2007

ANISOTROPIC POLARIZED LIGHT SCATTER AND MOLECULAR FACTOR COMPUTING IN PHARMACEUTICAL CLEANING VALIDATION AND BIOMEDICAL SPECTROSCOPY

Aaron Andrew Urbas
University of Kentucky, aaurba2@uky.edu

ABSTRACT OF DISSERTATION

Aaron Andrew Urbas

The Graduate School

University of Kentucky

2007

ANISOTROPIC POLARIZED LIGHT SCATTER AND MOLECULAR FACTOR COMPUTING IN PHARMACEUTICAL CLEANING VALIDATION AND BIOMEDICAL SPECTROSCOPY

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry at the University of Kentucky

By
Aaron Andrew Urbas

Lexington, Kentucky

Director: Dr. Robert A. Lodder, Professor of Chemistry

Lexington, Kentucky

2007

ABSTRACT OF DISSERTATION

ANISOTROPIC POLARIZED LIGHT SCATTER AND MOLECULAR FACTOR COMPUTING IN PHARMACEUTICAL CLEANING VALIDATION AND BIOMEDICAL SPECTROSCOPY

Spectroscopy and other optical methods can often be employed with limited or no sample preparation, making them well suited for *in situ* and *in vivo* analysis. This dissertation focuses on the use of a near-infrared spectroscopy (NIRS) and polarized light scatter for two such applications: the assessment of cardiovascular disease, and the validation of cleaning processes for pharmaceutical equipment.

There is a need for more effective *in vivo* techniques for assessing intravascular disorders, such as aortic aneurysms and vulnerable atherosclerotic plaques. These, and other cardiovascular disorders, are often associated with structural remodeling of vascular walls. NIRS has previously been demonstrated as an effective technique for the analysis of intact biological samples. In this research, traditional NIRS is used in the analysis of aortic tissue samples from a murine knockout model that develops abdominal aortic aneurysms (AAAs) following infusion of angiotensin II. Effective application of NIRS *in vivo*, however, requires a departure from traditional instrumental principles. Toward this end, the groundwork for a fiber optic-based catheter system employing a novel optical encoding technique, termed molecular factor computing (MFC), was developed for differentiating cholesterol, collagen and elastin through intervening red blood cell solutions. In MFC, the transmission spectra of chemical compounds are used to collect measurements directly correlated to the desired sample information.

Pharmaceutical cleaning validation is another field that can greatly benefit from novel analytical methods. Conventionally cleaning validation is accomplished through surface residue sampling followed by analysis using a traditional analytical method. Drawbacks to this approach include cost, analysis time, and uncertainties associated with the sampling and extraction methods. This research explores the development of *in situ* cleaning validation methods to eliminate these issues. The use of light scatter and polarization was investigated for the detection and quantification of surface residues. Although effective, the ability to discriminate between residues was not established with these techniques. With that aim in mind, the differentiation of surface residues using NIRS and MFC was also investigated.

KEYWORDS: near-infrared, biomedical spectroscopy, process analytical technology, polarized light scatter, cleaning validation

Aaron Urbas

March 15, 2007

ANISOTROPIC POLARIZED LIGHT SCATTER AND MOLECULAR FACTOR COMPUTING IN PHARMACEUTICAL CLEANING VALIDATION AND BIOMEDICAL SPECTROSCOPY

By

Aaron Andrew Urbas

Robert A Lodder

_____

Director of Dissertation

Robert B. Grossman

_____

Director of Graduate Studies

March 15, 2007

_____

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name	Date

DISSERTATION

Aaron Andrew Urbas

The Graduate School

University of Kentucky

2007

ANISOTROPIC POLARIZED LIGHT SCATTER AND MOLECULAR FACTOR COMPUTING IN PHARMACEUTICAL CLEANING VALIDATION AND BIOMEDICAL SPECTROSCOPY

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry at the University of Kentucky

By
Aaron Andrew Urbas

Lexington, Kentucky

Director: Dr. Robert A. Lodder, Professor of Chemistry

Lexington, Kentucky

2007

Acknowledgements

This work would not have been possible without the encouragement, guidance, support, and generosity of so many people. I would like to thank my advisor for his time and commitment. I am grateful for the freedom he extended in my scientific pursuits and for having confidence in my abilities when it was lost to me. I also would like to thank my doctoral committee for their time and helpful comments. I am greatly indebted to all my friends, whose kindness and support have been invaluable the past few years. My deepest thanks go to my family, who has always been there when I needed them. I would like to specially thank my mother and father. Without their enduring support and encouragement in this endeavor and throughout my life, none of this would have been possible.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

AAA	Abdominal Aortic Aneurysm
API	Active pharmaceutical ingredient
BEST	Bootstrap error-adjusted single sample technique
BSA	Bovine serum albumin
CV	Canonical variable
FDA	Food and Drug Administration
FN	False negative
FP	False positive
GMP	Good manufacturing practice
HDL	High density lipoprotein
HICI	Hyperspectral integrated computational imaging
HPLC	High performance liquid chromatography
ICS	International Chemometrics Society
IMS	Ion mobility spectrometry
IR	Infrared
ISP	Integrated sensing and processing
IVUS	Intravascular ultrasound
LDA	Linear discriminant analysis
LDL	Low density lipoprotein
LOOCV	Leave one out cross validation
MAESA	Multi-functional active excitation spectral analyzer
MFC	Molecular factor computing
MLR	Multiple linear regression
MOE	Multivariate optical element
MSC	Multiplicative scatter correction
MSD	Multi-dimensional standard deviation
NIPALS	Nonlinear iterative partial least squares
NIR	Near infrared
NIRS	Near infrared spectroscopy
OCT	Optical coherence tomography
OSP	Optical signal processing
PAC	Process analytical chemistry
PAT	Process analytical technology
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PLS	Partial least squares
PLS-DA	Partial least squares discriminant analysis
RBC	Red blood cell
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of performance
RSD	Relative standard deviation
SD	Standard deviation
SEC	Standard error of calibration

SEE	Standard error of estimation
SEM	Scanning electron microscopy
SEP	Standard error of performance
S/N	Signal / noise
SNV	Standard normal variate
SVD	Singular value decomposition
TN	True negative
TOC	Total organic carbon
TP	True positive
UV	Ultraviolet

Preface

Motivation

The development of novel analytical technologies that are rapid and robust is of interest to analysts across a number of different fields. In many cases the direct analysis of samples without prior sample preparation or even transportation to the laboratory bench is highly desirable. The continuous qualitative or quantitative monitoring of materials on a production line represents a prime example. In other cases, such as the *in vivo* analysis of vascular disease pathologies, it is absolutely essential. These situations require innovative approaches to instrumental development. At the same time, there are practical considerations such as instrumental complexity and cost effectiveness that constrain prospective solutions. This dissertation aims to demonstrate that relatively simple instrumentation coupled with or motivated by chemometric data analysis techniques can be successfully applied to two distinct problems of this nature. Spectroscopy and other optical techniques are ideally suited for *in situ* and *in vivo* applications, and the methods used to approach these problems are encompassed in this sphere.

There is tremendous need for novel instrumental techniques in the field of biomedical spectroscopy and imaging. Developing effective techniques for the early detection of potentially lethal vascular conditions is of critical importance. One objective of this work was to demonstrate the potential utility of near-infrared spectroscopy (NIRS) for analyzing vascular tissue. To this end, the use of NIRS was explored for the analysis of excised aortic tissues from a murine knockout model for abdominal aortic aneurysm, which builds upon an established foundation of NIRS for biomedical applications. The early detection of vulnerable atherosclerotic plaques is another area of critical importance where extensive research has been done using an array of modalities but no reliable means have yet been developed. The potential

effectiveness of near-infrared spectroscopy (NIRS) for this purpose has been previously investigated but *in vivo* application still remains to be demonstrated. There are significant limitations to the application of traditional instrumentation for this purpose. The necessity to employ small fiber optics limits the optical throughput. Of greater importance is the problem associated with data collection in a flowing blood stream environment. The unstable positioning of the optical delivery and detection device (i.e., catheter tip) greatly hinders the capacity to collect useful spectroscopic data.

A potential solution to these limitations can be found in the area of integrated sensing and processing (ISP). The aim of ISP is physical replication of multivariate data processing methods at the transducer level of instruments. Potential advantages of this methodology include the reduction of computational and data storage burdens, significantly reduced data collection times, and rugged instrumentation. In this work an ISP approach called molecular factor computing (MFC) was investigated as the basis for a NIRS-based catheter system for diagnosing vulnerable plaques. In MFC, the transmission spectra of chemical compounds are used to collect measurements directly correlated to the desired sample information. MFC has the potential to overcome the limitations described above in the application of NIRS for *in vivo* analysis. The utility of the very same methodology can be seen in pharmaceutical applications as well.

The field of process analytical technology (PAT) has experienced a surge in growth within the pharmaceutical industry as a result of new Food and Drug Administration regulations and initiatives. The primary goal in PAT is the incorporation of quality control throughout the entire manufacturing process to establish a "quality by design" paradigm. This goal is being realized by the application of at-line and in-line sensors at critical points in the manufacturing process capable of providing real-time qualitative/quantitative assessments during production.

The ideal PAT tools sought for these applications are simple, rugged, cost effective instruments. The same drawbacks with traditional instrumentation as described previously are also relevant in this environment, namely large data volumes, time consuming analysis and the need for rugged instrumentation suitable for a process environment. One area in pharmaceutical manufacturing where PAT solutions are being sought is in the validation of cleaning procedures. This is of critical importance for multi-product facilities to ensure that potentially dangerous cross contamination of active pharmaceuticals is avoided. The desire for rapid *in situ* methods for detecting surface residues is great since conventional approaches are time consuming and often result in lengthy downtimes for processing equipment. Toward this goal the use of light scatter and polarization was investigated for the detection and quantification of surface residues on glass and stainless steel surfaces. In these studies, the utilization of chemometric techniques demonstrated that effectiveness could be achieved with simple instrumentation. However, the ability to discriminate between residues was not established with these techniques. To address this issue, the ISP approach of MFC in the NIR spectral region was investigated for discriminating surface residues.

Content Summary

The organization of this dissertation was chosen to present the material in a fluid manner. As described above, the research in this dissertation was devoted to two distinct applications. There are three core sections representing the bulk of this document devoted to background material and each of the applications. An inclusive introductory chapter, covering all the material, was avoided for continuity purposes. There are, however, several parallel themes among the methods used in these applications: NIR spectroscopy, chemometrics, and MFC. The first

chapter and section in this dissertation serves as in introduction to these three subjects. The fundamental principles of NIR spectroscopy are covered first. Next, a presentation of the chemometric methods used throughout the work is given. Finally, a discussion of MFC concludes chapter 1. The next two sections of the dissertation are each devoted to one of the applications: biomedical spectroscopy and pharmaceutical cleaning validation. Each of these sections is prefaced by an introduction to the application followed by chapters covering the research pertaining to those applications.

Section 2 is devoted to biomedical spectroscopy, and the application of NIR spectroscopy as a potentially useful tool in the analysis of intravascular disorders is the focus. Chapter 2 presents research on the use of traditional NIR in the analysis of aortic tissue sections of a murine knockout model of abdominal aortic aneurysms. Presented in chapter 3 is preliminary work toward the development of a novel NIR catheter system based on MFC for *in vivo* diagnosis of vulnerable atherosclerotic plaques. Specifically, the differentiation of the plaque components collagen, elastin and cholesterol through red blood cell solutions is presented.

Section 3 is devoted to pharmaceutical cleaning validation, and the development of rapid methods for *in situ* determination of surface residues. Chapter 4 presents research on the use of laser light scatter to quantify protein residues on glass surfaces. Chapter 5 explores an extension of this work by examining the polarization of scattered light from a broadband source for the detection and quantification of surface residues on stainless steel surfaces. Finally, in chapter 6 the application of NIR spectroscopy and MFC for detecting residues on glass and stainless steel surfaces is presented.

Section I: Background

Chapter One – Near-Infrared Spectroscopy, Chemometrics and Molecular Factor Computing

Near-Infrared Spectroscopy

The discovery of the near-infrared (NIR) region of the electromagnetic spectrum can be traced to the astronomer William Herschel in 1800. However, its practical application was not realized for more than one and half centuries until the pioneering work of Karl Norris in the 1960's.[1, 2] Since then, with the introduction of novel instrumentation and data processing techniques (chemometrics), NIR spectroscopy has rapidly grown as a valuable analytical method for both qualitative and quantitative applications in a diverse range of fields that includes agriculture, food processing, pharmaceutics, fuel, remote spectral imaging and biomedical spectroscopy.[3, 4] The increasing popularity of NIR in these and other fields is in part a result of several advantages NIR spectroscopy offers over traditional analytical methods. These include: limited or no sample preparation/pretreatment; simultaneous multi-component determination; application in a wide variety of sample matrices; non-destructive analysis; penetration depth of radiation into samples; the ability to determine chemical as well as physical properties; and rugged instrumentation that can be implemented in-line. The following is a brief discussion of the principles of NIR spectroscopy and the reader is referred to several excellent sources for thorough treatments.[3-5]

Absorption bands in the NIR spectral region are dominated by harmonics and combinations of the fundamental vibrations of -OH, -CH, -NH, and –SH functional groups. The energy band from 780 – 2500 nm (12820 – 4000 cm^{-1}) is generally described as the NIR band,

but harmonics and combinations of the abovementioned functional groups are observed from approximately 700 to beyond 3000 nm. The fundamental occurrence and characteristics of NIR absorption bands can be understood by considering anharmonicity in molecular vibrations. For simple diatomic molecules in the gas phase, the fundamental molecular vibrations of can be determined using Hooke's Law (Equation 1.1):

$$\nu = \frac{1}{2\pi c}\sqrt{\frac{k\left(m_1 + m_2\right)}{m_1 m_2}} \qquad \textbf{1.1}$$

where ν is the vibrational frequency, c is the speed of light, k is a force constant, and m_1 and m_2 are the two masses. The simple harmonic oscillator model is a useful first approximation in understanding molecular vibrations. The potential energy as a function of atom displacement for a classical model (i.e., balls on a spring) can be approximated according to Equation 1.2:

$$V_r = \frac{1}{2}kr^2 \qquad \textbf{1.2}$$

where V_r is the potential energy, k is a force constant, and r is displacement. This equation describes the continuous curve of the harmonic potential in Figure 1.1 (A) as function of atom displacement from the equilibrium position (r_e). From quantum mechanics it is known that molecular vibrations do not assume a continuous energy profile but instead can only occupy discrete energy levels. For the simple quantum harmonic oscillator, these energy levels are given by Equation 1.3:

$$E_\upsilon = \left(\upsilon + \frac{1}{2}\right)h\nu \qquad \textbf{1.3}$$

where E_n is the energy associated with the υ^{th} quantum level, υ is the vibrational quantum number, h is Planck's constant, and ν is the fundamental vibrational frequency. This model implies that the energy difference between any two adjacent vibrational levels is always the same

as demonstrated with the discrete energy levels of the harmonic oscillator in Figure 1.1 (A). Furthermore, transitions between non-adjacent energy levels (i.e., $\Delta\upsilon = \pm 2,3...$) are forbidden in this model. Unfortunately the quantum harmonic oscillator model is not sufficient for real molecular vibrations and the determination of absorption bands in the NIR region is far from straightforward.

A more realistic approximation to describe molecular vibrations is based on the anharmonic oscillator. The potential energy of molecular vibrations in diatomic molecules can be described by the Morse function (Equation 1.4):

$$V_r = D_e\left(1 - e^{-a(r-r_e)}\right)^2 \qquad \textbf{1.4}$$

where V_r is the potential energy, D_e is the dissociation energy of the bond, r is the distance between atoms, r_e is the equilibrium bond distance, and a is a constant associated with the molecule. This equation describes the continuous curve of the anharmonic potential in Figure 1.1 (B) as a function of atom displacement from the equilibrium position (r_e). Again, this curve is representative of a classical balls-on-a-spring model. A quantum mechanics model of the Morse equation gives vibrational energy levels described by Equation 1.5:

$$E_\upsilon = h\nu\left(\upsilon + \frac{1}{2}\right) - \chi_m h\nu\left(\upsilon + \frac{1}{2}\right)^2 \qquad \textbf{1.5}$$

where χ_m is the anharmonicity constant of the vibration and the other terms are as described above. The discrete energy levels derived from this equation are shown for the anharmonic oscillator in Figure 1.1 (B). This model of molecular vibrations is much more realistic than the harmonic oscillator and actually agrees remarkably well with the fundamental vibrations of non-rotating diatomic molecules. Its utility is limited for more complicated molecules except as a conceptual tool. Several aspects of the quantum anharmonic oscillator model are of particular

significance, though. In contrast to the harmonic oscillator, the energy difference between adjacent vibrational levels is not constant and decreases with increasing υ. In addition, the model predicts transitions of $\Delta\upsilon = \pm 2$ and higher and it is these multi-level energy transitions that are the origin of NIR overtone bands. Finally, under the assumptions of the anharmonic model, vibrational modes are no long independent and can interact with each giving rise to combination bands, which are the other dominant feature found in the NIR region.

There are a number of additional factors that influence the occurrence and properties of NIR absorption bands. Other interactions such as intermolecular hydrogen bonding and dipole interactions alter the vibrational energy states, which result in the shifting of (and in some cases the generation of) NIR absorption bands. In addition, Fermi resonance plays an important role in combination bands with fundamental vibrations of very similar energies. This causes a NIR band to split into two bands of higher and lower energy instead of simply occurring at the sum of the two frequencies. The result of this complex array of interactions is characteristically broad absorption bands with frequencies occurring at imprecise multiples of the fundamentals. Figure 1.2 shows the typical locations of absorption bands from overtones and combinations of the most common functional groups that are active in the NIR spectral region. The probabilities for transitions corresponding to overtone and combination bands are significantly smaller than fundamental vibrations in the mid-IR and absorption bands are typically 10 – 1000 times weaker in comparison. This imposes considerable sensitivity limitations to NIR spectroscopy from the view of classical spectroscopy. However, this aspect of NIR is one of the reasons that make it a popular and useful analytical method, permitting the direct analysis of solids and liquids, which can be highly absorbing or scattering, with little or no sample preparation using transmission or reflectance measurements. Even for simple systems, though, NIR spectra are characterized by

broad, highly overlapping bands and classical spectroscopic analysis is very difficult and often impossible for these data sets. As a result, the practical application of NIR spectroscopy necessitates the use of multivariate statistical and mathematical modeling techniques, which is the focus of the field of chemometrics, the topic of the following section.

Chemometrics

There is no strict description of the field of chemometrics, but the International Chemometrics Society's (ICS) definition is reasonably comprehensive: "Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods." A vast array of techniques is encompassed by chemometrics for a diverse range of applications including, but not limited to, experimental design, optimization, calibration (regression), signal processing and pattern recognition. In this section, only the specific chemometric techniques used in this work will be covered. These techniques include several data preprocessing techniques and the following data analysis methods: multiple linear regression (MLR), principal component analysis (PCA), principal component regression (PCR), partial least squares regression (PLS), and linear discriminant analysis (LDA). Finally, several performance metrics for evaluating calibration and classification models will be presented. The discussions are more conceptually focused and only concise mathematical formulations of the algorithms are presented. When convenient, the application of these methods to spectroscopic data sets is introduced for clarification. The reader is referred to several excellent sources for rigorous presentations of these and other chemometric methods.[6-12]

Preprocessing Methods. As the name suggests, data preprocessing involves the manipulation of a data set prior to the application of other data processing procedures, the results of which are the principal goal. Many data preprocessing methods exist with a number of different objectives that include data cleaning (e.g., noise removal, outlier detection), normalization or scaling, transformations (e.g., estimating derivatives), and data reduction. Several preprocessing methods were used in this work and a brief discussion of them will be presented here.

Mean Centering and Autoscaling. Mean centering is one of the simplest preprocessing techniques and simply entails subtracting the mean value of each variable from the entire data set. Mean centering removes the constant background component from the data while leaving the underlying variation in the data intact. One result in relation to multiple regression models is that regression coefficients are easier to interpret and the intercept term is removed in models derived from mean centered data. Autoscaling (also known as z-scoring or standardizing) is a similar in that the variables are also mean centered, but, in addition, the variables are scaled to unit variance. Autoscaling is occasionally superior to mean centering when looking at a spectroscopic data set from a single instrument, but is often essential when integrating data from different methods with dramatically different variable ranges. Two of the statistical methods described below, PCA and PLS, require that the data be preprocessed by mean centering or autoscaling.

Multiplicative Scatter Correction (MSC). MSC is a preprocessing method developed to correct for spectral variations arising from path length and particle size differences by removing both additive and multiplicative effects from the data. This technique is widely applied in NIR spectroscopy.[13, 14] Due to the fact that NIR is often used to analyze samples (including solids and turbid liquids) with little or no pretreatment, path length and particle size differences are

common. Scatter correction is carried out with respect to the mean spectrum across the entire data set. For each spectrum, a least squares regression is performed based on Equation 1.6:

$$x_i = a + b\bar{x}_i \qquad \textbf{1.6}$$

where x_i is the i^{th} wavelength of the spectrum being corrected, $\bar{x}_i$ is the i^{th} wavelength of the mean spectrum, and a and b are regression coefficients representing the additive and multiplicative components, respectively. A single set of regression coefficients is computed over the entire spectrum, and the scatter correction is accomplished according to Equation 1.7:

$$x_{i,MSC} = \frac{(x_i - b)}{a} \qquad \textbf{1.7}$$

where $x_{i,MSC}$ is the scatter corrected value at the i^{th} wavelength. For an illustrative demonstration, a collection of NIR spectra from geological samples before and after MSC are presented in Figure 1.3.

Cubic Splines for Smoothing and Derivative Approximation. Splines are a popular mathematical tool for interpolation capable of modeling arbitrary functions by means of piecewise polynomials.[15, 16] Functions are subdivided into sufficiently small intervals so that relatively low degree polynomials provide good approximations. In addition, the polynomials are defined in such a way that they blend together smoothly, and the composite function (i.e., a spline) has several continuous derivatives. A specific type of spline, known as the cubic smoothing spline, was used in this work for the purposes of noise removal and derivative approximation. For a discrete valued function $y = f(x)$, the cubic smoothing spline (S) is defined to minimize W according to Equation 1.8:

$$W = \rho \sum_{i=1}^{n} \frac{[S(x_i) - y_i]^2}{\sigma_i^2} + (1-\rho)\int |S''(x)|^2 dx \qquad \textbf{1.8}$$

where n is the number of discrete data points, σ_i is the uncertainty associated with y_i, $S''(x)$ denotes the second derivative of S, and ρ is a "smoothing" parameter with a range of 0 to 1. The first term in Equation 1.8 is analogous to a squared error term (weighted by the uncertainty) as encountered in least squares fitting. The second term in Equation 1.8 is an integral over the square of the second derivative, which essentially quantifies the curvature of the spline. The two terms are contradictory and the tradeoff is determined by the smoothing parameter ρ. As ρ decreases the spline increasingly diverges from the data values and becomes smoother. Note that at the $\rho = 0$ limit, the first term drops out and the second term is minimized when the derivative is 0 across the range of x, which is a straight line. Presented in Figure 1.4 are examples of smoothed and second derivative spectra from a cubic smoothing spline fit to a synthetic spectrum of overlapping Gaussians with white noise added.

Multiple Linear Regression (MLR). MLR (also known as inverse least squares) is a straightforward extension of simple univariate linear regression and is used to generate a quantitative relationship between a group of predictor variables and a response:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon \qquad \textbf{1.9}$$

where y is the response, x_i are predictors, β_i are regression coefficients, and ε is the residual.[6, 9, 11] Note that additional terms, such as powers (x_i^k) or cross-terms ($x_i x_j$), can be included and the model remains linear even though the function may not be a straight line. For a data set of known responses and predictors, the linear model can be expressed in matrix form according to Equation 1.10:

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{\varepsilon} \qquad \textbf{1.10}$$

where $\mathbf{y}$ (m×1) is the column vector of responses, $\mathbf{X}$ (m×n) is the matrix of predictors, $\mathbf{b}$ (n×1) is the unknown column vector of regression coefficients, and $\boldsymbol{\varepsilon}$ (m×1) is the column vector of

residuals. For a unique solution to exist, m must be larger than n. Equation 1.11 provides the least squares estimate of the regression coefficients ($\hat{\mathbf{b}}$):

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \qquad \textbf{1.11}$$

where $\mathbf{X}^{\mathbf{T}}$ denotes the transpose of $\mathbf{X}$, and $\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}$ denotes the inverse of $\mathbf{X}^{\mathbf{T}}\mathbf{X}$. This solution is only attainable when $\mathbf{X}$ is of full rank. The equation is useful theoretically but in practice has poor numerical properties and more robust methods are generally used.

