 1071 (2005) 229.

[153] J. L. Hope, A. E. Sinha, B. J. Prazen, R. E. Synovec, "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 1086 (2005) 185.

[154] O. Fiehn, "Metabolomics – the Link Between Genotypes and Phenotypes" Plant Mol. Biol. 48 (2002) 155.

[155] R. Bro, H. A. L. Kiers, "A New Efficient Method for Determining the Number of Components in PARAFAC Models" J. Chemom. 17 (2003) 274.

[156] C. D. Brown, P. D. Wentzell, "A Modification to Window Target-Testing Factor Analysis Using a Gaussian Window" Chemom. Intell. Lab. Syst. 51 (2000) 3.

[157] D. Bylund, R. Danielsson, G. Malmquist, K. E. Markides, "Chromatographic Alignment by Warping and Dynamic Programming as a Pre-Processing Tool for PARAFAC Modeling of Liquid Chromatography-Mass Spectrometry Data" J. Chromatogr. A 961 (2002) 237.

[158] I. Garcia, L. Sarabia, M. Cruz Ortiz, J. Manuel Aldama, "Three-Way Models and Detection Capability of a Gas Chromatography-Mass Spectrometry Method for the Determination of Clenbuterol in Several Biological Matrices: the 2002/657/EC European Decision" Anal. Chim. Acta 515 (2004) 55.

[159] E. Comas, R. A. Gimeno, J. Ferre, R. M. Marce, F. Borrull, F. X. Rius, "Quantification from Highly Drifted and Overlapped Chromatographic Peaks Using Second-Order Calibration Methods" J. Chromatogr. A 1035 (2004) 195.

[160] Y. Y. Tam, J. P. Slovin, J. D. Cohen, "Selection and Characterization of Alpha-Methyltryptophan-Resistant Lines of Lemna-Gibba Showing A Rapid Rate of Indole-3-Acetic-Acid Turnover" Plant Physiology 107 (1995) 77.

[161] Y. Tozawa, H. Hasegawa, T. Terakawa, K. Wakasa, "Characterization of Rice Anthranilate Synthase Alpha-Subunit Genes OASA1 and OASA2. Tryptophan Accumulation in Transgenic Rice Expressing Mutant of OASA1" Plant Physiology 126 (2001) 1493.

[162] A. D. Wright, M. B. Sampson, M. G. Neuffer, L. Michalczuk, J. P. Slovin, J. D. Cohen, "Indole-3-Acetic-Acid Biosynthesis in the Mutant Maize Orange Pericarp, a Tryptophan Auxotroph" Science 254 (1991) 998.

[163] N. L. Glass, T. Kosuge, "Cloning of the Gene for Indole Acetic Acid-Lysine Synthetase From Pseudomonas Syringae Subsp. Savastanoi." J. Bacteriol. 166 (1986) 598.

[164] A. Walz, S. Park, J. P. Slovin, J. Ludwig-Muller, Y. S. Momonoki, J. D. Cohen, "A Gene Encoding a Protein Modified by the Phytohormone Indoleacetic Acid" Proc. Nat. Acad. Sci. 99 (2002) 1718.

[165] K. G. Gilbert, D. T. Cooke, "Dyes From Plants: Past Usage, Present Understanding and Potential" Plant Growth Reg. 34 (2001) 57.

[166] A. E. Sinha, J. L. Hope, B. J. Prazen, C. G. Fraga, E. J. Nilsson, R. E. Synovec, "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 1056 (2004) 145.

APPENDIX A

This appendix contains the Matlab functions written for the drug analysis portion of the work (Chapter 5).

