INFRARED MICROSCOPY AND LIQUID

CHROMATOGRAPHY APPLIED TO PROBLEMS IN

FORENSICS AND BIOANALYTICAL CHEMISTRY

By

KAUSHALYA SHARMA DAHAL

Bachelor of Science in Chemistry
Tribhuvan University
Biratnagar, Nepal
2005

Master of Science in Physical Chemistry
Tribhuvan University
Kirtipur, Nepal
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2019

INFRARED MICROSCOPY AND LIQUID

CHROMATOGRAPHY APPLIED TO PROBLEMS IN

FORENSICS AND BIOANALYTICAL CHEMISTRY

Dissertation Approved:

Dr. Barry K. Lavine
_____
Dissertation Adviser
Dr. Ziad El Rassi
_____

Dr. Nicolas Materer
_____

Dr. Richard Bunce
_____

Dr. John Mintmire
_____

ACKNOWLEDGEMENTS

First of all, I would like to extend my heartfelt gratitude to my professor and research advisor, Dr. Barry K Lavine, who made this academic success possible through his excellent academic advice and support throughout my research work and doctoral study at Oklahoma State University.

My sincere thanks goes to my graduate advisory committee, Dr. Ziad El Rassi, Dr. Nicholas Materer, Dr. Richard Bunce, and Dr. John Mintmire for their guidance, advice, and valuable comments.

I also would like to thank my lab mates Francis, Isio, George and Haoran for keeping the team spirit during my tenure in the lab. I would also like to acknowledge my former lab mate Dr. Nuwan Parera for his guidance and helpful discussions.

My special thanks goes to my daughter, Minisha Khatiwada, for being a source of my happiness, and my heartfelt thanks to my husband, Ram Prasad Khatiwada, for being by my side and supporting me in every step of my academic career.

Finally, my special gratitude to my parents, Tanka Prasad Dahal and Hari Maya Dahal, for their love and emotional support. I would also like to acknowledge all of my family and friends for their love, tremendous support and encouragement during my doctoral study.

Name: KAUSHALYA SHARMA DAHAL

Date of Degree: JULY, 2019

Title of Study: INFRARED MICROSCOPY AND LIQUID CHROMATOGRAPHY APPLIED TO PROBLEMS IN FORENSICS AND BIOANALYTICAL CHEMISTRY

Major Field: CHEMISTRY

Abstract:

The main objectives of the research described in this dissertation are: (1) development of a methodology to prepare thin sections of automotive paint samples that involve exposing the edge of the sample using a microtome for infrared image analysis, (2) implementation of infrared library searching to match attenuated total reflection spectra obtained from an IR microscope to a transmission infrared spectral library, and (3) development of a reversed phase liquid chromatographic method to simultaneously detect renal cell carcinoma cancer biomarkers in urine. In one study, infrared imaging microscopy was investigated as a potential method to improve both the speed and accuracy of forensic automotive paint analysis. This is achieved by scanning across the layers of a paint sample using an infrared imaging microscope followed by decatenation of the image using alternating least squares. This approach, not only eliminates the need to hand section and analyze each paint layer separately, but also ensures that the IR spectrum of each layer is "pure" and not a mixture as a result of sampling too to the boundary between two paint layers a scalpel is used to hand section a paint chip to separate the individual layers. In a second study, an isocratic reversed phase liquid chromatographic method was developed to simultaneously detect and quantify three potential renal cell carcinoma biomarkers in urine which are present at relatively high concentrations: quinolinic acid, gentisic acid and 4-hydroxybenzoic acid. Using a methanol or butanol in water mobile phase with 0.6% acetic acid and a Zorbax C_{18} column, baseline resolution for creatinine and the three biomarkers in synthetic urine was achieved when water rich mobile phases were used. Detection of creatinine and the three biomarkers in urine using the proposed method was limited by background from the urine matrix for the later eluting compounds and from the dead marker for the earlier eluting compounds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The research described in this dissertation is directed towards three specific goals: (1) development of a methodology to prepare thin sections of an automotive paint sample (that involve exposing the edge of the paint sample using a microtome) for IR image analysis, (2) application of a prototype pattern recognition assisted infrared library search system to match attenuated total reflection spectra from an IR microscope to a transmission infrared spectral library, and (3) development of a reversed phase liquid chromatographic method to simultaneously detect renal cell carcinoma cancer biomarkers in urine. The significance of this research lies in the development of new methods to address problems of widespread interest in the areas of forensic science and bio-analytical chemistry.

Infrared imaging microscopy was investigated as a potential method to improve both the speed and accuracy of forensic automotive paint analysis. One way to minimize the time necessary for data collection is to collect concatenated IR data from all paint layers in a single analysis. This is achieved by scanning across the cross-sectioned layers of the paint sample using an FTIR imaging microscope equipped with an imaging detector, which can perform a complete scan of all layers in less than one hour. Once the

data has been collected, it can then undergo decatenation using chemometrics to obtain a "pure" IR spectrum of each paint layer. This approach, not only eliminates the need to hand section and analyze each paint layer separately, but also ensures that the final spectrum for each layer is "pure" and not a mixture as a result of sampling too to the boundary between two paints using a scalpel when hand sectioning the paint chip to separate the individual layers. Minimizing the probability of collecting a mixed spectrum results in a time savings as well as objectively ensuring that only "pure" spectra from each layer have been collected and are used in subsequent searches. By integrating the proposed imaging experiment with a prototype pattern recognition IR library searching system previously developed by our research group, the forensic examination of automotive paint is facilitated in terms of both speed and accuracy.

An isocratic reversed phase liquid chromatographic (RPLC) method was developed to simultaneously detect and quantify three potential renal cell carcinoma (RCC) biomarkers in urine: quinolinic acid, gentisic acid and 4-hydroxybenzoic acid. These three compounds are present at relatively high concentrations in urine. Generated by the kidney, urine has considerable value as a diagnostic biological fluid because it contains large amounts of metabolites and the noninvasive nature which samples can be collected makes it a favored bio-sample. As cancer develops in a patient, there is an additional energy demand on the patient's body, and cancer cells activate specific metabolic pathways to compensate for this energy demand. Therefore the patient's metabolic profile as reflected in his/her urine will often contain reliable biomarkers for early cancer detection.

Using a methanol or butanol in water mobile phase with 0.6% acetic acid and a

Zorbax C_{18} column, baseline resolution for creatinine and the three RCC biomarkers in synthetic urine was achieved. The concentration of creatinine was closely monitored as it corresponded to renal dilution so its levels should be monitored in any urinary biomarker analysis. Better resolution was achieved for the separation of these four compounds when water rich mobile phases were used. Detection of creatinine and the three biomarkers in urine using the proposed RPLC method was limited by background from the urine matrix for the later eluting compounds and from the dead marker for the earlier eluting compounds. The results of this study are presented in Chapter 4 of the dissertation.

This dissertation is divided into five chapters. The first chapter is the introduction which provides an overview of the research problems pursued in this dissertation. Chapter 2 provides the necessary theory and background about the multivariate analysis methods used to analyze the infrared images of automotive paint chips described in Chapter 3 of this dissertation. Chapter 4 describes the RPLC method developed for the detection of potential RCC biomarkers in urine. A summary of the results obtained in this dissertation research are outlined in Chapter 5.

CHAPTER II

CHEMOMETRICS

2.1. INTRODUCTION

Chemometrics is a branch of analytical chemistry that uses mathematical, statistical, and other methods of formal logic to determine by indirect means the properties of substances that otherwise would be very difficult to measure [2-1]. The actual term, "chemometrics," was first coined in 1975 by Bruce Kowalski in a letter that appeared in the Journal of Chemical Information and Computer Science [2-2]. Motivation for the development of this subfield of analytical chemistry at the time was simple enough - an increase in the number and sophistication of chemical instruments triggered interest in the development of new data analysis techniques that can extract information from large arrays of chemical data that are being routinely generated. Furthermore, the placement of chemical instrumentation under computer control paved the way for the development of algorithms to resolve overlapping peaks and improve instrument calibration. Much of the growth in the field of chemometrics that has occurred continues to be driven by the press of too much data. Chemometrics continues and will continue to play an ever important role in the extraction of information from spectroscopic data.

In this chapter, techniques based on principal component analysis are discussed. A summary of the techniques used in the studies described in this dissertation are presented in the following sections. Special emphasis is placed on the application of these techniques to problems in spectral pattern recognition.

2.2. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is probably the oldest and best known of the techniques used in multivariate analysis [2-3]. The overall goal of PCA is to reduce the dimensionality of a data set, while simultaneously retaining the information that is present in the data. Dimensionality reduction or data compression is possible with PCA because chemical data sets are often redundant. This redundancy is due to correlations among the measurement variables, which diminish the information content of the data [2-4]. Consider a set of samples characterized by two measurements, X_1 and X_2. Figure 2.1 shows a plot of these data in a 2-dimensional measurement space, where the coordinate axes (or basis vectors) of this measurement space are the variables X_1 and X_2. There appears to be a relationship between these two measurement variables, which suggests that X_1 and X_2 are correlated. Fixing the value of X_1 limits the range of values for X_2. If X_1 and X_2 were uncorrelated, the enclosed rectangle in Figure 2.2 would be fully populated by the data points. As information can be defined as the scatter of points in a measurement space, correlations between measurement variables decrease the information content of this space. The data points, which are restricted to a smaller region of the space due to correlations among the variables, can even reside in a subspace for

highly correlated measurement variables. This is shown in Figure 2.2. X_3 is perfectly

correlated with X_1 and X_2 since X_1 plus X_2 equals X_3. Hence, the six sample points are

constrained to a plane even though each point is described by three measurements.
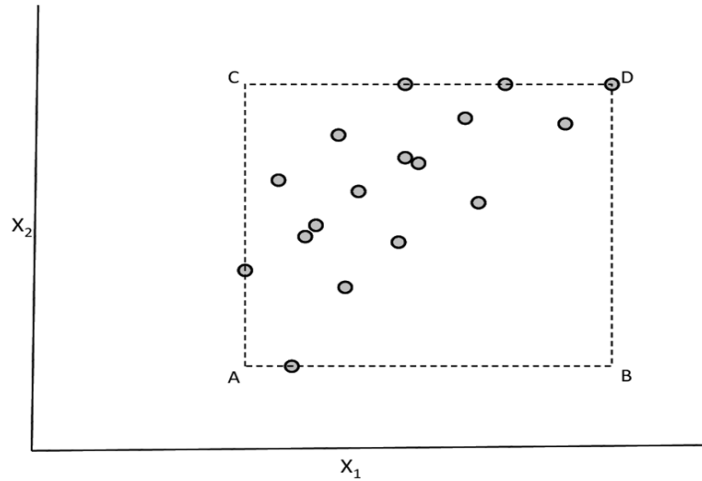
Figure 2.1. Seventeen hypothetical samples projected onto a 2-dimensional measurement space defined by the measurement variables X_1 and X_2. The vertices, A, B, C, and D, of the rectangle represent the smallest and largest values of X_1 and X_2. (Adapted from *NBS J. Res.*, 1985, 190(6), 465-476)

$$X = \begin{pmatrix} 0.5 & 0.5 & 1.0 \\ 1.9 & 0.7 & 2.6 \\ 2.0 & 1.0 & 4.0 \\ 0.3 & 1.8 & 2.1 \\ 1.9 & 1.7 & 3.6 \\ 1.2 & 0.2 & 1.4 \end{pmatrix}$$
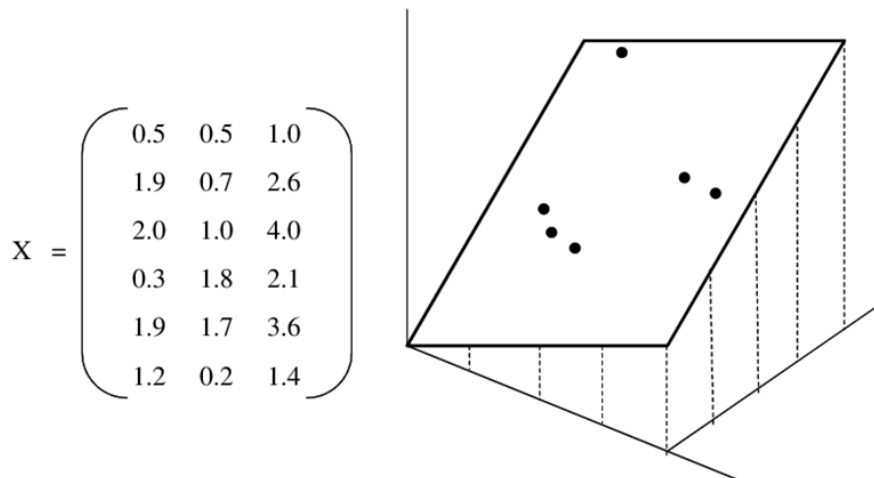
Figure 2.2. Six hypothetical samples projected onto a 3-dimensional measurement space. Because of strong correlations among the 3 measurement variables, the data points reside in a 2-dimensional subspace of the original measurement space. (Adapted from *Multivariate Pattern Recognition in Chemometrics*, Elsevier Science Publishers, Amsterdam, 1992)

Variables that are highly correlated are said to be collinear. High collinearity between variables is a strong indication that a new set of basis vectors can be found that is better at conveying the information present in data than axes defined by the original measurement variables. This new basis set is linked to variation in the data. The principal components of the data define the variance-based axes of this new coordinate system. Each principal component is perpendicular to the other. A measure of the amount of information conveyed by each principal component is the variance of the data explained by the principal component. The variance explained by each principal component is expressed in terms of its eigenvalue. For this reason, principal components are usually arranged in order of decreasing eigenvalue. The expectation is that the most informative principal component is the first and the least informative is the last. The maximum number of principal components that can be extracted from the data is the smaller of either the number of samples or number of measurements in the data set, as this number defines the largest number of independent variables in the data.

The principal components of the data are extracted from the data matrix using the singular value decomposition algorithm [2-5]. The decomposition of X (n x p) is shown in Equation 2.1, where T (n x f) is the score matrix which has the coordinates of the samples in the principal component space, P' (f x p) is the transpose of the loading matrix which defines the relationship between the original measurement variables (e.g., the wavenumbers) and each principal component, and E (n x p) is the residual matrix, where n is the number of samples or spectra in the data set and p is the number of points in each spectrum. Because of correlations among the measurement variables, f (which is the

number of principal components necessary to represent the spectral data) is smaller than p and often f is only two or three.

$$X = 1x_{mean} + TP' + E \qquad\qquad (2.1)$$

The data matrix is usually centered about the mean. This is accomplished by subtracting the mean of the variable from each entry in the corresponding column of the data matrix. Mean centering of a data matrix adjusts the means of each column in the matrix to zero. To ensure that each variable has equal weight in the analysis, the data are usually auto-scaled. In other words, after mean centering, each entry of a column is divided by the standard deviation of the column. Autoscaling adjusts the value of the measurements such that each variable has a mean of zero and a standard deviation of one. Autoscaling removes inadvertent weighting of the variables that otherwise would occur due to differences in magnitude among the measurement variables.

The distribution of the spectra in the p-dimensional space by plotting the columns of the score matrix against each other. Using principal component score plots, relationships among samples in the data can be uncovered, e.g., identifying outliers present in data and detecting similarities and differences among groups of samples, since each principal component captures a different source of information, i.e., variation in the data.

2.3. Genetic Algorithms for Pattern Recognition

In this section, the methodology used for the classification of infrared spectral data from an infrared imaging microscope is discussed. The premise underlying the

methodology used for classification in this dissertation is straight-forward. All classification methods perform well when the problem is simple. By identifying the appropriate features, a "hard" problem is transformed into a "simple" one. To ensure the selection of all relevant features, it is best to use a multivariate approach to feature selection. The approach used in the dissertation research is based on the idea of identifying a set of measurement variables that optimize the separation of the classes in a plot of the two or three largest principal components of the data. Because principal components maximize variance, the bulk of the information conveyed by these variables is about differences between the classes in the data. Using this approach to variable selection, an eigenvector projection of the data is formulated that discriminates the sample classes by maximizing the ratio of between- to within-group variance. This approach to feature selection avoids overly complicated solutions, which do not perform as well on the prediction set because of over-fitting. Although a principal component plot is not a sharp knife for discrimination, if we have a principal component plot that shows clustering, then our experience is that we will be able to predict robustly using this set of features. Furthermore, the principal component plot displays the variability between large numbers of samples and show the major clustering trends present in the data.

The approach to feature selection described in the previous paragraph is implemented using a genetic algorithm [2-6]. The plot of the two or three largest principal components of each feature subset, which is used by the fitness function of the genetic algorithm acts as an information filter, as sets of spectral features are selected based on their principal component plots, with a good principal component plot generated

9

by features whose variance or information is primarily about differences between the classes or groups. This restricts the search to sets of features with these attributes, thereby reducing the size of the search space.

To track and score the principal component plots generated by the genetic algorithm during each generation, class and sample weights, which are an integral part of the fitness function, are computed (see Equations 2.2 and 2.3), where CW(c) is the weight of class c (with c varying from 1 to the total number of classes in the data set), and SWc(s) is the weight of samples in class c. Class weights sum to 100, and the sample weights for samples from a particular class sum to a value equal to its class weight.

$$CW(c)=100\frac{CW(c)}{\sum_c CW\ (c)} \tag{2.2}$$

$$SW(s) = CW(c)\frac{SW(s)}{\sum_{s\in c} SW(s)} \tag{2.3}$$

Each PC plot generated for each feature subset is scored using the K-nearest neighbor classification algorithm [2-7]. For each sample in the training set, Euclidean distances are computed between it and the other samples which are represented as points in the principal component plot. These distances are arranged from the smallest to the largest, and a poll is taken of the sample's K_c nearest neighbors. For the most rigorous classification of the data, K_c is assigned a value corresponding to the number of samples in the class to which the sample is a member. The number of K_c nearest neighbors with the same class label as the data point in question, the so-called sample hit count, SHC(s), is computed $(0 < SHC(s) < Kc)$ for each sample. It is then a simple matter to score the

principal component plot (see Equation 2.4) representing each feature subset. First, the contribution to the overall fitness score by each sample in class 1 is computed. SHC for each sample in the class is divided by K_c, multiplied by SW(s), and this calculation which is repeated for the other samples in the class is summed to yield the contribution of this class to the overall fitness score. This same calculation is repeated for the samples in the other classes with the fitness score of each class summed to yield the overall fitness score, *F(d)*.

$$F(d) = \sum_c \sum_{s \in c} \frac{1}{K_c} \times SHC(s) \times SW(s) \qquad (2.4)$$

The fitness function of the pattern recognition GA is able to focus on those samples and/or classes that are difficult to classify by boosting their weights over successive generations. To boost the sample and class weights, it is necessary to compute the sample hit rate (SHR), which is the mean value of SHC/Kc over all the feature subsets comprising the population of potential solutions that are generated in a particular generation (see Equation 2.5), and the class-hit rate (CHR), which is the mean sample hit rate of all samples in a particular class (see Equation 2.6). ϕ in Equation 2.5 is the number of chromosomes in the population, and AVG in Equation 2.6 is the average or mean value.

$$SHR(s) = \frac{1}{\emptyset} \sum_{i=1}^{\emptyset} \frac{SHC_i(s)}{K_c} \qquad (2.5)$$

$$CHR_g(c) = AVG\left(SHR_g(s) : \forall_{s \in c}\right) \qquad (2.6)$$

11

During each generation, class and sample weights are adjusted using a perceptron algorithm [2-8] (see Equations 2.7 and 2.8) with the momentum, P, set by the user and with g + 1 being the current generation and g being the previous generation. Classes with a lower class hit rate are boosted more heavily than those classes that score well.

$$CW_{g+1}(s) = CW_g(s) + P(1 - CHR_g(s))) \qquad (2.7)$$

$$SW_{g+1}(s) = SW_g(s) + P(1 - CHR_g(s))) \qquad (2.8)$$

Boosting is crucial to the successful operation of the pattern recognition GA as it modifies the fitness landscape by adjusting the values of both the class and sample weights in each generation. This allows the pattern recognition GA to learn and assists it to obviate the problem of premature convergence to a local optimum. Thus, the fitness function of the pattern recognition GA is changing as the population evolves towards an optimal solution. Further details about the operation of the pattern recognition GA can be found elsewhere [2-9 to 2-13].

2.4. Multivariate Curve Resolution Using Alternating Least Squares

The goal of multivariate curve resolution (MCR) is to perform a mathematical decomposition of spectral data of a mixture into its contributions by each component of the mixture [2-14, 2-15]. Data (obtained from an image) can be arranged in the form of a data matrix X (n x p), where the n rows represent each spectrum recorded and the p columns represents the absorbance value of each wavelength comprising the spectra. MCR decomposition of X is performed using Equation 2-9, where each row of C is the

concentration profiles of a constituent, each column of S represents the spectrum of a

constituent and each row of E (i.e., the residual matrix) is the so-called error associated

with each spectrum that is not explained by the product of C and S^T.

$$X = CS + E \qquad (2.9)$$

The matrix, X, can be decomposed using noniterative or iterative MCR methods.

In the studies discussed in this dissertation, alternating least squares (ALS) [2-16, 2-17],

an iterative method, was used to decatenate the infrared image data. ALS solves

Equation 4.1 iteratively in two constrained least squares steps (see Equations 4.2 and

4.3). To perform ALS, an initial estimate of C (using the appropriate boundary condition)

is provided by the user. Using this estimate of C, an estimate of S is computed. Using

the estimate of S, C is computed. From the product of C and S, an estimate of the

principal component analysis (PCA) reproduced data matrix, X_{PCA}, is calculated. This

process is repeated until convergence as defined by a specific threshold value for a user

specified criterion indicative of improvement in the fit for S and C has been met [2-18].

The use of the PCA reconstructed data matrix, instead of the original data matrix,

stabilizes the calculations and reduces the noise in the concentration and spectral matrices

computed by ALS.

$$S^T = (C^TC)^{-1}\,(C^TX_{PCA}) \qquad (2.10)$$

$$C = (X_{PCA}S^T)\,(SS^T)^{-1} \qquad (2.11)$$

The actual values selected by the user for the initial estimates and the number of principal components for the data matrix are crucial to ensure the success of ALS. As for the boundary conditions, these constraints work towards defining the shape of each concentration and spectral profile in the bilinear model imposed on the data (see Equation 2.9). Examples of constraints frequently used in ALS include non-negativity, unimodality, and closure. The use of the PCA reconstructed data matrix, instead of the original data matrix, stabilized the calculations and reduced the noise in the concentration and spectral matrices.

REFERENCES

2-1.   B. K. Lavine and J. R. Workman, "Chemometrics: Past, Present, and Future," in B. K. Lavine (Eds.) Chemometrics and Chemoinformatics, ACS Symposium Series 894, Oxford University Press, 2005.

2-2.   B. R. Kowalski, "Letter to the Editor," J. Chem. Inf. Comput., Sci., 1975, 15, 201.

2-3.   I. T. Jolliffee, "Principal Component Analysis", Springer-Verlag, New York, 1986.

2-4.   J. Mandel, "The regression analysis of collinear data," J. Res. NBS, 1985, 90(6), 465-476.

2-5.   G. Golub, C. Van Loan, "Matrix Computations", Johns Hopkins University Press, Baltimore, 1971.

2-6.   D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley Publishing Co., Reading, MA 1989.

2-7.   J. T. Tou and R. C. Gonzalez, "Pattern Recognition Principles", Addison-Wesley Publishing, Reading, MA, 1974.