For a typical spectroscopic data set we are often confronted with the situation of having significantly more variables than samples ($m<n$). One solution to this dilemma is the selection of a subset of variables to constrain n to be smaller than m. This approach is used frequently in practice but is not without drawbacks. In spectroscopic data, for instance, the absorbance values in a spectrum at multiple wavelengths tend to vary together with changing constituent concentration. This effect is known as collinearity, and causes instability in the mathematical solution. Another drawback is that the removal of variables from the model discards potentially useful information. Finally, determination of an optimal subset of variables, especially when thousands may be available, presents considerable difficulties. An alternative to the problem of too many variables involves factor-based approaches for dimensionality reduction. One such approach is principal component analysis, a discussion of which follows.

Principal Component Analysis (PCA). PCA is a mathematical procedure involving a linear transformation of multivariate data from one coordinate system into another and is typically used as a dimensionality reduction tool for preprocessing and exploratory data analysis.[6, 9-11, 17] The traditional method of calculating the PCs involves the eigenvalue decomposition of the covariance matrix of the data matrix. A more computationally efficient algorithm for computing

the PC transformation is based on the singular value decomposition (SVD). Let **X** denote an $m \times n$ real valued matrix. There exist matrices **U**, **S**, and **V** such that:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}} \qquad \textbf{1.12}$$

where **U** ($m \times m$) is an orthonormal matrix with columns composed of the *left singular vectors*, **S** ($m \times n$) is the matrix of *singular values*, $\mathbf{V}^{\mathbf{T}}$ ($n \times n$) is an orthonormal matrix with columns composed of the *right singular vectors*. If $m > n$, typically only the first n columns of **U** are calculated and **S** is and $n \times n$ diagonal matrix of singular values. By convention the singular values are sorted by magnitude with the largest located in $\mathbf{S}_{1,1}$ (upper left corner). An important aspect of the SVD that pertains to PCA is that $\mathbf{X}^{(k)}$, defined in Equation 1.13:

$$\mathbf{X}^{(k)} = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{s}_i \mathbf{v}_i^{\mathbf{T}} \qquad \textbf{1.13}$$

where $\mathbf{u}_i$, $\mathbf{s}_i$, and $\mathbf{v}_i$ are the i^{th} column vectors of the **U**, **S**, and **V** matrices is the closest rank-k matrix to **X** in a least squares sense. If a data matrix **X** of m observations (rows) and n variables (columns) is mean centered with respect to the columns, there is a direct correspondence to principal components calculated from the covariance matrix. In particular, each column vector u_i represents a PC score, the corresponding singular value $s_{i,i}$ is proportional to the variance captured by that PC, and the corresponding row vector v_i , typically referred to as a loading, is the quantitative relationship between the original variables and the PC. The first PC accounts for the maximum amount of variation in the data set. Successive PCs account for the maximum amount of remaining variation in the data set with the constraint that each of these PCs are orthogonal to the previous ones.

For high dimensional data sets it is often the case that the intrinsic dimensionality is far less than the number of variables, and is generally related to the complexity of the system. For spectroscopic data sets this can be loosely related to the number of components contributing to

the spectrum. For a simple mixture of non-interacting species following Beer's Law ideally across the spectral range, the number of significant PCs would be equivalent to the number of components (as long as concentrations were not highly correlated) and the remaining PCs would be attributable to random variations (e.g., noise). However, this is very rarely the case with real data and often the exact composition of a sample matrix is unknown to being with. There are several of methods for estimating the number of significant PCs, such as the scree test, but it is often necessary to use a trail-and-error approach or validation in practice. Principal component regression (PCR) is used several times in this work and refers to PCA applied as a preprocessing step before MLR. The reduced dimensionality and orthogonality of the principal component space addresses the issues of underdetermined systems and collinearity in MLR.

Partial Least Squares (PLS). The goal of PLS is analogous to PCR and MLR, namely to build a quantitative model for a response from a group of predictors (i.e., Equation 1.9). It is a factor based method similar to PCR, however, the approach to factor extraction from data sets differs. In PCR, a data set is decomposed into a set of scores and associated loadings independently of the responses. The scores are then regressed against the responses in order to obtain the calibration. In contrast, PLS extracts factors from the data set that are also relevant to responses. Specifically, PLS performs a iterative simultaneous decomposition of the data matrix and vector of responses with factors extracted to explain as much of the covariance between them as possible. The decomposition arrives at the factorizations in the following equations:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \qquad \textbf{1.14}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathbf{T}} + \mathbf{F} \qquad \textbf{1.15}$$

where $\mathbf{X}$ ($m \times n$) is the matrix of predictors, $\mathbf{Y}$ ($m \times k$) is the vector of responses, $\mathbf{T}$ ($m \times p$) and $\mathbf{U}$ ($m \times p$) are matrices of scores (referred to as latent variables or latent vectors), $\mathbf{P}^{\mathbf{T}}$ ($p \times n$) and $\mathbf{Q}^{\mathbf{T}}$

(p×k) are the matrices of loadings for **X** and **Y**, respectively, and **E** (*m×n*) and **F** (*m×k*) are residual matrices for **X** and **Y**, respectively. A variable number (*p*) of latent vectors are extracted from the data, and *p* is generally determined from a cross-validation or external validation procedure. There are two main variants of PLS algorithms: PLS1, which calculates separate PLS models for each constituent, and PLS2, which calculates a single model for multiple constituents. A number of different algorithms have been developed for extracting latent variables from data set, but the classic approach is called the nonlinear iterative partial least squares (NIPALS) method.[18-23] The NIPALS and other PLS algorithms are more elaborate than PCA and, for the sake of brevity, the mathematical formulation will not be presented.

Due to the fact that PLS extracts information directly related to both the predictor and response data it is often described as being more straightforward in application than PCR. After the decomposition of the data in PCR, an optimal subset of variables for the prediction of responses must still be determined. However, with conscientious application, similar calibration results can be obtained from both methods. In this work, PLS was used as a discriminant method in one study. PLS discriminant analysis (PLS-DA) is essentially the application of the traditional PLS method except that binary classification data is used in place of a continuous response variable.

Linear Discriminant Analysis (LDA). LDA is a statistical tool used to generate linear models for the discrimination of two or more groups from multivariate data.[6, 24] LDA is categorized as a supervised classification method as it requires that class information of the samples be known to generate a model. The formal goal of LDA is to maximize the objective function J_W, defined in Equation 1.16:

$$J_W = \frac{\mathbf{W}^{\mathbf{T}}\mathbf{S}_B\mathbf{W}}{\mathbf{W}^{\mathbf{T}}\mathbf{S}_W\mathbf{W}} \quad \textbf{1.16}$$

where $\mathbf{W}$ is the matrix of discriminant function vectors and $\mathbf{S_B}$ and $\mathbf{S_W}$ are the between-class and within-class scatter matrices, respectively. For a data set with c classes, n_i samples in each class, and N total samples over all classes, these matrices are defined in the following equations:

$$\mathbf{S}_B = \sum_{i=1}^{c} (\bar{\mathbf{x}}_i - \boldsymbol{\mu})(\bar{\mathbf{x}}_i - \boldsymbol{\mu})^T \quad \textbf{1.17}$$

$$\mathbf{S}_W = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \quad \textbf{1.18}$$

$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij} \quad \textbf{1.19}$$

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^{c} n_i \bar{\mathbf{x}}_i \quad \textbf{1.20}$$

where $\bar{\mathbf{x}}_i$ denotes the mean of the i^{th} class, $\mathbf{x}_{ij}$ denotes the j^{th} sample of the i^{th} class, and $\boldsymbol{\mu}$ denotes the global mean over all classes. A cursory examination of the objective function in Equation 1.16 reveals that discriminant functions in LDA are sought to maximize between class scatter while minimizing within class scatter. The solution to Equation 1.16 can be found by formulating and solving it as a generalized eigenvalue problem:

$$\mathbf{S}_B\mathbf{w}_k = \lambda_k \mathbf{S}_W \mathbf{w}_k \quad \textbf{1.21}$$

where $\mathbf{w}_k$ is the k^{th} discriminant function (eigenvector), with associated eigenvalue λ_k. For a data set with c classes, there is a maximum of c-1 possible discriminant functions (i.e., non-zero eigenvalues) in the solution and their significance is associated with the magnitude of the eigenvalue. Solving Equation 1.21 requires $\mathbf{S}_W$ is to be non-singular, which necessitates that at least as many samples are present as predictors in the model. In addition, collinearity in

predictors can lead to an unstable solution. Similar to MLR, these aspects pose significant limitations when the number of predictors is larger than the number of samples, as often the case with spectroscopic data. As with MLR, the application of PCA prior to LDA provides a solution, and that approach was used in this work.

LDA results in a collection of discriminant functions, also known as canonical variables (CV), that provide a new multi-dimensional space optimized for the discrimination of two or more classes. An example of synthetic two-class bivariate data is presented in Figure 1.5, which shows the contrast between the principal components and canonical variable derived from the same data. The decision criterion to determine whether samples belong to a particular class is not established by the discriminant functions. A potential solution to this problem is to define rigid linear decision functions between classes. Presented in Figure 1.6 are canonical variables derived from LDA of a synthetic 3 class data set. The dotted lines represent various possible linear decision functions for subdividing the discriminant function space into classes. The choice of these functions is fairly arbitrary and an infinite number of collections exist that could effectively separate the data points defined by the 3 classes. The problem with this methodology can be seen when several unknown samples, represented by '**X**'s in the figure, are considered, which do not appear to belong to any of the 3 classes. These samples may be classified into one or more of the classes based on the linear decision functions. This dilemma can be avoided if we define a classification boundary as demonstrated by the solid line for Class 1 in the figure. The conventional method to accomplish this is to determine a multi-dimensional standard deviation (MSD) based on Mahalanobis distance.[25] Although effective, there are shortcomings to Mahalanobis distance, which include the assumption that the data are normally distributed and computational instability when the number of dimensions is near the number of samples. An

alternative approach to determining MSDs that avoids these issues is the BEST algorithm, which is discussed in the following section.

Bootstrap Error-adjusted Single-sample Technique (BEST). The BEST algorithm is based on the statistical technique of bootstrapping, which is broadly categorized as a resampling method for statistical inference.[26-30] Bootstrapping is used for estimating a distribution by sampling with replacement from the original sample usually for the purposes of estimating uncertainties for population parameters (e.g., mean, median, regression coefficients). Advantages of bootstrap methods include the lack of parametric assumptions and efficacious results with limited numbers or samples or when parametric inference is extremely difficult or impossible. In the BEST algorithm, the bootstrap is used to estimate multi-dimensional standard deviations of a sample population.[31-34] Let **P** represent all possible samples of a particular population from an *n*-dimensional space **R** (rows of **P** represent samples, and columns of **P** represent independent variables). **P*** is a discrete realization of **P** based on a calibration set **T** ($m \times n$) of *m* samples that is chosen once to represent as nearly as possible all the variations present in **P**. **P*** is calculated using a bootstrap procedure and has parameters **C** ($1 \times n$) and **B** ($k \times n$), where **C** = E(**P**) is the expected value of **P** (i.e., mean) and **B** is the Monte Carlo approximation to the bootstrap distribution with *k* bootstrap replicates. The number of bootstrap replicates (*k*) included in **B** is somewhat arbitrary, but should be significantly larger than the number of samples. In this work, a range of 2000-5000 was used for *k*. The steps up to this point comprise the calibration phase of the BEST algorithm.

The estimation of the MSD of a new test spectra **X** ($1 \times n$) is found by first mapping the rows of **B** on to $\overrightarrow{\mathbf{CX}}$, the vector connecting **C** and **X**, by orthogonal projections. A directionally dependent skew adjustment is made along $\overrightarrow{\mathbf{CX}}$ based on a comparison of the expectation value **C**

and the median of **T** projected onto $\overrightarrow{\mathbf{CX}}$. The integral over the hyperspace **R** is calculated from the center of **P** outward and the result is an asymmetric standard deviation along the hyperline connecting **C** and **X**. Equation 1.22 defines the standard deviation in the direction of **X**, and equation 1.23 defines the standard deviation in the opposite direction:

$$+\vec{\sigma}\left|\frac{\int_{0}^{+\sigma}\int_{R}\mathbf{P}^{*}\rightarrow\left(\overrightarrow{\mathbf{CX}}\right)}{\int_{R}\mathbf{P}^{*}\rightarrow\left(\overrightarrow{\mathbf{CX}}\right)}=0.34\right. \qquad \textbf{1.22}$$

$$-\vec{\sigma}\left|\frac{\int_{0}^{-\sigma}\int_{R}\mathbf{P}^{*}\rightarrow\left(\overrightarrow{\mathbf{CX}}\right)}{\int_{R}\mathbf{P}^{*}\rightarrow\left(\overrightarrow{\mathbf{CX}}\right)}=0.34\right. \qquad \textbf{1.23}$$

Performance Metrics. The effectiveness of the modeling methods just described can be assessed using various metrics. In general, metrics useful for evaluating regression models have limited applicability for classification models and vice versa. The measures used in this work for these two types of predictive models are described here.

Regression Models. For multiple regression models, regression coefficient significance was determined using T-tests and overall model significance using F-tests. Predictive performance of regression models was estimated using a leave-one-out-cross-validation (LOOCV) procedure where a sample is removed from the data set entirely, the model is calibrated on the remaining samples and then used to predict the sample left out. This process is repeated for each sample in the data set. The LOOCV predictions were assessed by examining the coefficient of determination (r^2) as well as the standard error of estimate (SEE) and standard error of prediction (SEP), defined as:

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}} \quad \textbf{1.24}$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_{i,LOO})^2}{n-1}} \quad \textbf{1.25}$$

where y_i denotes the known or reference value, $\hat{y}_i$ denotes the estimate of y_i in the regression model, $\hat{y}_{i,LOO}$ denotes the predicted value of y_i from the LOOCV routine, n is the number of samples, and k is the number of variables in the regression model. A significant different between SEE and SEP typically indicates over-fitting in the regression, and F-tests are performed to determine if the SEE and SEP differed significantly.

Classification Models. As opposed to the continuous response of a regression model, classification model predictions are often representative of a binary decision (i.e., the sample belongs to a class or does not). To evaluate the performance of classification models, the three metrics defined below were used.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{1.26}$$

$$Precision = \frac{TP}{TP + FP} \quad \textbf{1.27}$$

$$Recall = \frac{TP}{TP + FN} \quad \textbf{1.28}$$

where TP denotes *true positive*, TN denotes *true negative*, FP denotes *false positive*, and FN denotes *false negative*. These metrics are somewhat complementary in nature and the collective result provides a good indication of model effectiveness. *Accuracy* is a measure of overall effectiveness taking all outcomes into account, *precision* measures of the effectiveness of the model when any class assignments are designated, and *recall* measures the effectiveness of the

model to classify samples only with respect to the correct class. These values are reported as percentages in this work, as is customary.

Molecular Factor Computing

The following is a concise introduction to the molecular factor computing (MFC) concept and the reader is referred to several sources for a comprehensive treatment.[35-37] Modern hyperspectral imaging techniques are able to collect unprecedented amounts of data in short periods of time. Reduction of these volumes of data from physical fields to high-level useful information is accomplished by multivariate data analysis techniques such as principal components analysis and partial least squares regression. Depending on instrument and sample complexity, the complete data collection and analysis process can be computationally intensive and time consuming. Integrated sensing and processing (ISP) is one approach receiving increased attention to simplify both instrumentation and computational analysis in these and other applications. In ISP, the data collection phase is no longer a passive constituent of the multivariate analytical process. Instead, the acquisition of useful information about the samples being analyzed is incorporated into the physical data collection phase. The nature of optical spectroscopy makes it an excellent technique for exploiting of the ISP paradigm. The ISP approach to optical spectroscopy involves tailoring the optical spectrum to encode high-level information about the samples in the sensing stage. The current availability and ongoing development of optical materials and instrumentation provides numerous approaches for practical application. Potential advantages of this methodology include smaller data volumes, shorter collection and analysis time, simpler and more rugged instrumentation and high optical throughput.

In traditional multivariate calibration methods, information is extracted from raw physical fields (i.e., spectra, chromatograms, etc.) by way of statistical and mathematical modeling techniques such as principal component (PC) and partial least squares (PLS) regression.[38, 39] Many of these techniques are factor-based where a weighting process of the raw data is utilized to reduce the original variables to a small number of informative components. In optical spectroscopy, the factor approach of these multivariate methods can be emulated during data collection by weighting the intensity of different wavelengths of light across a broad optical band. Theoretical treatments toward this objective can be found in the literature.[35, 40-42] One of the first methods devised to apply this methodology, termed optical signal processing (OSP), employed a grating to disperse wavelengths into a spatial dimension followed by a variable optical transmission mask to pass weighted spectroscopic information, with the integrated intensity collected by focusing the light transmitted through mask onto a single element detector.[41] A more recent approach for the practical application of this technique is based on the fabrication of thin film solid-state interference filters, called molecular optical elements (MOEs), which are carefully designed to transmit a weighted optical signal across a specified wavelength band. Several practical applications of this methodology in near-infrared and UV-VIS spectroscopy have been reported.[40, 43-45] Although quite effective, one drawback to this method is the high cost of filter manufacture, which requires the deposition of many alternating layers of high- and low-refractive index materials of precise thicknesses. The complexities in filter design and economics of the thin-film deposition techniques involved result in a nearly equivalent cost to produce one application-specific filter or hundreds. This translates into a cost effective solution when the assay has a sufficient number of end-users, but otherwise it may be cost prohibitive.

Alternatively, molecular absorption filters can be used as mathematical factors in spectral encoding to create a factor-analytic optical calibration in a high-throughput spectrometer, which we term molecular factor computing (MFC). In MFC, the transmission spectrum of the molecules used as filters replicate the weighting procedure of factor-based multivariate signal processing methods, such as principal-component (PC) and partial-least-squares (PLS) regression. One or more molecular filters are used to produce detector signals that are highly correlated to the desired sample information. The collection of this small number of detector responses, termed molecular factors, for a particular sample make up a "spectrum" that can then be processed using traditional multivariate methods to derive the desired information. This approach provides a multiplex advantage (versus a dispersive instrument) by simultaneously collecting a broad spectrum of light using a single detector, as well as an optical throughput advantage because it does not require a slit to achieve resolution. In addition, a filter-based approach provides a more rugged construction compared to complex scanning instruments with many moving parts. Potential applications of this methodology are abundant. Our group has previously demonstrated its application in the NIR spectral region for the determination of ethanol concentrations in water.[36, 37]

We will now revisit factor based multivariate methods, PCA in particular, data to help clarify MFC theory. As applied to spectroscopic data sets, these methods aim to transform information contained in the multidimensional space defined by the individual wavelengths (i.e., the Cartesian coordinate system) into a more efficient or useful application specific factor space. The coordinates of a spectrum in this new space are generally referred to as scores, and for a linear transformation such as PCA, they are related to the original spectrum by the relationship:

$$Score = f_1 x_1 + f_2 x_2 + \ldots + f_n x_n \qquad \textbf{1.29}$$

where x_i is the intensity (e.g., absorbance, transmittance) at the i^{th} wavelength, f_i is the coefficient quantifying the contribution of the i^{th} wavelength to the score, and n is the total number of wavelengths in the spectrum. The determination of the coefficients (f) in Equation 1.29 for factor based methods is largely dependent upon the desired outcome. For example, PCA extracts factors based on the amount of variance accounted for in the spectral data set, while PLS extracts factors based on the amount of covariance between spectra and the property(s) of interest. It is also possible to weight each wavelength in a spectrum optically using the absorbance of molecules, which is the aim of MFC. Ignoring source intensity and detector response, the coefficients (f_i) in Equation 1.29 are based on the transmission spectrum of the components used as molecular filters. The "scores" in this case, are simply the voltage recorded on the detector by integrating the total light through the sample and filter over a broad wavelength band.

Referring again to PCA, the transformation from spectral data to scores can be represented in matrix notation as:

$$\mathbf{S} = \mathbf{XF} \qquad \textbf{1.30}$$

where $\mathbf{S}$ ($m \times p$) is the matrix of scores with each row corresponding to a sample, $\mathbf{F}$ ($n \times p$) is the matrix of loadings (coefficients from Equation 1.29), with each column corresponding to a principal component, $\mathbf{X}$ ($m \times n$) is the matrix of spectral data, with each row corresponding to a sample spectrum, and m, n, and p are the number of samples, wavelengths and PCs, respectively. In relation to the SVD presentation of PCA, $\mathbf{S}$ and $\mathbf{F}$ here correspond to $\mathbf{US}$ and $\mathbf{V}$, respectively, in Equation 1.12. In MFC, the counterpart to $\mathbf{F}$ is a matrix of the transmission spectra of the filter compounds. Note that in PCA, the $\mathbf{X}$ matrix and can be composed of either transmission/reflectance or absorbance values and is preprocessed in some fashion (mean centered at least). In contrast, MFC is constrained to transmission/reflectance values and raw

spectra as there are no optical analogs to collecting direct absorbance measurements or preprocessing the sample spectrum. MFC scores are also a function of the emission spectrum of the light source and the response of the detector, and a more realistic formulation of equation 1.29 for MFC would be:

$$Score = I_1 R_1 f_1 x_1 + I_2 R_2 f_2 x_2 + \ldots + I_n R_n f_n x_n \qquad \textbf{1.31}$$

where I_i and R_i are the intensity of the light source and response of the detector, respectively, at i^{th} wavelength, and the other terms are as described previously. Similarly, equation 1.30 can be adapted to take these factors into account if the data matrix **X** is transformed by the element-by-element scaling of each spectrum with the emission spectrum (I) and detector response (R) vectors.

The next step in the development of an MFC method is to determine the necessary filter molecules to achieve the desired result. One possibility is to select filter molecules with aim of replicating the weights in a regression vector derived from PCR modeling or other multivariate techniques. There are several reasons why this approach is not a practical solution. In order to model a principal component, the loading must be divided into complementary segments representing the positive and negative coefficients. Each of these filters is composed of one or more bands where transmission is desired separated by bands where complete absorption is required (i.e., bands where transmission is desired in the complementary filter). Using molecular species for absorption, this is an especially difficult task. Moreover, while loading vectors from multivariate methods have "spectrum like" characteristics, they are far more abstract, and emulating these features with molecular absorption is complex. Simulations during the preliminary development of the MFC technique demonstrated that for a relatively narrow spectral region (i.e., a few hundred nanometers or less), it would be feasible to approximate a

regression vector with molecular absorption; however, the number of chemical components for a reasonable representation would be excessively high in many cases.

Another approach to filter selection in MFC is objective based. In any multivariate statistical method, the ultimate goal is the end result, such as quantitative prediction of an analyte or classification of a sample. Applying PCR and PLS methods to the same data set can produce statistically equivalent models in terms of predictive performance with different regression vectors. The utility of an MFC method results from the scores collected from a sample. The effectiveness of a given set of molecular filters can be estimated by processing (e.g., building a regression or classification model) simulate scores based on equation 1.31. Filter selection is then directed by searching a library of transmission spectra for components that provide the optimal performance. In this work, a library of NIR transmission spectra from 1923 compounds (John Wiley & Sons, Inc.) was used to simulate MFC filters. Ideally, we want the fewest number of filters that provides acceptable performance for the desired application. The development of filter selection algorithms to achieve this purpose for regression and linear discriminant models was the focus of other research in this group.[36, 37] The method used in this work involved the pre-filtering of the library using genetic algorithms followed by a step-wise selection procedure. The instrumentation used in this work for MFC data collection will be covered in later chapters.

Chapter One Figures

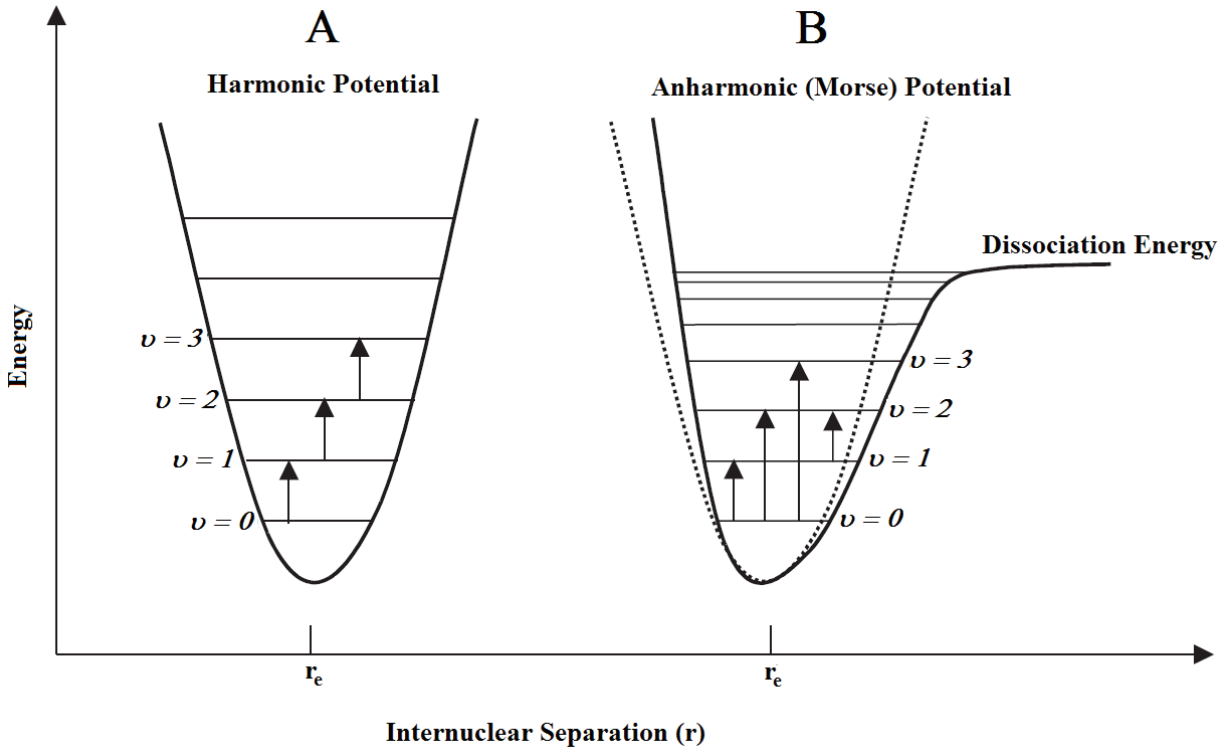

Figure 1.1: Potential energy functions representative of the simple harmonic (**A**) and anharmonic (**B**) oscillators.