TFA function:

```
function
[rd,tspec,cindex]=targetsegp_new(cube,library,theta,index_range);

% Sarah E.G. Porter and Sarah C. Rutan
% October 20, 2005 - last updated
%
% USAGE:       [rd,tspec,cindex,r,s]=targetsegp(cube,library);
%              -OR- [rd,tspec,cindex]=targetsegp(cube,library);
% INPUTS:
%     cube     This is the three dimensional data array.  Number of
%              columns is the number
%              of wavelengths in the spectra. Number of rows is the
%              number of time points
%              in the chromatogram. Number of slices is the number of
%              chromatograms to be
%              analyzed.
%
%     library This is a matrix containing the known spectra of the
%               components to be searched.
%              Each column should represent a known component
%
%     theta      threshold value for theta
%
%     index_range   this is the ranges of the data you would like to
%                    search
%                   (optional inupt)
%
%
% OUTPUTS:
%     rd       The relative deviation of the test spectra from the
%              library spectra. This
%              is the result of the target testing algorithm.
%
%     tspec    These are the predicted spectra based on the abstract
%              loadings of the unknown
%              chromatogram.
%
```

```
%
%     cindex  This is an index matrix indicating which components were
%             selected by the
%             algorithm.  A 1 indicates that the component is present,
%             while a 0 indicates
%             that a component is absent.
%
%    %%% ALS outputs, do not use if ALS is not needed:
%     r       This is the resolved retention profiles resulting from
%             ALS
%             analysis.
%
%     s       This is the resolved spectral profiles resulting from ALS
%             analysis.


[nrows,ncol,nslices]=size(cube);
ncomp=size(library,2);
rd=[];
tspec=[];
cindex=[];
r=[];
s=[];
if nargin==3; index_range=[1 nrows]; end;
if size(library,1)~=size(cube,2) | size(index_range,3)~=size(cube,3);
    error('Size of inputs is not correct!')
end

% clear t tspec rd r s;
for y=1:size(index_range,1)
    for m=1:nslices
        [u,s,v]=svd(cube(index_range(y,1,m):index_range(y,2,m),:,m),0);
%v is the abstract loadings (spectra) of the test chromatogram (L)

        eigs=(diag(s)).^2;          % eigenvalues of the data matrix

        %rank determination (see also rankdet.m)
        for i=1:size(eigs,1)-1;
            rsd=(sum(eigs((i+1):end))/(index_range(y,2,m)-
index_range(y,2,m)*(size(eigs,2)-i))).^0.5;
            if rsd<.6;
                ranknum(m)=i;
                break
            end

        end
        for n=1:ncomp;
            %T=pinv(L)L*  T is the transformation matrix
            %Lhat=LT       Lhat is the predicted spectra

tspec(:,n)=v(:,1:ranknum(m))*(pinv(v(:,1:ranknum(m)))*library(:,n));
```

```
rho=(library(:,n)'*tspec(:,n))/(norm(library(:,n))*norm(tspec(:,n)));
            rd(n,y,m)=acos(rho)*(180/pi);
        end
    end
end


%Component index
cindex=rd;
for x=1:numel(cindex)
    cindex(x)=cindex(x)<theta;
end
```

WTTFA function:

```
function [theta]=wttfa(data,window_size,ncomp,st,time)

%SEGP 4/1/2005
%usage: [theta]=wttfa(data,window_size,ncomp,st,time)
%Function performs "boxcar" window target testing factor analysis
% (WTTFA) on spectrochromatograms.
%
%Inputs:
%     data         can be 2-or 3-way data.  Spectra should be in rows
%                  and chromatograms should be in columns.
%
%     window_size  size of the window to be used. Usually should
%                  correspond
%                  with the baseline width of a peak in the chromatogram
%
%     ncomp        number of components to be considered (overestimate)
%
%     st           Spectral "library" contains known spectra to be
%                  tested
%
%     time         (optional input) time axis for plots
%
%
% Outputs:
%       theta       angle between the library spectra and the spectra
%                   in the chromatogram.
%
% Reference: Lohnes, M.T. et al, Anal. Chim. Acta, 389 (1999), 95-113

ntime=size(data,1);
nwindows=size(data,1)-window_size;
nlib=size(st,2);
nsamples=size(data,3);


theta=NaN(nlib,ntime,nsamples);
for y=1:nsamples;
    for i=1:nlib;
        for n=1:nwindows;
            [u,s,v]=svd(data(n:n+window_size-1,:,y),0);
            shat=v(:,1:ncomp)*v(:,1:ncomp)'*st(:,i);
            rho=st(:,i)'*(shat/(norm(st(:,i))*norm(shat)));
            theta(i,round(n+0.5*window_size),y)=acos(rho)*(180/pi);
        end
    end

        if nargin==5
            figure(y)
            subplot(211); plot(time,theta(:,:,y));
            subplot(212); plot(time,data(:,:,y));
```

```
        else
            figure(y)
            subplot(211); plot(theta(:,:,y)');
            subplot(212); plot(data(:,:,y)');
        end
end
```