2-8.   P. D. Wasserman, "Neural Computing", Van Nostrand Reinhold, New York, 1989.

2-9.   B.K. Lavine, C.G. White, T. Ding, M.M. Gaye, D.E. Clemmer, Wavelet Based Classification of MALDI-IMS-MS Spectra of Serum N-Linked Glycans from Normal Controls and Patients Diagnosed with Barrett's Esophagus, High Grade Dysplasia, and Esophageal Adenocarcinoma, Chemom. Intel. Lab. Syst**.,** 2018, 176, 74-81.

2-10.   B. K. Lavine, K. Nuguru, and N. Mirjankar, "One Stop Shopping - Feature Selection, Classification, and Prediction in a Single Step," **J. Chemometr.,** 2011, 25, 116-129

2-11.   B. K. Lavine and C. E. Davidson, "Multivariate Approaches to Classification Using Genetic Algorithms," in: Brown S, Tauler R, Walczak R (eds.) Comprehensive Chemometrics, 2009, 3, pp. 619-646 Oxford: Elsevier.

2-12.   J. Karasinski, S. Andreescu, O. A. Sadik, B. Lavine, and M. N. Vora, "Multiarray Sensors with Pattern Recognition for the Detection, Classification, and Differentiation of Bacteria at Subspecies and Strain Levels," **Anal. Chem**., 2005, 77(24), 7941-7949.

2-13.   B. K. Lavine and A. J. Moores, "Genetic Algorithms for Pattern Recognition Analysis and Fusion of Sensor Data," in Pattern Recognition, Chemometrics, and

Imaging for Optical Environmental Monitoring, K. Siddiqui and D. Eastwood (Eds.), Proceedings of SPIES, 1999, pp. 103-112

2-14.   A. de Juan, and R. Tauler, "Soft Modeling of Analytical Data," in Encyclopedia of Analytical Chemistry: Instrumentation and Applications, Wiley, New York, 2000.

2-15.   A. Malik, A. de Juan, and R. Tauler, "Multivariate Curve Resolution: A Different Way to Examine Chemical Data," in B. K. Lavine, S. D. Brown and K. Booksh, 40 Years of Chemometrics – From Bruce Kowalski to the Future," ACS Symposium Series 1199, Oxford University Press, 2015.

2-16.   M. R. Garrido, F. X. Rius, and M. S. Larrechi, "Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) Applied to Spectroscopic Data from Monitoring Chemical Reaction Processes, Anal. Bioanal. Chem., 2008, 390, 2059-2066.

2-17.   R. Tauler, B. Kowalski, and S. Fleming, "Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial Process," Anal. Chem., 1993, 65(15), 2040-2047.

2-18.   A. de Juan and J. Jaumot, and R. Tauler, "Multivariate Curve Resolution (MCR). Solving the Mixture Analysis Problem," Anal. Methods, 2014, 6(14), 4964-4976.

Chapter III

ATTENUATED TOTAL REFLECTION INFRARED IMAGING
MICROSCOPY FOR THE FORENSIC EXAMINATION OF
AUTOMOTIVE PAINT

3.1 Introduction

An automotive paint chip recovered from a vehicle or the clothing of a pedestrian

in a hit-and-run accident is often the only form of trace evidence at a crime scene involving

a vehicle related fatality. Furthermore, this paint chip cannot be compared to the vehicle

of a potential suspect as there were no witnesses to the crime. For this reason, each layer

of the intact paint chip is examined using both physical and chemical methods. Based on

previous studies performed by the Royal Canadian Mounted Police [3-1, 3-2], automotive

vehicles can be differentiated by make and model on the basis of color, layer sequence and

the chemical composition of each layer in the manufacturer's original paint system.   A

forensic paint data base known as the paint data query (PDQ) database was developed by

the Royal Canadian Mounted Police (RCMP) to make forensic comparisons of these paint

layers to vehicles sold in North America [3-3,3-4]. PDQ contains over 21,000 automotive

paint sample with 84,000 individual layers, with the information about the chemical composition of each layer determined from its IR spectrum collected using a high pressure diamond anvil cell [3-5].

Automotive paint is a complex multicomponent system applied to the frame of a vehicle to protect it from corrosion, photochemical degradation while providing the vehicle with the desired color and finish [3-6]. Automotive paint contains binders, pigments and additives. The binder is the polymer matrix that maintains the physical integrity of the paint and serves as a medium for the pigments and additives that are suspended in it. Acrylic melamine enamel is the most common binder used in the finish coat of North American vehicles [3-7]. A pigment, which is a powdered compound that is insoluble in the paint solvent, imparts color, luster, and opacity to the paint. Titanium dioxide (in the form of rutile) is the most common pigment used in automotive paint coatings. Other compounds such as kaolin, calcite (calcium carbonate), quartz (silicon dioxide) and barites (barium sulfate) are known as additives or extender pigments [3-7], have a refractive index similar to the binder, do not affect its optical properties but increase the bulk matter in the paint. Automotive assembly plants use a unique combination of binders and pigments for each layer of paint. It is this unique combination that allows forensic scientists to determine the make, line, model, and year of a vehicle from a paint chip recovered at the crime scene.

Clear Coat (OT2)   50 µm

Base Coat (OT1)   20 µm

Surfacer (OU1)   25µm

E-coat (OU2)   25µm
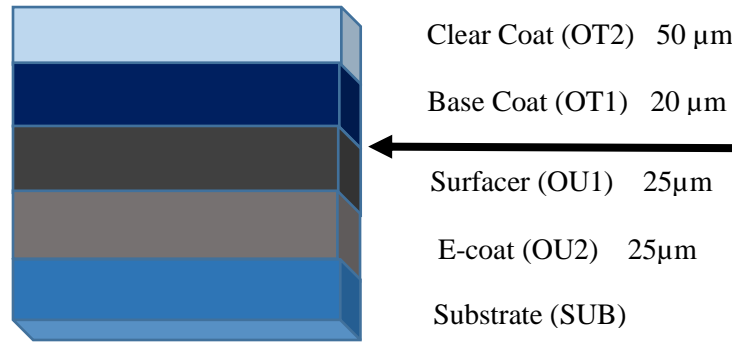
Substrate (SUB)

Figure 3.1 Modern automotive paint system

Modern automotive paint systems consist of four distinct layers: e-coat, primer surfacer, color coat and clear coat respectively from the substrate (see Figure 3.1). All four layers of paint contain binders and the color coat, surfacer-prime and e-coat contain both pigments and fillers [3-8]. The metal substrate to which the paint system is applied is pretreated with phosphate to protect it from corrosion and to enhance adhesion of the e-coat layer which is applied to the substrate by electrodeposition. The e-coat layer enhances the corrosion resistance of the substrate and provides elasticity and resistance against stone chipping as well as forming a bond with the surfacer primer layer. The function of the surfacer-primer layer is to provide support for the color coat layer, which imparts a specific color to the vehicle. The attributes of the surfacer-primer layer are defined by the choice of binders and resins used. The top most layer, the clear coat layer, is colorless, glossy and protects the color coat layer from UV radiation. The clear coat layer is the thickest layer (approximately 50µm in thickness compared to 25 µm or 20 µm for the other layers. Modern automotive clear coats are either acrylic melamine styrene polyurethane or acrylic melamine styrene [3-9]. Clear coats form a glossy and transparent coating which is in direct contact with the environment. For this reason, the clear coat layer must be resistant to abrasion, stone chipping, and ultraviolet light.

For the forensic analysis of automotive paints, many techniques have been developed to characterize automotive paint samples. Pyrolysis-gas chromatography-mass spectrometry [3-10 to 3-12] is the technique used to characterize the monomers comprising the binders in each layer after hand sectioning the paint chip. The binder decomposes into volatile fragments after applying thermal energy, and these fragments are separated by gas chromatography (GC) and then characterized by mass spectrometry (MS). The disadvantages of this technique is the lengthy time for analysis and the destruction of the sample. Furthermore, the mass of the sample must be at least 10 micrograms. Many paint chips recovered from the crime scene are often less than 10 micrograms. Other techniques for automotive paint analysis such as scanning electron microscopy coupled with energy dispersive spectrometry (SEM/EDS) [3-10] and micro- x-ray fluorescence (µXRF) [3-11, 3-12] are limited to elemental analysis. These techniques are complementary.

Raman Spectroscopy [3-13, 3-14] is another technique which has been used to characterize an automotive paint sample as it provides information about organic binders, inorganic pigments, and extenders present in each layer of the paint sample. However, fluorescence is often a problem, and there are no Raman automotive paint libraries for vehicle identification. Infrared (IR) spectroscopy [3-15 to 3-17] is at present the established method for the forensic analysis of automotive paint in North America. IR spectroscopy can characterize binders, extenders, organic and inorganic pigments. However, sample preparation can be lengthy as each layer must be hand sectioned using a sharp scalpel. Furthermore, sampling each layer with a sharp knife or scalpel too close to the boundary between layers can result in an IR spectrum that is a mixture of two paint layers. As a result, the accuracy of an infrared library search can be affected. Not having

a "pure" IR spectrum of each layer prevents the forensic paint examiner from developing an accurate hit-list of potential vehicles after performing a search of the unknown paint sample against IR spectra in the PDQ database.

The study described in this chapter attempts to address many of the problems previously encountered by workers in forensic paint analysis. It is proposed to decrease the time necessary for data collection (compared to the current method of hand sectioning and analyzing each layer separately by FTIR or some other method) by collecting IR data from all layers of paint in a single analysis. This is accomplished by scanning across the edge of a paint sample (whose layers have been exposed by a microtome) using an FTIR microscope equipped with an imaging detector. A complete scan of a paint sample can be performed in one hour. After the data has been collected, it can undergo decatenation using multivariate curve resolution methods to obtain a "pure" IR spectrum of each layer. This approach, not only eliminates the need to analyze each layer separately, but also ensures that the final spectrum of each layer is "pure" and not a mixture. Minimizing the probability of collecting a mixed IR spectrum will result in a considerable time savings as well as ensuring that only "pure" IR spectra from each layer are collected and used in subsequent searches of the PDQ database which will reduce the number of hits and increase the accuracy of a library search.

3.2. Attenuated Total Reflection (ATR) Infrared Imaging Microscope

An attenuated total reflection FTIR imaging microscope consists of an FTIR spectrometer coupled to an optical microscope (see Figure 3.2).

FTIR - ATR
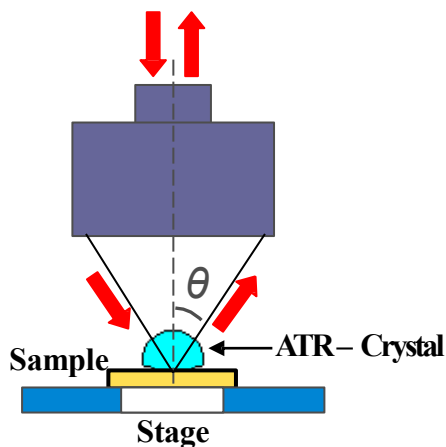
Figure 3.2 Attenuated total reflection infrared microscope

The microscope utilizes reflecting optics and spherical surfaces adapted to infrared radiation in order to minimize optical aberrations [3-18]. The spatial resolution that can be achieved by an ATR imaging microscope is given by Equation 3.1 where d is the spatial resolution, λ is the wavelength, and NA is the numerical aperture of the microscope objective, which is defined in Equation 3.2, where θ is the half-angle of the inverted cone of illumination entering the objective, and n is the refractive index of the imaging medium. (The half-angle of the cone is used in this equation since only half of the inverted cone of illumination is passed to the sample.) In a transmission microscope, the numerical aperture is equal to 0.6 (as n = 1 and the half-angle of the cone is 36.87^0) and the spatial resolution can be approximated by the wavelength of the radiation. For an ATR imaging microscope using a germanium crystal, (n = 4 and the numerical aperture equals 2.4), the spatial resolution is one-fourth of the wavelength of the incident radiation.

$$d = \frac{1.22\lambda}{2(NA)} \qquad (3.1)$$

$$NA = n_1 \sin\theta \qquad (3.2)$$

Due to the magnification provided by the Germanium ATR crystal, the size of the sample image interrogated by the IR microscope is only one-fourth the size of the corresponding sample image for the transmission microscope. For example, a 25 μm x 25 μm map image which is typically used in transmission infrared microscopy as it is the minimum size needed to obtain enough infrared energy [3-19] corresponds to 6.25 μm x 6.25 μm field of view in ATR infrared microscopy. This is a significant advantage over transmission when analyzing a laminated structure such as an automotive paint chip whose layer thicknesses are typically 20 μm.

3.3. Sample Preparation and Infrared Image Analysis

Thirty-eight automotive paint samples (see Table 3.1) from six manufacturers (Chrysler, Ford, General Motors, Honda, Nissan, and Toyota), which represent 80% of the vehicles driven in North America, were obtained from the RCMP. The 38 automotive paint samples were collected from vehicles spanning a six-year range (2000-2006). The automotive paint samples were cleaned with methanol in order to remove dirt or particulate matter. Each paint sample was peeled away from the metal substrate by bending it using pliers. To cross section a paint sample, the chip was placed between two small polyethylene slabs (see Figure. 3.3) and is then positioned horizontally in the sample holder

of the microtome (Reichert-Jung 2050). A stainless steel blade at 90 degrees to the exposed

edge of the paint sample was used to section the sample. After each paint sample was

cross-sectioned, the blade is either repositioned or replaced with a new blade to ensure that

a thin cross section (approximately 5 to 7 µm thick) cut by the microtome contained all

four layers. The distance between the sample and blade which is crucial to obtain the

desired thickness of the cut was determined empirically.

Table 3.1. Automotive Paint Samples Analyzed by Infrared Microscopy

PDQ Number	Manufacturer Line/Model	PDQ Number	Manufacturer Line/Model
UAZP00401	Chrysler / Durango	UAZP00331	GM / Chevrolet Suburban
UAZP00412	Chrysler / RAM	UAZP00436	GM / Chevrolet Tahoe
UAZP00421	Chrysler / JBT	UAZP00499	GM / Pontiac Bonneville
UAZP00451	Chrysler /CND	UAZP00565	GM / Pontiac Lucerne
UAZP00569	Chrysler / RAM	UAZP00271	GM / Chevrolet Tahoe
UAZP00600	Chrysler / Neon	UAZP00507	GM / GMC Trailblazer
UAZP00342	Ford / Focus	UAZP00336	GM /Chevrolet Suburban
UAZP00404	Ford / EPR	UAZP00337	GM /Chevrolet Tahoe
UAZP00467	Ford / Escapade	UAZP00501	GM / GMC Yucatan
UAZP00596	Ford / Mustang	UAZP00503	GM / GMC Silverado
UAZP00477	Ford / Mustang	UAZP00567	GM / Chevrolet Malibu
UAZP00729	Honda / CRV	UAZP00381	Toyota / Camry
UAZP00730	Honda / Civic	UAZP00313	Toyota / Camry Solara
UAZP00277	Honda / Odyssey	UAZP00733	Toyota / Camry
CONT00726	Honda / Pilot	UAZP00561	Toyota / Tacoma
CONT00736	Honda / Accord	UAZP00484	Toyota / Highlander
UAZP00385	Nissan /Sentra	UAZP00731	Nissan /Murano
UAZP00440	Nissan / Altima	UAZP00527	Nissan / Altima
UAZP00745	Nissan / Titan	UAZP00537	Nissan / Pathfinder
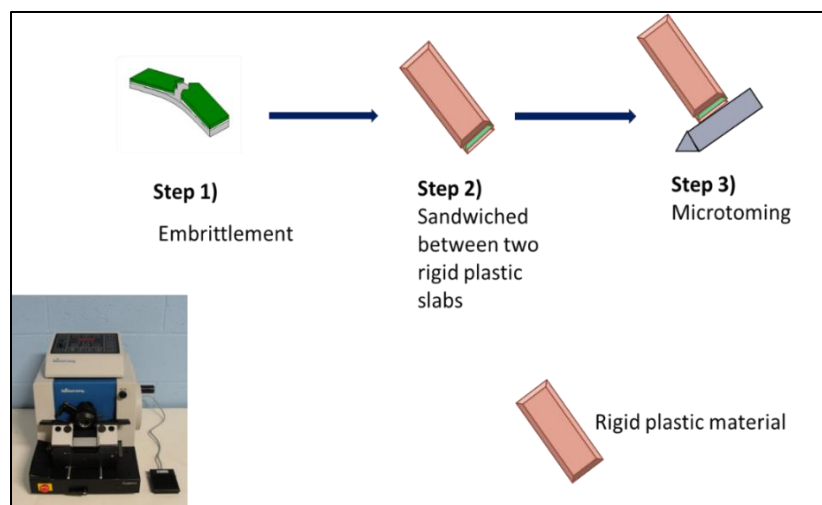
Figure 3.3. Steps involved in cross-sectioning an automotive paint chip.

The ATR sampling accessory is sensitive to changes that occur on the surface of the paint sample and the degree of contact between the sample surface and the ATR imaging crystal. Streak marks on the surface of the sample due to the microtome blade would deliver uneven contact between the sample and the ATR imaging crystal. This problem was addressed by adjusting the position of the sample in the microtome to ensure that the microtome only moved perpendicular to the laminated paint sample. In addition, low profile microtome blades had been previously used to section the sample with homebuilt spacers inserted in the microtome to obtain the optimum exposure of the microtome blade. These spacers caused difficulties in sectioning of the sample due to gaps between the blade and blade holder. This problem was obviated using high profile microtome blades which ensured the generation of high quality sections representative of the four paint layers and a smooth paint surface crucial for better surface contact with the internal reflection element of the ATR.

ATR infrared imaging of the cross-sectioned paint chip initially proved to be problematic due to difficulties encountered in the alignment of the paint sample on the germanium crystal. Using a curved tweezer to support the sample holder, a small needle (with a bent tip) to maneuver the thin paint sample and a video camera to guide its placement, alignment of the thin paint sample with the midpoint of the germanium ATR crystal was achieved, which ensured complete coverage of the crystal by the sample when a pressure of 10-15 psi was applied.   The noise level of IR spectra for thin samples was low due to the higher intensity of the IR beam. When the paint samples were properly aligned, differences between spectra of the same paint sample in PDQ versus that of the infrared imaging microscope did not markedly affect the quality of the spectral library match.

For ATR imaging, the tip accessory and the linear array detector was initially used but this approach was discontinued because the size of the aperture could not be adjusted to accommodate a thin sample. Furthermore, the sample was damaged (i.e., cracked) as a result of the tip moving in and out of the sample which in turn affected the alignment of the sample on the crystal.   By contrast, the aperture and step size of the single MCT detector, which were not fixed, could be changed to smaller values. Because the analysis of paint samples whose layer thickness was in the range of 20 to 50 microns (see Figure 3.1) requires a small step size and small aperture, the single MCT detector which performs a point by point mapping was selected for ATR infrared imaging analysis. The 25 micron x 25 micron aperture and 10 micron step size yielded the best results. To locate the sample for imaging, the linear array detector was used to select the region of the sample for point by point mapping as it provided a faster analysis of a large area image map.   The amount

of time required for an analysis was approximately 1 hour because two line maps must be generated: one for the sample and the other is for the background. A large number of scans (approximately 128 per spectrum) is required to generate each line map in order to achieve a comparable signal to noise level.

With a small aperture, the noise in the ATR spectra was high around the carbonyl band. This was attributed to changes that occurred in the amount of atmospheric water and carbon dioxide in the environment during collection of the sample and background spectra. This occurred because the system is open to the environment. As it is not possible to purge the system with nitrogen, we were able to mitigate the problem by placing a plastic cover over the ATR imaging accessory. Furthermore, the resolution of the imaging microscope was set to $8cm^{-1}$ to reduce the noise as $4 \ cm^{-1}$ yielded noisier spectra which affected the quality of the library matches.

3.4 ATR Correction Algorithm for the PDQ Library

To perform library searching on the infrared spectra recovered from the ATR imaging microscope, it will be necessary to transform the PDQ infrared transmission spectral library into an ATR infrared library using a correction algorithm previously developed by Lavine and co-workers [3-20, 3-21]. Due to differences in the optical configuration and the process used to measure IR absorption, ATR infrared spectra differ from transmission spectra in their band shapes, relative intensities and peak frequencies. This hinders library searching as a precise comparison of ATR spectra to transmission spectra in the PDQ spectral library is precluded by these differences.

The correction algorithm addressed the issues contributing to spectral distortion including the lower relative intensities at higher wave-numbers due to the lower depth of penetration of the evanescent wave, and the peak broadening and wave-number shifts due to anomalous dispersion. The surface reflection phenomenon of the incident beam at the boundary between the sample and the IRE is characterized by Fresnel's equations. To calculate the reflectance of the p- and s- polarized light, the first step is to compute the optical constants of the sample (i.e., the n- and k- indices) at each wavelength from the transmission spectrum. First the k– indices are computed using Equation 3.3, where $A(v)$ is the absorbance value of the transmission spectrum as a function of wavenumber (v), and d is the sample thickness [3-22]. The complex refractive index, which is a function of both the real and imaginary components, are related through the Kramers-Kronig relationship [3-23]. Once the imaginary part (k-index) of the refractive index has been calculated by Equation 3.3, the n- index is computed by Kramers-Kronig integration of the k- index, see Equation 3.4 where P is the principle value of the integral and $n(\infty)$ is the refractive index at high wavenumber where there is no absorbance, which serves as an anchor value. After the complex refractive index of the sample has been computed, the n- and k- indices are provided to Equations 3.5 and 3.6 to calculate Fresnel's reflection coefficients [3-24] for p- and s-polarized light at each wavenumber where n_0 and $\tilde{n}_1$ are the refractive indices of the IRE and the sample, and θ_0 and θ_1 are the angles formed between the incident beam and the normal in the IRE and in the sample. The reflectance at each wavelength in the ATR spectrum is computed using Equation 3.7.

$$k(v) = \frac{2.303A(v)\lambda}{4\pi d} \quad or \quad k(v) = \frac{2.303A(v)}{4\pi v d} \tag{3.3}$$

$$n(v_a) = n(\infty) + \frac{2}{\pi}P\int_0^\infty \frac{vk(v)}{v^2 - v_a^2}dv \tag{3.4}$$

$$r_s = \frac{n_0\cos(\theta_0) - \tilde{n}_1\cos(\theta_1)}{n_0\cos(\theta_0) + \tilde{n}_1\cos(\theta_1)} \tag{3.5}$$

$$r_p = \frac{\tilde{n}_1\cos(\theta_0) - n_0\cos(\theta_1)}{\tilde{n}_1\cos(\theta_0) + n_0\cos(\theta_1)} \tag{3.6}$$

$$R = \frac{(R_p + R_s)}{2} = \left(|r_p|^2 + |r_p|^2\right) \tag{3.7}$$

The ATR correction algorithm requires as input the angle of incidence of the infrared beam. Due to the large variation in the angle of the incident beam of the Cassegrain reflectors used in the IR microscope and the size of the aperture used in the analysis, the effective angle of incidence for a single spectral point is not a single value but is a range of values. Therefore, this range of values must be determined in order to perform the ATR correction for the PDQ library spectra. To determine this range of incident angles, the mid-point of the ATR imaging accessory (germanium hemisphere) was used. The mid-point of

the ATR internal reflection element (IRE) was located using the video image and the IR energy distribution of the internal reflection element of the ATR. The clear coat layer from a paint chip (SCC118) was transferred to the sample holder in such a manner as to ensure that the entire area of the crystal surface was covered by the clear coat paint sample. A single beam IR spectral image was collected in reflection mode. For the background, a single beam IR spectral image was collected without the sample on the ATR crystal and the FTIR image was calculated by dividing the single beam spectral image by the background spectral image. The IR spectrum at the mid-point of the ATR crystal was used to ascertain the range of the incident angles. The transmission IR spectrum of the corresponding clear coat paint sample (SCC118) from the PDQ library served as the IR spectrum to be matched by the correction algorithm. Using the ATR simulation algorithm, simulated ATR spectra were generated for incident angles from 20^0 to 50^0 at 0.1^0 increments. This range included the angles for the Cassegrain reflectors.