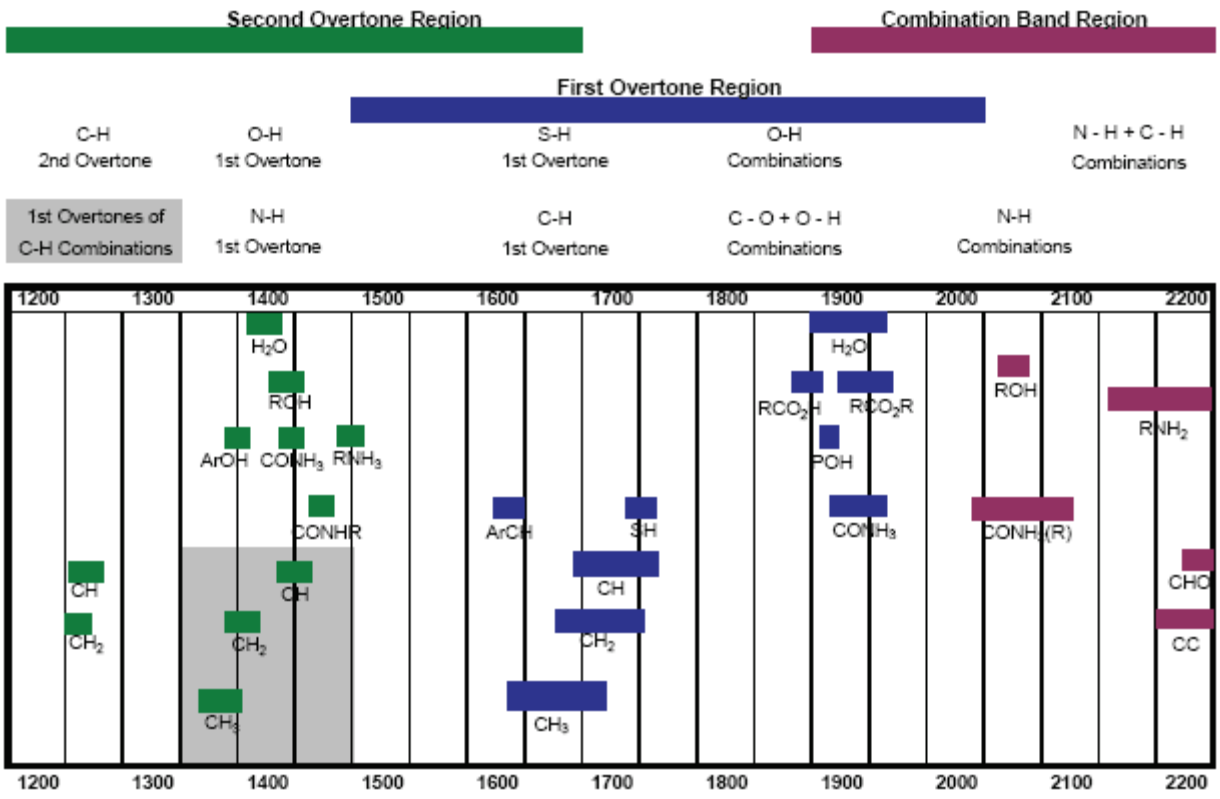

Figure 1.2: NIR functional group/spectrum correlation chart. This chart serves as an approximate guide for NIR band assignments by functional group.

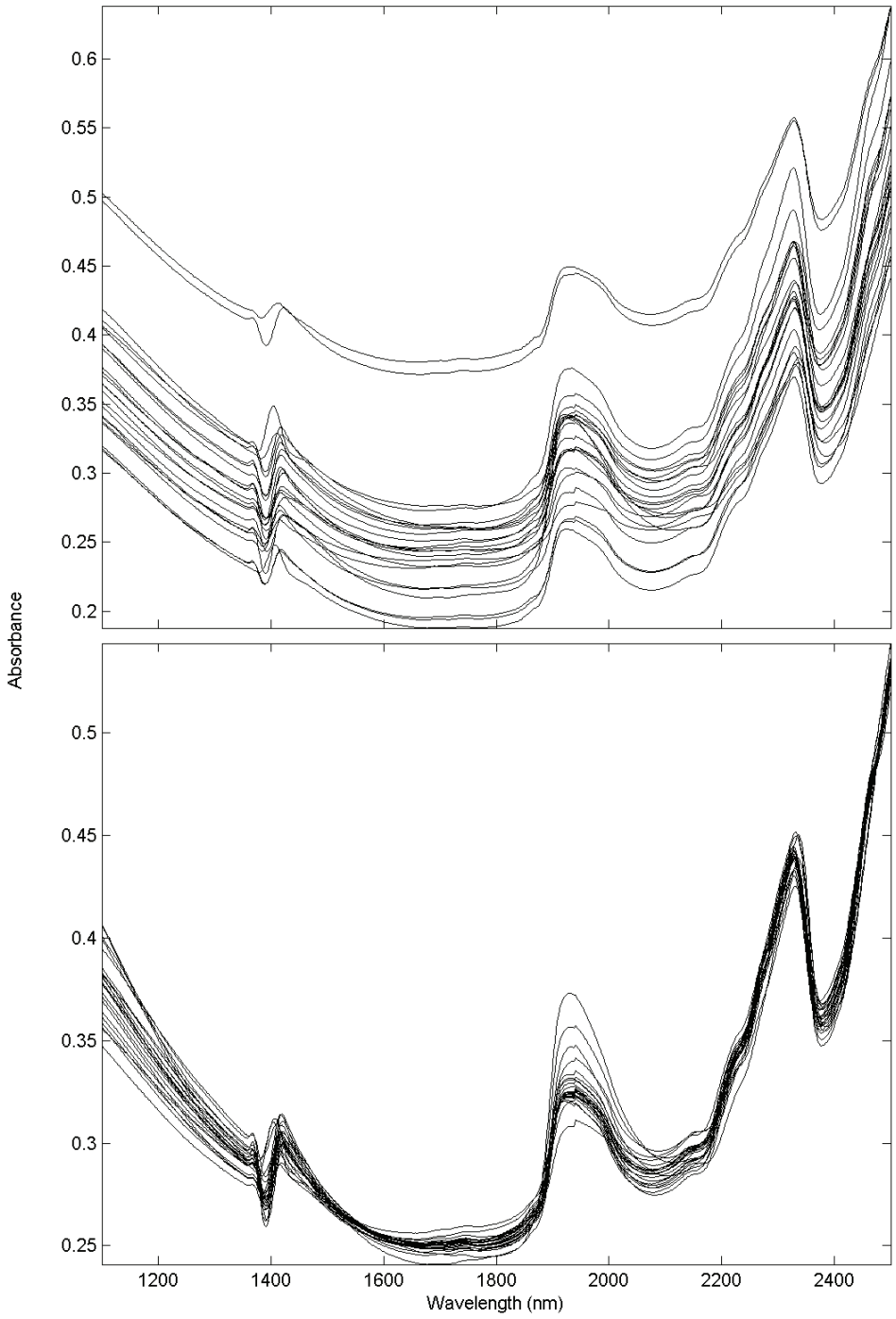

Figure 1.3: An example of a NIR spectral data set of geological samples before (top panel) and after (bottom panel) multiplicative scatter correction (MSC).

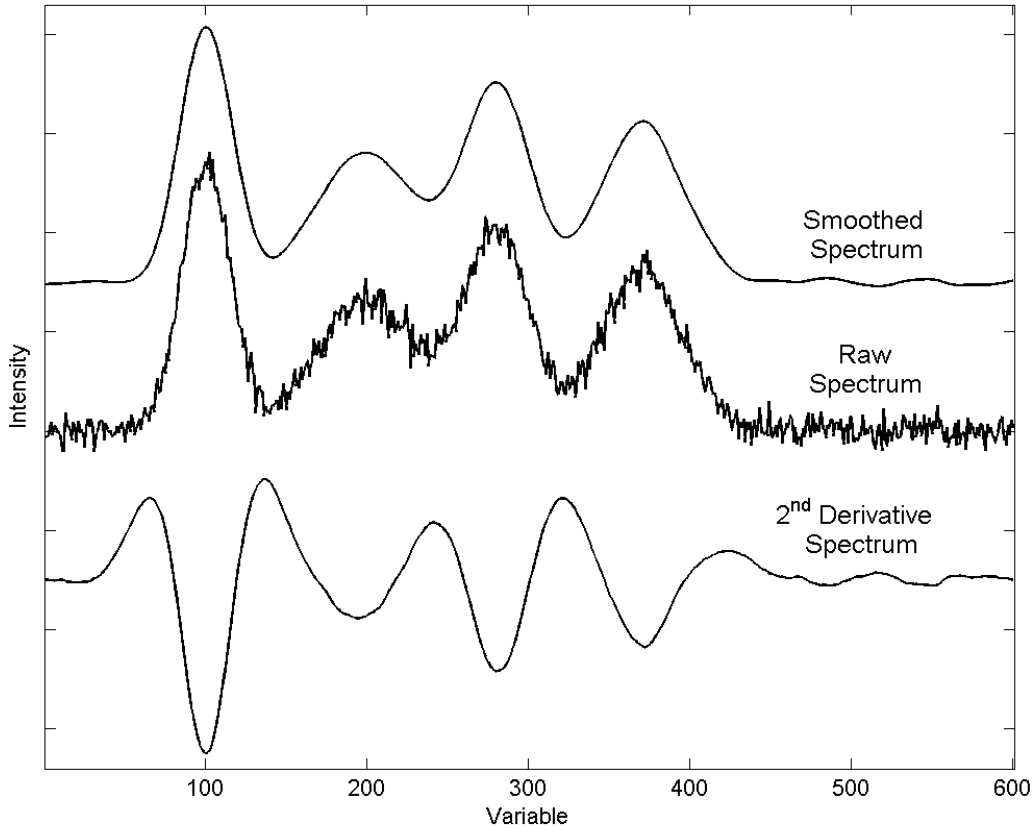

Figure 1.4: An example of smoothing and second derivative approximation using a cubic smoothing spline on a synthetic spectrum composed of overlapping Gaussians with white noise added.

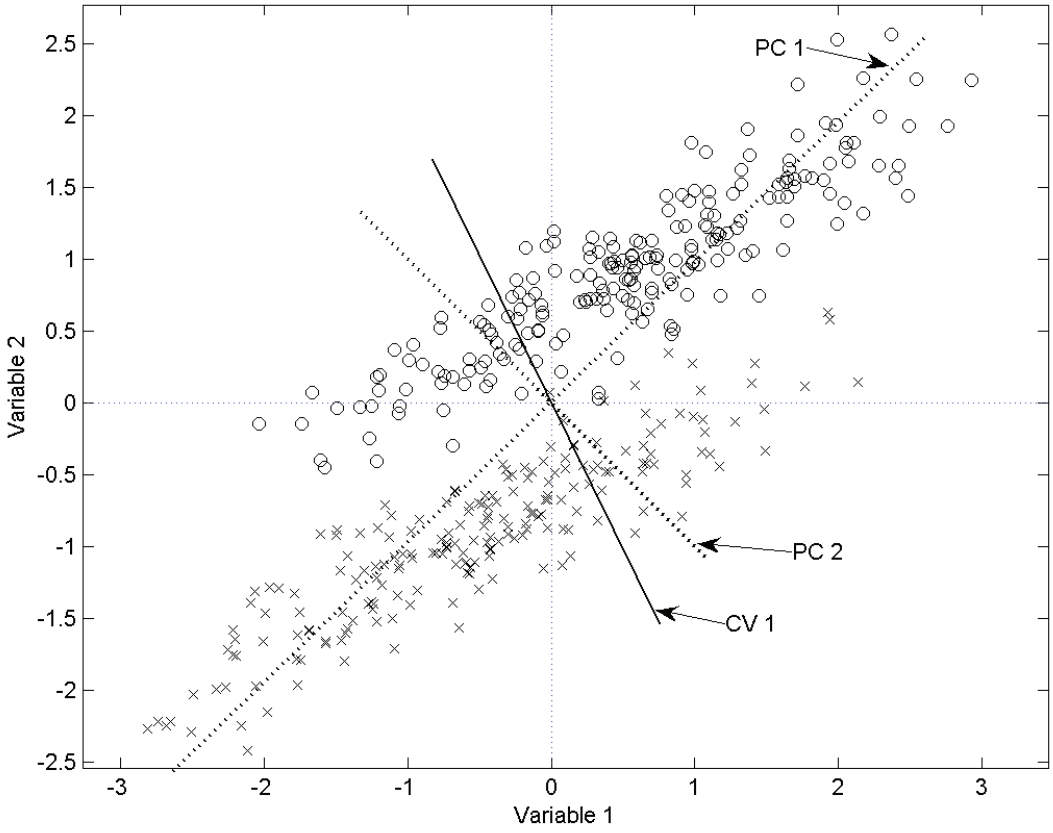

Figure 1.5: An example of synthetic two class (designated by '×' and 'o') bivariate data demonstrating the contrast between principal components and the canonical variables for identical data. Principal components are defined to maximize the variance captured while canonical variables are defined based on class separation.

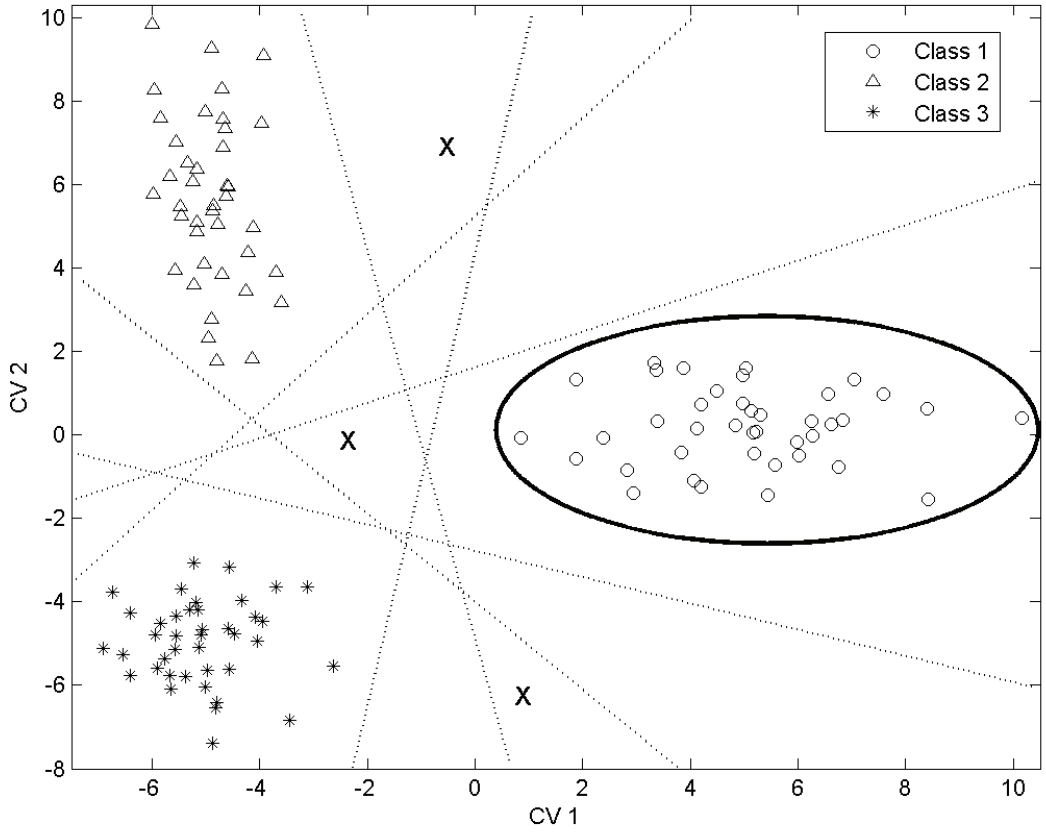

Figure 1.6: An example of canonical variables derived from LDA of a synthetic 3 class data set. Dotted lines represent possible linear decision lines for class separation. The **X**'s represent potential test samples that do not appear to belong to any of the 3 sample populations. A decision boundary based on a multi-dimensional standard deviation (MSD), represented by the solid ellipse, provides a better alternative.

Section II: Biomedical Spectroscopy

Atherosclerosis is a chronic inflammatory process with complications that are the leading cause of death in western societies, claiming the lives of approximately 450,000 annually in the U.S. alone.[46-48] Extensive research has been done to determine the complex pathophysiology of atherosclerosis, although mechanisms for various aspects are still being elucidated. A number of chemical and structural changes in the vessel wall during the development and progression of atherosclerosis have been identified. Among these are an increase in retained lipoproteins[49, 50], and their subsequent oxidation[49, 51, 52] in the subendothelial matrix. Recruitment of monocytes and the eventual development of lipid-laden macrophages (foam cells) are another hallmark of the early atherosclerotic process.[53, 54] Proliferation and phenotypic changes in smooth muscle cells are seen as well.[55, 56] The advanced atherosclerotic lesion may be characterized by accumulation of extracellular lipid, development of a lipid rich necrotic core, formation of a fibrous cap, and calcification. Although not completely understood, the vulnerability of atherosclerotic plaques to rupture has been studied extensively.[57-60] The thin-cap fibroatheroma is the most widely recognized vulnerable plaque, and is loosely characterized by a thin fibrous cap (< 65 μm) that is rich in inflammatory macrophages and covers a large lipid pool.[57, 59, 61, 62]

Collagen and elastin are major structural components of vessel walls. Changes in the quantity and distribution of these components in the vessel walls take place during vascular remodeling in the development of atherosclerosis and other vascular disorders, such as aneurysm. The local content of collagen and elastin correlates with histological and biochemical properties of normal aorta, atherosclerotic plaques and infiltration of inflammatory cells.[63-68] Collagen degradation in the fibrous cap of atherosclerotic plaques is implicated in vulnerability

by reducing the mechanical strength of the cap, which leads to plaque rupture. By inference, elastin deficiency can aid in locating atherosclerotic plaques with fibrous caps, which are composed primarily of smooth muscle cells and collagen.

Abdominal aortic aneurysms (AAAs) represent potentially life-threatening conditions that occur in up to 10 percent of the aged populations in industrialized nations. An aneurysm is broadly defined as a permanent localized dilatation of an artery. AAAs arise due to substantial remodeling of the extracellular matrix and are frequently accompanied by atherosclerosis. They may be manifested by catastrophic rupture, signs of pressure on other viscera or an embolism originating in the aneurysmal wall, but most are asymptomatic. Collagen and elastin have been widely implicated in aneurysm formation, progression, and rupture. The most prevalent structural modification associated with human AAAs that has been reported is a reduction in elastin concentration in the aortic wall.[69-75] A common histochemical observation in these studies has been the gross alteration in the organization and deposition of elastin in aneurysmal walls. Significant correlations between reduced elastin concentration and AAA diameter have been observed.[72] Alternatively, other studies have shown that reduction in elastin concentration is essentially complete prior to dilation is AAAs.[74, 76] One proposed mechanism for reduced elastin concentrations is degradation or loss brought about by elastolysis.[69-71, 74] Other work has reported that elastin content in the vessel walls of AAAs actually increases.[75, 77] In these studies, a 2.5-fold increase in elastin content was found in AAAs versus normal aortic samples of equal length. This increase, however, was accompanied by a significantly greater increase in total matrix proteins, which suggests that reduction in elastin content is at least in part due to dilution. These results and work by others suggest that an important mechanism in AAA formation is the regulation of matrix macromolecule synthesis.[75, 77, 78]

Increased collagen concentration is another matrix modification that has been widely observed in human AAAs.[71-73, 75, 79] Modifications in collagen organization and deposition have been correlated to rupture in human AAAs.[70, 72, 74, 79] It is widely believed that organization and deposition of elastin plays a fundamental role in aneurysm formation, but the causal or compensatory relationship between collagen and AAAs is still being elucidated. Although there are noticeable differences in the findings of these studies, it is evident that increases in the collagen to elastin ratio are a general observation in AAAs.

Chapter Two – Near-Infrared Spectrometry of Abdominal Aortic Aneurysm in the ApoE$^{-/-}$ Mouse

Introduction

A number of murine animal models have been developed in recent years to mimic atherosclerosis[80] and abdominal aortic aneurysm[81] (AAA) in humans. Techniques for monitoring the onset, progression, and regression of these processes in murine models could provide valuable pathophysiological insights into the disease processes. In addition, these analytical methods may be useful in assessing the effectiveness of possible treatments.

Diffuse reflectance Near-Infrared (NIR) Spectroscopy has proven to be a useful technique for identifying chemical content of biological species.[82] NIR spectroscopy is based on the absorbance of light by organic molecules and allows detailed analysis of chemical composition. NIR can provide rapid, simultaneous, multicomponent, non-destructive chemical analysis of biological tissues. Little or no sample preparation is required, and molecular information as well as gross biological and physical properties can be derived from the spectra. Biological applications of NIR spectroscopy include monitoring systemic and cerebral oxygenation and identifying plasma constituents including glucose, total protein, triglycerides, cholesterol, urea, creatinine, and uric acid.[83-87] Our group has reported on the use of NIR spectroscopy to classify human aortic atherosclerotic plaques and to identify cholesterol, HDL, and LDL in arterial wall samples.[88-90]

This work describes preliminary research to determine the feasibility of using near-infrared spectroscopic methods to monitor atherosclerosis and aneurysm in mice aortas. The physiological alterations described above for these vascular disorders provide a potential means

to monitor them *in vivo*. The ultimate goal of this research is the eventual development of a novel fiber optic probe capable of near-infrared spectral analysis *in vivo*.

Materials and Methods

Mice. The animal model has been detailed elsewhere[91], and a brief description will be presented here. Infusion of angiotensin II (Ang II) into mature apolipoprotein E-deficient (*apoE*$^{-/-}$) mice promotes an increase in the severity of aortic atherosclerotic lesions and the formation of abdominal aortic aneurysms. The formation of aneurysms in the animal model is independent of arterial blood pressure and lipoprotein profiles; however, it requires the hyperlipidemic background and is dependent on Ang II dose and gender (males develop aneurysms at a greater incidence than females). Female *apoE*$^{-/-}$ mice (backcrossed 10 X into the C57BL/6J background) were obtained from The Jackson Laboratories (Bar Harbor, ME). All mice were maintained under barrier conditions. Water and normal laboratory diet were available *ad libitum*. All procedures involving animals were approved by the Institutional Animal Care and Use Committee at the University of Kentucky.

Ang II Infusion. Alzet osmotic minipumps (Model 2004; ALZA Scientific Products, Mountain View, CA) were implanted into *apoE*$^{-/-}$ mice (n = 6) at 6 months of age. Pumps were filled either with saline vehicle (control group, n = 2) or solutions of Ang II (Sigma Chemical Co., St. Louis, MO) that delivered (subcutaneously) either 500 (medium dose group, n = 1) or 1000 (high dose group, n = 3) ng/min/kg of Ang II for 28 days. At this point, the mice were sacrificed.

Tissue Preparation. Aortic tissue was removed from the ascending aorta to the ileal bifurcation and placed in 4% paraformaldehyde in PBS overnight at room temperature. The intimal surface was exposed by a longitudinal cut through the inner curvature down the whole length of the

aortic tree. Each aortic sample was cut into 6 roughly equal sections resulting in 36 total samples.

Instrumentation and Spectra Collection. NIR spectra were collected with a FLEX (Bran and Luebbe, Elmsford, NY) spectrophotometer. Each sample was placed on an ordinary microscope slide for analysis. The samples were not immersed in solution during spectra collection. The instrumental setup allowed for the entire tissue sample to contribute to the NIR signal. NIR reflectance measurements were made at 20 wavelengths between 1445 and 2350. Absorbance values were obtained as log(1/R). Continuous spectra were obtained by a cubic spline fitting routine (Speakeasy Computing Corp., Chicago, IL) from the absorbance values. The spectral data were scatter corrected prior to data analysis. Spectra from a control group and a high dose group tissue sample are presented in Figure 2.1 for comparison. The chemical composition of the tissue samples between groups is similar, and as a result the gross appearances of the two spectra are similar.