Rank determination by several methods:

```
function [ranknum,output]=rankdet(cube,method);
%Rank determination by relative standard deviation method.
%
%Usage: [ranknum,output]=rankdet(cube,method);
%  2- or 3-way data can be input.  If 3-way data is used, output
% 'ranknum'
%  will be a vector containing the rank determination for each slice of
%  the cube.
%
%   Inputs:
%           method      use a 1 for RSD (residual standard deviation) -
%                       generally use 1 only if you know the relative
%                       error in the data.
%
%                       use a 2 for IE (imbedded error)
%                       use a 3 for IND (Malinowski indicator function)
%
%   Outputs:
%             The first output is always ranknum
%             The second output (output) depends on your choice of
%             method - it
%             will be a list of the values of the function for
%             increasing
%             numbers of components.
%
%Reference: Malinowski, E. R. "Factor Analysis in Chemistry" 2nd
%ed.1991, Wiley & Sons.
%
%
%Edited by SEGP 2/24/05 to remove break in 'for' loop
%Edited by SEGP 12/13/05 to change SVD to 'econ' mode
%Edited by SEGP 1/6/06 to allow choice of rank determination method

nslices=size(cube,3);
for m=1:nslices
    [u,s,v]=svd(cube(:,:,m),'econ');


    c=size(s,1);
    r=size(cube,1);
    eigs=(diag(s)).^2;

    %Relative Standard Deviation Method
    if method==1;
        for i=1:c-1;
            rsd(i)=(sum(eigs((i+1):c))/(r*(c-i))).^0.5;
            if rsd(i)<.15;
rsd(i)=0;
            end
```

```
        end
        output=rsd;
        [val,ranknum]=min(rsd)
        ranknum(m)=ranknum-1;
    end


    %Imbedded error function
    if method==2;

        for n=1:c-1;
            ie(n)=((n*sum(eigs((n+1):c)))/(r*c*(c-n)))^0.5;
        end
        [val,ranknum(m)]=min(ie)
        output=ie;
    end

    %Malinowski Indicator Function
    if method==3;
        for n=1:c-1;
            ind(n)=(((sum(eigs((n+1):c)))/(r*(c-n)))^0.5)/((c-n)^2);
        end
        output=ind;
        [val,ranknum(m)]=min(ind)
    end

end
```

Total drug analysis (Chapter 5):

This script gives the commands and programs used for the analysis of all of the chromatograms discussed in chapter 5.

```
%NEATEST_drug_notes.m
index=0;
[riobs,ricorr]=rt_index1(tr,rtsprim,rtssec,ribook);
% [riobslib]=ri_pred(ribook,rtsprim,rtssec);
[index_range]=find_time_ranges(time,tr,.05);
for win=[4 8 12];
    for theta=[10 7.5 5];
        index=index+1;

[rd,tspec,cindex]=targetsegp_new(data,carrlib,theta,index_range);
%         results=drug_results(cindex,list,riobs,ribook,win);

[results,chrom(index,:),peak(index,:)]=drug_results_fixed_window(cindex
,ricorr,ribook,win,true);
        results_riobs{index}=results;
    end
end
```

APPENDIX B

This appendix contains the Matlab functions written for the 2D-LC-DAD portion of the work (Chapter 6).