To compute the range of IRE angles for the ATR germanium hemisphere, simulated spectra for a given range were uploaded into OMNIC. The average spectrum for the range of IRE angles was calculated and compared to the experimentally obtained ATR spectrum (after spectral reconstruction using multivariate curve resolution) at the mid-point of the imaging ATR crystal. (The mid-point was selected because it was a convenient and well defined location on the crystal.) This process was repeated by changing the range of angles until the calculated average spectrum was similar to the mid-point spectrum of the imaging ATR crystal. The range of incident angles for the simulated spectra most similar to the mid-point of the crystal was 21.7^0-50^0. This experiment was repeated for three additional

paint samples to validate the result. For all samples, the range of incident angles was in agreement with 21.7^0-50^0.

Studies were also undertaken to better understand the relationship between the position of the paint sample on the crystal and the resulting ATR spectra. Both vertical and horizontal line maps were generated for a single layer (clear coat) paint sample (UAZP00331). The clear coat paint layer was positioned on the sample holder to ensure the Ge ATR crystal was completely covered with the clear coat. Two line maps, one vertical and one horizontal, were generated to investigate the resulting spectra as a function of the position of the paint sample on ATR crystal. Absorbance changes in the low wavenumber region were observed with spectra collected in the middle of the crystal versus spectra collected at the edge of the crystal suggesting that the range of the actual incident angle is a function of the physical location of the sample on the Ge ATR crystal. The range of the incident angle for different locations on the crystal is summarized in Table 3.2.

For Table 3.2, simulated spectra with all possible incident angles were generated. Spectra for the lower angles (20^0 - 32^0) were generated in 0.1 increments whereas spectra for the higher angles (32^0 - 50^0) were generated in 1^0 increments. A 1^0 increment was used to compute spectra for the higher incident angles as there was no significant difference between these simulated spectra when a 0.1 increment was used in the simulation. All possible incident angle ranges were computed using the ATR simulation program to convert transmission IR spectra to ATR spectra, with the average IR spectrum computed for each incident angle range. These average simulated ATR spectra were placed in a library. Pure clear coat IR spectra calculated from a spectral line map using ALS were searched against this library of average simulated IR spectra using OMNIC to determine

31

the optimum incident angle range, which varied as a function of the specific paint sample used. Absorbance changes in the low wavenumber region that were observed may also be attributed to differences in the contact angle between the crystal and the paint sample.

Table 3.2. Range of Incident Angles as a Function of Position on the Ge Crystal

Direction from the Midpoint		Left/Up	Right/Down
Horizontal	Distance 1000 µm	24.7^0-30^0	24.2^0-30^0
	Distance 1500 µm	24.5^0-30^0	24.3^0-30^0
Vertical	Distance 1000 µm	24.6^0-30^0	24.8^0-30^0
	Distance 1500 µm	24.8^0-30^0	25.9^0-30^0

The range of incident angles estimated for the paint sample in the mid-point of the crystal was 21^0 - 38^0. (Although the ATR simulation algorithm used 21.7^0-50^0 as the range for the incident angle, changing the range to 21^0 - 38^0 did not improve the quality of the simulations for library matching.) When the sample position was changed from the mid-point of the crystal, the range of the angles narrowed as shown in Table 3.2. The data, which is summarized in Table 3.2, suggests that the range of the incident angles is dependent upon the position of the paint sample on the germanium crystal. The range of incident angles for the paint sample is narrower as the paint sample is moved away from the mid-point of the Ge. However, the noise level is also greater as the distance from the mid-point of the crystal increases, which can be attributed to lower IR beam intensity.

3.5 Results and Discussion

3.5.1 Multivariate Curve Resolution

To demonstrate the advantages and drawbacks of the proposed method, three automotive paint samples (from different "makes", lines, and models) were selected for an

32

initial sample cohort. These three samples because of their small size (3 mm in length) are angled and are also representative of paint chips recovered from the crime scene of a vehicle-vehicle collision or a hit-and-run accident where injury or death to a pedestrian has occurred. Unlike a transmission IR imaging experiment, an angled paint chip will not create problems as positioning the sample in the center of the crystal ensures that a representative line map capturing all four paint layers is obtained. The best results are always obtained using the longest possible diagonal that fully bisects the automotive paint chip. This type of cut through the line map maximizes the number of spectra collected at the boundary between layers, thereby providing sufficient information about changes in the IR spectra of automotive paint as a function of its composition and position on the line map.

Four separate ALS models were computed for each sample to account for the ambiguities associated with an underdetermined system: a four-component model, a six-component model, a ten-component model, and a fifteen-component model. All thirty-five components were used to ascertain the pure IR spectra of each paint layer from the reconstructed IR spectra of the cross sectioned paint sample. Because the IR spectra of the clear coat, color coat, surfacer-primer, and e-coat layers are distinctive, the thirty-five IR spectra could readily be divided into four groups based on the identity of the layer. The separate ALS models used to reconstruct each paint sample improved the quality of the library matches obtained for each layer as compared to using a single ALS model to reconstruct the IR spectra. The region used for decatenation of the IR spectra comprising each line map by ALS was $4000cm^{-1}$ - $680cm^{-1}$.

Prior to the application of ALS, atmospheric carbon dioxide and water were suppressed in the spectra using air as the background. The presence of CO_2 in the sample spectra was also suppressed by directly interpolating between 2280 cm^{-1} and 2400 cm^{-1}, and the presence of water vapor in the spectrum was diminished by collecting the IR spectra at 8 cm^{-1} resolution which had the effect of decreasing the apparent absorption of IR radiation by the narrow water absorption lines.

Figures 3.4 to 3.6 show the ALS reconstructed IR spectra of the clear coat, surfacer-primer, and e-coat layers for UAZP00596 (Ford/Mustang), UAZP00600 (Chrysler/Neon), and UAZP00484 (Toyota/Highlander). Although the IR spectrum of the color coat layer could also be reconstructed by ALS, the large variation in its IR spectrum due to the presence of pigments in this layer, not to mention the relatively poor quality of the spectra as the IR signal is obscured by metal and pearlescent effect flakes that are present in this layer and preclude its use in a library search. For the spectral reconstructions by ALS, the initial estimates of the score and loading matrices are crucial for rotating these two matrices towards a correct solution. In this study, a varimax extended rotation previously developed by Lavine and coworkers [3-25 and 3-26] was applied to the IR spectra extracted from the line maps to compute the initial estimates of the concentration (score) and spectra (loading) matrices. First, the data were preprocessed (normalized to constant row sum, followed by range scaling and normalization to unity length) to identify IR spectra where the proportion of a particular layer is maximized relative to the other layers. Principal component analysis was then applied to the preprocessed data followed by a varimax rotation which utilized the extremum points to rotate the score and loading matrices towards a meaningful

solution. The IR spectrum extracted from a line map for each layer by ALS was a good match for the IR spectrum of the same paint sample in the PDQ library.
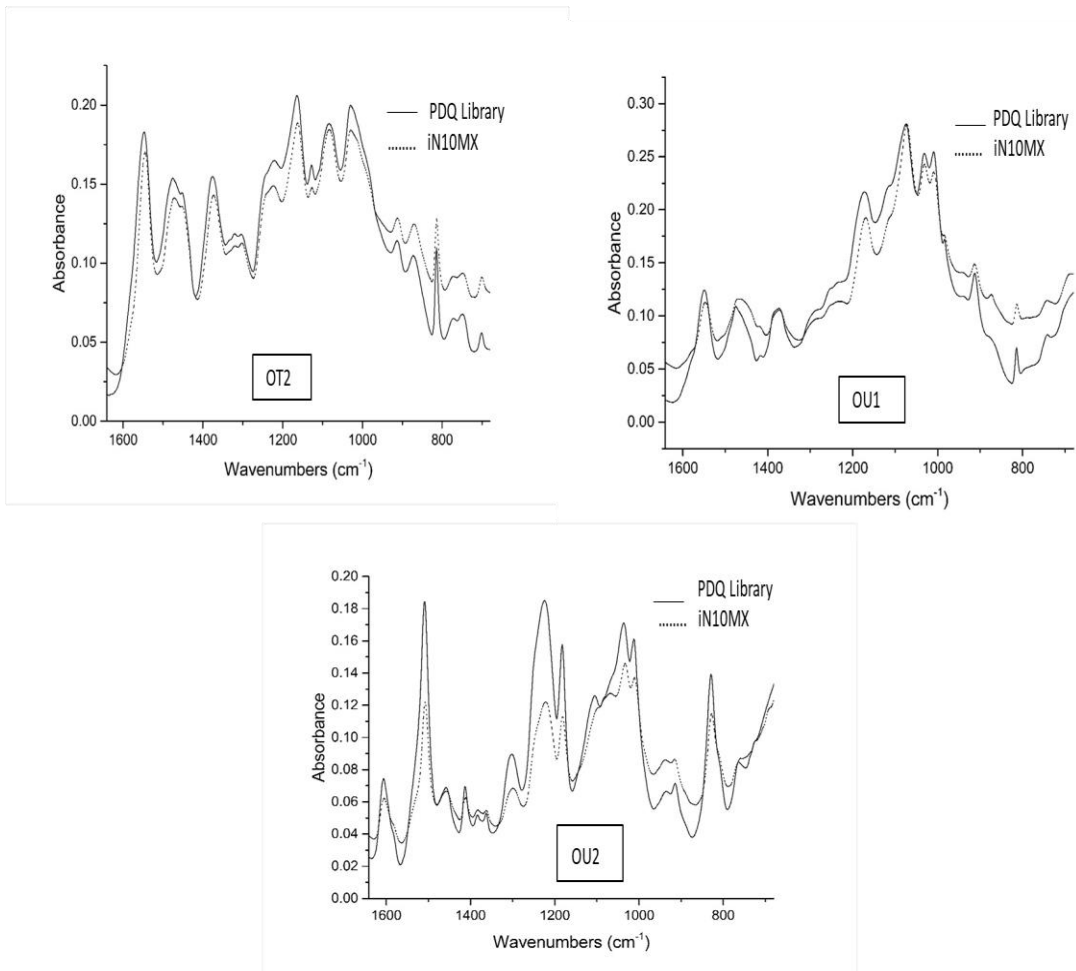
Figure 3.4. ALS reconstructed IR spectra of the clear coat (OT2), surfacer-primer (OU1), and e-coat (OU2) layers for UAZP00596 (Ford/Mustang)
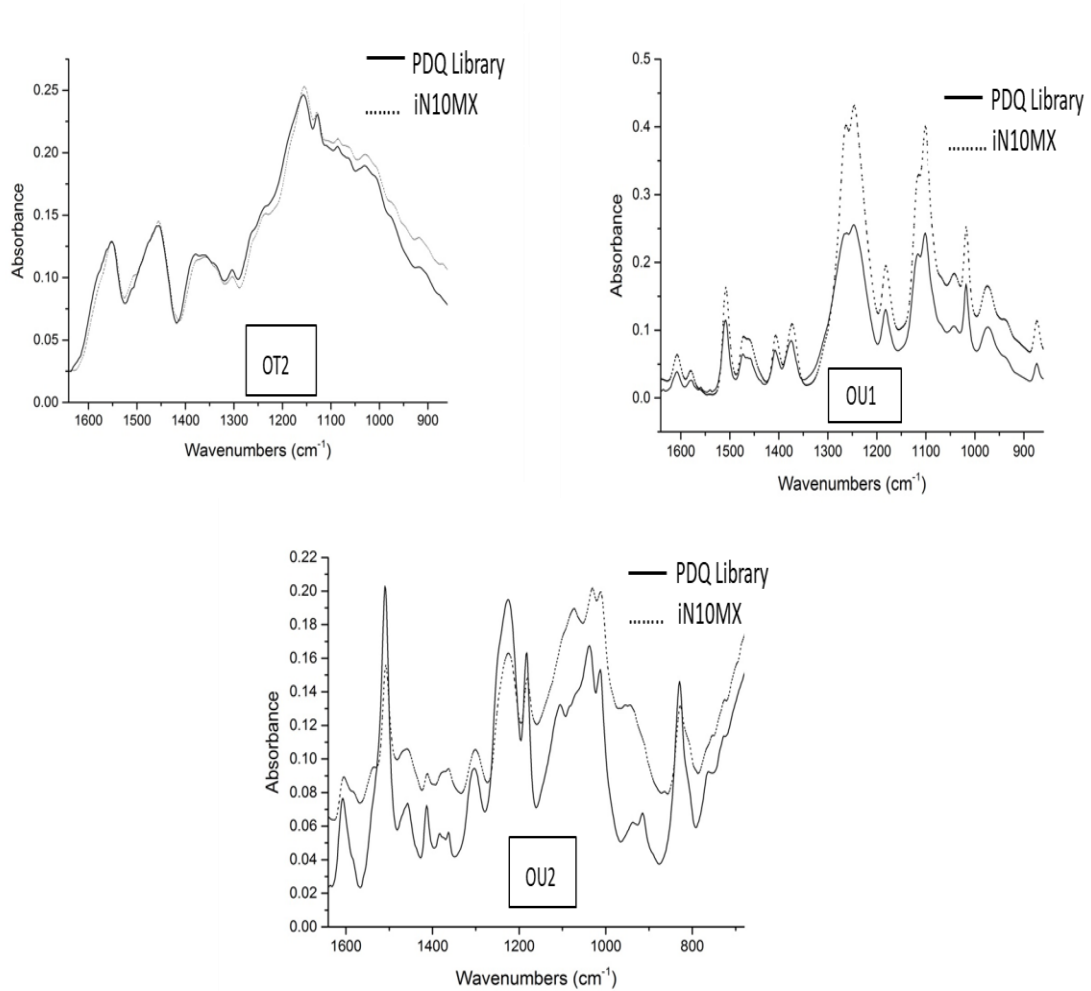
Figure 3.5. ALS reconstructed IR spectra of the clear coat (OT2), surfacer-primer (OU1), and e-coat (OU2) layers for UAZP00600 (Chrysler/Neon)
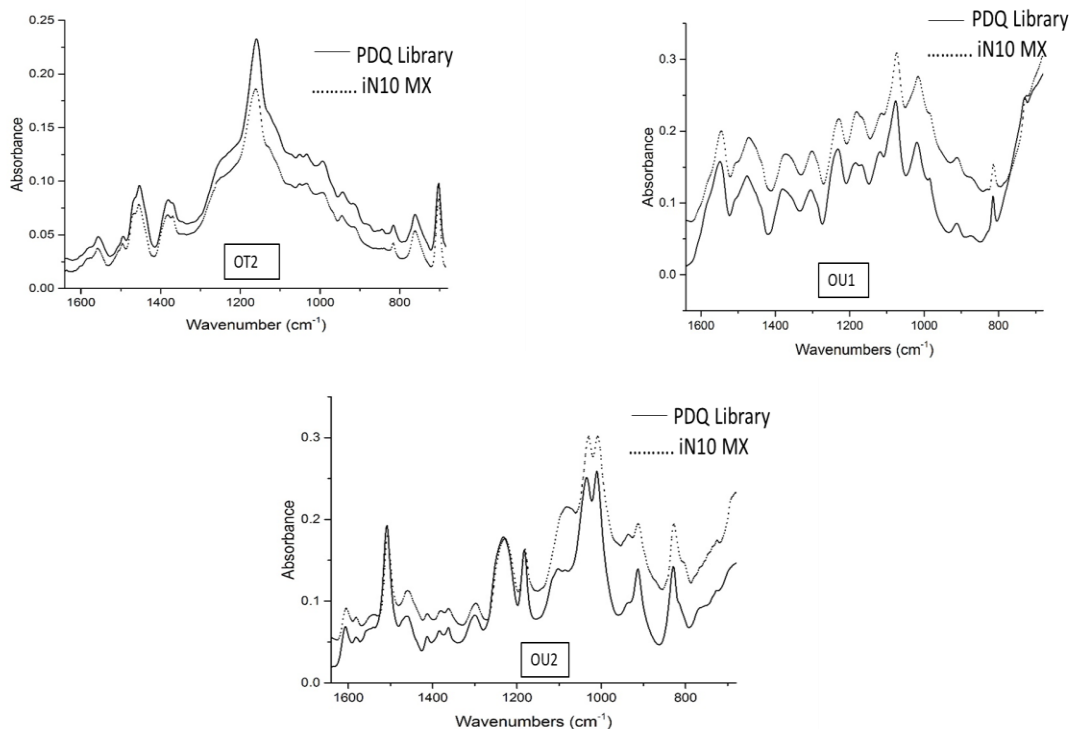
Figure 3.6. ALS reconstructed IR spectra of the clear coat (OT2), surfacer-primer (OU1), and e-coat (OU2) layers for UAZP00484 (Toyota/Highlander)

The next step was to perform a library search of these reconstructed IR spectra against the Ford, Chrysler or Toyota libraries (see Table 3.3) for the three samples comprising the cohort. For each automotive paint sample, the library that was searched corresponds to the same manufacturer (e.g., Chrysler) and production year (e.g., 2000-2006) as the automotive paint **s**ample (e.g., UAZP00600) from which the reconstructed IR spectra were obtained. The Ford, Chrysler, and Toyota libraries were transformed to ATR libraries using the correction algorithm discussed in the previous section of this chapter. All searches were restricted to the spectral region between 1641 cm^{-1} and 860 cm^{-1} which in our previous studies [3-5] was shown to contain information about the make, line and model of the vehicle. Outside of this region, the IR spectra only contained C-H stretching

bands and the carbonyl band that were found not to be discriminating. The library searches were performed against the Ford, Chrysler, and Toyota IR libraries for the clear coat, surfacer-primer, and e-coat layers using OMNIC (Themo-Nicolet) configured as correlation for search type and Happ-Genzel for apodization. The correct line and model of the vehicle was always in the top five entries of the search hit-list for each layer (see Table 3.4).

Table 3.3. Ford, Chrysler and Toyota Spectral Libraries

Manufacturer	Clear Coat	Surfacer-Primer Layer	E-Coat Layer
Ford	477	474	470
Chrysler	467	467	448
Toyota	269	308	298

Table 3.4 Search Results for UAZP00596, UAZP00600 and UAZP00484

PDQ Number	OT2		OU1		OU2	
	Position Hit	HQI	Position Hit	HQI	Position Hit	HQI
UAZP00596	4	95.54 %	1	92.89 %	4	95.78 %
UAZP00600	2	95.81 %	4	96.92 %	1	97.30 %
UAZP00484	4	94.21 %	5	91.49 %	4	90.67 %

For each search, only the top five entries of the hit list were reported as the identity of the unknown is expected to be captured in the top five hits for a successful search using the hit quality index (HQI) to rank the library spectra. A library search where either the actual paint sample or the correct line and model of the vehicle is not listed in the top five-

hits is judged to be an unsuccessful search. For every spectral match, the HQI value was greater than 90%. HQI values greater than 90% are indicative of high quality spectral matches [3-27].

ALS reconstructions of each paint layer for a larger sample cohort were also undertaken. The reconstructed IR spectra from thirty-eight automotive paint samples (see Table 3.1) were subject to an OMNIC library search performed in a manner similar to the search undertaken for the small data cohort to better gauge and understand the quality of the MCR spectral reconstructions obtained using ALS. Table 3.5 summarizes the OMNIC library search results. For the clear coat and surfacer primer layers, 36 of 38 paint samples were correctly matched (i.e., the correct line and model were present in the top five entries of the hit-list), whereas only 30 of the 38 samples were correctly matched for the e-coat layer. A visual examination of the two clear coat IR spectra that were not correctly matched revealed the presence of a peak in the PDQ library spectrum of one paint sample (i.e., UAZP00567, see Figure 3.7) that was not present in its reconstructed IR spectrum. This can probably be attributed to variations in the pressure applied by the diamond anvil cell to the sample resulting in inconsistent absorption, especially if the sample is thin. As for the other clear coat IR spectrum (i.e., UAZP00507, see Figure 3.8) that was not correctly matched, a doublet that was present in the PDQ spectrum (with 4 cm^{-1} resolution) appears as a singlet in the reconstructed ATR spectrum (8 cm^{-1} resolution) due to the lower resolution of the IR microscope.

Table 3.5   OMNIC Library Search Results for Thirty-eight Sample Cohort

PDQ Number	Manufacture	Clear coat		Surfacer		E-coat	
		Hit[1]	HQI	Hit[1]	HQI	Hit[1]	HQI
UAZP00401	Chrysler	1	97.42	3	96.19	3	91.22
UAZP00412	Chrysler	2	96.16	3	96.39	2	94.99
UAZP00421	Chrysler	1	92.96	1	97.32	1	91.71
UAZP00451	Chrysler	1	95.44	2	87.86	1	96.26
UAZP00569	Chrysler	1	97.14	3	95.30	1	90.30
UAZP00600	Chrysler	2	95.81	4	96.92	1	97.30
UAZP00342	Ford	1	95.50	5	92.98	1	94.53
UAZP00404	Ford	4	93.07	5	94.26	3	96.10
UAZP00467	Ford	1	95.45	1	91.71	3	93.86
UAZP00596	Ford	4	95.54	1	92.89	4	95.78
UAZP00477	Ford	4	95.99	1	87.27	1	86.41
UAZP00331	General Motors	1	97.72	2	94.82	1	91.27
UAZP00436	General Motors	1	98.66	2	91.77	2	92.66
UAZP00499	General Motors	5	96.42	1	93.13	-	90.41
UAZP00565	General Motors	5	97.05	1	95.69	-	91.29
UAZP00271	General Motors	2	97.46	1	95.71	-	90.41
UAZP00507	General Motors	-	95.02	-	85.35	1	96.02
UAZP00336	General Motors	5	98.51	2	95.87	-	92.86
UAZP00337	General Motors	2	98.36	3	94.38	-	88.21
UAZP00501	General Motors	4	96.56	4	90.01	-	92.53
UAZP00503	General Motors	4	97.05	5	93.27	-	94.69
UAZP00567	General Motors	-	95.79	-	91.94	-	92.13
UAZP00729	Honda	2	95.78	1	92.04	3	91.30
UAZP00730	Honda	1	96.19	2	90.36	2	94.21
UAZP00277	Honda	1	91.27	1	92.01	2	93.22
CONT00726	Honda	1	97.50	1	93.40	1	96.96
CONT00736	Honda	2	97.20	2	94.79	5	93.81
UAZP00385	Nissan	1	94.70	1	86.47	1	95.24
UAZP00440	Nissan	1	97.38	1	95.35	2	93.52
UAZP00745	Nissan	2	91.11	2	96.91	2	96.61
UAZP00731	Nissan	2	94.54	1	92.56	1	91.98
UAZP00527	Nissan	2	91.06	1	91.52	2	95.13
UAZP00537	Nissan	5	94.24	2	92.28	4	96.16
UAZP00381	Toyota	1	93.63	2	89.74	1	91.78
UAZP00313	Toyota	5	95.22	1	94.89	1	91.85
UAZP00733	Toyota	3	90.95	3	96.20	1	98.04
UAZP00561	Toyota	4	94.20	2	93.01	1	93.81
UAZP00484	Toyota	4	94.21	5	91.49	4	90.67

[1]The HQI value for the OT2, OU1 and OU2 layers without a ranking for hit was determined by comparing the spectrum of the sample searched against the spectrum of the same sample in PDQ. Although the HQI value for OT2, OU1, and OU2 layers of these paint samples is greater than 90.00, these samples did not appear in the hit list as OMNIC only provides the user with a list of the top ten matches.