Collagen to elastin ratios for tissue samples were obtained by freeze-fracture scanning electron microscopy (SEM) by Industrial Analytical Services Incorporated (Leominster, MA). Detailed information on this technique can be found in the literature.[92, 93] To determine collagen to elastin ratios three SEM images were obtained for each sample corresponding to the inner and outer membranes and a cross section of the aortic wall.

Data Analysis. Analytical software was written in Mathematica 4.1 (Wolfram Research, Inc., Champaign, IL). Principal components analysis (PCA) and principal components regression (PCR) were used to analyze the data. A brief explanation of these techniques will be presented here, but a detailed description can be found in the literature.[94, 95] PCA transforms a large number of correlated variables into a new set of uncorrelated factors, reducing the dimensionality

of the data set by a linear transformation of the coordinate system. The principal components (PCs) are structured so that the first few describe most of the variation in the original variables. The first PC contains information from the constituent(s) that contributes most to the total spectral variation in the data set. The second PC is orthogonal (i.e., uncorrelated) to the first and describes the maximum amount of remaining variation after removal of the first PC. Additional PCs, orthogonal to the previous ones, describe progressively smaller contributions to the spectral variation. In general, the first few PCs contain the desired information, and the remaining PCs describe insignificant constituents and noise.

PCR was used to construct calibration models to predict Ang II dose. PCR is an inverse least squares regression technique that fits the PC scores of the calibration data to the dependent variables. A leave-one-out cross validation routine was used in this work to assess the capability of predicting Ang II dose in mice aortas with NIR spectra.

Results and Discussion

The spectra were investigated first by PCA to examine the relationships between the samples. The first two PCs accounted for approximately 89% of the variation in the data set, and a plot of PC 1 vs. PC 2 is presented in Figure 2.2. As seen in the figure, the spectral separation among the different groups is significant. Also evident, is the relatively minor variation between spectra within each group. This is true between mice, as well, in the control and high dose groups, which contain 2 and 3 mice aortas, respectively. Since the medium dose group contained samples from a single mouse, this relation can not be established. The ability to predict Ang II dose from the NIR spectra was examined using PCR. The SEP provides a global estimate of the prediction capabilities of the method. The SEP for Ang II dose depends on the number of PCs

used in the model. The first two PCs provided the optimum prediction models, and the SEP for Ang II dose by PCR was 37 ng/min/kg (SEE 26 ng/min/kg, $r^2 = 0.99$, $f_{0.01}$ significance).

Collagen to elastin ratios in the samples were estimated using SEM to see if a correlation could be found between the NIR spectra and a histological marker. The range of collagen to elastin ratios obtained for the samples was 1.4 to 4.5 and is shown versus Ang II dose in Figure 2.3. A trend apparent in the data is a general increase in collagen to elastin ratio with increasing Ang II infusion ($r^2 = 0.85$). Comparison of this data with the NIR spectra illustrates that the within group variation is greater in the collagen to elastin ratios than in the NIR spectra. Two factors that likely contribute to the observed discrepancies are the limited scope of the SEM imaging, and the fact that collagen and elastin are not the only components contributing to the NIR spectra. The entire aortic sample contributes to the NIR spectra, including significant contributions from lipids. SEM sampling was limited to three SEM images from the inner and outer membranes and a single cross section of the vessel were obtained.

This work presents a macroscopic study of arterial changes in mice aorta by NIR spectroscopy. Figures 2.4 and 2.5 demonstrate that collagens and elastin have distinctive near-IR spectra. Collagens I and III are the principal collagens of the aorta. Figure 2.2 suggests that diffuse alterations occur in the aortic walls of Ang II infused *apoE*$^{-/-}$ mice, and that these changes can be observed with NIR spectroscopic analysis of intact tissue. The better correlation between the near-IR spectra and Ang II dose than between the near-IR spectra and the collagen-to-elastin ratio suggests that there is more occurring in the aortas than collagen/elastin changes and the near-IR spectra detect these additional chemical changes. This suggestion is not surprising given that near-IR spectra show contributions at some level from virtually every organic compound in any tissue. Though the data are limited, they also indicate that the NIR identifiable

characteristics of the aorta are consistent between mice within the same treatment group (this cannot be established with respect to the medium dose group, however, because the tissue samples came from a single mouse). Together, the results support the development of a catheter and extended-use fiber-optic implant for use in the blood vessels of mice and suggest wavelengths that might be important to monitor.

In an effort to determine which spectral changes in the aortas were associated with collagens I and III and elastin composition changes, a set of sample mixtures of collagens I and II and elastin was prepared using pure lyophilized standards. Figure 2.6 shows the composition of each of the prepared sample standards, with the pure collagen I (C1) standard in one corner of the triangle, the pure collagen III (C3) in another corner of the triangle, and finally the pure elastin in the remaining corner of the triangle. The concentrations of each constituent in the standard mixtures were set at 0, 25, 50, 75 or 100% of each lyophilized protein. The vertexes represent all possible combinations of mixtures in the percentages given (a total of 15 mixtures including the pure corner standards). The center (i.e., group mean, or GM in the Figure 2.6) would represent a mixture of one-third of each protein, but this sample was not actually prepared in the set.

The reflection spectra of the 15 mixtures were compared to the reflection spectra of the 36 aorta sections be mean-centering the spectra of the mixtures and the spectra of the aortas. The difference spectra between each standard sample spectrum and the mean spectrum of the strandard samples were calculated. Likewise, the difference spectra between each aorta section spectrum and the mean spectrum of the aorta sections were also calculated. The aorta difference spectra were then averaged for the control, low-dose, and high-dose Ang II groups. Finally, the

difference spectra of the standards and the averaged aorta sections were the correlated using equation 2.1:

$$r = \frac{\sum s_1 s_2}{\sum s_1^2 \sum s_2^2} \qquad \textbf{2.1}$$

where s_1 is a spectrum from the set of standards and s_2 is a spectrum from the set of aorta sections.[96]

The correlations between the average control group spectrum and each standard sample spectrum ranged between ± 0.97. The highest correlation was with the 50, 50, 0% (collagen I, collagen III, elastin) standard. The correlations between the average low-Ang II dose group spectrum and each standard sample spectrum ranged between ± 0.99. The highest correlation of the average spectrum of the low-Ang II dose group was with the pure elastin (0, 0, 100%) standard. The correlations between the average high-dose group spectrum and each standard sample spectrum ranged between ± 0.87. The highest correlation was with the pure collagen I (100, 0, 0%) standard. These correlations suggest that much of the variations in aorta spectra with Ang II dose can be attributed to changes in collagen and elastin composition. The locations of the mean control (C), low-dose (L), and high-dose (H) Ang II aortic sections are shown in figure 2.6. The H, L, and C points are interpolated to the location of maximum correlation to the standards. The spectra of the most highly correlated standard samples, and each treatment group is shown in Figure 2.7. The correlation between the aortic sections and the standard mixtures is highest for the control and low-dose Ang II groups in Figure 2.7. The observation that the changes in the spectra of the standard mixtures to not add up perfectly to the changes found in the aortas also suggests that there is more occurring in the aortas than collagen/elastin changes and the near-IR spectra detect these additional chemical changes. The majority of these

additional changes appear to be in the C−H stretching overtones that usually correspond to an increase in trans or saturated lipids.

This study is limited in several respects. The small number of mice (6) used as the source of the 36 aorta sections limits the observable variation in the data set. The lack of detailed histological data for the samples prevents the association of the spectra with specific tissue pathologies and comparison of pathologies between samples. Previous work with the animal model demonstrated an increased severity of atherosclerosis and the formation of aortic aneurysms with Ang II infusion.[91] It was assumed that similar alterations occurred in this population contributing to the spectral trends observed in the data set. Studies with larger, more diverse populations are necessary to better investigate the spectral trends and variation between mice.

Conclusion

Near-IR spectra are distinctive for proteins in the blood vessel wall (specifically collagens and elastin). The ability of near-IR spectrometry to collect useful spectra in aqueous environments may make it useful for proteomics in vivo. The results of this study suggest that NIR spectroscopy is a potentially useful technique for investigating vascular changes and protein composition associated with abdominal aortic aneurysm in a mouse model of the disease. These results support an expanded study in the future to correlate NIR spectra with chemical compositions and histological features in mice aortas.

Chapter Two Figures

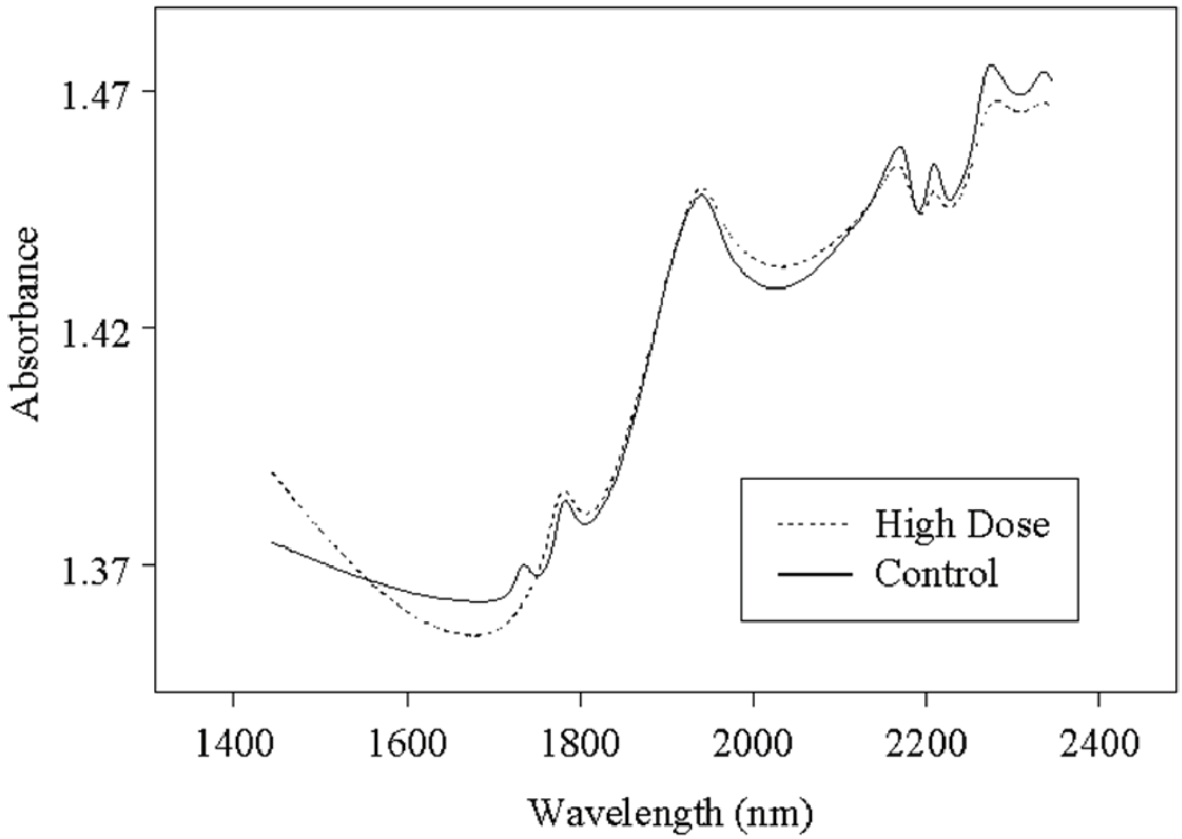

Figure 2.1: Examples of the near-IR spectra obtained from aorta sections of a high Ang II dose (1000 ng kg^{-1} min^{-1}) mouse and control mouse.

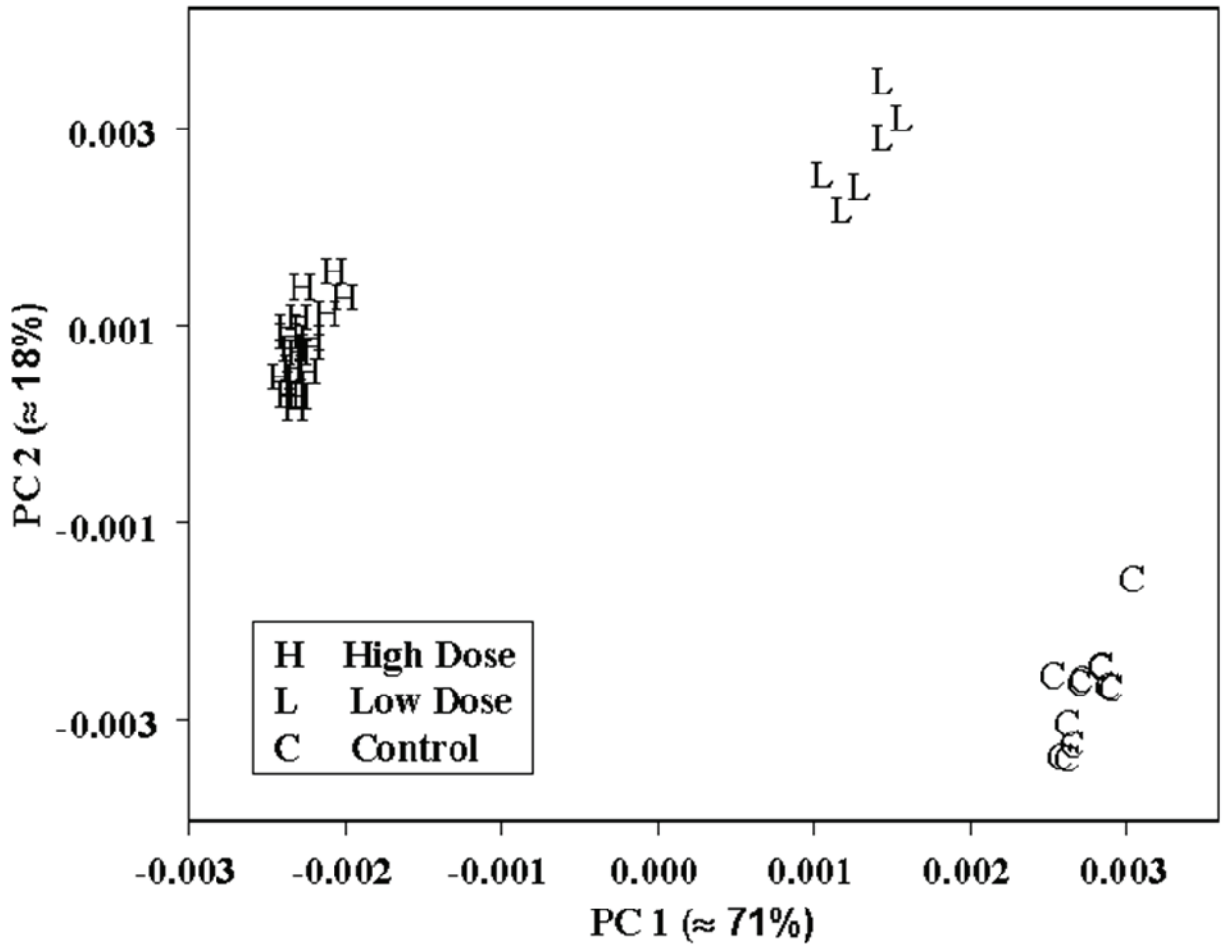

Figure 2.2: First two principal components of the near-IR spectra of the aorta data set. H denotes the spectra from aortic sections of the high-dose (1000 ng kg^{-1} min^{-1} Ang II, n = 18 samples), L denotes the spectra from aortic sections of the low-dose mice (500 ng kg^{-1} min^{-1} Ang II, n = 6 samples), and C denotes the spectra from aortic sections of the control mice (saline, n = 12 samples).

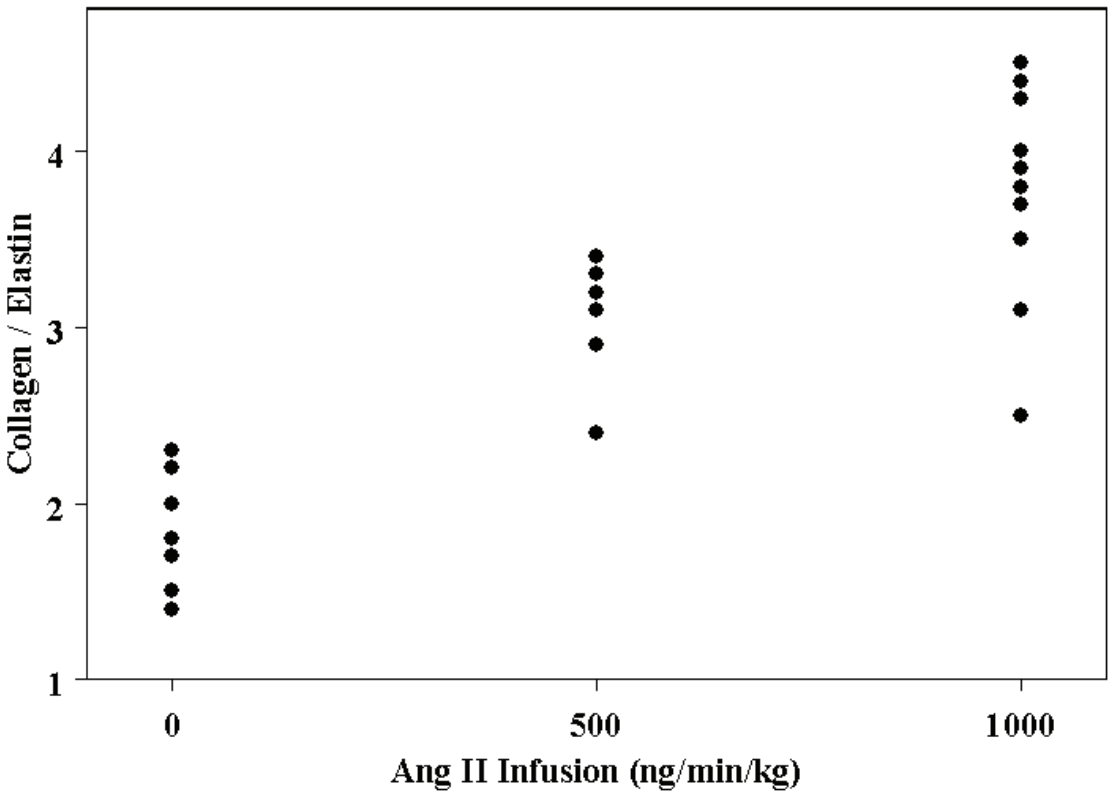

Figure 2.3: Collagen to elastin ratios versus Ang II dose in mice aortas determined by SEM analysis.

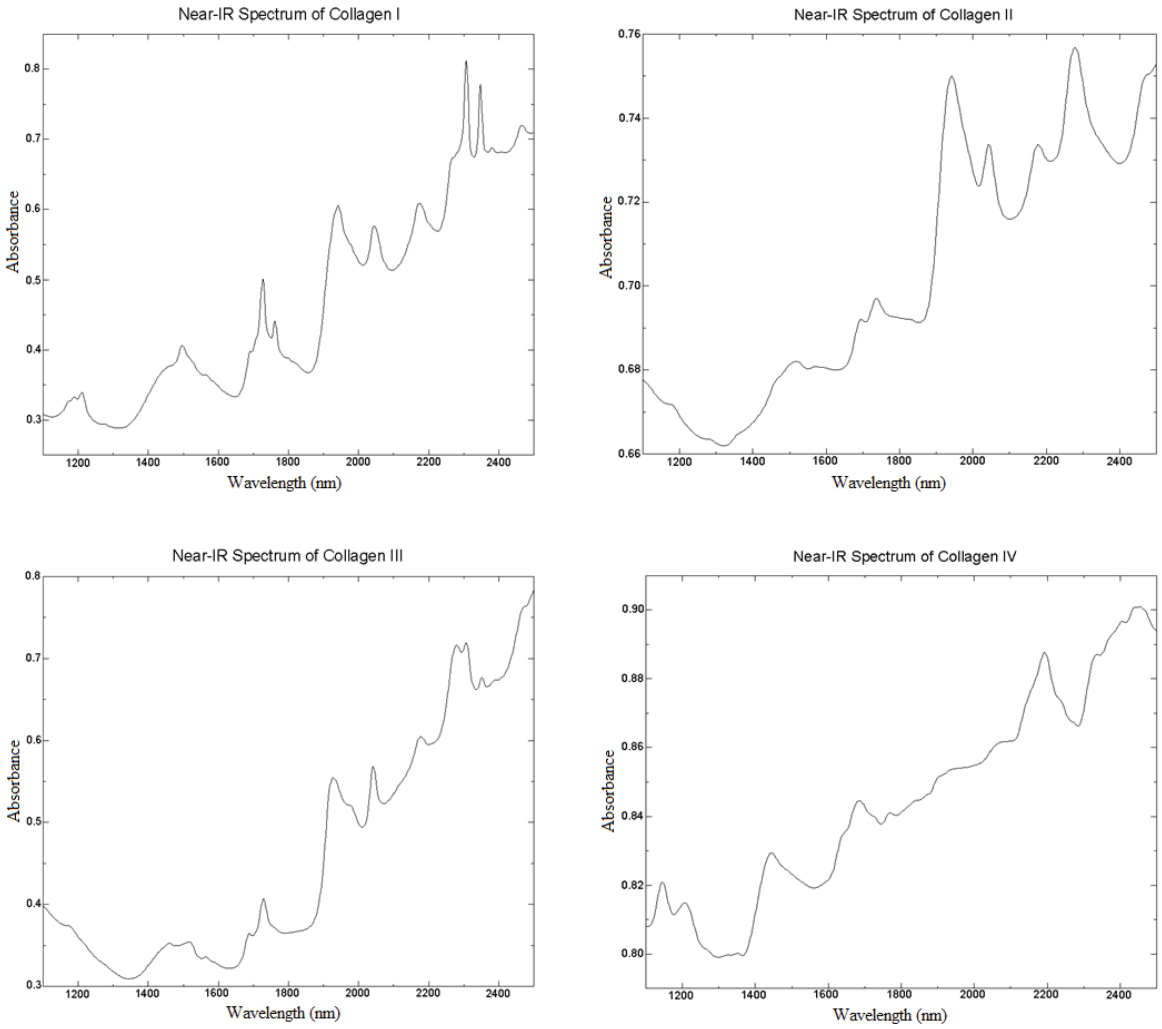

Figure 2.4: Near-IR spectra of collagen standards. Collagen I and collagen III are the principal collagens of the aorta. The spectra of the four collagens are distinctive, suggesting that simultaneous multicomponent analysis of the collagens is possible.

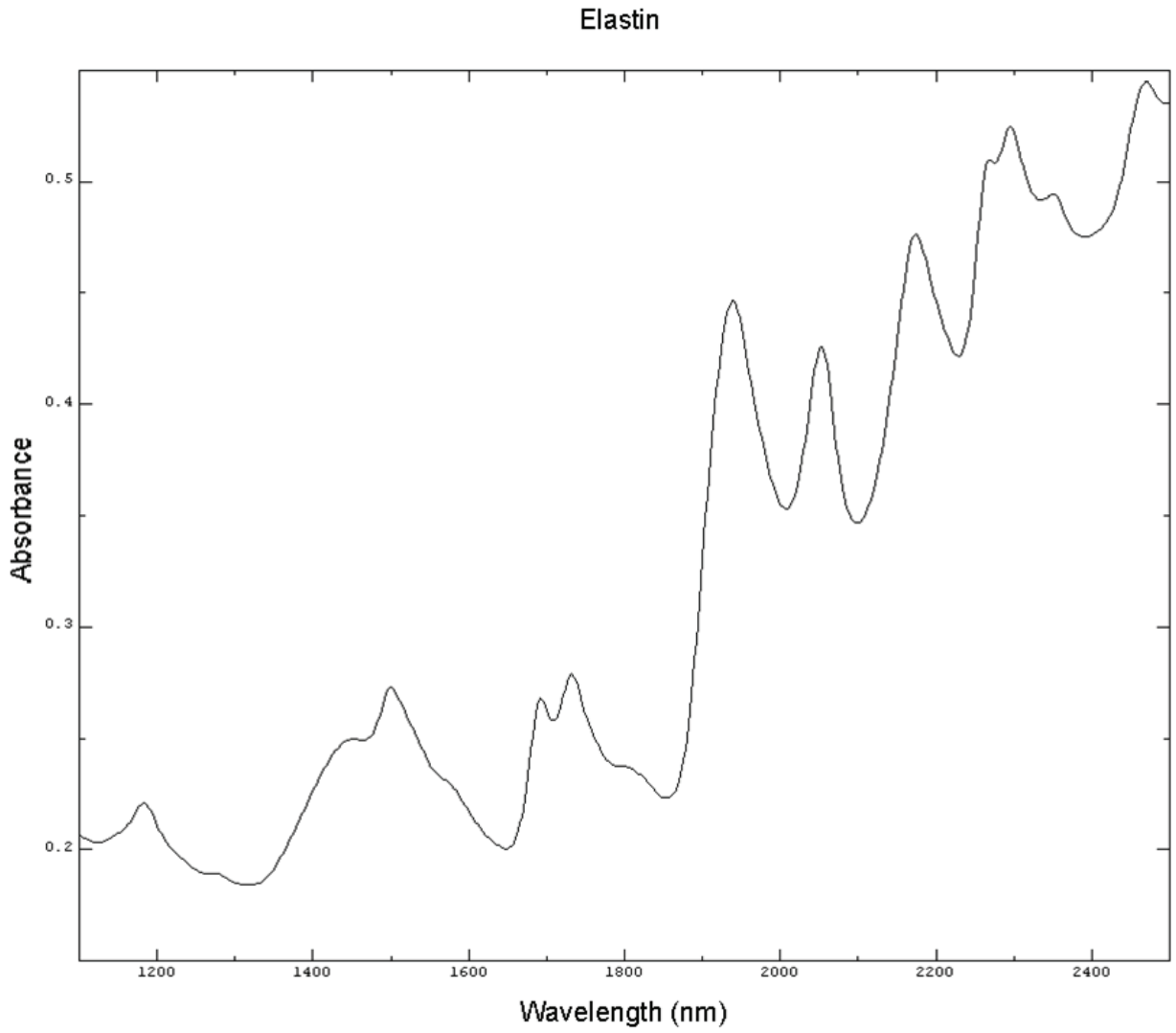

Figure 2.5: Near-IR spectrum of elastin standard. Elastin has a unique spectrum that differentiates it from the collagens.