These commands show the procedures applied for the quantitative PARAFAC analysis:

```
% 2DLC_Notes_AC2006.m
%% Load the files, interpolate, and reshape into a three-way array

load f0550
[indoles]=align2(f0550,.35,29.75,.0006666666667);
clear f0550

load f0551
[wt8]=align2(f0551,.35,29.75,.0006666666667);
clear f0551

load f0552
[wt6]=align2(f0552,.35,29.75,.0006666666667);
clear f0552

load f0553
[homo4]=align2(f0553,.35,29.75,.0006666666667);
clear f0553

load f0554
[homo10]=align2(f0554,.35,29.75,.0006666666667);
clear f0554

load f0556
[blank]=align2(f0556,.35,29.75,.0006666666667);
clear f0556

%These are the files that got a couple of extra wavelengths collected
homo4=homo4(:,1:75);
wt6=wt6(:,1:75);
wt8=wt8(:,1:75);

%% Reshape into a four-way array

X1=reshape(blank,1,525,84,75);
X2=reshape(homo10,1,525,84,75);
X3=reshape(homo4,1,525,84,75);
X4=reshape(indoles,1,525,84,75);
X5=reshape(wt6,1,525,84,75);
X6=reshape(wt8,1,525,84,75);
X=[X1;X2;X3;X4;X5;X6];
clear X1 X2 X3 X4 X5 X6 blank homo10 homo4 indoles wt6 wt8
```

```
% you should have only 1 variable in the workspace now - X

%% Make a rankmap for the summed data to determine sections using FSIW-
%EFA

time=[.35:.000666666667:29.75];  %time over both chrom dimensions (30
%min - 44100 points)
time1=[0:0.04:20.9999];          %time over "short" chrom dimension (21
%sec.)
time2=[.36:.35:29.75];           %time over "long" chrom dimension (30
%min.)
waves=[200.5:2:350];             %wavelength axis

X=permute(X,[2,3,4,1]);
Xsum=sum(X,4);
[svr,svspsum]=fsiwefa(Xsum,4,2,75,time1,time2);

%% Make the sections.

% a and b are the boundaries of the 2nd dimension chromatography (mode
% 1 - 525 points)
% and c and d are the boundaries of the 1st dimension chromatography
% (mode 2 - 84 points)
a=25
b=275
c=4
d=12
section2=X(a:2:b,c:d,:,:);
%%%%%%%%%%%%%%


%% Initial set-up
% From here - each section is already made and can be loaded to carry
% out the rest of this script.
load section2
a=25; b=275; c=4; d=12;  %This sets up the boundaries for the section.
s=2                              %What section?
nexp=6;                         % number of experiments in the section

%% Rank Determination

const=[2 2 2 2];                 %What constraints for PARAFAC?
Options(1)=.01;                  %What options for PARAFAC?
Options(6)=100;

%reshape and permute the data to evaluate rank in each dimension
eval(['chunk=[piece' num2str(s) '];']);
[w,x,y,z]=size(chunk);
X1=reshape(chunk,w,x*y*z);
X2=permute(chunk,[2,1,3,4]);
X2=reshape(X2,x,w*y*z);
X3=permute(chunk,[3,1,2,4]);
```

```
X3=reshape(X3,y,w*x*z);
X4=permute(chunk,[4,1,2,3]);
X4=reshape(X4,z,w*x*y);

% SVD on each dimension
for y=1:3
    eval(['[U,S' num2str(y) ',V]=svd(X' num2str(y) ',''econ'');'])
end

%Calculate the 90% variance rank
for y=1:3
    eval(['S=S' num2str(y) ';'])
    S=diag(S);
    expvar=(S./sum(S))*100;
    cumvar=cumsum(expvar);
    %
    ex=90 %how much of the variance do you want explained (in percent)?
    for n=1:length(cumvar);
        if cumvar(n)>=ex;
            cumvar(n)=0;
        end
    end
    [val,ncomp(y)]=max(cumvar)
end

clear S* X* w x y z U V cumvar ex expvar val
nfactors=max(ncomp);

[Factors,It,ssX] = parafac(chunk,nfactors,Options,const);

%Evaluate the results of the PARAFAC model
figure('Name', ['',num2str(size(Factors{1},2)),' component model,
Section ',num2str(s) ''])
subplot(224)