Figure 3.7 UAZP00567 OT2 layer: reconstructed versus PDQ IR library spectrum
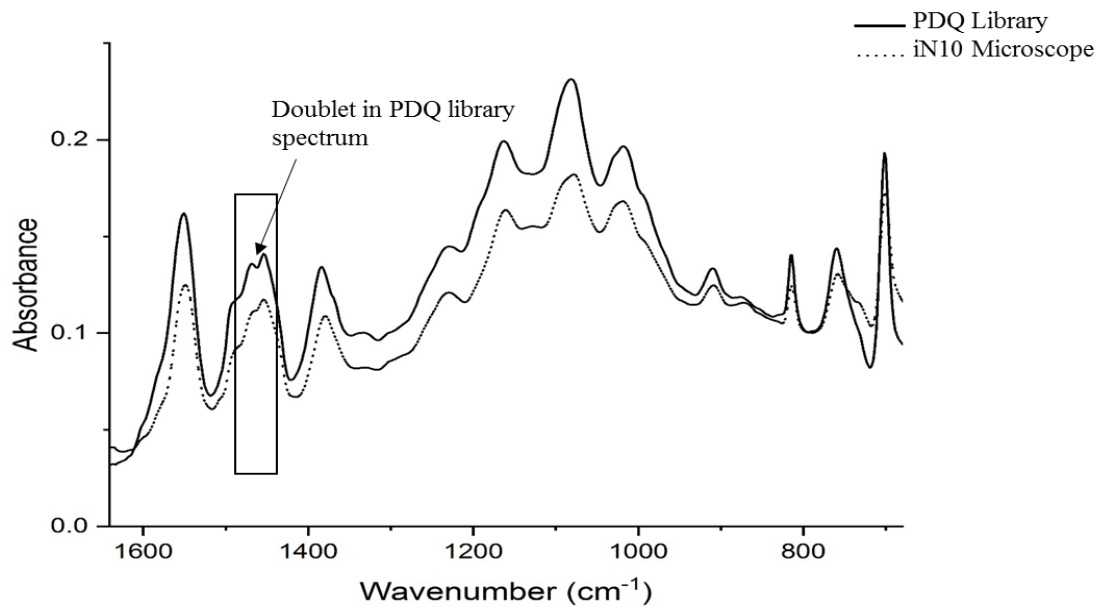
Figure 3.8 UAZP00507 OT2 layer: reconstructed versus PDQ IR library spectrum

As for the two IR spectra of the surfacer-primer layer that were not correctly matched by the General Motors paint library, one paint sample (i.e., UAZP00507, see Figure 3.9) was represented by a doublet around 1300 cm^{-1} in its PDQ library spectrum (4 cm^{-1} resolution) which appeared as a singlet in its reconstructed ATR spectrum (8 cm^{-1} resolution) presumably due to the lower resolution of the IR microscope. Furthermore, its reconstructed IR spectrum also exhibited a peak shift of approximately 16 cm^{-1} around 1200 cm^{-1} when compared to the corresponding PDQ library spectrum of the same paint sample in the General Motors Library. The other paint sample (i.e., UAZP00567, see Figure 3.10) exhibited an extra peak in its PDQ library spectrum that probably can be attributed to variations in the pressure applied by the diamond anvil cell to the sample. Furthermore, a singlet in the reconstructed IR spectra of the OU1 layer as opposed to a doublet in the corresponding PDQ library spectrum is probably due to the lower resolution of the iN10 microscope (8 cm^{-1} versus 4 cm^{-1}).
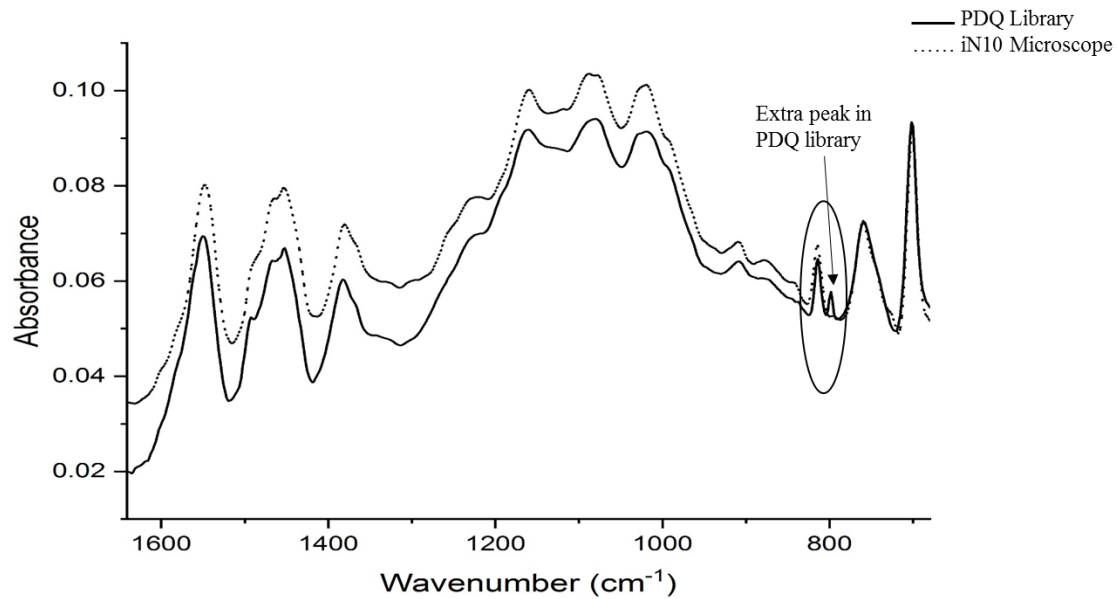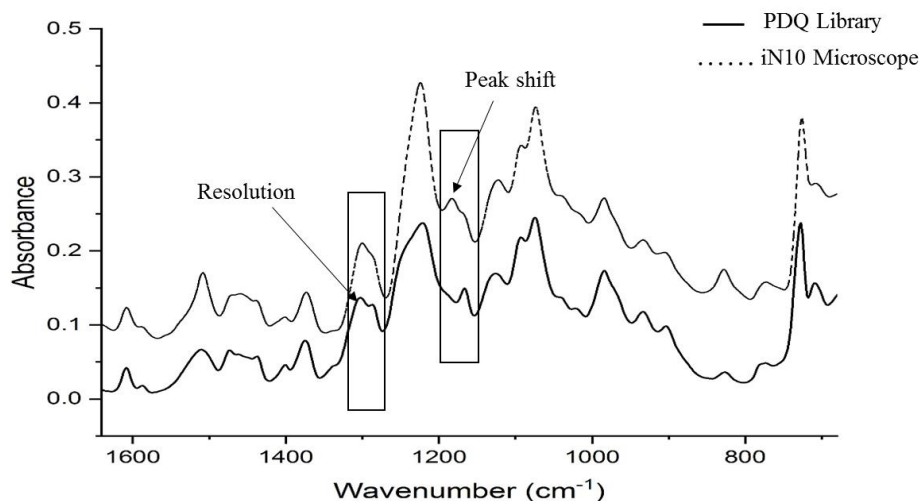

Figure 3.9 UAZP00507 OU1 layer: reconstructed versus PDQ IR library spectrum
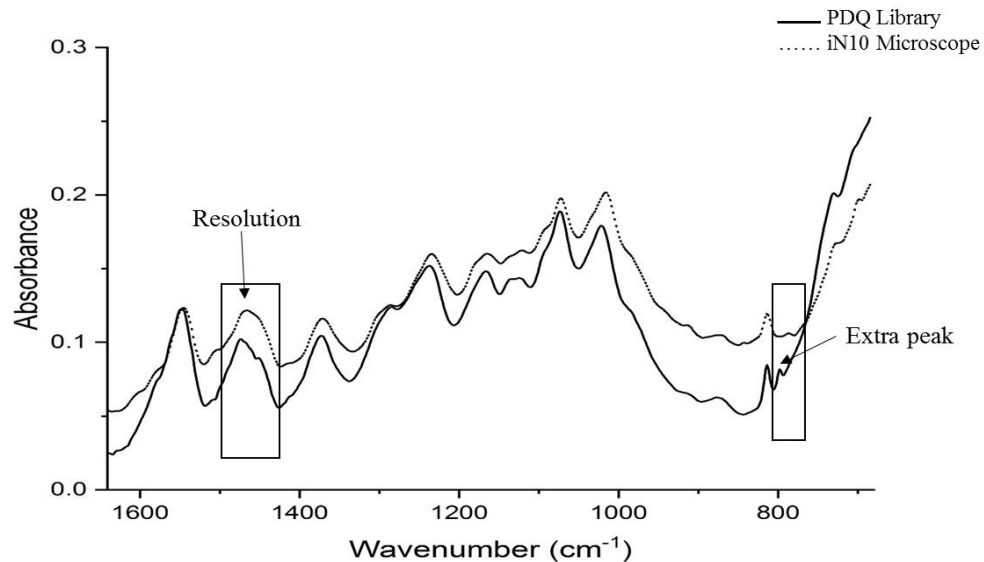
Figure 3.10 UAZP00567 OU1 layer: reconstructed versus PDQ IR library spectrum

For the OU2 layer, the 8 General Motors reconstructed IR spectra that were not correctly matched by OMNIC exhibited a peak shift at around 1100 cm^{-1} when compared to the corresponding IR spectra of the same paint samples in the PDQ library (see Figure 3.11). This can be attributed to the interaction of the OU2 layer of each of the 8 General Motors paint samples with the germanium crystal. Evidence in support of this assertion is shown in Figure 3.12. The ATR spectrum of UAZP00499 from the iN10 microscope for the pure OU2 layer is compared to the PDQ library spectrum of the OU2 layer of the same paint sample. The peak shift (germanium hemisphere versus diamond transmission cell) is approximately 7 cm^{-1}. To ensure that only the OU2 layer of UAZP00499 contributed to the ATR spectrum, the OU2 layer from a paint chip (that was not cross sectioned) was placed in direct contact with the germanium crystal. The thickness of the OU2 layer relative to the penetration depth of the evanescent wave ensured that only the OU2 layer was interrogated by the IR beam. Evidently, the large number of mismatched General

Motors paint samples is attributable to the unique interaction of the OU2 layer of the

General Motors paint samples with the germanium hemisphere crystal of the ATR

sampling accessory.
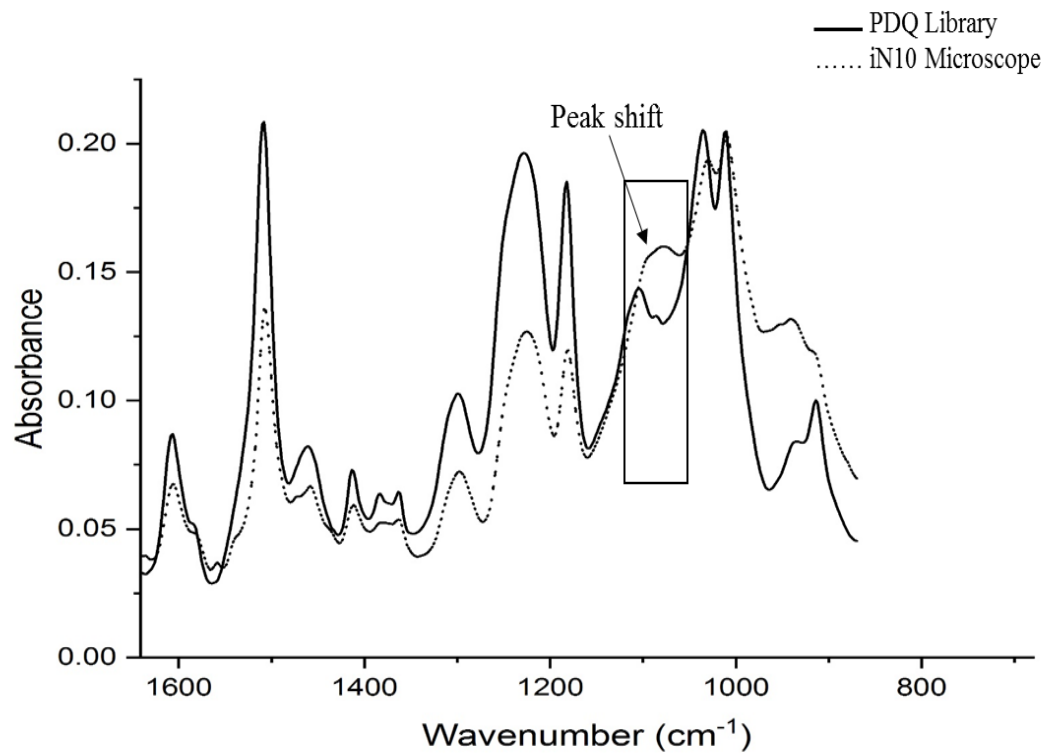
Figure 3.11 UAZP00499 OU2 layer: reconstructed versus the PDQ library spectrum.
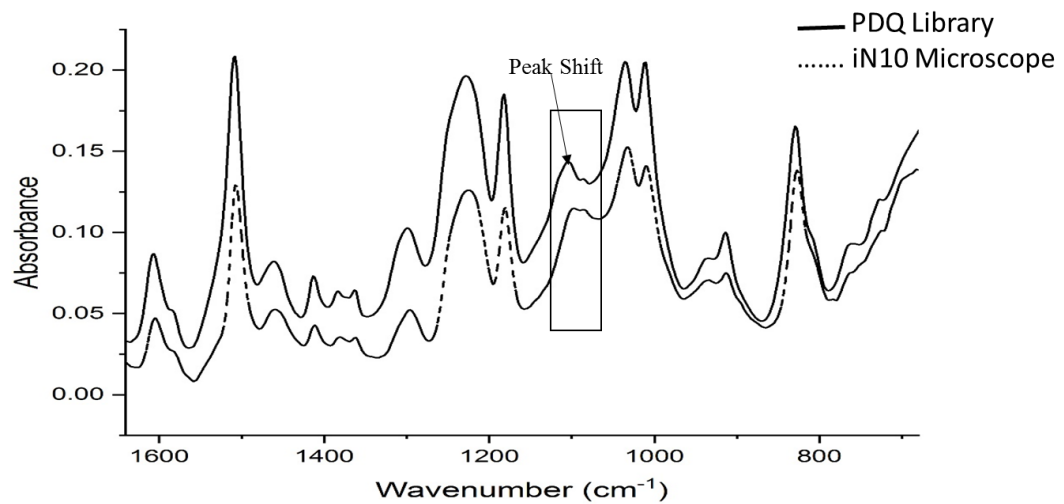
Figure 3.12 UAZP00499 OU2 layer: pure OU2 versus the PDQ library spectrum

The ALS reconstructions indicate that IR spectra of the clear coat, surfacer-primer, and e-coat layers can be collected in a single analysis from a multi-layered automotive paint chips using ATR infrared imaging microscopy. Decatenation of the IR spectral images as represented by line maps can be achieved using VER/ALS to obtain a pure IR spectrum of each layer. Nevertheless, the results of the library search for the e-coat layer raises questions as to the discrimination power of these ALS reconstructions for vehicle identification. In the next section, the evidentiary information content of these spectral reconstructions is assessed using a pattern recognition assisted infrared library search system.

.

3.5.2   Search Prefilters for Pattern Recognition Assisted Infrared Library Searching

The information content of the ALS reconstructed clear coat, surfacer-primer, and e-coat IR spectra of each of the 38 paint samples was assessed using a prototype pattern recognition library search engine previously developed for multiple automotive paint layers consisting of prefilters developed from the clear coat, surfacer-primer and e-coat layers [3-28 to 3-30]. The search prefilters (i.e., discriminants) developed from 1652 OEM paint systems spanned six manufacturers (19 assembly plants for General Motors, 25 assembly plants for Ford, 15 assembly plants for Chrysler, 6 assembly plants for Honda, 6 assembly plants for Nissan, and 5 assembly plants for Toyota) within a limited production year range (2000-2006). The six manufacturers selected for this study account for 80% of the vehicles purchased in North America.

The transmission spectra used to develop the search prefilters were first transformed into ATR spectra using the correction algorithm prior to search prefilter development. For search prefilter development, each corrected PDQ infrared library spectrum was normalized to unit length. The discrete wavelet transform [3-31] using the 8sym6 mother wavelet (Symlet wavelet family, sixth smallest filter size, eighth level of decomposition) was applied to the fingerprint region (1641 cm^{-1} to 680 cm^{-1}) of each layer. The Symlet 6 mother wavelet was selected because the shape of its scaling function closely matched the shape of the bands comprising the IR spectra of each automotive paint layer. Three sets of wavelet coefficients, one for each layer, were concatenated to form the data vectors used in search prefilter development. Wavelet coefficients from the lower levels of the

decomposition were retained, resulting in 3426 wavelet coefficients per paint sample (i.e., 1142 coefficients for each layer: clear coat, surfacer–primer, and e-coat).

A hierarchical classification scheme was used to identify the "make" and model of the vehicle from the reconstructed IR spectra of the paint samples comprising the larger data cohort. A search prefilter was developed to differentiate automotive paint samples by "make' using the transformed PDQ library spectra. For each "make", search prefilters were developed to identify the assembly plant (and hence the line and model of the vehicle) from FTIR spectra of the OEM paint systems.

The approach taken for developing a search prefilter system for automotive manufacturer is depicted in Figure 3.7. A hierarchical classification scheme was implemented by exploiting the linear separability of the classes (i.e., assembly plants or specific automotive manufacturers) comprising the training set, which consisted of the concatenated wavelet transformed IR spectra of 1652 OEM paint systems. A principal component plot of the 3426 wavelet coefficients of the entire spectral cohort was examined for sample clustering. Concatenated IR spectra from six Chrysler assembly plants was observed to form a tight sample cluster that appeared to be linearly separable from the other spectra in the principal component plot. The first prefilter separated these six Chrysler assembly plants from the other assembly plants of the six automotive manufacturers including the nine Chrysler assembly plants that remained with the data cohort. A genetic algorithm for pattern recognition analysis was used to identify wavelet coefficients characteristic of the six Chrysler assembly plants. To develop the first prefilter, the mutation rate of the GA was set at 0.4 and the number of chromosomes at 10,000. After 200 generations, the pattern recognition GA identified 19 wavelet coefficients for this two-

way classification problem. These 19 coefficients were identified by sampling key feature subsets, scoring their PC plots, and tracking those classes and/or spectra that were difficult to classify. The boosting routine of the pattern recognition GA used this information to steer the population to an optimal solution.

A principal component plot of the sample cohort (without the concatenated IR spectra of the six Chrysler assembly) revealed additional sample clustering. Prefilter 2 was developed using the pattern recognition GA to separate the two Chrysler assembly plants and four General Motors assembly plants from the cohort of IR spectra of the remaining General Motors, Chryslers, Honda, Nissan, Ford and Toyota paint samples. This process of discrimination followed by sample removal from the data cohort was repeated for the remaining IR paint samples. Prefilter 3 isolated two Chrysler assembly plants from 4 General Motors assembly plants in the six assembly plants identified by Prefilter 2. Prefilter 4 was developed to separate 3 Chrysler assembly plants from the remaining General Motors, Ford, Nissan, Toyota, Honda and Chrysler samples. All of the remaining paint samples for General Motors were separated from the other vehicle manufacturers using Prefilter 5. Prefilter 6 separated Toyota from Honda, Nissan, Ford and the remaining three Chrysler assembly plants in the cohort. Prefilter 7 separated these three Chrysler assembly plants and Ford from Honda and Nissan. Subsequently, Prefilters 8 and 9 were developed to effectively discriminate Honda from Nissan and Ford from the 3 remaining Chrysler assembly plants respectively. Further details about the manufacturer search prefilter system used in this study can be found elsewhere [3-32, 3-33].
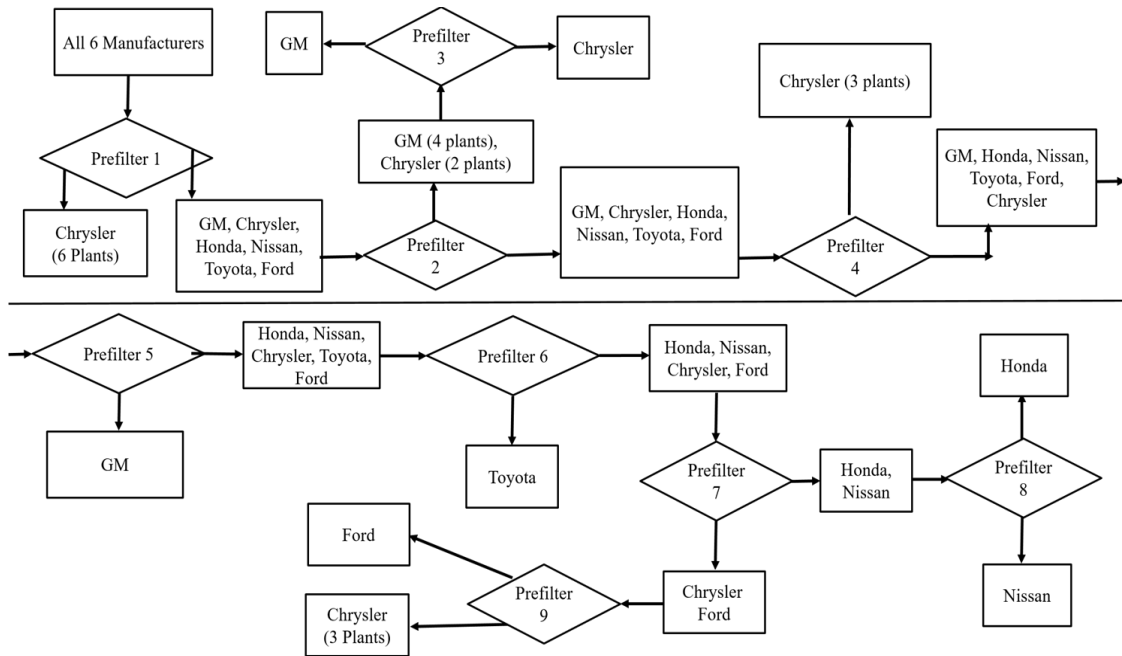
Figure 3.13. An overview of the manufacturer search prefilter system.