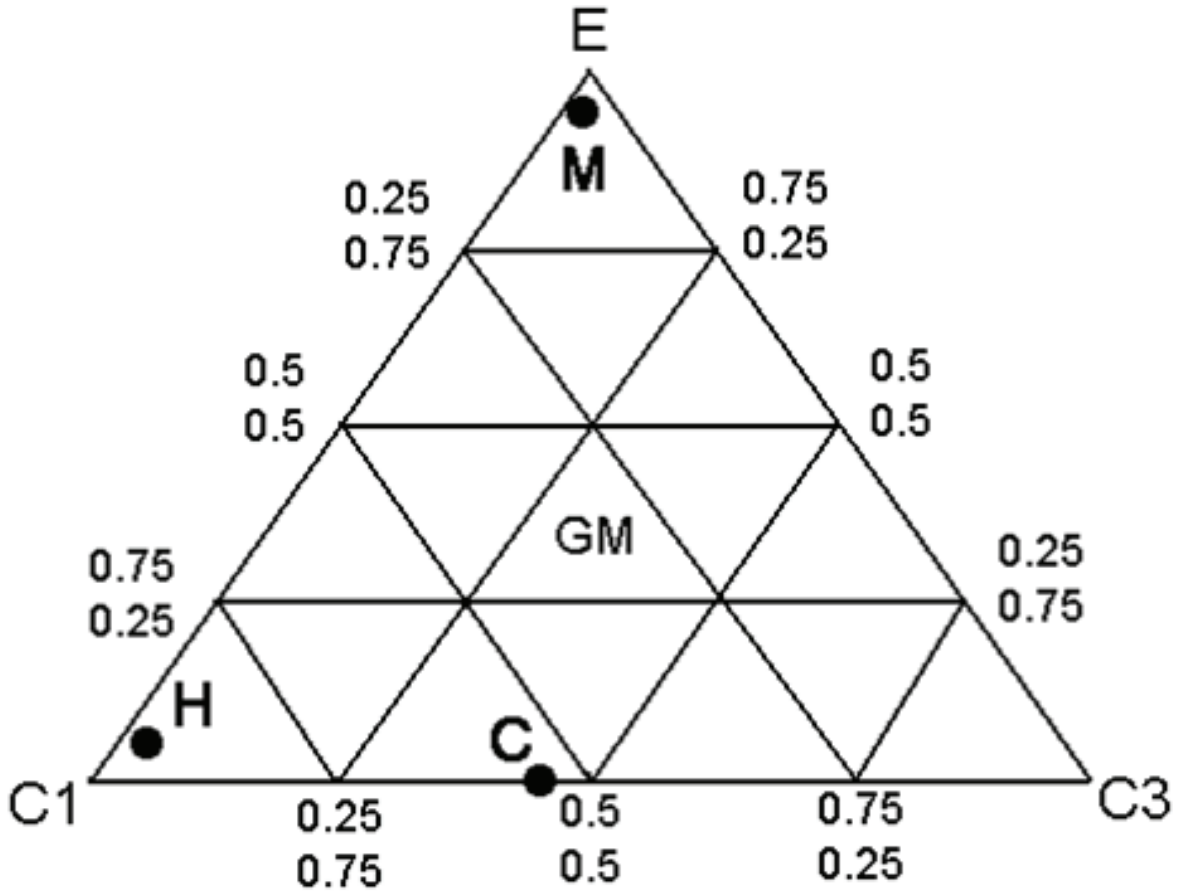

Figure 2.6: Diagram showing the composition of each of the prepared sample standards. The pure collagen I (C1) standard is in one corner of the triangle, the pure collagen III (C3) is in another corner of the triangle, and the pure elastin (E) in the remaining corner of the triangle. The concentrations of each of the three constituents in the standard mixtures were set at 0, 25, 50, 75, or 100% of each lyophilized protein. The vertexes in the triangle represent all possible combinations of protein mixtures in the percentages given (a total of 15 mixtures including the pure corner standards). GM, the group mean, is the center of the triangle and represents a mixture of 1/3 of each protein.

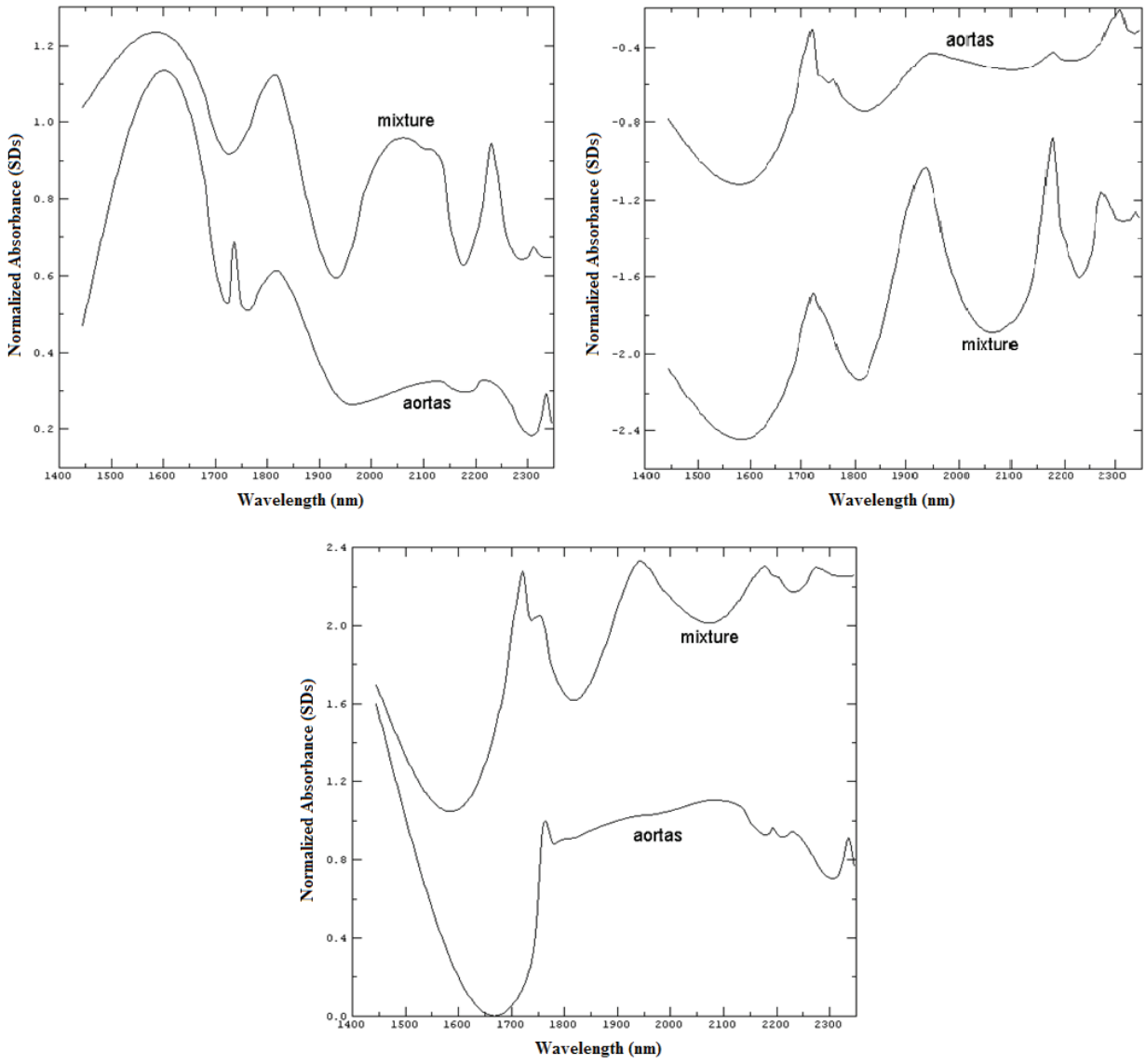

Figure 2.7: Spectra of the most highly correlated standard mixture of collagens and elastin and mean aorta spectrum for each treatment group: (top left) control group (saline infusion); (top right) low dose, 500 ng kg^{-1} min^{-1} Ang II; (bottom) high dose, 1000 ng kg^{-1} min^{-1} Ang II. The spectra in each treatment group are distinctive. The correlation between the aortic sections and the standard mixtures is the highest for the control and low dose Ang II groups.

Chapter Three – Molecular Factor Computing Near-Infrared Spectroscopy for Differentiating Cholesterol, Collage and Elastin through Red Blood Cell Solutions

Introduction

Extensive research has been conducted in the past two decades to develop methods for *in vivo* vascular imaging with a variety of aims including improved understanding of disease pathologies, monitoring therapeutic intervention, and early detection of potentially life threatening localized disorders (e.g., aneurysms, atherosclerotic plaques). One area that has received particular attention has been the early detection of vulnerable atherosclerotic plaques prior to vascular thrombosis and occlusion that lead to acute ischemic events.[97-106] The primary distinction between imaging modalities is whether they are invasive or noninvasive techniques. The ideal imaging system would be noninvasive and capable of structural and biochemical characterization of all atherosclerotic plaques; unfortunately, no such technique exists to date. The general consensus is that, for the near future, an effective system will require two or more complementary techniques. Among noninvasive techniques, magnetic resonance imaging (MRI) has received the most attention as a potentially useful tool in vulnerable plaque detection.[107-112] However, much of the research has been done *in vitro* and *in vivo* studies remain in the experimental stages. Trends point toward the need for an invasive approach (i.e., catheter) for effective plaque characterization.[112]

A number of invasive catheter-based methods are currently in use or under investigation. The leading technology at the present time is intravascular ultrasound (IVUS), which is based on the transmission and detection of high frequency sound waves.[98, 99, 113-118] IVUS has been shown to effectively distinguish between lipid core, fibrous cap, surface thrombus and calcification.

Extensive studies have been conducted on IVUS for plaque characterization and the determination of optimal therapeutic intervention but so far it has not been demonstrated to reliably detect unstable plaque or specific chemical changes. Optical coherence tomography (OCT) has also received a great deal of attention for plaque characterization.[119-122] OCT is an analogous to IVUS but uses an infrared source as opposed to sound. Tissue penetration is not as great as IVUS, but OCT can provide up to a two order of magnitude improvement in spatial resolution. OCT has been shown to useful in the detection of lipid pools and determination of fibrous cap thickness, but does not detect specific alterations in local elastin and collagen composition. Though powerful, the drawbacks to IVUS and OCT are primarily that the methods do not provide specific biochemical information, such as inflammatory cell content or collagen and elastin distribution, about the plaque surface.

Several approaches to address these issues have also been investigated but all are still in the developmental stages. Studies have shown that thermal heterogeneity of atherosclerotic plaques, attributed to activated inflammatory cells, are correlated with plaque disruption and thrombosis.[123-126] Thermography, which measures the local temperature of vascular tissue, has been proposed as a potential means for detection plaque vulnerability.[124, 127, 128] Spectroscopic methods have the potential to provide specific information about the chemical composition in atherosclerotic plaques due to the wealth of information obtainable from multivariate spectral data sets. Raman spectroscopy has been demonstrated as a useful technique for discriminating between a variety of plaque components *in vitro*[129, 130], and several studies have shown the potential for *in vivo* application.[131, 132] Fluorescence methods have also shown promise for detection of plaque components.[133] Specifically, *in vivo* studies have demonstrated the detection of macrophages and foam cells using this technique.[134-136] Near-infrared (NIR) spectroscopy has

been demonstrated as a useful tool for characterizing plaque components.[137, 138] In addition, the detection of cholesterol in aortas with a catheter system using NIR has also been reported.[139] The particular strength of the spectroscopic techniques just described is in their potential for qualitatively and quantitatively assessing biochemical constituents of plaques. However, the potential for three-dimensional, depth resolved imaging, as can be obtained with IVUS and OCT, of plaques *in vivo* is quite limited with current technologies. Their utility can be significantly enhanced when used in concert with other catheter-based instruments.

The following research represents preliminary work toward the development of a NIR-based catheter system using the novel optical encoding technique of molecular factor computing (MFC). Typical drawbacks to current spectroscopic based catheter systems is time consuming data collection and the need for the optical delivery system (i.e., fiber probe tip) to be in direct contact with the vessel wall. In this work we explore the use of MFC for differentiating plaque components (specifically cholesterol, collagen, and elastin) through red blood cell solutions. The potential advantages for this type of approach include significantly faster data collection and the ability to analyze the vessel wall without being in contact. This could provide a means for mapping vessel walls continuously instead of targeting a small number of specific points.

Materials and Methods

Traditional NIR Data Collection. The sample targets were created by encasing pure samples of cholesterol (Sigma-Aldrich, St. Louis, MO), collagen (Type I, Sigma-Aldrich) and elastin (Sigma-Aldrich) in cylindrical recesses (1 cm diameter, 2 mm deep) in anodized aluminum blocks. Ordinary microscope cover slips were permanently affixed to the blocks as windows to seal the samples inside. Traditional near-infrared diffuse reflectance spectra of the sample

targets were collected from 1100-2500 nm using a scanning monochromator with an external reflectance probe as described previously.[88, 90] Aliquots of rabbit red blood cell (RBC, 10%) solutions (Lampire Biological Laboratories, Pipersville, PA) were placed between the target and the window to simulate data collection through blood. RBC solutions were used in place of whole blood because the solutions are more stable and do not clot. The path length through RBC solutions was varied from 0.1 to 1.0 mm by placing the targets on a manual vertical translation stage (Melles Griot, Carlsbad, CA) while keeping the reflectance probe fixed. A schematic of the instrumental setup is presented in Figure 3.1. The resulting NIR spectra (n=26 for each component: cholesterol, collagen, elastin) were dominated by water absorption peaks and baseline offsets arising from the variable RBC solution path lengths. Two spectral regions (1150-1300 nm and 1640-1820 nm) were identified for discrimination of the samples by examining the second-derivative spectra. Figure 3.2 presents the mean second-derivative spectra from each group of samples. These two regions of the second-derivative spectra were more than sufficient to discriminate between the samples.

MFC Filter Selection. The details of molecular filter selection are described elsewhere[36, 37] and only a brief description is presented here. The chemicals used as molecular optical filters were found by searching a library of near-infrared transmission spectra from 1923 compounds (John Wiley & Sons, Inc., Hoboken, NJ). The library consisted of two spectra for each compound collected on slightly overlapping NIR regions, 952-1587 and 1388-2630 nm, with shorter path lengths used in the longer wavelength region. The two bands of interest in Figure 3.2, 1150-1300 nm and 1640-1820 nm, show markedly different degrees of reflectance is this system so it was decided to treat these separately in the MFC chemical selection routines. Molecular factor scores were simulated by calculating the dot products between transmission spectra from the

NIR library and the raw NIR reflectance spectra collected using the scanning monochromator. The simulated molecular factor scores allowed for a preliminary evaluation of the effectiveness of the chosen filter components. Note that two significant factors were not considered in these simulations, the spectrum of the broadband source and the detector response. A high degree of collinearity was observed in the simulated scores and was attributed to similarities in transmission spectra among chemicals and the dominant baseline differences observed in the raw reflectance data. To identify useful chemicals for MFC, the library was first filtered using a genetic-algorithm-linear-discriminant-analysis technique related to one described in the literature for variable selection for partial-least-squares regression.[140, 141] Approximately 200 compounds remained after filtering and a step-wise variable selection algorithm based on linear discriminant analysis was used to select eight MFC filters for differentiating the plaque chemical samples. The goal in the MFC filter-selection phase of this work was to develop a pathlength-independent model for differentiating the plaque chemical samples. The shorter wavelength region (1100-1350 nm) provided much better classification of the targets in the simulation, so chemicals from this region alone were selected. The disparity between the two regions was attributed to the significantly lower signal level through longer path lengths of the RBC solutions in the 1640-1820 nm band. Eight chemicals were selected for use as MFC optical filters: phenyl-hydrazine (PH), 2-dimethylbutyric acid (DMBA), 2,5-dihydrofuran (DHF), ethyl iodoacetate (EIA), tetramethylurea (TMU), thiopheneethanol (TE), 2,2-diethoxypropane (DEP), dicyclohexyl phthalate (DCP). Dicyclohexyl phthalate (DCP) was in powder form and dissolved in CCl_4 (0.5 g/mL). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO).

MFC Data Collection. A schematic of the instrumental setup used for MFC data collection is presented in Figure 3.3. A 250-W tungsten-halogen broadband source (Model 621, McPherson

Inc., Chelmsford, MA) with a 1000-nm long pass filter (Thorlabs, Newton, NJ) was used as the NIR source. The source beam was modulated with an optical chopper (Model SR540, Stanford Research Systems Inc., Sunnyvale, CA) before coupling to a 200-μm core optical fiber. The beam was exposed to the MFC optical filters using an aligned pair of collimating lenses (Model 74-VIS, Ocean Optics Inc., Dunedin, FL). A step-indexed cuvette tray was built in-house to permit the selection of cuvettes in the beam path. All cuvettes used were 1 cm path length optical glass. The sample targets were placed on a vertical translation stage as done for the traditional data collection described above. The termination end of the source fiber was fixed perpendicular to the target surface. A sealed 1 mm × 1 mm PbS detector was positioned adjacent to the fiber. Both the fiber termination and the PbS detector were submerged in RBC solutions as described above. For each of the eight MFC filters, the signal was integrated for three seconds. In this fashion, each sample factor "spectrum" consisted of 8 data points. Total data collection time for a single plaque chemical sample was approximately one minute. Measurements were taken with fiber tip-to-sample surface distances of 0 (fiber in contact with cover slip), 0.25 and 0.50 mm. Eighteen spectra were collected for each plaque chemical sample type (6 at each path length) giving a total of 54 MF spectra in the data set.

Data Analysis. All data analysis was performed using Matlab 7.0 (Mathworks, Inc., Natick, MA). Linear discriminant analysis (LDA) and principal component analysis (PCA) routines were written by the authors. PLS discriminant analysis (PLS-DA) was used for building classification models from the MFC data set. For the simulated and real MFC data sets, the data were preprocessed by autoscaling. The PLS-DA analysis was done using the PLS Toolbox (Ver. 3.5, Eigenvector Research, Inc. Wenatchee, WA) in Matlab. The performance of PLS-DA classification models was assessed by examining several metrics: *Accuracy*, *Precision*, and

Recall. In addition to the classification models for discriminating components, estimates of the detection limits for binary mixtures of these components were also calculated. The estimates were based on an extension of the BEST metric for sub-cluster detection with sample populations that has been described previously.[33, 34, 142-144]

Results and Discussion

Plots of the two significant canonical variables from LDA of the short wavelength region (1150-1300 nm) of the traditional NIR data (second-derivative spectra) and the simulated MFC data from the same region are presented in Figures 3.4 and 3.5, respectively. Figure 3.4 demonstrates that the samples can be clearly differentiated with traditional NIR data. The results from the simulated MFC data in Figure 3.5 demonstrate that, in principal, MFC should be able to achieve similar results using 8 chemical filters. Presented in Figure 3.6 are the two significant canonical variables for the actual MFC data collected through blood using the 8 chemical filters.

PLS-DA provided better classification accuracy than LDA for this data set and was therefore used for the subsequent processing of the data. PLS-DA models were built for classifying each type of sample based on the autoscaled MFC data. In all models, six latent variables provided the optimal classification accuracy. Due to the relatively small number of samples, the PLS-DA models were evaluated using a leave-one-out cross validation procedure. The cross-validation results are presented in a confusion matrix in Table 3.1. Several useful measures for evaluating the performance of the classification model are included in the table. The metrics for each classification model appear in the corresponding column of predictions in the confusion matrix for each sample type. In general, the cholesterol and collagen models performed fairly well while the elastin model was somewhat inferior.

It is apparent in Figures 3.4-3.6 that the separation of the three sample types observed in the simulation is diminished in the actual prototype instrument. There are several factors that likely contribute to the disparity between the simulated and the actual MFC performance. First, the instrumentation used for MFC data collection is still in the prototype stage. Electronic noise sources in instruments nearby were substantial, and relocation of the prototype soon to a new lab and/or a better choice of optical modulation frequency should improve the group separations. Although not comprehensively investigated, the MFC instrument showed more instrumental drift than the scanning monochromator used for traditional data collection. Rudimentary studies with the broadband source (e.g., 1000 nm filter in place but no MFC chemicals present) exhibited detector signal drift as high as 4% over one minute, which was the approximate time required to collect an MFC spectrum in this work. The drift problem was largely overcome using a double-beam design in the scanning monochromator system, and a similar approach should be used in the MFC spectrometer. To further increase signal in the monochromator system, the reflectance probe used an integrating sphere and had a 1 cm sample window aperture. In contrast, a 200-μm core fiber was used in the MFC instrument for light delivery, accompanied by a 1-mm^2 capture area from the PbS detector. The fiber probe dimensions in the MFC instrument were more consistent with use in a catheter than the dimensions of the monochromator instrument.

As noted in the Experimental section, several factors directly affecting MFC were not considered during the initial chemical selection phase. These included the broadband source emission spectrum, the PbS detector response, the pass-band spectrum of the 1000-nm long-pass filter, and the optical characteristics of the fiber used. The most significant deviation of the actual experiment from the theoretical model developed for the simulations resulted from the fact that a band pass filter was not available to isolate the 1100-1350 nm band of interest. With a

1000 nm long-pass filter the majority of the signal measured on the detector arises from this region. However, the contribution from light of longer wavelengths, primarily the region between the two strong water bands, is considerable. In addition, the radiance of the tungsten-halogen source and the sensitivity of the PbS detector are not flat across the near-infrared spectrum. The spectral output of the source was available, however, the response of the actual PbS detector used was not so an estimate was used based on typical PbS detectors. Ultra-low OH optical fiber was used and the transmission spectrum is essentially flat across the near-infrared region employed here. For comparison with the data collected, more realistic MFC scores were simulated by a convolution of the following: source spectral radiance, estimated detector sensitivity, transmission spectra of the eight MFC filters and the reflectance spectra of the samples over the 1100-2200 nm range. LDA was applied to these scores and the two significant canonical variables are presented in Figure 3.7. Comparison of Figures 3.5 and 3.7 demonstrates that the efficacy of the model suffers with the updated data set of simulated MFC scores, which agrees better with the experimental MFC data.

The experimental MFC data were then analyzed to estimate the limits of detection through blood for each component in binary mixtures of two components. For a particular component, this was performed by translating the sample population mean of another component towards this sample population's mean until the two clusters could not be differentiated using the BEST sub-cluster detection algorithm. The relative distribution about the mean of the sample population being translated was maintained in this procedure. The estimates of the detection limits in binary mixtures determined by this procedure are given in Table 3.2. For a specific element in this table, the reported detection limit is the percent composition of the column component that can be detected in a binary mixture with the component in the corresponding

row. Note that the table elements are not necessarily symmetric (i.e., the estimate of the amount of cholesterol detectable in collagen is not the same as collagen in cholesterol.)

Conclusion

This work demonstrated the potential of molecular factor computing near-infrared spectroscopy for differentiating biological components through a complex sample matrix. The ability to discriminate cholesterol, collagen and elastin samples through red blood cell solutions of varying pathlength with good accuracy using this method was established. This work supports the continued development of a spectroscopic catheter based system for the *in vivo* analysis of vulnerable atherosclerotic plaques and other vascular diseases.

Chapter Three Tables

Table 3.1: Classification results of the MFC experimental data from PLS-DA.

True Class	**Predicted Class**		
	Cholesterol	Collagen	Elastin
Cholesterol	18	0	5
Collagen	0	16	5
Elastin	3	3	15
Accuracy	94.4%	90.7%	75.9%
Precision	85.7%	84.2%	60.0%
Recall	100%	88.9%	83.3%

Table 3.2: Estimated detection levels of each component in binary mixtures by MFC.