bar(Factors{4}); axis tight;
colormap lines;
subplot(222)
plot(time1(a:2:b),Factors{1},'LineWidth',2); axis tight; xlabel('Time
(sec)')
subplot(221)
plot(time2(c:d),Factors{2},'LineWidth',2); axis tight; xlabel('Time
(min)')
subplot(223)
plot(wave,Factors{3},'LineWidth',2); axis tight; xlabel('Wavelength
(nm)')

%% fALS

eval(['piece=piece' num2str(s) ';'])

% uses PARAFAC output to make an appropriate initial guess for fALS and
```

```
% reshape the data into an augmented three-way array.  fALS is done
% with non-negativity only initially.
[piecea,IG]=ig_from_parafac(piece,Factors,[1,2,3,4],1);
[ALS_Factors{1},ALS_Factors{2},ALS_Factors{3},ALS_Factors{4},ALS_Factor
s{5},ALS_Factors{6}]=als4d_segp(piecea,IG,100,.01,1,2,1,0,1);

%Evaluate the results of the non negative fALS results:
figure('Name', ['',num2str(size(ALS_Factors{1},2)),' component model,
Section ',num2str(s),''])  %change number of components
subplot(3,1,1)
plot(time1(a:2:b),ALS_Factors{1}(:,:,1),'LineWidth',2);
title('Optimized 2nd Dimension Chromatograms'); axis  tight
subplot(3,1,2)
plot(time2(c:d),ALS_Factors{2}(:,:,1),'LineWidth',2); title('Optimized
1st Dimension Chromatograms'); axis tight
subplot(3,1,3)
plot(ALS_Factors{3},'LineWidth',2); title ('Optimized augmented
dimension'); axis tight

%Decide which components should be unimodal, and set the unimod
% constraint appropriately, then recalculate an initial guess.
unimod=[1 0 1 1 1 0 1 1;1 0 1 1 1 0 1 1];
three_way_ig=ig_recalc(ALS_Factors{1},ALS_Factors{3});
[ALS_Factors{1},ALS_Factors{2},ALS_Factors{3},ALS_Factors{4},ALS_Factor
s{5},ALS_Factors{6}]=als4d_segp(piecea,three_way_ig,100,.01,1,2,1,unimo
d,1);

% Evaluate the results again.  If more factors are required, go back to
% Line 120, and change nfactors to redo PARAFAC model.
figure('Name', ['',num2str(size(ALS_Factors{1},2)),' component model,
Section ',num2str(s),''])  %change number of components
subplot(3,1,1)
plot(time1(a:2:b),ALS_Factors{1}(:,:,1),'LineWidth',2);
title('Optimized 2nd Dimension Chromatograms'); axis  tight
subplot(3,1,2)
plot(time2(c:d),ALS_Factors{2}(:,:,1),'LineWidth',2); title('Optimized
1st Dimension Chromatograms'); axis tight
subplot(3,1,3)
plot(ALS_Factors{3},'LineWidth',2); title ('Optimized augmented
dimension'); axis tight

%% Final PARAFAC Analysis

%When satisfied with the fALS results, reshape the results back into a
%4-way array that can be used to initiate the quadrilinear PARAFAC
%algorithm.
ncomp=size(ALS_Factors{1},2);
nwave=size(ALS_Factors{3},1)/nexp;

ALS_Factors{3}=reshape(ALS_Factors{3},nwave,nexp,ncomp);
ALS_Factors{3}=permute(ALS_Factors{3},[1,3,2]);
```

```
ALS_Factors{7}=squeeze(sum(ALS_Factors{3},1))';       %gives the
% relative concentration profiles
ALS_Factors{8}=mean(ALS_Factors{3},3);
OldLoad{1}=ALS_Factors{1}(:,:,1);
OldLoad{2}=ALS_Factors{2}(:,:,1);
OldLoad{3}=mean(ALS_Factors{3},3);
OldLoad{4}=squeeze(sum(ALS_Factors{3},1))';
[Factors,it,err]=parafac(piece,size(OldLoad{1},2),[],const,OldLoad);
figure('Name', ['',num2str(size(Factors{1},2)),' component model,
Section ',num2str(s),''])
subplot(224)
bar(Factors{4}); axis tight;
subplot(222)
plot(time1(a:2:b),Factors{1},'LineWidth',2); axis tight; xlabel('Time
(sec)')
subplot(223)
plot(time2(c:d),Factors{2},'LineWidth',2); axis tight; xlabel('Time
(min)')
subplot(221)
plot(wave,Factors{3},'LineWidth',2); axis tight; xlabel('Wavelength
(nm)')
```