To demonstrate the operation of the manufacturer search prefilter system, infrared spectra of the clear coat, surfacer primer, and e-coat layers of a General Motors paint sample (UAZP00436 - Chevrolet Tahoe) were passed through the manufacturer search prefilter system. Prefilter 1 (see Figure 3.14) assigned UAZP00436 to the six automotive manufacturers including the nine Chrysler assembly plants that remained in the sample cohort. UAZP00436 was then passed to Prefilter 2 (see Figure 3.15) which assigned the sample to the data cohort that spanned all six manufacturers. Prefilter 4 (see Figure 3.16) again assigned UAZP00436 to the sample cohort that spanned all six manufacturers. Finally, UAZP00436 was passed to Prefilter 5 (see Figure 3.17) which assigned

UAZP00436 to the General Motors sample cluster. Thus, the "make" of the vehicle was

General Motors.
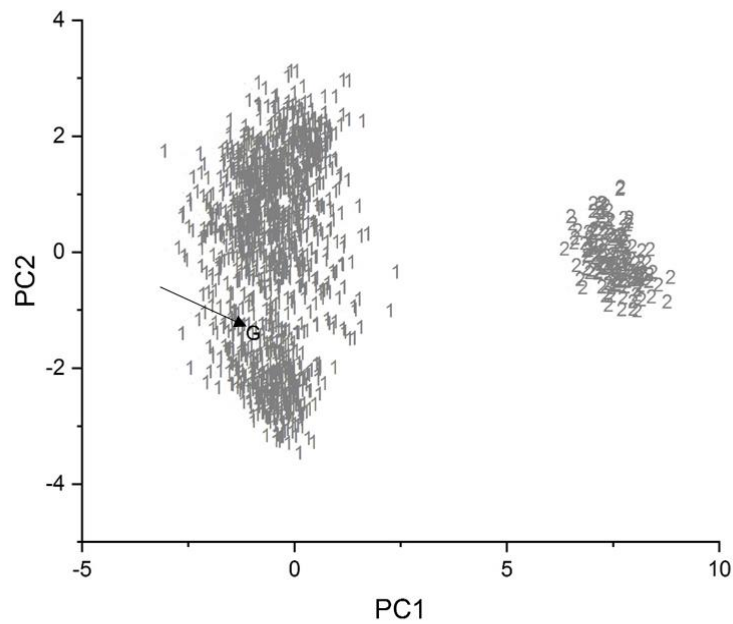
Figure 3.14. Projection of UAZP00436 onto the PC plot of Prefilter 1 defined by the 19 wavelet coefficients identified by the pattern recognition GA. Training set: 1 = General Motors, Chrysler, Honda, Nissan, and Toyota; 2 = Chrysler (6 assembly plants). Validation Set: G = UAZP00436 (Chevrolet Tahoe)
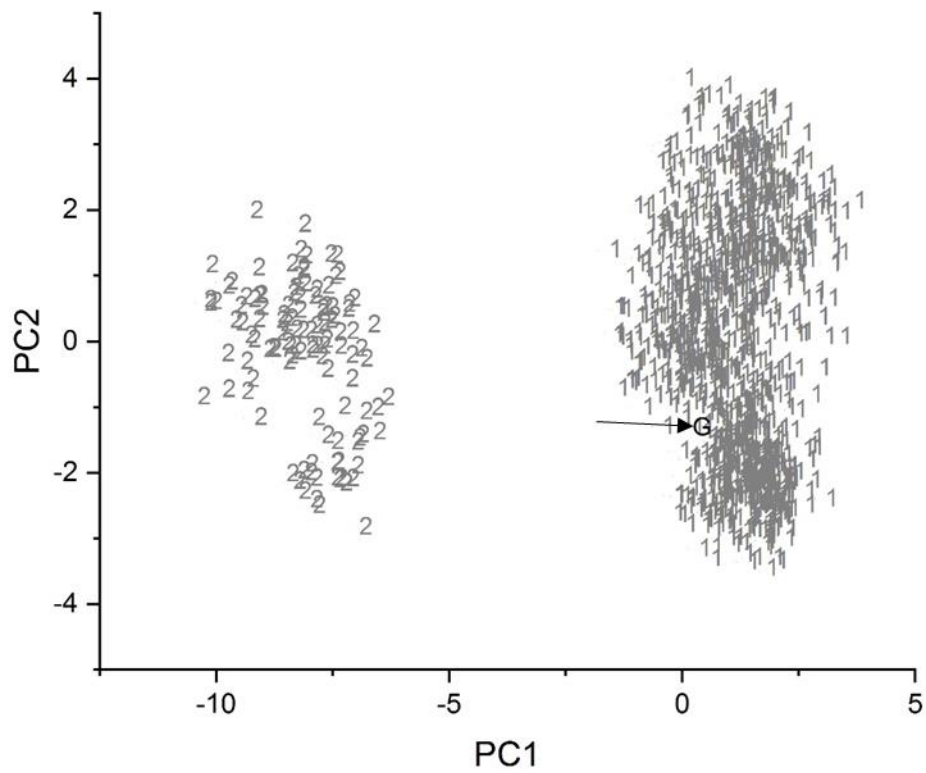
Figure 3.15. Projection of UAZP00436 onto the PC plot of Prefilter 2 defined by the 45 wavelet coefficients identified by the pattern recognition GA. Training set: 1 = General Motors, Chrysler, Honda, Nissan, and Toyota; 2 = Chrysler (2 assembly plants) and General Motors (4 assembly plants). Validation Set: G = UAZP00436 (Chevrolet Tahoe)
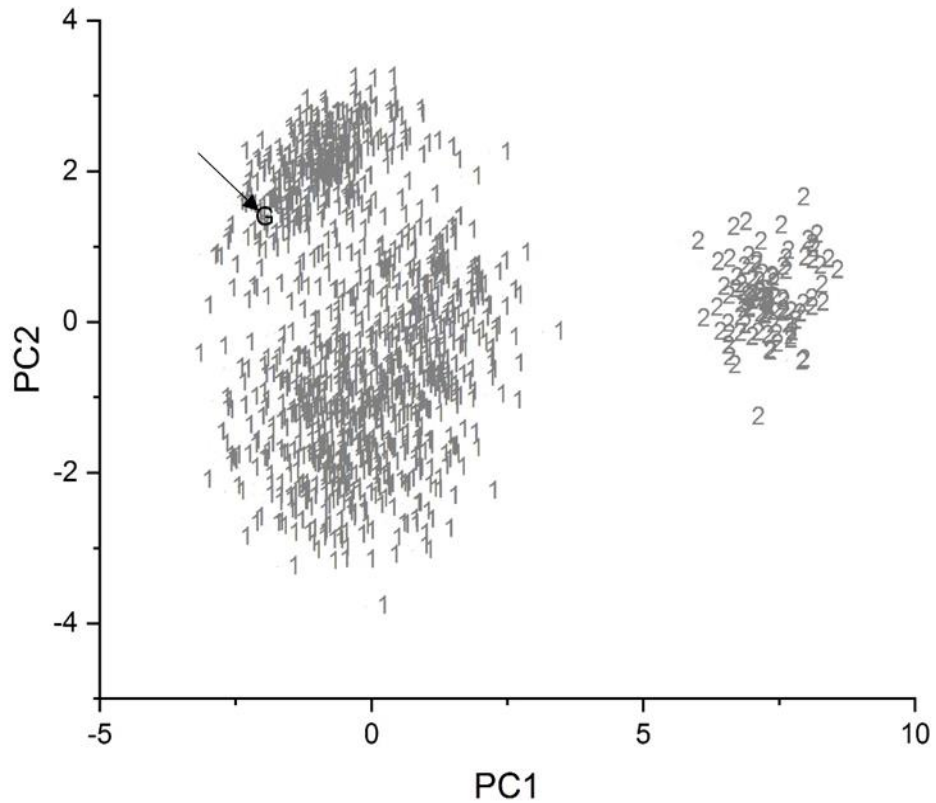
Figure 3.16. Projection of UAZP00436 onto the PC plot of Prefilter 4 defined by the 44 wavelet coefficients identified by the pattern recognition GA.   Training set: 1 = General Motors, Chrysler, Honda, Nissan, and Toyota; 2 = Chrysler (3 assembly plants). Validation Set: G = UAZP00436 (Chevrolet Tahoe)
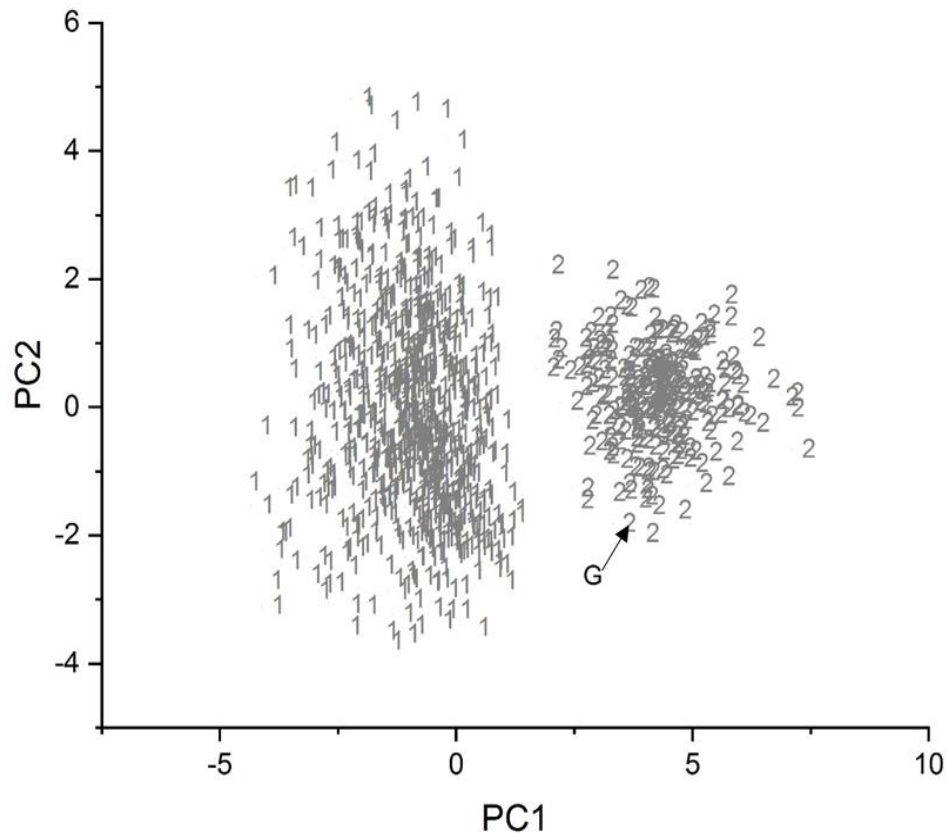
Figure 3.17. Projection of UAZP00436 onto the PC plot of Prefilter 5 defined by the 22 wavelet coefficients identified by the pattern recognition GA. Training set: 1 = Chrysler, Ford, Honda, Nissan, and Toyota; 2 = General Motors (all assembly plants). Validation Set: G = UAZP00436 (Chevrolet Tahoe)

Table 3.6 summarize the results obtained from the manufacturer search prefilter system for identifying the "make" of the vehicle from the wavelet transformed IR spectra of the clear coat, surfacer-primer and e-coat layers of the automotive paint samples. Twenty-six of the thirty-eight paint samples were correctly classified as to the "make" of the vehicle. The samples that were misclassified by the manufacturer search prefilter system spanned all six manufacturers.

Table 3.6. Manufacturer Search Prefilter System Results

PDQ Number	Manufacturer	Search Prefilter Output
UAZP00412	Chrysler	Chrysler
UAZP00421	Chrysler	Chrysler
UAZP00401	Chrysler	Chrysler
UAZP00600	Chrysler	Chrysler
UAZP00342	Ford	Ford
UAZP00404	Ford	Ford
UAZP00596	Ford	Ford
UAZP00477	Ford	Indeterminate
UAZP00436	General Motors	General Motors
UAZP00507	General Motors	General Motors
UAZP00331	General Motors	General Motors
UAZP00729	Honda	Honda
UAZP00277	Honda	Honda
CONT00726	Honda	Chrysler/Ford
CONT00736	Honda	General Motors
UAZP00730	Honda	Indeterminate
UAZP00440	Nissan	Nissan
UAZP00731	Nissan	Nissan
UAZP00527	Nissan	Nissan
UAZP00537	Nissan	Nissan
UAZP00733	Toyota	Toyota
UAZP00561	Toyota	Toyota
UAZP00484	Toyota	Toyota
UAZP00336	General Motors	General Motors
UAZP00501	General Motors	General Motors
UAZP00337	General Motors	General Motors
UAZP00499	General Motors	General Motors
UAZP00271	General Motors	General Motors
UAZP00503	General Motors	General Motors
UAZP00565	General Motors	Indeterminate
UAZP00567	General Motors	General Motors
UAZP00569	Chrysler	Indeterminate
UAZP00451	Chrysler	General Motors
UAZP00467	Ford	Indeterminate
UAZP00385	Nissan	General Motors
UAZP00745	Nissan	General Motors
UAZP00381	Toyota	General Motors
UAZP00313	Toyota	General Motors

The twelve automotive paint samples misclassified as to "make" by the manufacturer search prefilter system shared a common set of attributes. First, peak shifting was observed in the OU1 layer throughout the entire IR spectrum for 11 of the 12 cross sectioned paint samples: CONT00726, UAZP00730, UAZP00385, UAZP00745, UAZP00337 UAZP00451, UAZP00569, EAZP00467, UAZP00477, CONT00736, and UAZP00313. In 3 of the 12 paint samples, UAZP00467, CONT726 and UAZP00385, there is an extra peak in the reconstructed IR spectrum of the OU1 layer that is not present in the corresponding library spectrum. Two paint samples, UAZP00467 and UAZP00385, show a singlet instead of a doublet for the reconstructed IR spectrum of the OU1 layer due to $8cm^{-1}$ resolution in the reconstructed IR spectrum versus $4\ cm^{-1}$ resolution in the PDQ spectrum. As for the OU2 layer, each of the 12 paint samples exhibited peak shifting throughout the entire IR spectrum

After the "make" of the vehicle was identified by the manufacturer search prefilter system, the line and model of each vehicle were identified from the wavelet transformed IR spectra using a two-step process. First, the assembly plants of each manufacturer were divided into "plant" groups by applying cluster analysis to the fingerprint region of the average IR spectrum of the clear coat layer which served as a prototypical data vector to represent the formulation used by each assembly plant. Second, each plant group was separated into its constituent assembly plants using the wavelet transformed IR spectra of the clear coat, surfacer-primer and e-coat layers to develop a discriminant. Further details on the application of these search prefilters to OEM paint systems can be found elsewhere [3-28 to 3-30].

To demonstrate the operation of the assembly plant search prefilters, two Chrysler paint samples (UAZP00600 and UAZP00412) correctly identified as to "make" by the manufacturer search prefilter system were analyzed using the Chrysler search prefilter. First, the two paint samples were correctly assigned to their respective plant group, which is Plant Group 11 as these two paint samples originated from the Belvidere and Saltillo assembly plants (see Figure 3.12). For Plant Group 11, the associated search prefilter correctly assigned the paint samples to their respective assembly plants (see Figure 3.13). Because the assembly plant is indicative of the line and model of the vehicle, it is clear from this example that ALS reconstructed IR spectra of the layers can provide information about the "make", line, and model of the vehicle from which the paint sample originated.

Table 3.7 summarize the results obtained from the assembly plant search prefilters for identifying the assembly plant of the vehicle from the wavelet transformed spectra of the clear coat, surfacer-primer and e-coat layers of the automotive paint samples. Of the twenty-six paint samples that were correctly classified as to the "make of the vehicle, nineteen of the twenty-six paint samples were correctly assigned to their respective assembly plant.

Figure 3.18. Projection of UAZP00600 (Neon) and UAZP00412 (RAM) onto the PC plot of the Chrysler search prefilter for Plant Group. UAZP00600 and UAZP00412 were obtained from vehicles that were manufactured at the Belvidere and Saltillo assembly plants, which comprise Plant Group 11. 11 = Plant Group 11, 12 = Plant Group 12, C1 = UAZP00600, and C2 = UAZP00412.
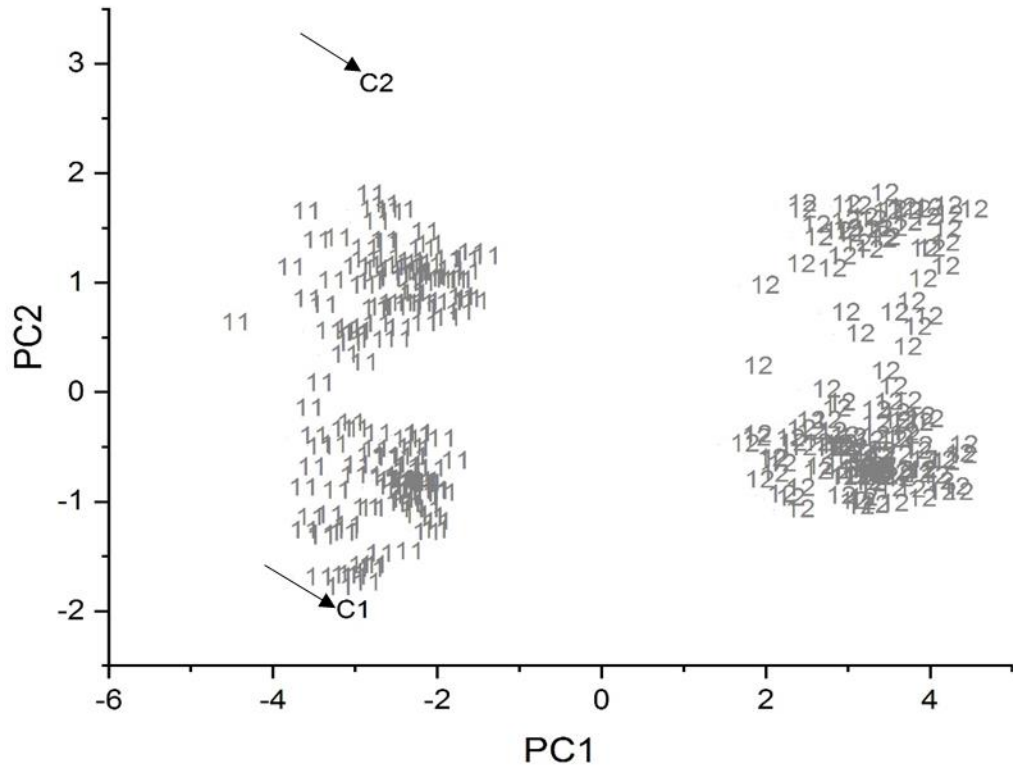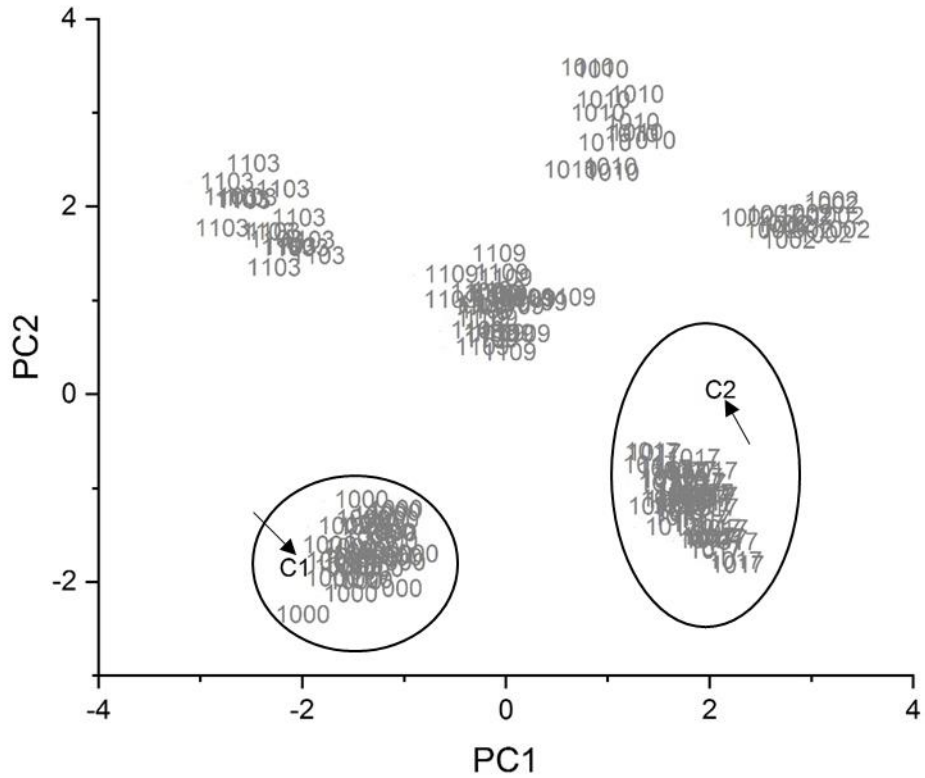
Figure 3.19. Projection of UAZP00600 (Neon) and UAZP00412 (RAM) onto the PC plot of the Chrysler search prefilter for assembly plant. UAZP00600 and UAZP00412 were obtained from vehicles manufactured at the Belvidere and Saltillo assembly plants. C1 = UAZP00600, C2 = UAZP00412, 1000 = Belvidere, 1002 = Bramalea/Brampton, 1010 = Toledo, 1017 = Saltillo, 1103 = Dodge Main, 1109 = St. Louis

Table 3.7. Assembly Plant Search Prefilter Results

PDQ Number	Manufacturer	Assembly Plant	Search Prefilter Output
UAZP00436	General Motors (GM)	[1]Arlington (ARL)	GM / ARL
UAZP00499	General Motors (GM)	[2]Orion (ORI)	GM / ORI
UAZP00331	General Motors (GM)	[3]Silao (SIL)	GM / SIL
UAZP00507	General Motors (GM)	[1]Oklahoma City (OKL)	GM / OKL
UAZP00567	General Motors (GM)	[1]Lansing (LAN)	GM / LAN
UAZP00600	Chrysler (CHR)	[5]Belvidere (BEL)	CHR / BEL
UAZP00412	Chrysler (CHR)	[5]Saltillo (SAL)	CHR / SAL
UAZP00421	Chrysler (CHR)	[6]Toledo (TOL)	CHR / TOL
UAZO00342	Ford (FOR)	[7]Hermosillo (HER)	FOR / HER
UAZP00596	Ford (FOR)	[7]Hermosillo (HER)	FOR / HER
UAZP00277	Honda (HON)	[8]Alliston, Canada	HON / Alliston
UAZP00729	Honda (HON)	[8]Sayama, Japan	HON / Sayama
UAZP00440	Nissan (NIS)	[9]Smyrna	NIS / Smyrna
UAZP00731	Nissan (NIS)	[9]Kyushu, Japan	NIS / Kyushu
UAZP00527	Nissan (NIS)	[9]Smyrna	NIS / Smyrna
UAZP00537	Nissan (NIS	[9]Canton	NIS / Canton
UAZP00733	Toyota (TOY)	[10]Georgetown	TOY / Georgetown
UAZP00484	Toyota (TOY)	[10]Japan	TOY / Japan
UAZP00561	Toyota (TOY)	[10]Fremont	TOY / Fremont
UAZP00336	General Motors (GM)	[4]Janesville (JAN)	GM/ARL
UAZP00501	General Motors (GM)	[4]Janesville (JAN)	GM/Freemont
UAZP00337	General Motors (GM)	[4]Janesville (JAN)	GM/Indeterminate
UAZP00271	General Motors (GM)	[4]Janesville (JAN)	GM/Indeterminate
UAZP00401	Chrysler (CHR)	[11]Newark (NEW)	CHR/Indeterminate
UAZP00404	FORD	[7]Louisville (LOU)	Ford/Dearborn

[1]Plant Group 1 of General Motors contains assembly plants Arlington, Oklahoma City, Lansing, Doraville, Fort Wayne, Moraine, Oshawa, and Pontiac

[2] Plant Group 2 of General Motors contains assembly plants Orion, Bowling Green, Hamtramck, and Wilmington.

[3]Plant Group 5 of General Motors contains assembly plants Silao, Ramos Arizpe, Spring Hill, and St. Therese

[4]Plant Group 4 of General Motors consists only of the Janesville assembly plant.