Component	**Estimated Component Detection Level**		
	Cholesterol	Collagen	Elastin
Cholesterol	X	19%	23%
Collagen	16%	X	29%
Elastin	20%	29%	X

Chapter Three Figures

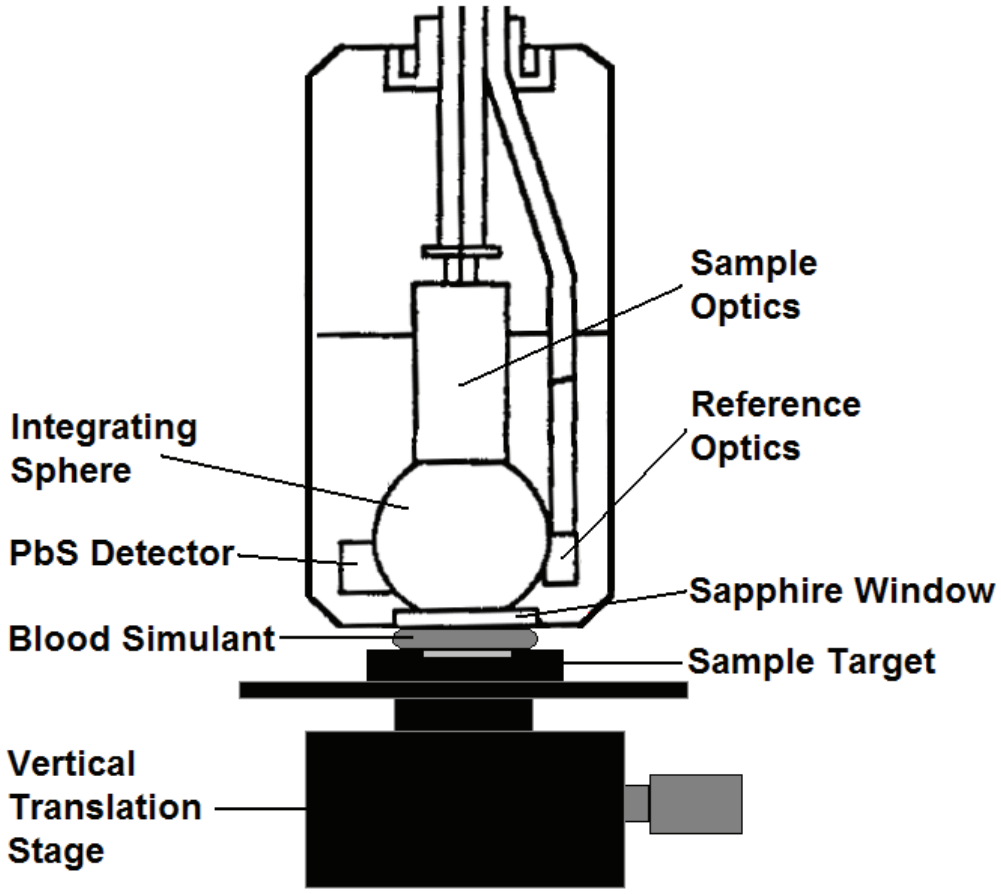

Figure 3.1: Instrumental setup for traditional near-infrared spectroscopic analysis.

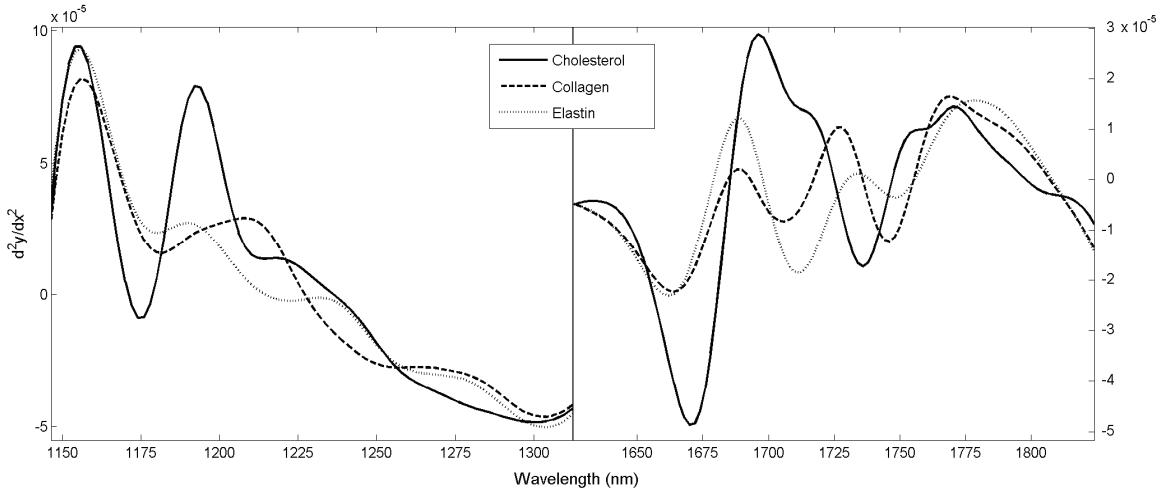

Figure 3.2: Mean second-derivative spectra of cholesterol, collagen and elastin through red blood cell solutions. The spectral regions shown here (1150-1300 nm: left panel, 1625-1825 nm: right panel) allow for differentiation between the components with solutions of red blood cells in the optical path.

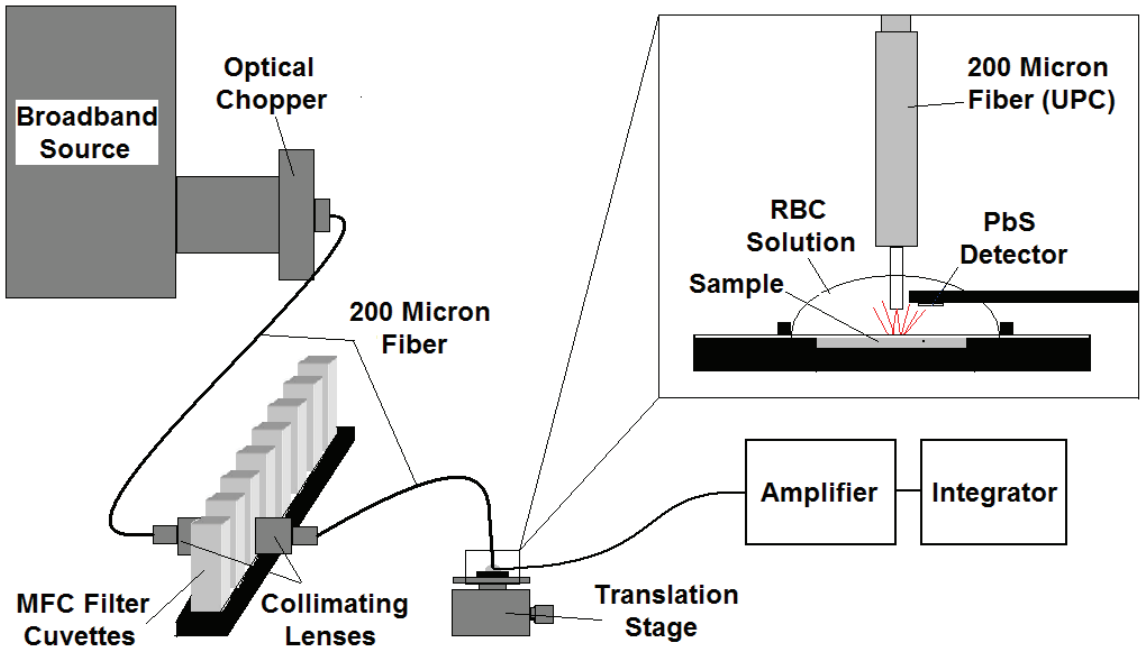

Figure 3.3: Instrument schematic of the molecular factor computing near-infrared spectrometer used in this work.

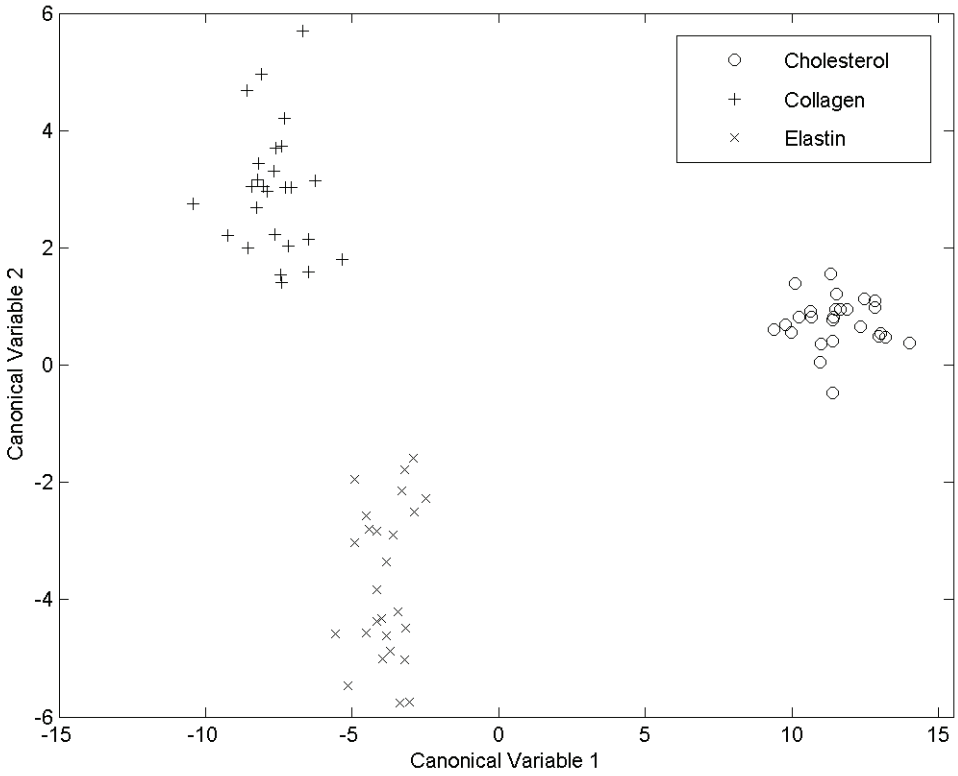

Figure 3.4: Canonical variables from the second-derivative near-infrared spectra of the sample targets in the 1150-1350 nm range.

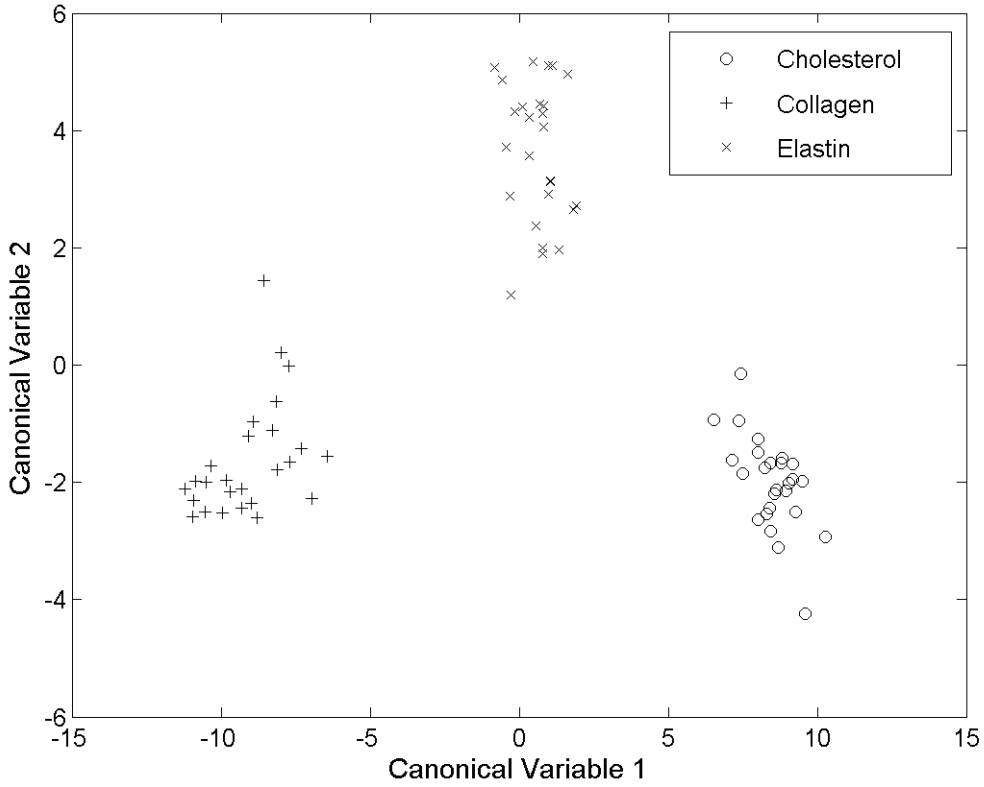

Figure 3.5: Canonical variables from the simulated MFC data of the eight chemical optical filters. This simulated data assumes an optical bandpass of 1100-1350 nm.

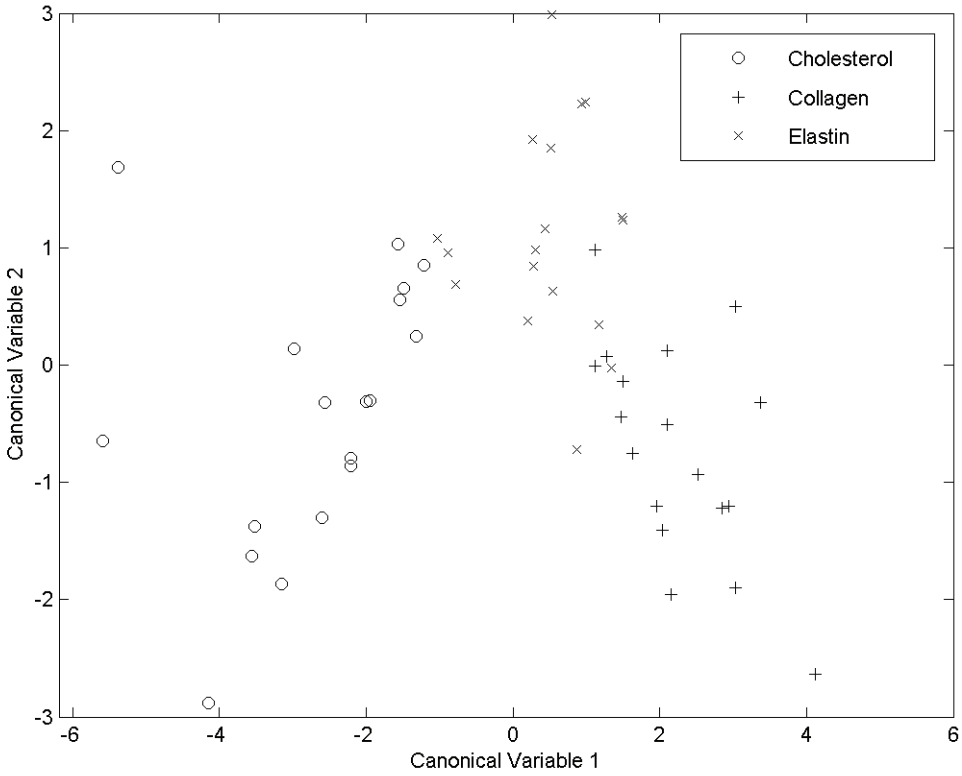

Figure 3.6: Canonical variables from the experimental MFC data using the eight selected chemical optical filters.

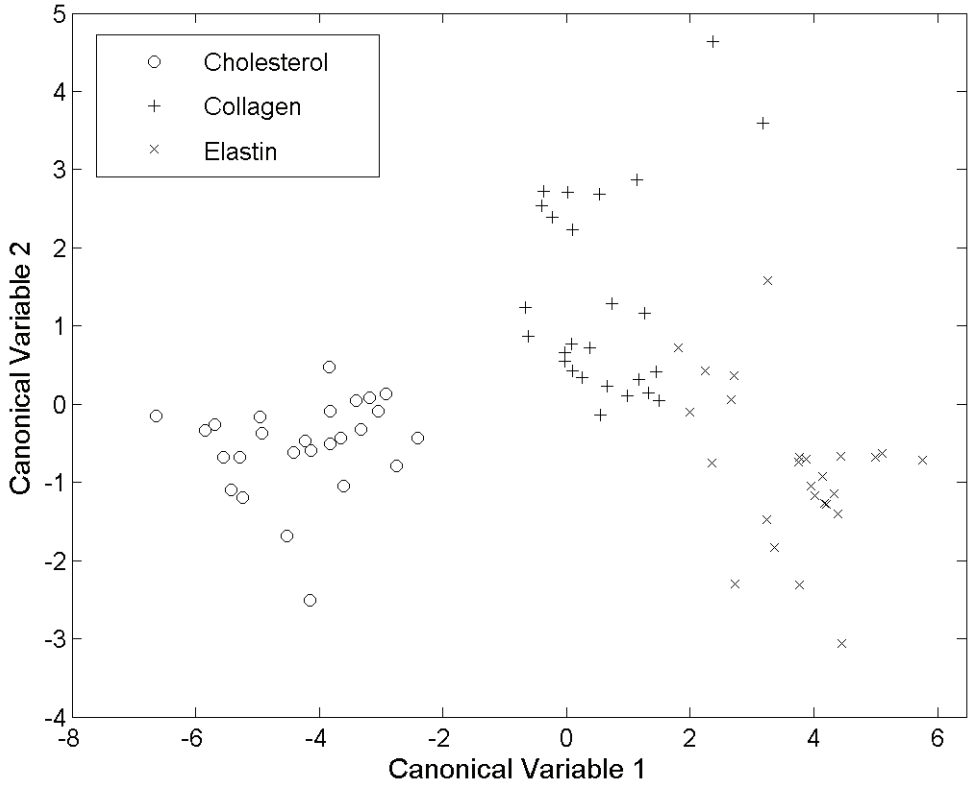

Figure 3.7: Canonical variables from the updated simulated MFC data to account for the extended optical pass band (1100-2200 nm), the spectral irradiance of the light source and the detector response.

Section III: Pharmaceutical Cleaning Validation

Cleaning validation is a methodology used to ensure the effectiveness of cleaning procedures to remove residues from manufacturing equipment to predetermined levels of acceptability and represents a critical aspect in the operation of multi-use pharmaceutical facilities. The primary goal of cleaning validation is to eliminate cross contamination between active pharmaceutical ingredients (APIs) that can lead to potentially harmful consequences for the drug recipients, but the removal of other substances such as solvents, micro-organisms, detergents and excipients is also of concern. Acceptable residue limits are generally based on the potency/toxicity of the API and are typically specified in mass/area dimensions ($\mu g/cm^2$ being customary). The conventional approach to cleaning validation utilizes either direct (swabs) or indirect (rinse matrix) methods to sample surface residues. This is followed by analysis using a traditional analytical technique, with HPLC based methods making up the majority. Although effective, these methods are not without shortcomings that include cumbersome manual sampling, incomplete analyte recovery (from both surfaces and swabs), repeatability issues, difficult method validation (particularly with rinse matrix methods) and time consuming analysis. The last of these, analysis time, is of particular consequence as down times during changeover periods can be up to several days while equipment is being cleaned and validated.

There has been a renewed interest in the development of new cleaning validation methods that are rapid and economical while maintaining or improving effectiveness compared to conventional approaches. This interest has been prompted in part by an increased attentiveness from regulatory bodies[145-147] and the encouragement of innovation brought about by the FDA's process analytical technologies (PAT) initiative.[148, 149] The FDA broadly defines

PAT as "*a system for designing and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes for raw and in-process materials and processes, with the goal of ensuring final product quality*". The principal goal in PAT is the incorporation of quality control throughout the entire manufacturing process to establish a quality by design paradigm. In concert with the PAT initiative, the FDA relaxed it's historically strict constraints on acceptable quality control methods (>80% of pharmaceutical assays are HPLC based) in an effort to facilitate PAT integration. The pharmaceutical industry has viewed the PAT initiative as a means of achieving better quality assurance but at the same time productivity improvements and reduction in operating costs. As a result, there has been a surge in the promotion of and investment in new technologies to accomplish these goals. Several strategies for the realization of PAT objectives have been particularly emphasized: the development of rapid process analytical chemistry (PAC) techniques for in-line and at-line measurements; utilization of multivariate statistical tools (chemometrics); and real-time information management systems for process control.

Several analytical methodologies have been proposed as PAC approaches to pharmaceutical cleaning validation. One method that has grown in popularity in recent years is total organic carbon (TOC) analysis.[150-152] In addition to faster analysis time, TOC is a non-specific method that allows for simpler method development compared to HPLC. The non-specific nature of TOC necessitates a worst-case interpretation of the outcome, though, where all carbon content is assumed to derive from the API or cleaning agents even when harmless excipients may be present. In addition, TOC requires an aqueous solvent that limits its applicability. Current applications of TOC typically involve surface swabbing for residue sample, but it is increasingly being promoted for in-line use, where automated processing by

sampling from a rinse matrix is performed. More recently, ion mobility spectrometry (IMS) has been introduced as a potentially useful method for cleaning validation applications.[153-155] IMS offers a rapid analysis time compared with HPLC without sacrificing chemical specificity as in TOC. While the selectivity achieved by IMS is not as good as that of HPLC, it may be sufficient for most cleaning validation applications, but this remains to be demonstrated in specific applications. IMS requires surface swabbing, which requires an additional solvent extraction step, or a rinse matrix for residue sampling. In general, the uncertainties associated with the residue sampling steps in cleaning validation protocols are significantly greater than the inherent uncertainty associated with the analytical methods.

An ideal validation technique for cleaning procedures would be an automated, rapid, *in situ*, multi-component analysis of the surface. The limitations and uncertainties associated with the swab and rinse matrix sampling techniques could be overcome with this type of instrumentation. Photonic methods have the potential to provide this type of assay. Recent work using mid-IR spectroscopy has resulted in the development of a fiber optic-based spectrometer capable of direct spectroscopic surface analysis by way of a grazing-angle reflectance sampling head.[156-159] Reported assessments of the device have demonstrated low detection limits for surface contamination as well as improved accuracy compared with a traditional swab-based HPLC approach. Drawbacks with this approach include cumbersome instrumentation and cost. The following three chapters detail efforts toward the development of competing *in situ* cleaning validation technologies. The use of light scatter and polarization for detecting and quantifying surface residues on glass and stainless steel surfaces is explored. While effective, chemical selectivity with these methods is limited. To address this issue, the use of NIR spectroscopy and MFC is also investigated.

Chapter Four – Quantification of Albumin (BSA) on Glass by Laser Scatter for Cleaning Validation

Introduction

This work represents initial investigations to develop a multi-functional spectral analyzer capable of illuminating a spot or area on a surface and examining the returned light for multiple spectroscopic phenomena. Multiplexing spectroscopic techniques offers the potential to create an accurate surface scanning instrument capable of multi-component analysis with wide applicability in cleaning validation and other areas. This idea was partly inspired by a similar device, a multi-functional active excitation spectral analyzer (MAESA), developed for materials characterization on planetary probes.[160] The focus of the present work is the prediction of Bovine Serum Albumin concentrations on the surface of glass by examining the scatter from an incident HeNe laser beam (632.8 nm).

Lord Rayleigh determined that the irradiance of the scattered radiation is given by Equation 4.1:

$$(E_{sc})_\theta = \frac{8\pi^4 \eta_0^2 (\alpha')^2 (1+\cos 2\theta) E_0}{\lambda^4 d^2} \qquad \textbf{4.1}$$

where α' is the polarizability of the particle in m^3, λ is the wavelength of incident radiation, θ is the angle between the incident and scattered ray, E_0 is the incident beam irradiance, and d is the distance from the center of the scattering to the detector, and η_0 is the refractive index of the solvent. For measuring the scatter from molecules in solution, the substitution in Equation 4.2 can be made:

$$\alpha \approx \frac{M\eta_0\left(\frac{\partial \eta}{\partial c}\right)}{2\pi N} \qquad \textbf{4.2}$$

where M is the molecular weight, N is Avogadro's number, and $(\partial\eta/\partial c)$ is the refractive index change of the solution with concentration. For a sample with concentration, c, and N_c/M particles per unit volume, substitution into Equation 4.1 yields Equation 4.3 for the scattered radiation per unit volume:

$$(E_{sc})_0^{'} = \frac{2\pi^2\eta_0^2\left(\frac{\partial \eta}{\partial c}\right)^2(1+\cos 2\theta)E_0 cM}{N\lambda^4 d^2} \qquad \textbf{4.3}$$

where the variables are defined as above. The system described by Equation 3 is different from that studied in this work, but provides insight into the factors involved in scattering at surfaces. For cleaning validation applications, several of the terms in Equation 3 are particularly important. The solvent in this case has been evaporated, and the volume is nearly two-dimensional. At low BSA concentrations, most of the interactions in a scattering event are with the glass surface. As the concentration increases, the probability of interactions with adjacent BSA molecules increases. This situation can viewed in a simple way as a molecule in a "solvent" of like molecules. In this view, $(\partial\eta/\partial c)$ is still applicable and introduces a quadratic dependence on concentration for scattered radiation. This phenomenon was observed and as a result quadratic regression models were used, providing significantly better predictive models. The M term is also important, and shows that the amount of light scatter increases with molecular weight. Finally, the λ^4 term exemplifies the inverse dependence on wavelength and the fact that sensitivity to particle detection can be increased by using shorter wavelength light.