This section shows the procedure followed for the qualitative WTTFA analysis:

```
%wttfa_notes_AC2006
[theta10]=wttfa(cube,10,6,pureindole);

load indolerts
load twodtimes
load corn_labels
colors='bgrcmyk';
for exp=1:6       %which slice of the cube
    for comp=1:6  %which component
        [rtindex,rtindex_time,thetabox]=getrts(cube,theta10,time,5,26);
        figure(exp); title(labels(exp));
        plot(indolerts(:,1),indolerts(:,2),'ko')
        hold on;

      contour(time2,time1,reshape(thetabox(comp,:,exp),525,84),1,
      colors(comp))
    end
end
```

This function gets the retention times from the output of WTTFA:

```
function
[rtindex,rtindex_time,thetabox]=getrts(theta,window_size,time);

% SEGP
% usage:
%[rtindex,rtindex_time]=getrts(theta,sizewindow,expt,comp,time);
%        or [rtindex]=getrts(theta,sizewindow,expt,comp);
% Inputs:
%         theta        output from wttfa
%         window_size  size of window used in wttfa
%         time         the time axis
%
% Outputs:
%         rtindex     contains the indeces of the peaks found
%
% Last updated: 6/2/2005

%     numcomp=size(theta,1);
%     ntimes=size(theta,2);
%     nexp=size(theta,3);


[nrows,ncol,nslices]=size(theta);
ncomp=size(library,2);

thetabox=theta;
for m=1:nslices;
```

```
    for n=1:ncomp;
        for y=1:nrows;
            if thetabox(n,y,m)<7
                thetabox(n,y,m)=1;
            else thetabox(n,y,m)=0;
            end
        end
    end
end


clear peaksize peakenv startpoint endpoint rtindex rtindex_time

rtindex=0;
rtindex_time=0;

for m=1:nslices; %number of experiments
    for n=1:ncomp;    %number of components in the library
        index=0;
        startpoint=0;
        endpoint=0;

        for y=1:nrows-1;   %number of time points

            if thetabox(n,y,m)==0 & thetabox(n,y+1,m)==1;
                index=index+1;
                startpoint(index)=y+1;
            end
        end

        index=0;
        for y=1:nrows-1;
            if thetabox(n,y,m)==1 & thetabox(n,y+1,m)==0;
                index=index+1;
                endpoint(index)=y;
            end
        end

        peakenv=[startpoint;endpoint];
        peaksize=peakenv(2,:)-peakenv(1,:);
%%%%%%%%%%%%%
        index=0;
        for x=1:size(peakenv,2)
            if peaksize(x)>window_size      %peak size should match
% window size
                index=index+1;
                rtindex(n,index)=round(median(peakenv(:,x)));
            end
        end
    end %end of ncomp loop
end
```

```
% if rtindex==0;
%     disp('No peaks found for this component!')
% %     continue
% end
rtindex_time=zeros(ncomp,size(rtindex,2));
for n=1:ncomp
    for x=1:size(rtindex,2)
        if rtindex(n,x)~=0;
            rtindex_time(n,x)=time(rtindex(n,x));
        end
    end
end
```