[5]Plant Group 11 of Chrysler contains assembly plants Belvidere, Saltillo, Bramalea/Brampton, Toledo (subplant), and Toluca

[6]Plant Group 12 of Chrysler contains assembly plants Toledo (subplant), Bloomington, Dodge Main, Sterling Heights, St. Louis, and Windsor

[7]Plant group 23 of Ford contains assembly plants Hermosillo, Kansas City, Saint Thomas Talbotsville and Louisville.

[8]Plant Group 31 of Honda contains assembly plants Alliston, East Liberty, Lincoln, Marysville, Suzuka and Sayama.

[9]Plant Group 41 of Nissan contains assembly plants Oppma, Smyrna, Kyushu Hiratsuka, and Canton.

[10]Plant Group 51 of Toyota contains assembly plants Cambridge, Georgetown, Japan, Princeton and Fremont

[11]Plant Group 13 of Chrysler contains assembly plant Jefferson North and Newark

The failure to correctly predict the "make' and model of all thirty-eight automotive paint samples using the proposed ATR infrared imaging method can probably be attributed to the alignment of the cross sectioned paint samples on the germanium (hemisphere) crystal in the ATR sampling accessory. Peak shifts throughout the entire IR spectrum of both the OU1 and OU2 layers (compared to the ATR corrected spectra from the PDQ library) were observed for those paint samples that were not correctly classified by OMNIC or the pattern recognition assisted infrared library search system. The range of incident angles used by the simulation algorithm to transform the PDQ transmission library spectra into an ATR spectral library is based on the cross sectioned paint sample being positioned at the mid-point of the germanium crystal. If the position of the sample is changed from the mid-point, the range of incident angles used by the simulation algorithm becomes narrower. As the simulation algorithm is using a larger range for cross sectioned paint samples that are misaligned from the mid-point of the crystal, the net effect is comparable to a change in the angle of incidence for the IR spectra of the OU1 and OU2 layers obtained by a single bounce ATR sampling accessory.

References

[3-1]   Bishea, G. A.; Buckle, J. L.; Ryland, S. G., *International forensic automotive paint database*. SPIE: 1999; 3576.

[3-2] Beveridge A, Fung T, MacDougall D. Use of infrared spectroscopy for the characterization of paint fragments. *In Forensic Examination of Glass and Paint Analysis and Interpretation*, Caddy B (Ed.). Taylor and Francis: NY, 2001; 220–233

[3-3] Buckle, J. L.; MacDougall, D. A.; Grant, R. R., PDQ—Paint Data Queries: The History and Technology Behind the Development of the Royal Canadian Mounted Police Forensic Laboratory Services Automotive Paint Database. *Canadian Society of Forensic Science Journal* **1997,** *30* (4), 199-212.

[3-4] Cartwright, N. S.; Cartwright, L. J.; Norman, E. W. W.; Cameron, R.; MacDougall, D. A.; Clark, W. H., A Computerized System for the Identification of Suspect Vehicles Involved in Hit and Run Accidents. *Canadian Society of Forensic Science Journal* **1982,** *15* (3-4), 105-115.

[3-5] Fasasi, A.; Mirjankar, N.; Stoian, R.-I.; White, C.; Allen, M.; Sandercock, M. P.; Lavine, B. K., Pattern Recognition-Assisted Infrared Library Searching of Automotive Clear Coats. *Appl. Spectrosc.* **2015,** *69* (1), 84-94.

[3-6] Streitberger,H. J.; Dsosel, K., Automotive paints and coatings ,2nd edition, WILEY-VCH Verlag Gmbh & Co. KGaA, 2008

[3-7] Rylan, S.G.; Suzuki. E.M.; Analysis of Paint Evidence, Forensic Chemistry Handbook, L. Kobilinsky (Ed.), John Willy & Sons, Inc., 2012, 131-139

[3-8] G. Pettis (Editor), Automotive Paints and Coatings, VCH Publication, New Yok, 1995.

[3-9] Lavine, B. K.; Fasasi, A.; Mirjankar, N.; Sandercock, M.; Brown, S. D., Search prefilters for mid-infrared absorbance spectra of clear coat automotive paint smears using stacked and linear classifiers. *Journal of Chemometrics* **2014,** *28* (5), 385-394.

[3-10] Zięba-Palus, J.; Zadora, G.; Milczarek, J. M.; Kościelniak, P., Pyrolysis-gas chromatography/mass spectrometry analysis as a useful tool in forensic examination of automotive paint traces. *Journal of Chromatography A* **2008,** *1179* (1), 41-46.

[3-11] Gothard, J., Evaluation of Automobile Paint Flakes as Evidence. *Journal of Forensic Sciences* **1976,** *21* (3), 636-641.

[3-12] Saferstein, R.; Manura, J., Pyrolysis Mass Spectrometry—A New Forensic Science Technique. *Journal of Forensic Sciences* **1977,** *22* (4), 748-756.

[3-13] Ferreira, K. B.; Oliveira, A. G. G.; Gomes, J. A., Raman spectroscopy of automotive paints: Forensic analysis of variability and spectral quality. *Spectroscopy Letters* **2017,** *50* (2), 102-110.

[3-14] Gelder, De.; Vandenabeele, J.; Govaert, P.; Moens, F., Forensic analysis of automotive paints by Raman spectroscopy. *Journal of Raman Spectroscopy* **2005,** *36* (11), 1059-1067.

[3-15] Lv, J.; Zhang, W.; Liu, S.; Chen, R.; Feng, J.; Zhou, S.; Liu, Y., Analysis of 52 automotive coating samples for forensic purposes with Fourier transform infrared spectroscopy (FTIR) and Raman microscopy. *Environmental Forensics* **2016,** *17* (1), 59-67.

[3-16] Griffiths, P.; de Haseth, J.A. Fourier Transform Infrared Spectrometry. New York: John Wiley and Sons, 2007. pp. 327-328

[3-17] Rodgers, P. G.; Cameron, R.; Cartwright, N. S.; Clark, W. H.; Deak, J. S.; Norman, E. W. W., The Classification of Automotive Paint by Diamond Window Infrared Spectrophotometry Part II Automotive Topcoats and Undercoats. *Canadian Society of Forensic Science Journal* **1976,** *9* (2), 49-68.

[3-18] Prati, S.J., Sciutto, E.; Mazzeo, G. R., *New Advances in the Application of FTIR Microscopy and Spectroscopy for the Characterization of Artistic Materials.* Accounts of Chemical Research, 2010. **43**(6): p. 792-801.

[3-19] Bhargava, R.L., I. W. , *Fourier Transform Mid-infrared Spectroscopic Imaging: Microspectroscopy with Multichannel Detectors.* Spectrochem. Analysis Using Infrared Multichannel Detectors, Blackwell Publishing, 2005, p. 1–24.

[3-20] Lavine, B.K.; Fasasi, A.; Mirjankar, N.; Nishikida, K.; Campbell, J. "Simulation of Attenuated Total Reflection Infrared Absorbance Spectra − Applications to Forensic Analysis of Automotive Clearcoats," Appl. Spec., 2014. 68(5), 608-615.

[3-21] Perera, U. D. N.; Nishikida, K.; Lavine, B. K. "Development of Infrared Library Search Prefilters for Automotive Clear Coats from Simulated ATR Spectra," Appl. Spectrosc**.** 2018. 186: 662-669.

[3-22] A. G. Marshall, Biophysical Chemistry: Principles, Techniques, and Applications, John Wiley & Sons, NY 1978.

[3-23] R. M. A. Azzam, N. M. Bashira, Ellipsometry and Polarized Light, Amsterdam, The Netherlands, Elsevier BV 1977.

[3-24] F. Wooten, Dispersion Relations and Sum Rules, in F. Wooten (editor) Optical Properties of Solids, New York, Academic Press, NY 1972, pp. 173-185

[3-25] Lavine, B. K., Ritter, J. P.; Voigtman, E. "Multivariate Curve Resolution in Liquid

Chromatography – Resolving Two Way Multicomponent Data Using a Varimax Extended Rotation," Microchemical J., 2002. 72(2): 163-178

[3-26] Lavine, B.K.; Davidson, C.E.; Ritter, J. P.; Westover, D.; Hancewicz, T. "Varimax Extended Rotation Applied to Multivariate Spectroscopic Image Analysis," Microchem. J., 2004. 76: 173-180.

[3-27] M. Boruta, "Spectral Library Matching," Spec., 2012. 27:1-6.

[3-28] Lavine, B. K.; White, C. G.; Allen, M. D.; Weakley, A. "Pattern Recognition Assisted Infrared Library Searching of the Paint Data Query Database to Improve Investigative Lead Information from Automotive Paint Trace Evidence," Applied Spec., 2017. 71(3): 480-495.

[3-29] Lavine, B. K.; White, C. G.; Allen, M. D. "Forensic analysis of automotive paints using a pattern recognition assisted infrared library searching system: Ford (2000-2006)," Microchem. J., 2016. 129: 173-183.

[3-30] Lavine, B. K.; White, C. G.; Allen, M. D.; Fasasi, A.; Weakley, A. "Evidential Significance of Automotive Paint Trace Evidence Using a Pattern Recognition Based Infrared Library Search Engine for the Paint Data Query Forensic Database," Talanta, 2016. 159: 317-329.

[3-31] J. S. Walker, A Primer on Wavelets and their Scientific Applications. Chapman & Hall, Boca Raton, FL 1999.

[3-32] Lavine, B. K.; Collin G. White, C. G.; Ding, T. "Library Search Prefilters for Vehicle Manufacturer to Assist in the Forensic Examination of Automotive Paints," Appl. Spec., 2018. 72(3): 476-488.

[3-33] F. Kwofie, "Vibrational Spectroscopy and Chemometrics Applied to the Forensic Analysis of Automotive Paints and Edible Oils," PhD Dissertation, July 2019.

Chapter IV

ANALYSIS OF GENTISIC ACID AND RENAL CELL CARCINOMA BIOMARKERS USING REVERSED-PHASE LIQUID CHROMATOGRAPHY WITH WATER RICH MOBILE PHASES

4.1 Introduction

Renal cell carcinoma (RCC) is the third most reported form of genitourinary cancer. RCC is caused by the malignancy of the renal tubules of the kidney [4-1]. According to published reports, RCC occurs in patients of all ages and has the highest mortality rate among patients with urological tumors [4-2]. The low survival rate of RCC patients is due the asymptomatic nature of this disease in its initial stages. Symptoms characteristic of RCC (e.g., blood in the urine, weight loss, fever, lethargy, and abdominal pain, low back pain, lump on the belly or back and night sweat) occur only during the terminal or metastatic stage [4-3], which limits the time frame for treatment. Clearly, both early detection and treatment are crucial to improve survivor rates. For this reason, advanced medical imaging techniques such magnetic resonance imaging [4-4], contrast-enhanced ultrasonography [4-5], and positron emission tomography-computed tomography [4-6] are used for the diagnosis of RCC. These imaging techniques are accurate and effective. However, patients with RCC require continued surveillance of small indolent renal masses, which is costly, and patients may suffer from side effects due to radiation exposure. Gonzalez and coworkers [4-7] have reported that approximately 29,000 cases of cancer occur each year in the United States due to radiation exposure from radiologic imaging. For all of these

reasons, routine screening of the general population using imaging techniques is not practical.

Urine, which is generated by the kidneys, is considered a valuable diagnostic biological fluid as it contains large amount of metabolites. As a cancer develops, the cancer cells activate specific metabolic pathways to supply the additional energy demand to the patient's body. During this process, the change in the metabolites profile may present potential biomarkers that are reliable for the early detection of RCC in the patient's urine. Furthermore, the noninvasive nature in which urine is collected and its availability makes it an ideal biological fluid for cancer prescreening in its early stages.

For RCC patients, Kim and coworkers [4-8] have reported that quinolinic acid, 4-hydroxybenzoic acid, and gentisic acid levels in urine are elevated compared to urine from controls. In Kim's study, urine samples were obtained from 29 RCC patients and from 33 patients with other urological conditions. Ultrahigh-performance liquid chromatography/tandem mass spectrometry and gas chromatography/mass spectrometry were used in Kim's study to perform the analysis. Ma and coworkers [4-9] developed an analytical procedure for the simultaneous determination of quinolinic acid, gentisic acid and 4-hydroxybenzoic acid in urine using liquid chromatography/tandem mass spectrometry. Both studies utilized expensive instrumentation that is generally not available in most clinical laboratories due to the high cost of instrument maintenance and per sample analysis.

In this study, a mobile phase of 5% methanol in water and 0.1% butanol in water acidified with 0.6% acetic acid percolating through a Zorbax C_{18} column were evaluated as a potential method to separate and detect RCC biomarkers in urine. In two previous studies, Lavine and coworkers [4-10, 4-11] investigated both short chain and medium chain-length alcohols as organic modifiers for enhancing selectivity in RPLC using a set of six vanillin compounds as retention probes. Greater selectivity in the separation of the vanillin compounds was obtained when hydrophobic alcohols such as butanol or pentanol were used as mobile phase modifiers. In another study, Lavine and co-workers [4-12] demonstrated that water-rich mobile

phases have advantages over methanol/water or acetonitrile/water mixtures for the separation of water-soluble and weakly retained compounds including improved separation of congeners and better tuning of HPLC separations. Using a water-rich mobile phase, an isocratic method to detect and quantify creatinine, quinolinic acid, gentisic acid and 4-hydroxybenzoic in synthetic urine has been developed and is reported in this chapter. Although creatinine is not an RCC biomarker, it is included in the study to account for the renal dilution of the urine. Since the concentration of creatinine corresponds to urine dilution, its levels must be closely monitored in any urinary biomarker analysis study. The biomarker-to-creatinine concentration ratio will always give more meaningful results than the concentration of a biomarker alone [4-13, 4-14].
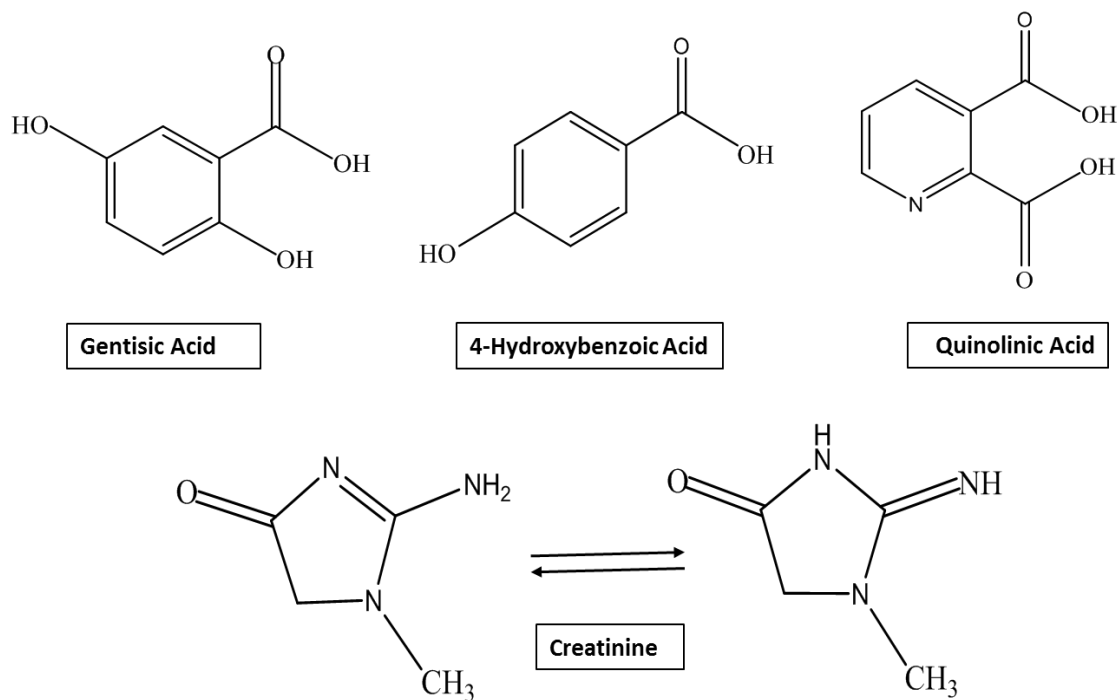
Figure 4.1 Structures of RCC biomarkers

4.2 Experimental

Creatinine, quinolinic acid, gentisic acid, and 4-hydroxybenzoic acid were obtained from Sigma Aldrich and used as received. Glacial acetic acid from Pharmaco (Auckland, New Zealand), methanol and butanol from Fisher Scientific were obtained and used as received. Doubly distilled water was prepared using a Barnstead Nano-Pure II System (Barnstead International, Dubuque, IA). Submicron filtered HPLC grade water obtained from Thermo-Fischer, USA.

Stock solution (2000 ppm) of creatinine, quinolinic acid, gentisic acid and 4-hydroxybenzoic acid were prepared by weighing and dissolving the required amount of compounds in doubly distilled water. These compounds are the probes for the separation and the structures are given in Figure 4.1. Standard solution (200 ppm) of each compounds were prepared by the dilution of stock solution using 0.6% acetic acid. These solutions were used to determine the retention time of each compound. A standard solution of the mixture of these four compounds (200 ppm each) was prepared and used throughout to study the solvent strength of the mobile phases investigated. For the calibration curves, standards of this four component mixtures of 5 ppm, 10 ppm, 25 ppm, 50 ppm, 70 ppm, 100 ppm 250 ppm, 500 ppm, 700 ppm, and 1000 ppm were prepared by diluting the 2000 ppm stock solution using 0.6% acetic acid in water.

Water-rich mobile phases used in these experiments were prepared by using methanol and butanol as organic modifiers. To prepare each mobile phase, transfer pipette were often employed because of the small volume of organic modifiers used. Each mobile phase was filtered under vacuum using 0.45 µm pore size filters (Varian, Walnut Creek, CA, USA). Varian Nylon 66 membrane filters were used to remove particulate matter from the mobile phase. All mobile phases were vacuum degassed prior to use.

For preparation of the urine samples, synthetic urine (Flinn Scientific) was diluted five-fold with HPLC grade water containing 0.6% acetic acid and then spiked with the appropriate

amounts of creatinine, quinolinic acid, gentisic acid, and 4-hydroxybenzoic acid to yield the corresponding 20 ppm (low concentration), 100 ppm (mid-level concentration) and 700 ppm (high level concentration) test mixtures of the RCC biomarkers.

The RPLC studies were performed using a Varian High-Performance Liquid Chromatograph equipped with a Shimadzu column oven, ProStar reciprocating pump, photodiode array detector and Galaxie Chromatography Workstation Software (Version 1.8.504.1). All RPLC separations were performed on a Zorbax 5µm Eclipsed-XBD C_{18} 80Å column (150 × 4.6 mm) at ambient temperature. Each mobile phase was equilibrated with the stationary phase by percolating the mobile phase for approximately 12 hours through the column at a flow rate of 1.0 mL/min prior to sample injection to ensure reproducible column wetting. In some instances, runs were continued from the previous day using the same mobile phase as the solvent mixture stored in the column. In these instances, only one hour was necessary for equilibration. A flow rate of 1.0 mL/min was used throughout the study because of the desire to develop an inexpensive isocratic RPLC method to detect RCC biomarkers in urine.

The sample injection volume used was 5 µL. The dead time, as determined by injecting 0.1M KNO_3 solution onto the Zorbax column with the flow rate of the mobile phase fixed at 1.0 mL/min, was 1.41 minutes for all methanol in water and 1.51 minutes for all butanol in water mobile phases investigated. These values were used in all k' calculations. All k' values reported in this study were averages of triplicate determinations. Deviations for individual k' values were always less than 1%.

Liquid chromatography/mass spectrometry (LC-MS) analysis of the urine sample and 4-hydroxybenzoic acid were also performed in this study using a Shimadzu 2010EV LC-MS equipped with electrospray ionization operated in the negative mode. All LC-MS separations were performed at 0.3 mL/min using a Shimadzu Premier C_{18} 3µm column (100 x 4.6 mm) operated at ambient temperature with 5% methanol water (0.6% acetic acid) and 0.1% butanol water (0.6% acetic acid) mobile phases.

4.3 Results and Discussions

      A series of chromatograms were run to demonstrate the advantages of using methanol and butanol as the organic mobile phase modifier to separate a test mixture of creatinine, quinolinic acid, gentisic acid, and 4-hydroxybenzoic acid on a Zorbax Eclipsed-XBD C_{18} column. Figure 4.2a and 4.2b show the chromatograms of the test mixture obtained from the best separation with methanol in water (5% methanol in water mobile phase with 0.6% acetic acid) and butanol in water (0.1% butanol in water with 0.6% acetic acid) as the mobile phase. The concentration of the test mixture was 200 ppm and the separation of the mixture by both mobile phases was performed at the ambient temperature. The elution order of the four compounds comprising the test mixture can be correlated to their solubility in water and to the log of the Octanol-Water Partition coefficient (Log P) value (see Table 4.1).

Table 4.1. Log P and Water Solubility Values of the Test Mixture Compounds

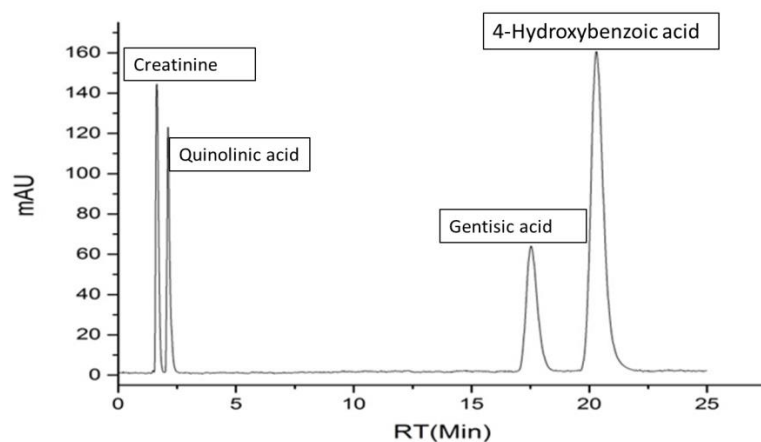| Compounds	Water Solubility (mg/ml)	Log P
Creatinine	12.9	-1.76
Quinolinic Acid	11	0.15
Gentisic Acid	5	1.42
4-hydroxybenzoic acid	5	1.67

Figure 4.2a. Chromatogram of the RCC test mixture using 5% methanol in 0.6% acetic acid as the mobile phase
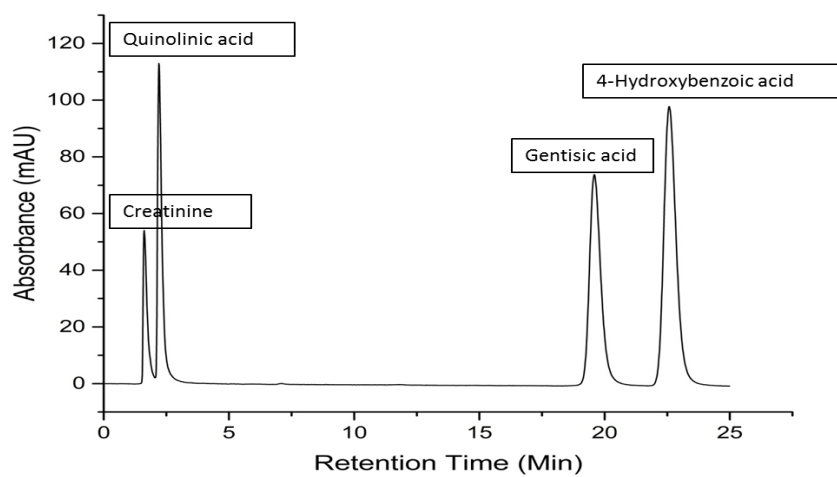

Figure 4.2b. Chromatogram of the RCC test mixture using 0.1% butanol with 0.6% acetic acid as the mobile phase.