Materials and Methods

Materials. The Bovine Serum Albumin (BSA) used in this work was obtained from Sigma-Aldrich (St Louis, MO). The microscope slides used were Teflon printed 24.4 by 16.7 mm single oval well slides obtained from Electron Microscopy Sciences (Fort Washington, PA). All other reagents used were of analytical grade.

Sample Preparation. A series of dilute solutions of varying BSA concentration were prepared in volumetric flasks by dissolving dry BSA in deionized water. For each slide sample, a 200-μL aliquot was taken from the appropriate BSA solution and spread as evenly as possible into the well on the slide surface and allowed to dry. In this fashion, a series of slides were prepared with known average surface BSA concentrations. A total of 11 slides were prepared with surface BSA concentrations ranging from 0-20 $\mu g/cm^2$.

Instrumentation. A schematic of the instrumental setup used in this work is presented in Figure 4.1. A 10-mW helium-neon (HeNe) laser with a 632.8 nm nominal wavelength (Spectraphysics, Eugene, OR) was used as the source to generate light scatter from the slide samples. A white image plane was placed in the path of the scattered light to provide a fixed background for imaging. The laser and image plane were separated by approximately 1 m with the slides placed 0.3 m from the laser. A spatial filter was placed in the beam path before interaction with the slide surface to eliminate stray light output from the source. An aperture in the image plane allowed the bulk of the laser light (i.e., the unscattered portion of the light) to pass through to a beam dump. This provided a significant improvement in sensitivity to scattered photons by limiting the dynamic range of the scattering signal to be captured by the A/D. An Olympus D-520 Zoom digital camera (Olympus Optical Co., LTD., Tokyo, Japan) was used to collect the light scattering images and was fixed at approximately 45° to the image plane and 0.7 m away.

Data Collection. Light scatter images were collected in a raster fashion from the slide surfaces by mounting the slides on a translation stage while keeping the rest of the instrumentation fixed. The translation stage provides precise, reproducible locations and allows direct comparison of light scatter images from different locations on the slide surface and between different slides. A preliminary study was carried out to optimize the angle of the laser with respect to the slide surface. For this investigation, the angle of the slide in the path of laser (i.e., θ in Figure 4.1) was varied with twenty images collected at each position. Only three samples (0, 0.80, and 20.0 $\mu g/cm^2$) were used for this initial study due to large volumes of data generated. For quantitative studies, thirty images were collected from each slide. All images were collected as 1600 X 1200 pixel JPEGs. Examples of light scatter images from two slides are presented in Figure 4.2.

Data Analysis. Due to computer memory constraints, images were resampled to a smaller size for some investigations. The effect of image size (i.e., resolution) on prediction accuracy was investigated and will be discussed below. Prior to analyzing images, one-dimensional data vectors were obtained by extracting the red channel from the reconstructed RGB bitmaps and aligning rows of the image end to end. An example of the data vector extracted for the light scatter image in panel A of Figure 4.2 is presented in Figure 4.3. The image was resized to 400 X 300 pixels prior to data extraction. To reduce data dimensionality, principal axis transformations were performed for quantitative analysis of BSA concentrations on the slide surfaces. As mentioned in the introduction, quadratic fitting models were used for regression. A thorough treatment of these techniques can be found elsewhere.[94, 95] For the angle optimization experiments, image group separation was examined by the Bootstrap Error-adjusted Single-sample Technique (BEST). This method is used to determine multi-dimensional distances of

individual samples in standard deviations from a population. A thorough treatment of this technique can be found in the literature.[31-33, 161]

Results and Discussion

Laser Beam Angle. The angle (θ) of the laser beam to the slides surface was investigated to maximize the between-group separation of light scatter images from three of the slide samples (0, 0.80, and 20.0 μg/cm^2). The result of this process is an improvement in sensitivity. Light scatter images were collected on the three slides at 90, 60, 45, 30, and 15°. Angles below 15° could not be investigated because of significant interfering light scatter caused by the edges of the Teflon printed wells on the slides. Group separation was examined by finding the average BEST distance between the image groups for each slide. BEST distances were calculated by treating the images from one slide as a population and finding the average distance to the individual images of the other groups.

The results of this analysis are presented in Table 4.1. The column labels identify the slides being compared, where the first slide is treated as the population and the distances are found to each of the images from the second slide. For example, "Blank vs. Low" gives the average BEST distances of the 0.80 μg/cm^2 slide images to the group of blank slide images. Note that the treatment of each group as a population will generate two sets of distances between each of the slide samples at each angle. Although not universal, there is a trend toward increased separation as the angle decreases. The two exceptions to this occur when finding distances from the 0.80 μg/cm^2 (Low) slide groups. These show an initial decline in average BEST distance to the other samples as the angle decreases. Below 45°, though, the group separation increases again. The most important observation drawn from the table is that the average separation

reaches a maximum in all but one of the comparisons at 15°. In the one exception to this (Blank vs. High), the separation is more than adequate at 15°. This angle was selected for the subsequent quantitative BSA determinations.

Image Resolution. The effect of image resolution on the BSA prediction accuracy was studied by finding the standard error of predictions (SEP) at a number of different image sizes. Images were resized individually and all thirty images for each sample were averaged. BSA concentrations were predicted for each slide by cross-validation using principal component regression. The SEP versus image size is presented in Figure 4.4. As seen in the figure, the prediction accuracy remains relatively stable down to an image size of 25 X 19. Here the prediction error begins to increase and reaches a maximum when the mean pixel value is used for each image. For all of the multi-dimensional data sets (i.e., all image sizes except the mean pixel value) two principal components were significant in the regressions. These results demonstrate that knowledge of the spatial distribution of the light scatter significantly improves prediction of BSA concentration on the slide surfaces with the instrumental setup used in this work. The improvement when the spatial distribution of light scatter is considered at sufficiently high resolution is approximately an order of magnitude.

Studying the effect of image size as done here was an attempt to simulate data collection at lower resolution. The block-averaging used here to resize images is a crude method with respect to image efficacy in a visual sense. Several other image resizing methods were also investigated, including nearest-neighbor and interpolation routines. The prediction accuracy for all of these methods was comparable with the block-averaging method at higher image resolutions but was significantly worse in general at resolutions below 25 X 19. Regardless of the method employed, it was clear that the initial image resolution of 1600 X 1200 was

considerably greater than necessary. More complex systems, with multiple analytes present, will likely require higher resolution imaging in conjunction with other spectroscopic techniques to achieve acceptable precision and accuracy.

Number of Points Sampled on Glass Surface. In the investigation of image resolution, all thirty images from each slide sample were averaged to generate a single representative BSA-distribution image for each sample. Prediction accuracy was next investigated as a function of the number of light-scattering images averaged to generate each BSA-distribution image. For this study, images were resized by the block-averaging method to a resolution of 800 X 600 pixels. A variable number of images were drawn at random without replacement from the groups of thirty images. These sub-samples were averaged for each slide and the SEP was found by cross-validation using principal component regression. To gauge the uncertainty in the SEP, the procedure was repeated 10 times for each sub-sample size. The number of images drawn was varied from 1 to 25. The average SEP versus the number of images in the sub-samples is presented in Figure 4.5. Note that the SEP is also included for the average of all thirty images. No error bars are present for this point, however, because there is only one possible sample. As evident in the figure the SEP, as well as the uncertainty in the SEP, decreases as the number of images averaged increases. These results are expected because of the likelihood of uneven BSA distribution on the slide surfaces, so averaging light-scattering images from more locations on the slide's surface will generate a better representative BSA distribution image. The conclusions drawn from Figure 4.5 are in part confounded by the fact that the samples are all drawn from the same populations. As the number of images sampled increases, the probability that images appear in multiple replications increases. This repetition has a more significant influence on the

observed trend of decreasing uncertainty in the SEP as the sample size increases than on the observed concurrent trend of decreasing SEP.

The uneven BSA distribution on the surfaces can be visualized by examining the light scatter images in relation to their collection points on the slide surface. To reduce complexity, the average red pixel value for each image was used to provide a rough estimate of the BSA present at a given location on the slide surface. These values were arranged in a grid according to their sampling location. A smooth two-dimensional spline surface was fit to the data to approximate the distribution of BSA in the slide well. Figure 4.6 is an example of this for the slide sample with an average surface BSA concentration of 4.8 $\mu g/cm^2$. The values of the spline surface approximation at the sampling locations (black dots in the figure) are within 3% of the actual values. The figure demonstrates the irregular distribution of BSA on a slide and exemplifies the need to sample numerous locations in order to get an accurate prediction for the average concentration.

Prediction of total BSA concentration on each slide surface using the averages of all thirty light-scattering images gives ng/cm^2 errors. Figure 4.7 is a graph of the true versus predicted BSA concentration using all light-scattering images. To further illustrate the conclusions from the earlier section on image resolution, the predicted concentrations are plotted for both 800 X 600 (black) and 50 X 37 (red) pixel image sizes. The differences between the results for the two resolutions are imperceptible in the graph for several of the samples, and quite small for the others. For the 800 X 600 images, the SEP and average SEE are 0.154 and 0.114 $\mu g/cm^2$, respectively (cross validated using the F-test at $p = 0.05$). For the 50 X 37 images, the SEP and average SEE are 0.135 and 0.098 $\mu g/cm^2$, respectively (cross validated using the F-test at $p = 0.05$).

The major limitation to this work is that BSA was the only component present on the slide surfaces. Plans for future work include analyzing multi-component systems containing APIs, excipients, and cleaning agents. It is highly unlikely that single wavelength light scatter will be sufficient to characterize such systems completely, and therefore development of a hyperspectral MAESA device will be pursued. Other spectroscopic techniques including polarization modulation and luminescence will be studied in addition to light scatter. These techniques should help in the analysis of contaminants on other surfaces such as stainless steel.

Conclusion

This work demonstrated that examination of laser light scatter can be used to quantify BSA concentrations on glass surfaces. Analysis of the spatial distribution of light scatter provided improved prediction accuracy compared to the prediction by the amount of light scatter alone for the instrumental setup employed in this work. When time is limited, if faced with a choice between collecting higher resolution scattering images or sampling more locations on a glass surface, the time is better spent sampling more locations on the surface. In addition, the sensitivity to low concentrations of BSA can be enhanced by decreasing the incident angle of the laser beam with respect to the glass surface.

Chapter Four Tables

Table 4.1: Mean BEST distances between image groups for selected samples at various laser beam angles to the slide surface.

Angle (°)	BEST Distances, Mean (±SD)					
	Blank vs. Low	Blank vs. High	Low vs. Blank	Low vs. High	High vs. Blank	High vs. Low
90	1.616 ± 1.074	3.299 ± 0.863	5.713 ± 4.158	6.927 ± 1.991	3.991 ± 2.876	2.768 ± 0.416
60	2.504 ± 1.377	5.366 ± 0.839	4.134 ± 2.548	5.457 ± 0.932	3.647 ± 0.870	3.158 ± 0.555
45	2.623 ± 1.180	6.062 ± 0.797	2.571 ± 2.103	4.262 ± 0.653	4.135 ± 0.887	3.773 ± 0.852
30	2.895 ± 0.750	6.479 ± 1.191	4.283 ± 3.061	6.265 ± 1.092	4.319 ± 0.475	3.551 ± 0.517
15	4.341 ± 2.120	5.888 ± 0.935	6.020 ± 3.540	7.105 ± 0.715	7.268 ± 1.120	5.709 ± 0.877

Chapter Four Figures

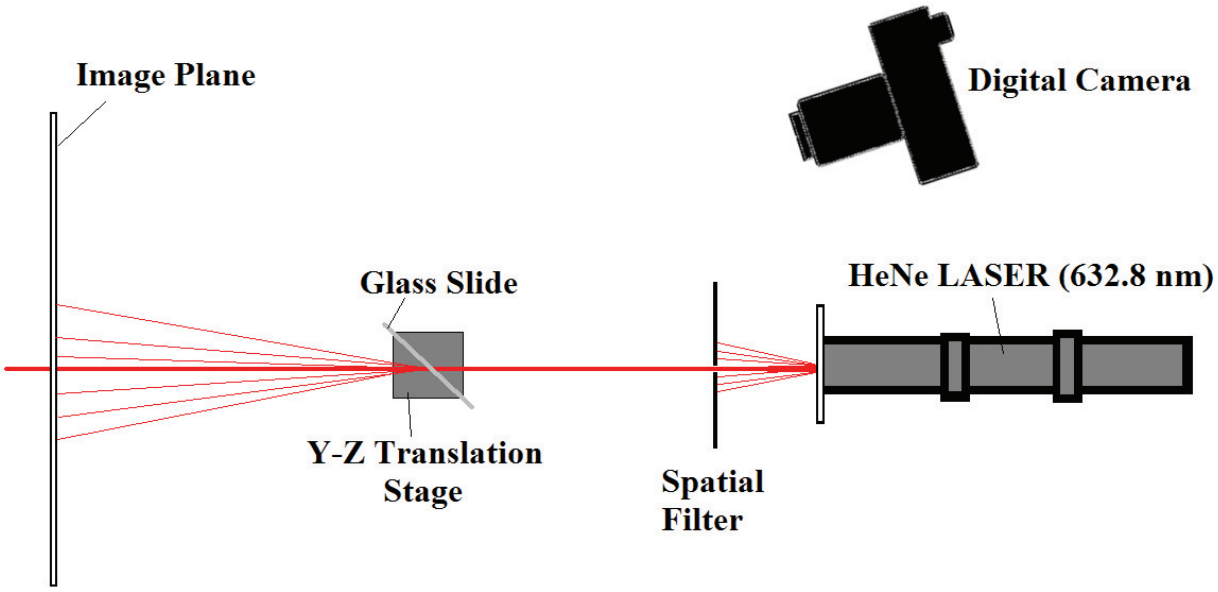

Figure 4.1: Instrumental setup (top view) used to collect light scatter images from slide surfaces.

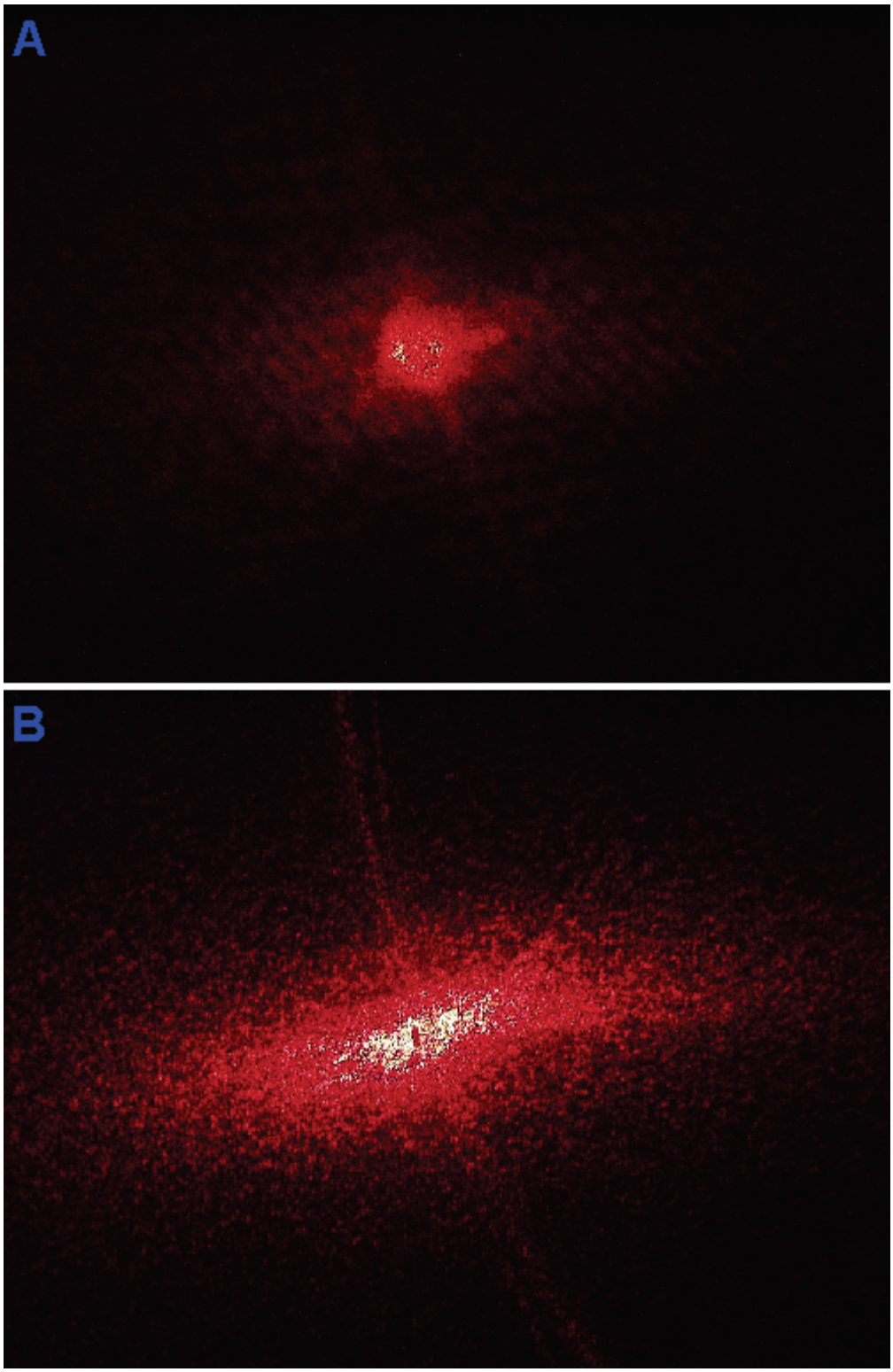

Figure 4.2: Examples of individual light scatter images collected from two slide samples: 0.8 μg/cm^2 (panel A) and 20.0 μg/cm^2 (panel B).

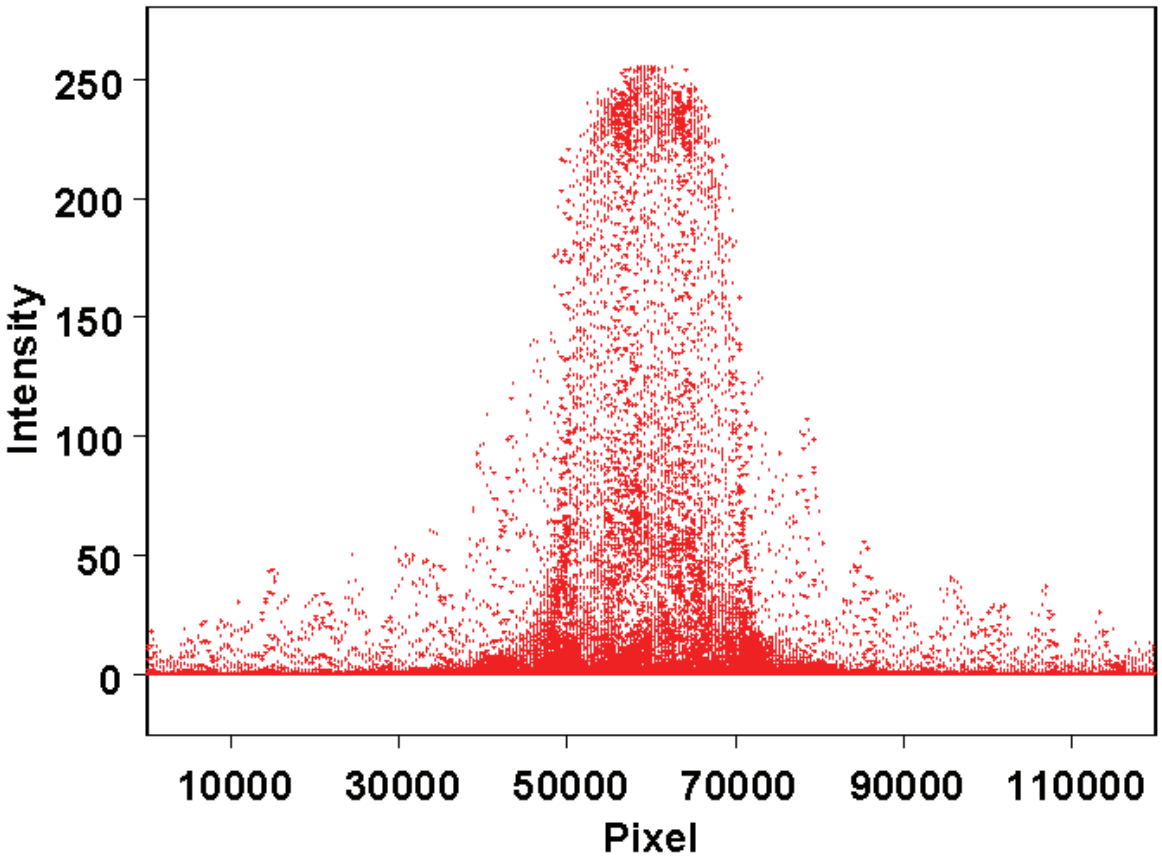

Figure 4.3: An example of a data vector extracted from a light scatter image by retaining only the red channel of an RGB image and representing the two-dimensional image as a vector. This data was extracted from the image shown in panel A of Figure 4.2.

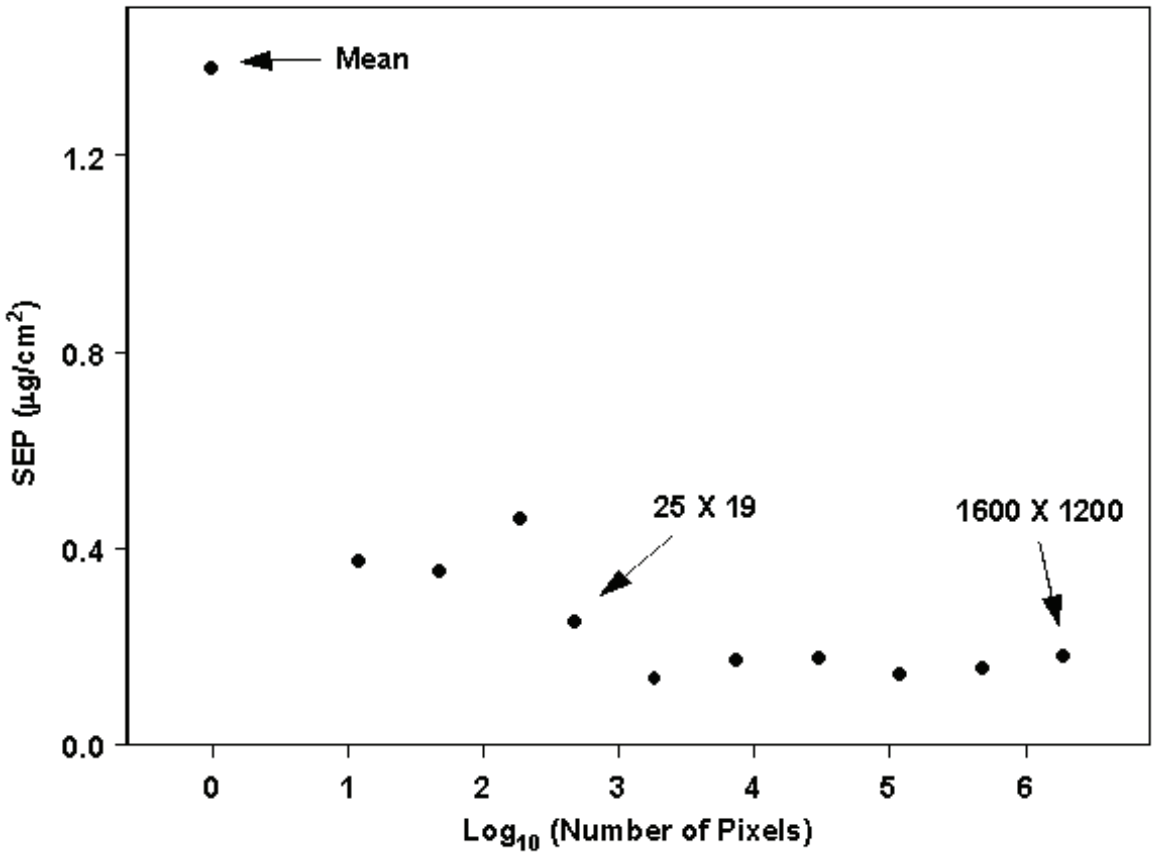

Figure 4.4: The SEP for BSA concentration as a function of the resolution of the light scatter images.