VITA

Sarah Porter (nee Sarah Elizabeth Graham) was born at the Fitzsimons Army Medical Center in Aurora, Colorado on November 7, 1976. She graduated from York High School in Yorktown, Virginia in 1995, and attended the University of Virginia where she obtained her B.S. degree in chemistry, with a minor in physics, in 1998. She moved to Richmond in the summer of 1998 and worked for Whitehall-Robins Healthcare (now Wyeth) in the Research and Development lab for two years while finishing her M.S. degree in criminal justice with specialization in forensic science at VCU. Upon completion of the master's program in December of 2000, she got a job with the Virginia Division of Forensic Science as a breath alcohol instructor, where she instructed police and other law enforcement officials in the operation of evidential breath testing equipment. She left the Division in January of 2002, and taught chemistry as an adjunct faculty member at John Tyler Community College. She enrolled in the chemistry graduate program at VCU in August of 2002 and joined Dr. Sarah Rutan's research group, where she has been ever since. She is the first author on three publications and has presented her research at numerous scientific meetings.

Publications

Porter, S. E. G., Stoll, D. R., Rutan, S. C., Carr, P. W., and Cohen, J. D. "Analysis of Four-way Two-dimensional Liquid Chromatography – Diode Array Data: Application to Metabolomics." Analytical Chemistry, 2006, 78, pp. 5559-5569.

Porter, S. E. G., Stoll, D. R., Paek, C., Rutan, S. C., and Carr, P. W. "Fast Gradient Elution Reversed-Phase Liquid Chromatography with Diode-Array Detection as a High Throughput Screening Method for Drugs of Abuse: II. Data Analysis." Journal of Chromatography A, 2006, 1137, pp. 163-172.

Porter, S. E. G., Keithley, R. B., and Rutan S. C. "Development of an *In-vitro* Incubation Procedure for Screening of CYP2D6 Intrinsic Clearance" Journal of Chromatography B, 2006, in press.

Presentations

Porter, S. E. G., Cantwell, M. T., and Rutan, S. C. "Evaluation of the Selectivity of Multi-way Analysis Methods." Oral presentation at the 10th International Conference on Chemometrics in Analytical Chemistry. September 13, 2006, Águas de Lindóia, Brazil.

Porter, S. E. G., Stoll, D. R., Carr, P. W., Cohen, J. R., and Rutan, S. C. "Data Handling in 2D-LC: Applications to Maize Metabolomics." Oral presentation at the Pittsburgh

Conference on Analytical Chemistry and Applied Spectroscopy. March 14, 2006, Orlando, FL.

Porter, S. E. G., Stoll, D. R., Carr, P. W., Cohen, J. R., and Rutan, S. C. "Data Handling in 2D-LC: Applications to Maize Metabolomics." Oral presentation at the ACS National Meeting. August 29, 2005, Washington, D.C.

Porter, S. E. G., Sanchez-Ponce, R., Thekkudan, D., Paul, S. N., and Rutan, S. C. "The Pharmacokinetics of Amphetamine Analogs: Analyzing Complex Mixtures Using LC/MS and Chemometric Methods." Poster presented at Daniel T. Watts Research Symposium, October 28, 2004, Richmond, VA.

Porter, S. E. G., Sanchez-Ponce, R., Thekkudan, D., Paul, S. N., and Rutan, S. C. "The Pharmacokinetics of Amphetamines: Analyzing Complex Mixtures Using LC/MS and Chemometric Methods." Poster presented at HPLC Symposium, June 14, 2004, Philadelphia, PA.