Tables 4.2 and 4.3 list the number of plates generated by the Zorbax C_{18} column for each compound in these two chromatograms calculated using the Foley-Dorsey method [4-15]. The number of theoretical plates obtained for gentisic acid and 4-hydroxybenzoic acid is consistent with the number of plates that one would expect to obtain for a C_{18} column 15 cm in length and packed with 5 μm particles. Although the extra column volume of this system was not measured, the low number of plates obtained for creatinine and quinolinic acid, which elutes with or near the dead marker, can probably be attributed to extra column band broadening. The excessive tailing of the creatinine and quinolinic acid peaks is suggestive of interactions with silanol groups.

Table 4.2 Number of Plates[a,b] Generated by 5% Methanol Water Mobile Phase.

Compounds	Retention time	Number of theoretical plates	Asymmetry values
Creatinine	1.61 minutes	1242	2.48
Quinolinic acid	2.15 minutes	1627	2.93
Gentisic acid	17.64 minutes	10,092	1.20
4-Hydroxybenzoic acid	20.86 minutes	10,634	1.20

[a] Plate calculation performed using the Foley-Dorsey method [15]
[b] Each plate calculation is an average of triplicate determinations.

Table 4.3 Number of Plates[a,b] Generated by 0.1% Butanol Water Mobile Phase

Compounds	Retention time	Number of theoretical plates	Asymmetry values
Creatinine	1.6 minutes	518	3.98
Quinolinic acid	2.2 minutes	938	2.77
Gentisic acid	19.7	9820	1.52
4-Hydroxybenzoic acid	22.7	10950	1.42

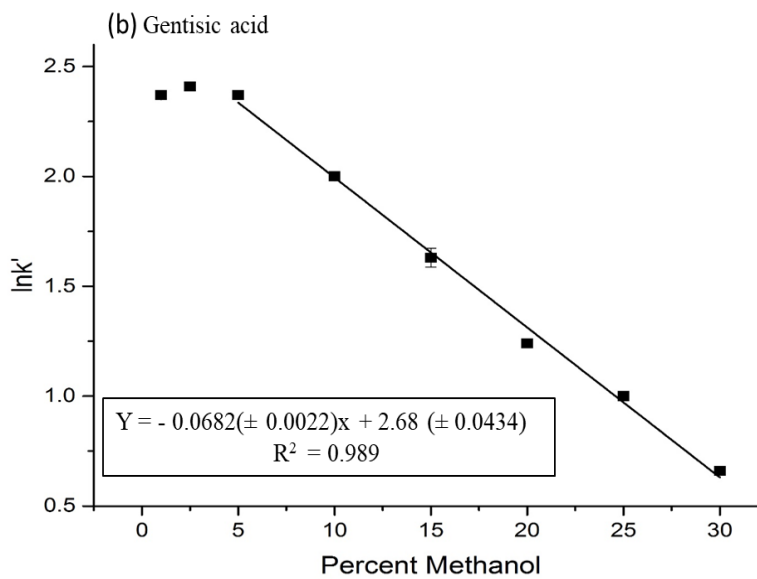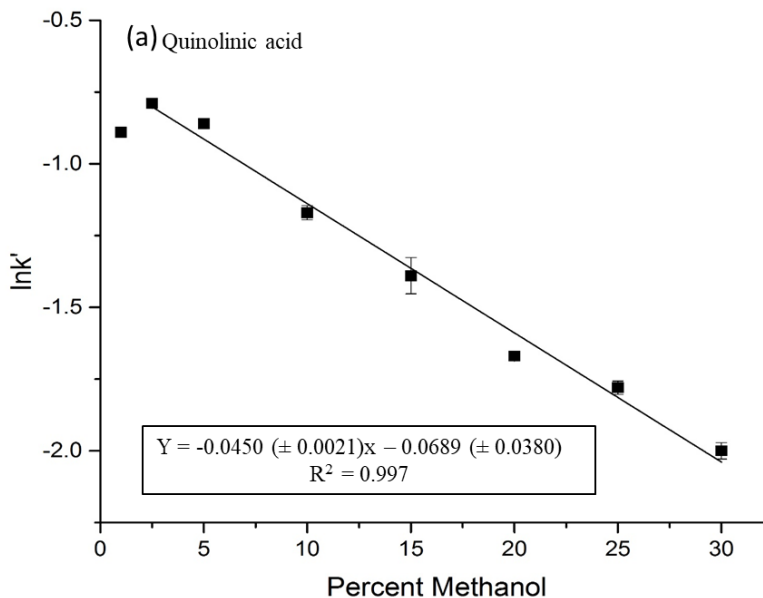[a] Plate calculation performed using the Foley-Dorsey method [4-15]
[b] Each plate calculation is an average of triplicate determinations.

Synder solvent strength plots [4-16] were constructed and analyzed using the four RCC compounds to obtain insight into the factors that may influence retention using Equation 1, where ϕ is the volume percentage of organic modifier in the mobile phase, B is a measure of the interaction of the solute with the mobile phase and $\ln k_w$ is the logarithm of the retention factor of the compound in a purely aqueous media. Both B and $\ln k_w$ are constants characteristic of the solute.

$$\ln k' = \ln k_w - B\phi \tag{4.1}$$

Figure 4.3 a), b) and c) show the plots of $\ln k'$ versus ϕ for quinolinic acid, gentisic acid and 4-hydroxybenzoic acid for the methanol in water mobile phases. Eight methanol-water mobile phases with 0.6% acetic acid were used to generate the plots: 30%, 25%, 20%, 15%, 10%, 5%, 2.5%, and 1% methanol in water. For all three compounds, the 1% run was excluded from the regression analysis and the 1% and 2.5% runs were excluded from the regression analysis for 4-hydroxybenzoic acid and gentisic acid as these points did not follow the trend expected for organic compounds from classical RPLC retention behavior. Exclusion of only the 1% run for quinolinic acid can probably be attributed to this compound being less hydrophobic than 4-

hydroxybenzoic acid and gentisic acid, which in turn would make it more accessible to the pore volume of the stationary phase when using very small amounts of methanol as the mobile phase modifier. Both B and ln k_w in Equation 1 for these three compounds can be correlated to their water solubility and log P values (see Table 4.1).

(a) Quinolinic acid

$$Y = -0.0450 \ (\pm 0.0021)x - 0.0689 \ (\pm 0.0380)$$
$$R^2 = 0.997$$

(b) Gentisic acid

$$Y = -0.0682(\pm 0.0022)x + 2.68 \ (\pm 0.0434)$$
$$R^2 = 0.989$$

(c) 4-hydroxybenzoic acid
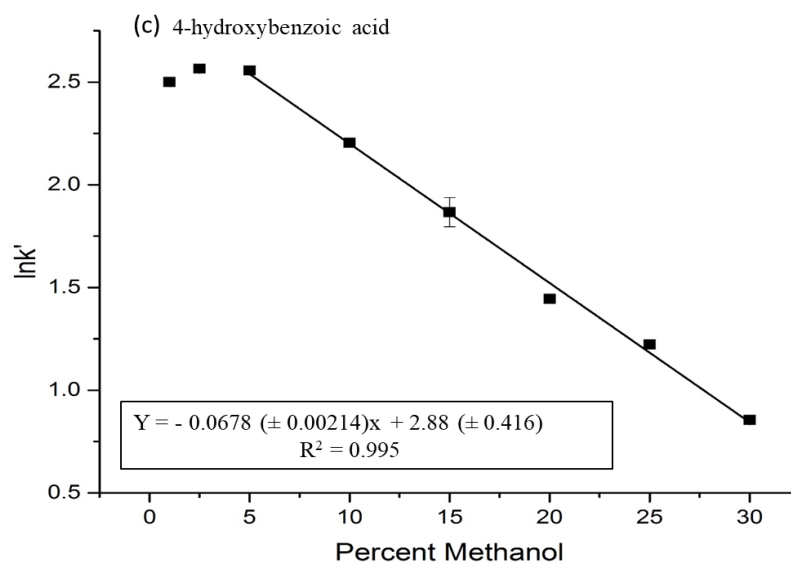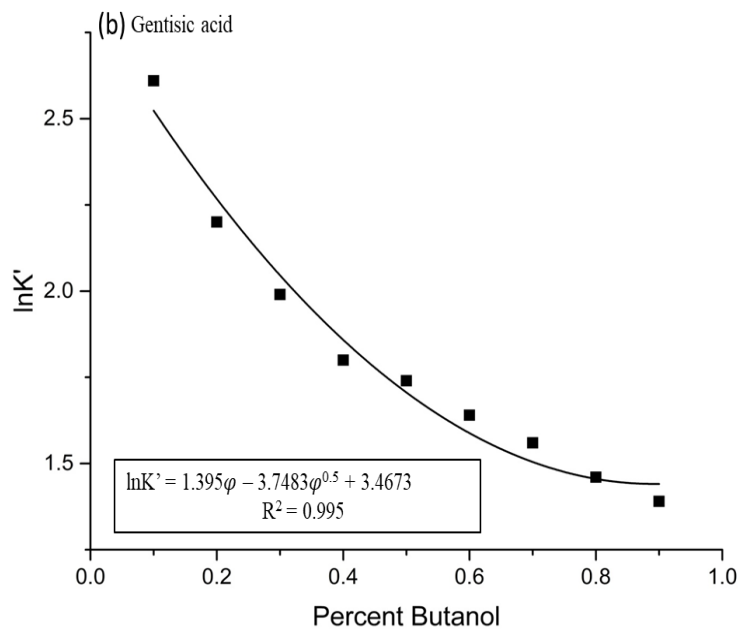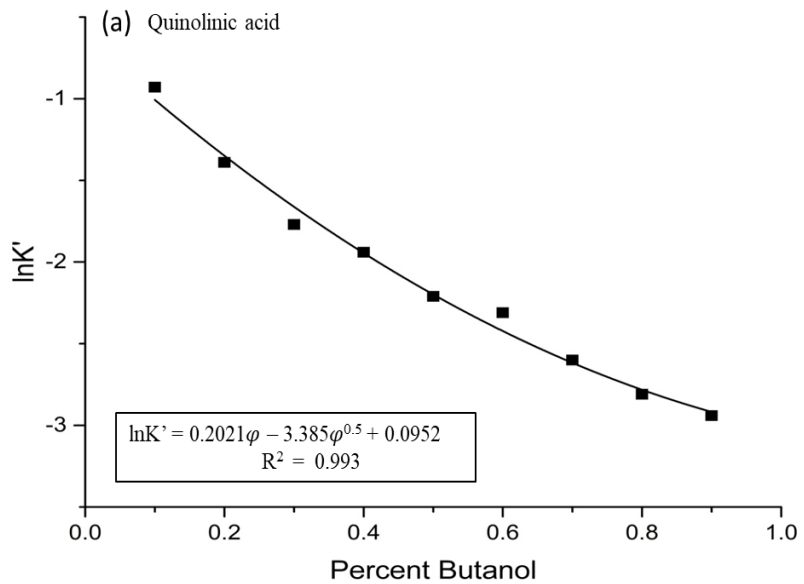
$Y = -0.0678 (\pm 0.00214)x + 2.88 (\pm 0.416)$
$R^2 = 0.995$

Figure 4.3. Plot of ln k' versus percent methanol for a) quinolinic acid b) gentisic acid c) 4-hydroxybenzoic acid

Figure 4.4 shows a plot of ln k' versus ϕ for quinolinic acid, gentisic acid, and 4-hydroxybenzoic acid. Creatinine was excluded from this analysis because it co-eluted with the dead marker. Nine butanol-water mobile phases were used to generate these plots: 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, and 0.9% butanol in water with 0.6% acetic acid. All three ln k' plots did not follow the linear trend expected for classical RPLC behavior. Therefore, each plot was fitted using Schoenmaker's solubility parameter model, which has been used by other workers to describe nonlinear behavior in Snyder solvent strength plots [4-17, 4-18]. An $E\phi^{0.5}$ term, where E is the regression coefficient, is added to Snyder's solvation model (see Equation 1) to describe the curvature at organic modifier concentrations less than 5%.

Extrusion of the mobile phase and the biomarkers from the particle pores [4-19, 4-20], where most of the C_{18} alkyl bonded phase is located, may occur when using a mobile phase such as a methanol/water or acetonitrile /water mixture with a low percentage of the organic modifier. However, dewetting of the column did not occur with water-rich mobile phases that used butanol

as the organic modifier. The small amount of butanol in the mobile phase solutions investigated

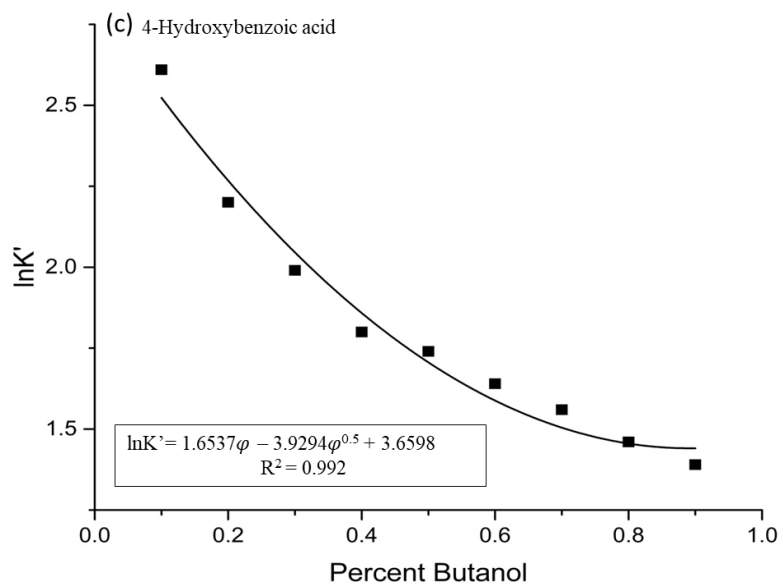is also responsible for the curvature in the Snyder solvent strength plots.

(a) Quinolinic acid

$$\ln K' = 0.2021\varphi - 3.385\varphi^{0.5} + 0.0952$$
$$R^2 = 0.993$$

(b) Gentisic acid

$$\ln K' = 1.395\varphi - 3.7483\varphi^{0.5} + 3.4673$$
$$R^2 = 0.995$$

(c) 4-Hydroxybenzoic acid

$$\ln K' = 1.6537\varphi - 3.9294\varphi^{0.5} + 3.6598$$
$$R^2 = 0.992$$

Figure 4.4. Plot of ln k' versus percent butanol for a) quinolinic acid, b) gentisic acid, and c) 4-hydroxybenzoic acid.

The calibration curve for creatinine, quinolinic acid, gentisic acid and 4-hydroxybenzoic acid were developed from standards prepared using doubly distilled water with 0.6% acetic acid. For creatinine, quinolinic acid, and gentisic acid, 232 nm was selected for the development of the calibration plots as this wavelength corresponded to an absorbance maximum, whereas 252 nm was selected for 4-hydroxybenzoic acid due to its high εb value. The calibration curve for each compound was prepared by ploting the average peak area of each standard against its concentration.

Figure 4.5 a), b), c) and d) are the calibration curves for creatinine, quinolinic acid, gentisic acid and 4-hydroxybenzoic acid using a 5% methanol water-rich mobile phase in 0.6% acetic acid. Calibration curves for both creatinine and quinolinic acid did not pass through the origin which can be attributed to the uncompensated background from the dead marker as these compounds eluted with the dead marker or near the dead marker, whereas the calibration curves

for both gentisic acid and 4-hydroxybenzoic acid, which are quite distant from the dead marker pass through the origin.
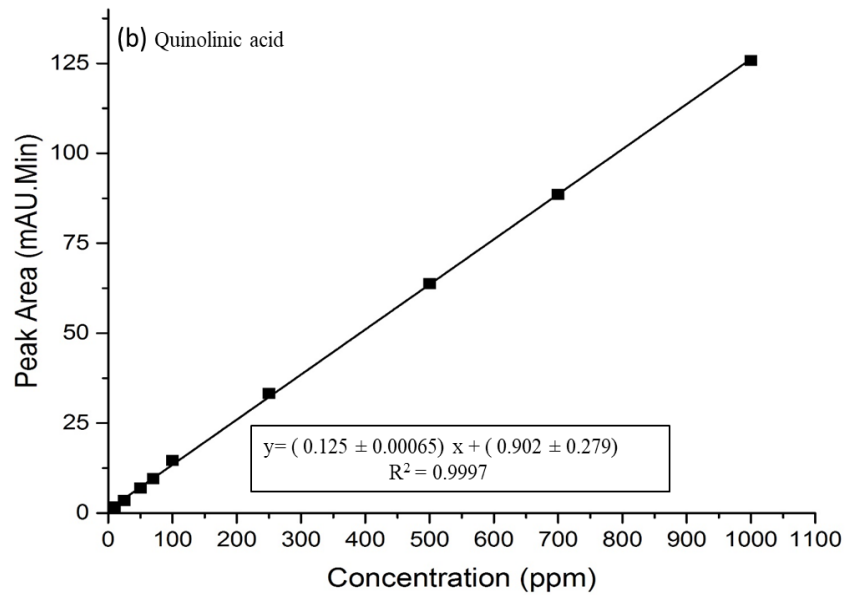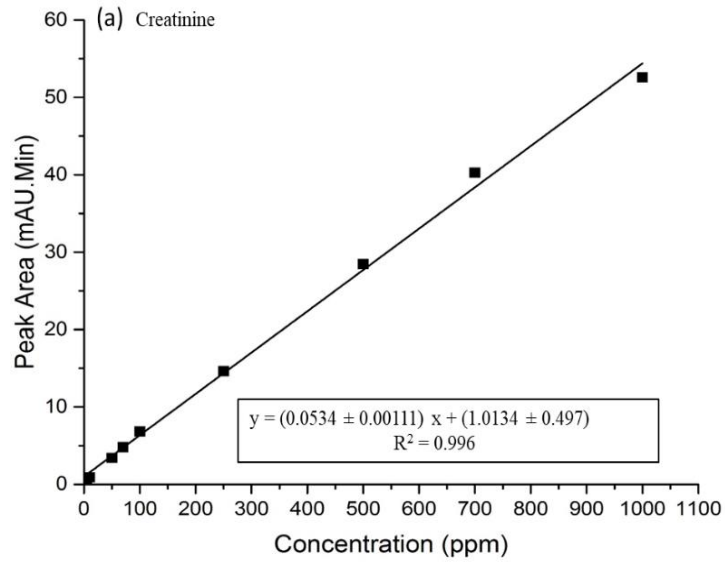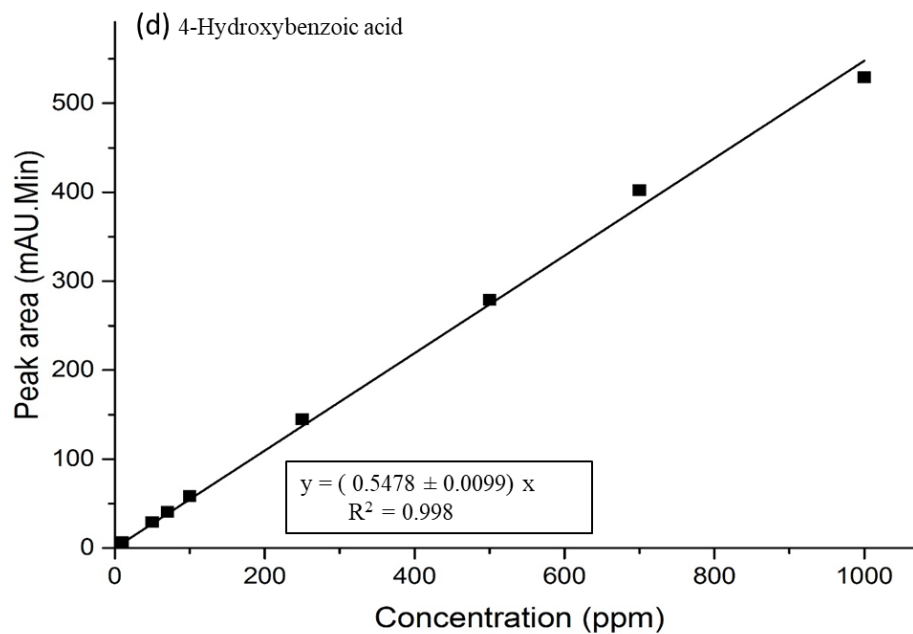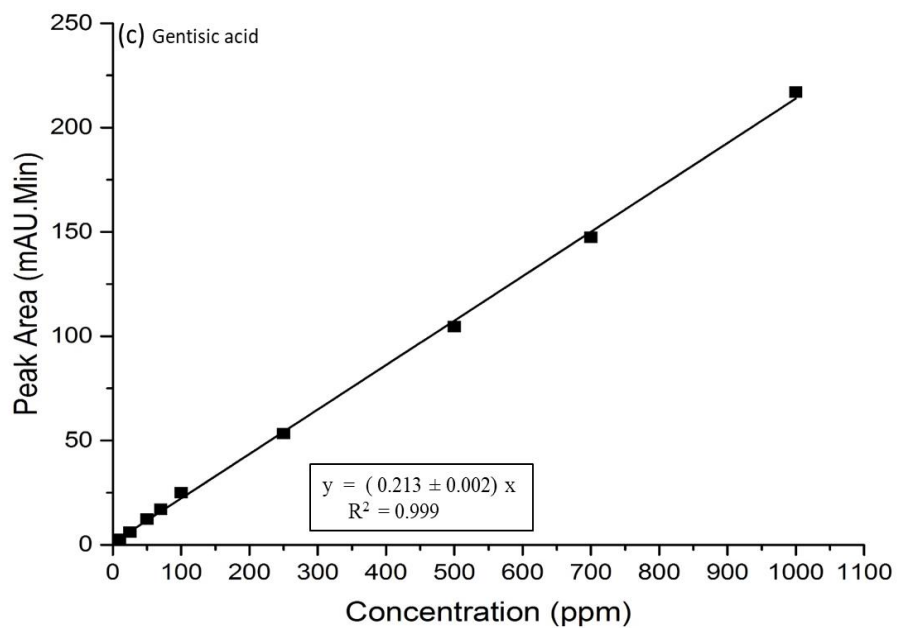
Figure 4.5 Calibration curve for a) creatinine b) quinolinic acid c) gentisic acid and d) 4-hydroxidebenzoic acid using 5% methanol in water with 0.6% acetic acid as the mobile phase.

Figure 4.6 a), b), and c) are the calibration curves for quinolinic acid, gentisic acid and 4-hydroxybenzoic acid obtained using 0.1% butanol in water with 0.6% acetic acid as the mobile phase. The calibration curve for creatinine was not plotted as it co-eluted with the dead marker. The linearity of these three calibration curves over the concentration range investigated (5 ppm to 1000 ppm) is excellent. Furthermore, the calibration curves for all three RCC biomarkers pass through the origin, which was not the case with methanol in water mobile phase.
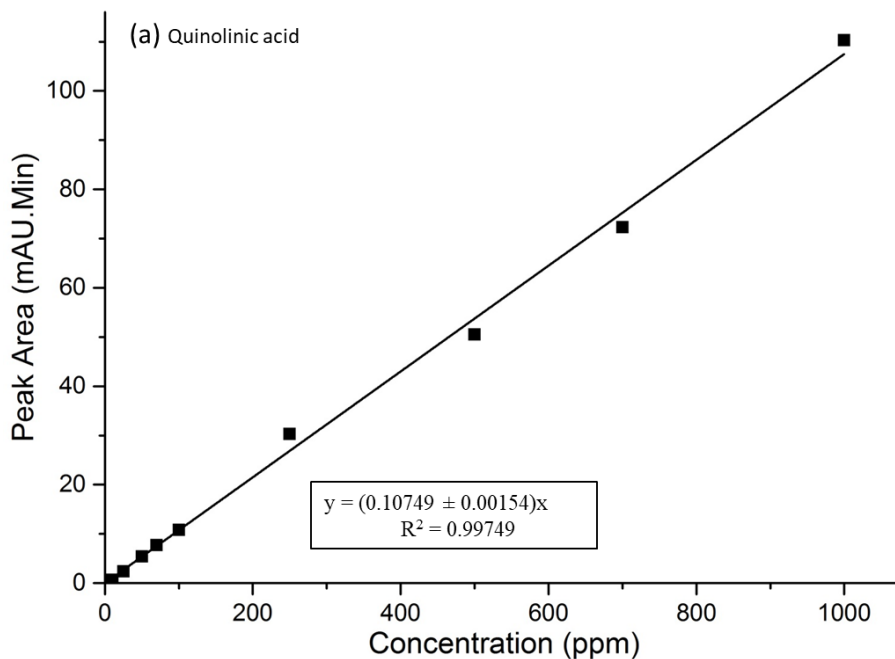

Figure 4.6. a) Calibration curve of quinolinic acid using 0.1% butanol with 0.6% acetic acid as the mobile phase.
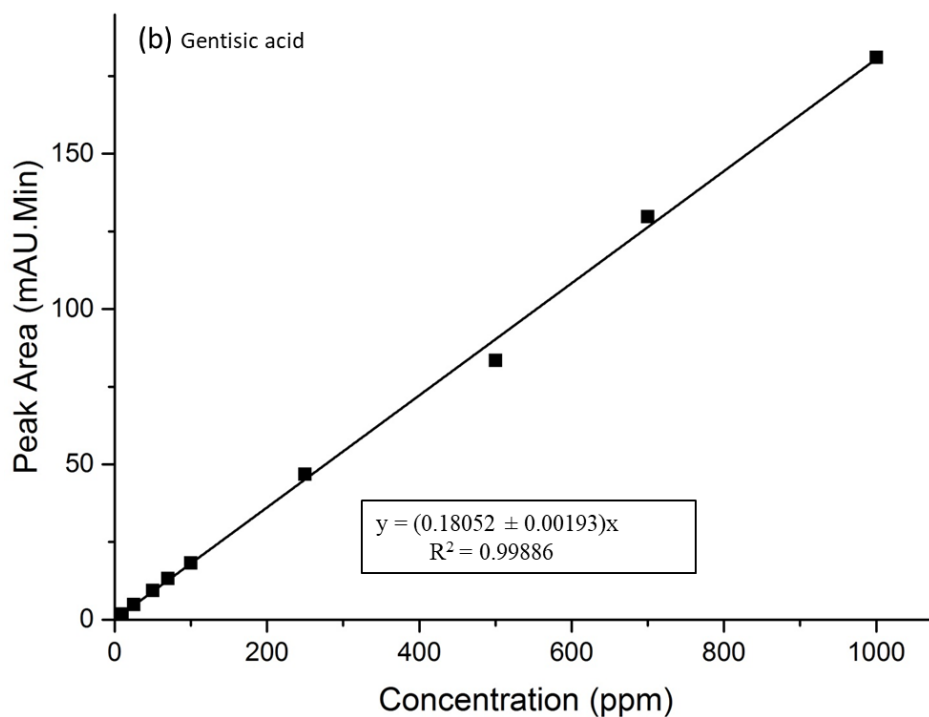
Figure 4.6 b) Calibration curve of gentisic acid using 0.1% butanol with 0.6% acetic acid as the mobile phase
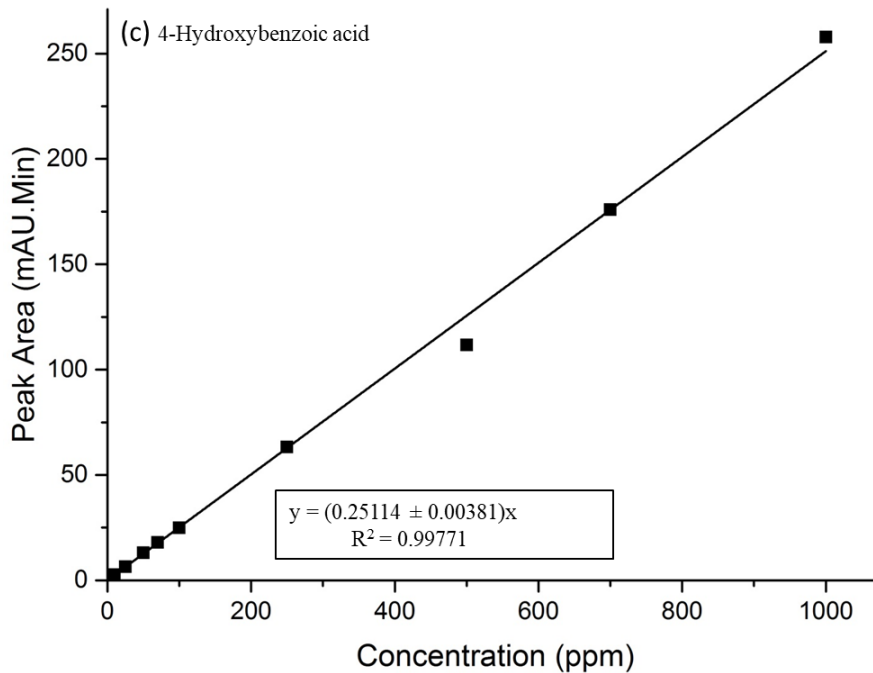

Figure 4.6 c) Calibraition curve of 4-hydroxybenzoic acid using 0.1% butanol with 0.6% acetic acid as the mobile phase

Three urine samples containing 20 ppm, 100 ppm, and 700 ppm of each test mixture component were prepared using synthetic urine to mimic the detection of the four RCC biomarkers by RPLC. Recovery test of these biomarkers as performed by using both 5% methanol in water with 0.6% acetic acid and 0.1% butanol in water with 0.6% acetic acid as the mobile phase. Each urine sample was run in triplicate, and the calibration curves (see Figure 4.5 a, b, c, d for methanol and Figure 4.6 a, b, and c for butanol) prepared from standards using doubly distilled water with 0.6% acetic acid, were used to quantify the amount of each biomarker in the three urine samples.

Table 4.4 lists the recovery and relative standard deviation obtained for each urine sample using 5% methanol. The relative standard deviations for these analyses generally decreased with increasing biomarker concentration. For all four biomarkers, the recovery was near 100% for the most concentrated urine sample. Both creatinine and 4-hydroxybenzoic acid were plagued by interference from the dead marker or the urine matrix, which would explain their higher reported recovery rates for the 20ppm urine sample. Quinolinic acid and gentisic acid, by comparison, showed the opposite trend – an increase in recovery rate with increasing concentration – this is a result of the dead marker or the urine matrix not interfering with the detection of these two compounds. The high relative standard deviation of quinolinic acid at 20 ppm indicates that the detection limit of this compound is being approached using the proposed RPLC method. For creatinine, quinolinic acid, and 4-hydroxybenzoic acid, it is evident that the background from the urine matrix or the dead marker will limit the detection of these compounds in urine.

Table 4.4. Relative Standard Deviation (RSD) and Recovery at Low, Medium and High Concentration of Spiked Urine for 5% Methanol.

Compounds	Low concentration (20 ppm)		Middle concentration (100 ppm)		High concentration (700 ppm)	
	RSD	Recovery	RSD	Recovery	RSD	Recovery
Creatinine	7.93 %	410 %	2.35 %	163 %	0.37 %	108 %
Quinolinic Acid	15.57 %	82.5 %	3.76 %	89 %	0.3 %	97 %
Gentisic Acid	4.37 %	90.7 %	4.27 %	92 %	8.08 %	97 %
[1]4-Hydroxybenzoic Acid	21 %	127 %	2.86 %	101 %	1.43%	105 %

[1]LC-MS analysis of the synthetic urine indicated that 4-hydroxybenzoic acid was a constituent. To correct for the background from the synthetic urine, we diluted the urine five-fold as was done for each sample in Table 4.4. Using the same isocratic HPLC method developed for the samples in Table 4.4, the peak area for 4-hydroxybenzoic acid in the synthetic urine was measured and then subtracted from the peak area for 4-hydroxybenzoic acid in the three urine samples.

Table 4.5 lists the recovery and relative standard deviation obtained for quinolinic acid, gentisic acid, and 4-hydroxybenzoic acid in each urine sample using 0.1% butanol water (0.6% acetic acid) mobile phase. Recovery tests using synthetic urine containing 20 ppm, 100 ppm, and 700 pm of each of these three RCC biomarkers were successful. Creatinine was not included for the recovery test as it is eluted with the dead marker. For 4-hydroxybenzoic acid, there was a contribution to its peak area from the synthetic urine matrix (see Figure 4.7). Based on the recovery tests, accurate detection of 4-hydroxybenzoic acid, gentisic and quinolinic acid was successful at medium and high biomarker concentrations for both 5% methanol and 0.1% butanol in water mobile phases with 0.6% acetic acid. Clearly, water-rich mobile phases (e.g., 0.1% butanol or 5% methanol) have advantages over more traditional hydro-organic solvent mixtures (e.g., 20% methanol in water with 0.6% acetic acid) for the analysis of water soluble and weakly retained compounds by liquid chromatography operated in the reversed phase mode.
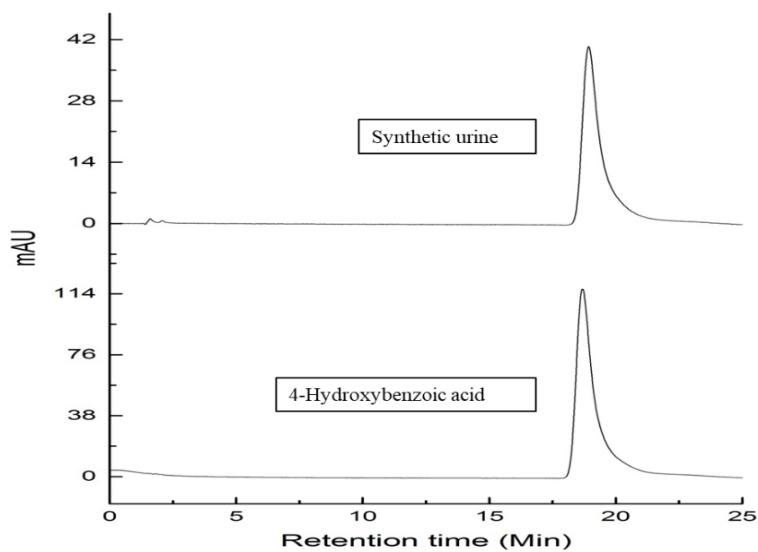
Figure 4.7 Chromatogram of synthetic urine (five-fold diluted with 0.6% acetic acid) and
4-hydroxybenzoic acid (0.6% acetic acid).

**Table 4.5. Relative Standard Deviation (RSD) and Recovery at Low, Medium and High
Concentration of Spiked Urine Using 0.1% Butanol**

	Low Concentration (20 ppm)		Medium Concentration (100 ppm)		High Concentration (700 ppm)	
Compounds	RSD	Recovery	RSD	Recovery	RSD	Recovery
Quinolinic acid	4.6 %	102 %	10.6 %	105%	4.8 %	109 %
Gentisic acid	16.4 %	88 %	10.3 %	105%	9.3 %	103 %
[1]4-Hydroxybenzoic acid	20.4 %	160 %	17 %	91 %	4.5 %	106%

[1]LC-MS analysis of the synthetic urine indicated that 4-hydroxybenzoic acid was a constituent.
To correct for the background from the synthetic urine, we diluted the urine five-fold as was done
for each sample in Table 4.5. Using the same isocratic HPLC method developed for the samples
in Table 4.5, the peak area for 4-hydroxybenzoic acid in the synthetic urine was measured and
then subtracted from the peak area for 4-hydroxybenzoic acid in the three samples listed in this
table.

References

[4-1]   Siegel, R. L.; Miller, K. D.; Jemal, A.; Cancer Statistics 2019, *Cancer J. Clin*., 2019, *69* (1), 7-34. DOI: 10.3322/caac.21551.

[4-2] European Network of Cancer Registries, Eurocim Version 4.0. European Incidence Database V2.3, 730 Entity Dictionary, 2001.

[4-3] A. Abbasi, T.V. Johnson, V.A. Master, "Minor Symptoms are Suggestive of Metastatic Renal Cell Carcinoma if C – Reactive Protein Remains High After Curative Nephrectomy," J Clin Oncol. 29(6) (2011) e133-134.

[4-4] Wang, C.; Yu, C.; Yang, F.; Wang, G.; Diagnostic Accuracy of Contrast-Enhanced Ultrasound for Renal Cell Carcinoma: A Meta-Analysis, *Tumor Biol*. 2014, *35* (7), 6343-6350. DOI: 10.1007/s13277-014-1815-2.

[4-5] Wagstaff, P. G.; Zondervan, P. J.; de la Rosette, J. J.; Laguna, M.P.; The Role of Imaging in the Active Surveillance of Small Renal Masses, Current Urology Reports 2014, *15* (3), 386. DOI: 10.1007/s11934-013-0386.

[4-6] Kim, K.; Taylor, S. L.; Ganti, S.; Guo, L.; Osier, M. V.; Weiss, R. H.; Urine Metabolomic Analysis Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer, *Omics: J. of Integ. Biol.* 2011, *15* (5), 293-303. DOI: 10.1089/omi.2010.0094.

[4-7] B.A. Gonzalez, M. Mahesh, K.P. Kim, M. Bhargavan, R. Lewis, F. Mettler, C. Land, "Projected Cancer Risks From Computed Tomographic Scans Performed in the United States in 2007,"Arch. Intern. Med., 169(22) (2009) 2071-2077.

[4-8] Kim, K.; Taylor, S. L.; Ganti, S.; Guo, L.; Osier, M. V.; Weiss, R. H.; Urine Metabolomic Analysis Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer, *Omics: J. of Integ. Biol.* 2011, *15* (5), 293-303. DOI: 10.1089/omi.2010.0094.

[4-9] Chen, S.; Burton, C.; Kaczmarek, A.; Shi, H.; Ma, Y.; Simultaneous Determination of Urinary Quinolinate, Gentisate, 4-Hydroxybenzoate, and α-Ketoglutarate by High-Performance Liquid Chromatography-Tandem Mass Spectrometry, *Anal. Methods*, 2015, *7* (16), 6572-6578. DOI: 10.1039/C5AY01643F.

[4-10] Lavine, B. K.; Ritter, J. P.; Peterson, S.; Enhancement of Selectivity in Reversed Phase Liquid Chromatography, *J. Chromatogr. A*. 2002, *946*, 83-90. DOI: 10.1016/s0021-9673(01)01520-5.

[4-11] Lavine, B. K.; Corona, D.T.; Perera, D.N.T.; Analysis of Vanilla Extract by Reversed Phase Liquid Chromatography Using Water Rich Mobile Phases, *Microchem. J*., 2012, *103*, 49-61. DOI: 10.1016/j.microc.2012.01.004.

[4-12] Ding, T.; Lavine, B. K.; Separation of Imidacloprid and its Degradation Products Using Reversed Phase Liquid Chromatography with Water Rich Mobile Phases, *J. Chromatogr. A*., 2011, *1218*, 9221– 9226. DOI: 10.1016/j.chroma.2011.10.057.

[4-13] Gamagedara, S.; Gibbons, S.; Ma, Y.; Investigation of Urinary Pteridine Levels as Potential Biomarkers for Noninvasive Diagnosis of Cancer, *Clin. Chim. Acta.* 2011, *412* (1-2) 120-128. DOI: 10.1016/j.cca.2010.09.015

[4-14] Gamagedara, S.; Shi, H.; Ma, Y.; Quantitative Determination of Taurine and Related Biomarkers in Urine by Liquid Chromatography-Tandem Mass Spectrometry, *Anal, Bioanal. Chem.* 2012, *402* (2) 763- 770. DOI: 10.1007/s00216-011-5491-4.

[4-15] Foley, J. P.; and Dorsey, J. D.; Equation for Calculation of Chromatographic Figures of Merit for Ideal and Skewed Peaks, *Anal. Chem.*, 1983, *55*, 730-737. DOI: 10.1021/ac00255a033
.

[4-16] Valkó. K.; Synder, L.R.; Glajch, J.L.; Retention in Reverse-Phase Liquid Chromatography as a Function of Mobile-Phase Composition, *J. Chromatogr. A*, 1993, *656* (1-2), 501-520. DOI:10.1016/0021-9673(93)80816-Q

[4-17] Schoenmakers, P. J.; Billet, H. A. H.; Tussen,R.; and De Galan, L.; Gradient Selection in Reversed Phase Liquid Chromatography, *J. Chromatogr.* 1978, *149*, 519-537. DOI: 10.1016/S0021-9673(00)81008-0.

[4-18] Schoenmakers, P. J.; Billet, H. A. H.; Tussen,R.; and De Galan, L.; Description of Solute Retention Over the Full Range of Mobile Phase Compositions in Reversed Phase Liquid Chromatography, *J. Chromatogr.* 1983, *282*, 107-121. DOI: 10.1016/S0021-9673(00)91596-6

[4-19] Walter, T. H.; Iraneta, P.; Capparella, M.; Mechanism of Retention Loss when C8 amd C18 Columns are Used with Highly Aqueous Mobile Phases, *J. Chromatogr*., 2005, *1075*, 177-183. DOI: 10.1016/j.chroma.2005.04.039.

[4-20] Bidlingmeyer, B. A.; Broske, A. D.; The Role of Pore Size and Stationary Phase Composition in Preventing Aqueous-Induced Retention Time Loss in Reversed Phase HPLC, *J. Chromatogr. Sci*., 2004, *41*, 100-106.

Chapter V

Conclusion

Thirty-eight automotive paint samples from six manufacturers (General Motors, Chrysler, Ford, Toyota, Nissan, and Honda) in a limited production year range (2000 to 2006) were obtained from the Royal Canadian Mounted Police. Each paint chip was placed between two small polyethylene plastic slabs, positioned horizontally in the sample holder of a microtome and cross sectioned using high profile microtome blades which ensured the generation of high quality thin sections representative of the four paint layers and a smooth paint surface crucial for effective surface contact with the internal reflection element of the ATR. Infrared imaging of the cross-sectioned paint chip was challenging due to difficulties encountered in the alignment of the chip on the germanium crystal. However, this problem was addressed by aligning the thin paint sample with the midpoint of the germanium ATR crystal using a curved tweezer to support the sample holder, a small needle (with a bent tip) to maneuver the thin paint sample and a video camera to guide the placement of the chip. This ensured complete coverage of the crystal by the sample when a pressure of 10-15 psi was applied. The noise level of IR spectra for thin samples positioned at the midpoint of the crystal was low due to the higher intensity of the IR beam.

For ATR infrared imaging, the single MCT detector, which performs a point by point mapping, was selected for ATR analysis. The 25 micron x 25 micron aperture and 10 micron step

size yielded the best results. The amount of time required for an analysis is approximately 1 hour because two line maps must be generated: one for the sample and the other for the background. After the line map for a sample has been collected, it is subject to decatenation using alternating least squares to obtain a "pure" IR spectrum of each layer. This approach to automotive paint analysis, not only eliminates the need to analyze each layer separately, but also ensures that the final spectrum of each layer is "pure" and not a mixture of adjacent layers.

The reconstructed IR spectra of the clear coat, color coat, surfacer-primer, and e-coat layers in each paint chip were searched against a forensic automotive paint library to identify the "make" and model of the vehicle from which the paint chip originated using OMNIC or a pattern recognition assisted infrared library search system. Twenty-three (clear coat layer) and twenty-four (surfacer-primer layer) of the thirty-eight paint samples were correctly identified by OMNIC as to the "make" and model of the vehicle. For the pattern recognition assisted infrared library search system, twenty-six of the thirty-eight paint samples were correctly identified as to manufacturer and nineteen of the thirty-eight were correctly identified as to the model of the vehicle.   Paint samples that were not correctly matched by OMNIC or the pattern recognition assisted infrared library search system exhibited peak shifting in all regions of the IR spectra of the surfacer primer and e-coat layers compared to the ATR corrected spectra of the same paint samples and layers that were obtained using a high pressure diamond cell. This peak shifting can be attributed to the positioning of the cross sectioned paint sample on the germanium crystal.   If all layers of the cross sectioned paint sample do not lie in the mid-point region of the germanium crystal, the effective angle of incidence will differ for those layers that are distant from the mid-point compared to those layers that lie directly on the mid-point. Due to the greater thickness of the clear coat layer (approximately 50 μm) compared to the surfacer-primer (25 μm) and e-coat

(25 μm) layers, it is not surprising that differences in the internal reflection angle of the IR beam for the surfacer-primer and e-coat layers can occur for misaligned cross sectioned paint samples. The misalignment probably occurred as a result of the displacement of the sample from the mid-point region of the crystal after the sample is raised to make contact with the crystal.

Quinolinic acid, 4-hydroxybenzoic acid, and gentisic acid in urine have been implicated as potential markers to detect renal cell carcinoma, which is the third most reported form of genitourinary cancer. Ultrahigh-performance liquid chromatography/tandem mass spectrometric and gas chromatography/mass spectrometric methods have been developed by previous workers to detect and quantity these three compounds, which are believed to exist in urine at elevated levels for patients in the early stages of renal cell carcinoma. The use of water rich mobile phases in RPLC, e.g., 0.1% butanol in water with 0.6% acetic acid, and a Zorbax C_{18} column has been investigated as a potential method to detect and quantify these three biomarkers in urine. Baseline resolution is achieved in the separation of these three compounds and creatinine when hydrophobic alcohols such as butanol are used as organic modifiers. This can be attributed to butanol partitioning into the bonded phase and providing a more extended surface with the net result of increasing the contact surface area of the stationary phase and the selectivity of the separation. Recovery tests for quinolinic acid, gentisic acid, and 4-hydroxybenzoic acid using synthetic urine at low, medium, and high concentrations were successful except for 4-hydroxybenzoic acid at low concentration due to the presence of 4-hydroxybenzoic acid in the synthetic urine matrix.

VITA

Kaushalya Sharma Dahal

Candidate for the degree of

Doctor of Philosophy

Thesis: INFRARED MICROSCOPY AND LIQUID CHROMATOGRAPHY
           APPLIED TO PROBLEMS IN FORENSICS AND
           BIOANALYTICAL CHEMISTRY

Major Field: Chemistry

Biographical:

   Education:

   Completed the requirements for the degree of Doctor of Philosophy in Chemistry
   at Oklahoma State University, Stillwater, Oklahoma in July, 2019.

   Completed the requirements for the Master of Science in Physical Chemistry at
   Tribhuvan University, Kirtipur, Nepal in 2007

   Completed the requirements for the Bachelor of Science in Chemistry major and
   Biology minor at Tribhuvan University, Biratnagar, Nepal in 2005

   Experience:

   Teaching / Research Assistant, Department of Chemistry, Oklahoma State
   University (Aug 2013 – July 2019), Chemistry Teacher, Araniko Higher
   Secondary School (Aug 2008- July 2010), Biratnagar, Morang Nepal, Chemistry
   Instructor, COBASS College (Aug 2010 – July 2013), Biratnagar, Morang, Nepal