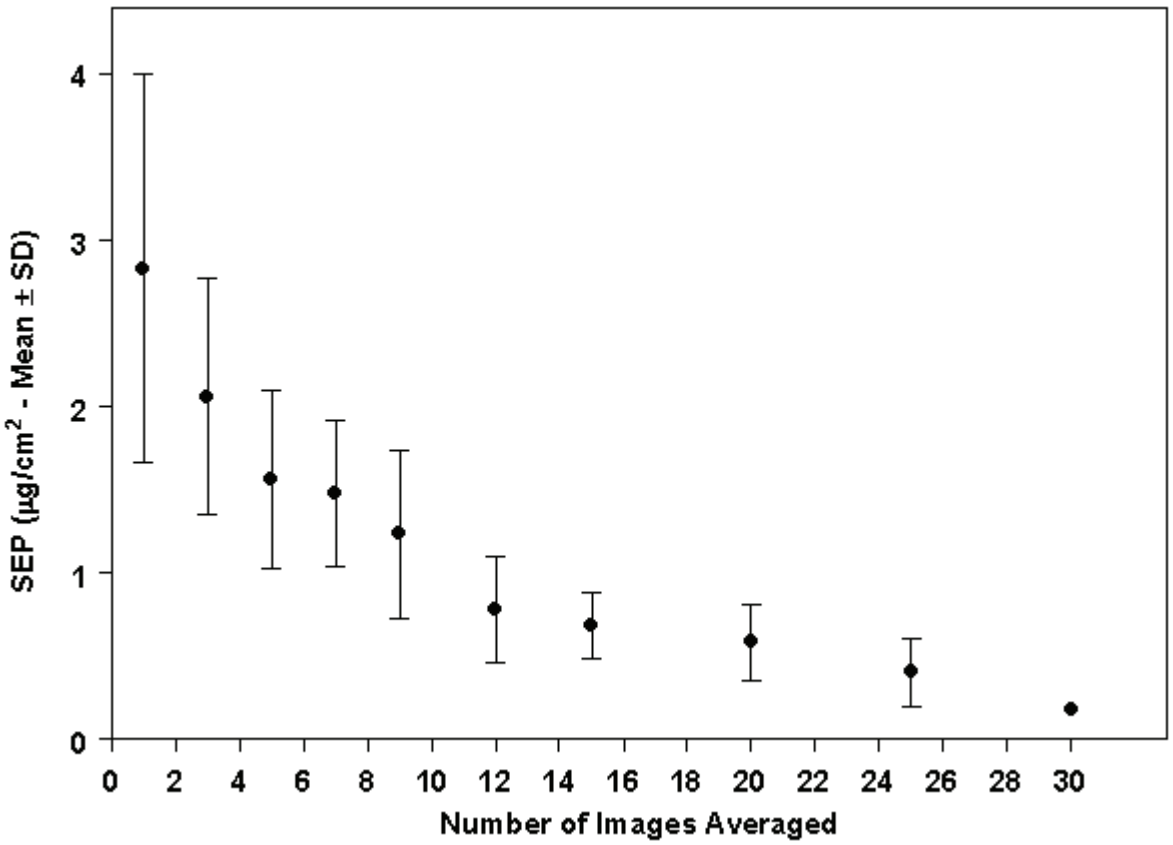

Figure 4.5: The SEP for BSA concentration as a function of the number of locations sampled to generate a representative average image for an individual slide.

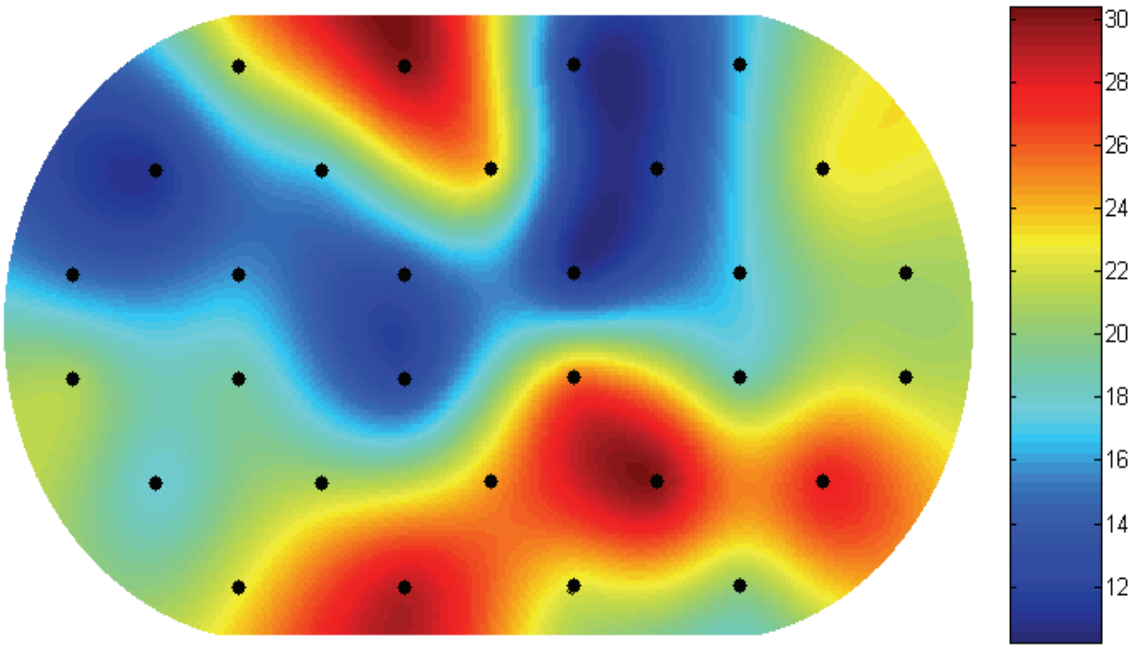

Figure 4.6: Estimation of the distribution of BSA in the well of a slide sample with an average surface BSA concentration of 4.8 μg/cm^2. Light scatter is indicate by color, increasing from blue (least scatter) to red (most scatter), and is estimated by finding the average red pixel value for each image. Increased scatter is interpreted as an approximation to BSA concentration at a given location.

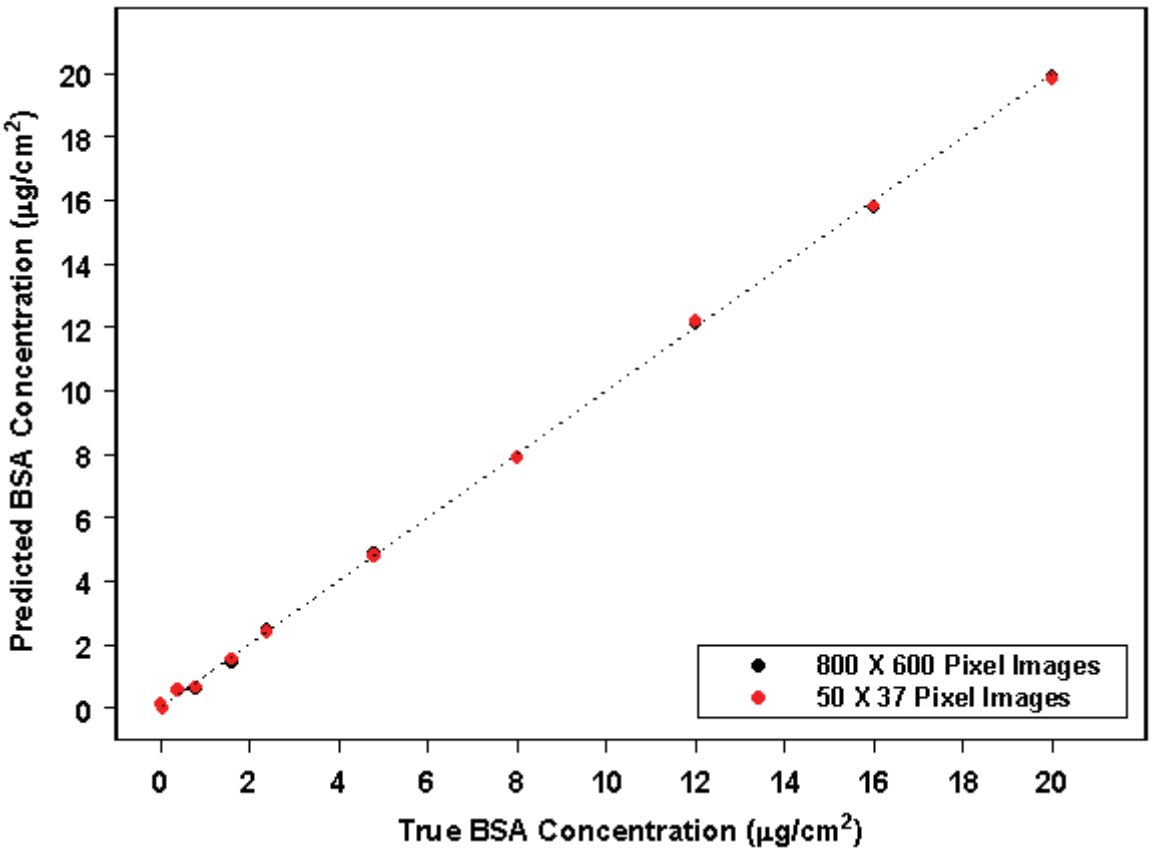

Figure 4.7: True versus predicted BSA concentration using the average of all 30 light-scattering images for each slide at two different image resolutions: 800×600 (black) and 50×37 (red). The dotted line corresponds to a perfect correlation.

Chapter Five – Anisotropic Polarized Light Scatter Imaging for Pharmaceutical Cleaning Validation

Introduction

The previous chapter demonstrated the use of laser light scatter for the detection of surface residues on glass. Although this method potentially offers detection limits at pharmaceutically relevant concentrations the utility of it quickly becomes limited when other sources of light scatter exist. The discrimination between contamination in the presence of surface imperfections, such as roughness or abrasions, is largely unachievable by examining light scatter alone. The previous investigation involved unblemished glass surfaces with surface contamination as the principal scattering source. This type of surface is not necessarily representative of a typical situation encountered in pharmaceutical cleaning validation. Polished stainless steel is the predominant material used in the manufacture of pharmaceutical processing equipment. A significantly greater amount of scatter is expected from these surfaces due to increased surface roughness and the lack of transparency compared to glass. In addition, scatter arising from abrasions must be taken into account in a realistic application.

Optical scattering has proven to be a valuable technique for the characterization of smooth surfaces such as mirrors, optics and silicon wafers.[162] In these applications, light scatter is primarily used to identify deviations from ideal conditions, locate defects and characterize surface roughness. More recent studies have shown that the polarization of scattered light can be used to identify the scattering source.[163-171] These investigations have been aimed at the discrimination between light scatter from surface roughness, subsurface defects and particle contamination on smooth surfaces. The applications for which these methods are being explored

include testing of optical materials, silicon wafers, patterned semiconductors and optical storage media. In general, the manufacturing tolerances associated with these surfaces are much more stringent in comparison with what is expected in pharmaceutical equipment but the results in the optics and electronics industry provide a foundation for the pharmaceutical objective. Mathematical treatments of polarized light scatter in these works are daunting and normally necessitate the use of simple theoretical systems to obtain predictions. Typical models of polarized light scatter from particles on surfaces involve models of single spherical or slightly non-spherical particles on smooth surfaces, with the additional constraints that the particles are sufficiently smaller than the wavelength of light and multiple photon interactions with particles are ignored. The resulting models generally exhibit dependencies on particle shape, size, refractive index and wavelength. Consequently, experimental investigations to substantiate these models are typically carried out using particles of uniform size and shape with monochromatic radiation incident on sufficiently small surface areas, which is usually accomplished with laser sources. Models that are more complex have been developed or extended for studying light scatter from surfaces with more intricate systems that incorporate irregular particle shapes as well as agglomerates, but polarization has generally been ignored until quite recently.[172-177]

This work represents a preliminary investigation into the use of polarized light scatter for pharmaceutical cleaning validation applications. The specific aim of this study is to establish its potential utility from an experimental standpoint using a system that is representative of a realistic cleaning validation application. The investigations here were motivated by studies demonstrating that the light scattered from *p*-polarized light incident on a surface at an oblique angle has a polarization that is dependent on the scattering source.[167-169, 178] In particular these studies have shown that the polarization of out-of-plane light scatter shows directional

dependency that can be used to differentiate between roughness, subsurface defects and particulate contamination on the surface. In this work, a simpler instrumental setup is employed than what is used in typical light scatter experiments. Collimated white light from a broadband source is used for illumination. To facilitate a quantitative determination of surface residues, a substantially greater surface area (a 6.5×4.6 cm ellipse) is illuminated than in traditional setups, where point illumination by laser sources is generally used. In the end, illuminating a larger area can significantly speed up the cleaning validation process. The scattered light is imaged with a digital camera to maintain the spatial dimensions of the surface. This instrumental setup introduces two compromises in the polarization data gathered. First, the use of broadband illumination confounds the wavelength information. Second, imaging of the surface results in light being collected in a significantly broader angular band compared with single element detectors. In spite of these compromises, the effectiveness of the instrumentation for residue detection is demonstrated.

Materials and Methods

Instrumental Setup. A schematic diagram of the instrumentation used in this work is presented in Figure 5.1. A Xe Arc Lamp with a UV cut-off filter was used as the broad band visible light source. Note that this choice was simply due to availability, and any suitable broadband visible light source (e.g., tungsten-halogen) should suffice. Linear polarizers were used at the source and imaging points (Edmund Optics, Barrington, NJ). For all investigations the angle and polarization of the incident light beam were fixed. Figure 5.1 shows that if the stainless steel plate defines the x-z plane (the z-plane being perpendicular to the page), the incident beam lay in the x-y plane at a 45º angle to the steel plate surface. The incident beam was polarized in the x-y

plane, perpendicular to the stainless steel surface (i.e., p-polarized). The beam diameter (circular) was approximately 4.6 cm resulting in an elliptical illumination spot on the surface with major and minor axes of 6.5 and 4.6 cm. Images were collected using a Nikon D70 digital camera (Nikon Inc., Melville, NY) and 38-75 mm zoom lens (Tamron Inc., Commack, NY). The f/stop was fixed at f/11 and ISO sensitivity to 100 for all investigations while the shutter speed was varied to control exposure. It was necessary to maximize depth of field in the images due to the orientation of the camera in relation to the steel surface. This f/stop was chosen to achieve good depth of field while maintaining a reasonable exposure time. The maximum exposure was 1 s for these investigations. Images were recorded in raw Nikon 12-bit format.

Surface Preparation. All surfaces used in this work were grade 318 near-mirror polished (grain-free) stainless steel plates. To produce more realistic surfaces that were similar to what might be encountered in pharmaceutical process equipment, defects were introduced using sand paper (various grit sizes) to score the surface. Clean surfaces were prepared by scrubbing the plates with a residue-free cleaner, thoroughly rinsing with deionized water followed by ethanol, and allowing the plates to dry. Aspirin residues (acetylsalicylic acid) on the plate surfaces were used for most of the investigations in this research. After a method was developed for detecting aspirin residue, the method was applied to additional residues including: ethanol-deposited acetaminophen and ibuprofen; powder samples of aspirin, acetaminophen and ibuprofen; fingerprints; an industrial chain lubricant (L716, Radiator Specialty Company, Charlotte, NC); and a mineral oil-based food grade lubricant (Petrol-Gel, McGlaughlin Oil Company, Columbus, OH). Aspirin, ibuprofen and acetaminophen were obtained from Sigma-Aldrich (St. Louis, MO). When reported, average surface concentrations were approximated by dividing the total

residue deposited by the average spot-size area. Additional details about the residue deposition procedures are given below.

Imaging Angle Optimization. During the preliminary angle optimization phase of this research, imaging was carried out at θ_r = 30, 60 and 90º in the plane of incidence. Although not depicted in the schematic, which portrays imaging in the plane of incidence (Φ_r = 0), out of plane imaging was performed at Φ_r = 15 and 30º above the beam plane at the same 3 angles. At each camera orientation, images were collected at 10º polarization phase steps from -90 to 90º relative to the incident polarization (19 total images). Due to the varying intensity of light at different imaging angles, shutter speeds of 1/8-1/2 s were used. Variation in shutter speed did not significantly detract from the objective, which was to examine the relative intensity between steel and residue. For this phase, 8 μL aliquots of a fixed concentration of aspirin in ethanol solution were pipetted on clean stainless steel plates giving average residue spot diameters of 2.3 cm (2.1 – 2.6 cm range) and approximate surface concentrations of 75 $\mu g/cm^2$. The relative intensity of scatter from the two sources was estimated by isolating the portions of the image covered by each. Figure 5.2 shows examples of isolated spots for images from two different angles (left panel: θ_r = 90º, Φ_r = 0º; right panel: θ_r = 30º, Φ_r = 30º) with intensity calculated by averaging pixels inside the red ellipse for aspirin and between the red and blue ellipses for steel. The values for each source were normalized by dividing the intensity over all analyzer angles by the maximum value. Note that this does not isolate the sources entirely because the steel underlying the aspirin residue contributes to the image. This will be discussed in more detail below.

Fixed Angle Imaging. The optimal imaging angle determined from the preceding investigation was θ_r = 30, Φ_r = 30º. In addition, imaging was performed at only two polarization angles, -90 and -60º relative to the source polarization, which corresponded to the minimum and maximum

relative intensities between the steel plates and aspirin residue. Note that the -90 and 90º polarizations are identical and correspond to crossed-polarizers. To increase signal-to-noise ratio, dynamic range and sensitivity seven images were collected at each polarization with different exposure times. A composite image was generated from these by averaging the intensity at each pixel over all exposures that exhibited an intensity in the range 160 – 3920 on a 12-bit scale (i.e., 0-4095), the contributions weighted to account for exposure. The difference between the two composite images was taken for further processing. In addition, the images were cropped to remove the edges of the illuminated spot on the surface due to intensity inconsistencies and distortions. The first aim of this investigation was a quantitative study of aspirin residue. Twelve μL aliquots of varying aspirin concentration were deposited on clean plates resulting in average spot diameters of 2.9 cm (2.4-3.3 range) and an approximate surface concentration range of 0.05 – 100 $\mu g/cm^2$. Data was collected on four replicate samples of nine concentrations in this range in addition to 30 clean samples giving a total of 66 samples.

The application of this method for detecting residues other than aspirin was carried out using the same imaging procedure. By an analogous ethanol deposition method as previously described, samples were prepared with residues of acetaminophen and ibuprofen with surface concentrations of 12.5 $\mu g/cm^2$. For the following samples, surface concentrations were unknown and the aim was merely to determine if the residues could be detected with this method. Dry samples (i.e., without solvent deposition) of these two compounds as well as aspirin were prepared by dispersing powder on the surfaces and "smearing" the residue with a dry Kimwipe (this was necessary to promote adhesion since the plates were vertically mounted). Similarly, samples with the lubricants were prepared by depositing the substances on Kimwipes first then smearing the surface. Lastly, plates with fingerprints were prepared.

Data Analysis. All data analysis was performed using Matlab 7 (Mathworks, Inc., Natick, MA). Two different regression models for the quantitative aspirin residue study were compared. The first method was a univariate least-squares regression using a pixel intensity summation above a designated threshold as the dependent variable. For the second method, principal component regression (PCR) models were built using the distributions of pixel intensity from the difference images. Finally, linear discriminant analysis (LDA) was used to investigate differentiation between clean samples and samples with aspirin residue. Thorough treatments of PCR and LDA can be found in the literature.[6, 17, 24]

Results and Discussion

The graphs of the normalized intensities for stainless steel (solid lines) and aspirin residue (dashed lines) at Φ_r = 0, 15 and 30º are presented in Figures 5.3, 5.4 and 5.5, respectively. Examination of the data from imaging in the plane of incidence (Figure 5.3) demonstrates that the trends in intensity from the scattered light of steel and aspirin residue are essentially the same for all three angles and provides no means to differentiate the scatter arising from the two sources. In contrast, the divergence of intensity from the two sources becomes readily apparent in the out of plane data at Φ_r = 15º and θ_r = 30º (see Figure 5.4). The trends for θ_r = 60 and 90º in this figure are not as symmetric as the in plane data but at θ_r = 30º the minimum for steel occurs at a -70º analyzer phase shift relative to the incident polarization while the apparent minimum for aspirin occurs at -80º. For out of plane imaging at Φ_r = 30º and θ_r = 30º (Figure 5.5) the disparity becomes even more pronounced with the minimum for steel occurring at a -60º analyzer phase shift while the apparent minimum for aspirin still occurs at -80º. The reference to apparent minima in relation to aspirin is used because the intensity measurements are

confounded by the contributions of both aspirin and the underlying steel. The combined effect of this is clearly seen as the minimum of the steel intensity shifts relative to aspirin, resulting in the profile of aspirin broadening and skewing. Furthermore, visual examination of the images suggests that the actual minimum for aspirin residue occurred at -90º (i.e., crossed polarizers) regardless of the imaging angle. To decrease data collection time and computational demands for further studies it was decided to reduce imaging to two polarizations.

A comprehensive scan of the polarization range results in considerable information redundancy and a pair of suitably chosen angles should provide sufficient discriminatory capability between the scatter sources. The angles were chosen by examining the intensity ratio between aspirin and steel across the polarization range. Presented in Figure 5.6 is the ratio of normalized intensities of aspirin to steel at $\Phi_r = 30º$. At $\theta_r = 30º$ we find that the -60 and -90º analyzer polarizations yield the maximum and minimum relative intensities across the data set ($\Phi_r = 0$ and 15º included) and these were selected for subsequent data acquisition.

Imaging with this polarization pair was performed on the 36 aspirin samples of various concentration and 30 clean samples. The normalized experimental pixel distributions from the difference images of this data set are presented in Figure 5.7. The abscissa has been cropped to [-400,400] to facilitate comparison but the full range was [-964,591]. The clean surface data resembles a negatively skewed extreme value distribution[179] while the aspirin residue shows this underlying distribution with the addition of a very broad, flat density toward a positive difference. This distinction between the distributions provides a rudimentary means for differentiation between the sources. Applying an appropriate threshold to the image difference, pixels attributed to scatter from steel can be removed. An exact threshold is somewhat arbitrary and the initial selection was zero, which corresponded to a cumulative distribution value of

0.9985 for the clean surface samples. A schematic of the application of this processing procedure to an individual sample is presented in Figure 5.8 (aspirin surface concentration of 5 $\mu g/cm^2$). Additional examples of images processed with this threshold method from clean surfaces and surfaces with different concentrations of aspirin residue are shown in Figure 5.9.

Two approaches to building regression models for aspirin surface concentration were evaluated. Data for the univariate method were generated by a summation of the pixel distribution above the defined threshold with each pixel count weighted by corresponding difference value. Other strategies as well as different thresholds were considered for univariate models but no significant improvements in results were found. For the PCR models, the raw pixel distributions were used as data vectors and the data set was mean-centered prior to principal components analysis. Eight PCs were included in the optimal model accounting for 98.1% of the total variation. The cross-validation results from these two methods are presented in Figure 5.10 for comparison. The global SEPs for the two models (7.86 and 3.36 $\mu g/cm^2$ for the threshold and PCR model, respectively) demonstrate the improved prediction offered by the PCR model but do not reveal the complete picture. If the predictions from the clean samples and aspirin residue samples are examined separately we find a trade-off between the two models with better predictions for the threshold model on the clean samples (SEP = 0.91 versus 1.98 $\mu g/cm^2$) but a 3-fold improvement for the PCR model on the residues (SEP = 3.81 versus 11.29 $\mu g/cm^2$).

The prediction results are supported by several observations. The threshold was defined specifically for the clean surface data so it could be expected to perform well on those samples. However, significant overlap between the collective distributions of clean and aspirin contaminated surfaces exists (see Figure 5.7) and thus an effective threshold necessitates the removal of useful information related to the contamination. The distributions of individual

samples reveal other trends not evident in the collective samples. Presented in Figure 5.11 are example distributions of several clean surface samples (top panel) and of the four replicate samples with 25 μg/cm^2 aspirin residue (bottom panel). The disparities in peak location and shape amongst clean samples likely account for the increased prediction error for the PCR model on these samples. From visual comparison there is a definite correlation between peak broadening and increasing surface deterioration, however the source of the inconsistency with peak location has not been elucidated at this point. The aspirin samples display much more variation than the clean surfaces. Because the examples are of the same surface concentration, the differences must be attributed to differences in distribution of the residue on the surface, particle sizes, or other film characteristics. The trends in these data provide a rationale for the relatively large number of PCs required (8) to build an effective calibration in spite of the fact that this is one-constituent chemical system.

Examination of the prediction results reveals that the limits of detection for the above regression methods were approximately 5 μg/cm^2. The actual concentration range investigated included samples 2 orders of magnitude less than this. To investigate if this was the true detection limit an LDA model was developed for differentiating clean from aspirin contaminated surfaces with the concentration information ignored. From the LDA model using 8 PCs the first discriminant function was significant ($p = 4.91 \times 10^{-19}$, we are limited to one discriminant function because only two groups are present), however the first two are presented in Figure 5.12 for graphical purposes. As seen in the figure the aspirin contaminated surfaces are differentiated from the clean surfaces with the exception of the 0.05 μg/cm^2 samples. The 2.5 and 0.5 μg/cm^2 samples that were previously below detection limits are resolved in this figure suggesting that it may be possible to reduce detection limits by an order of magnitude. This result is not

remarkable because the regression models are minimizing the global error in a least-squares fashion. The least-squares approach is biased against very low concentration samples when high concentrations are present.

The final investigation was somewhat cursory with the aim of determining if the method developed could detect other potential surface contaminants. Images were collected and processed using the simple threshold method for the samples described above in the Experimental section. Examples of the processed images for each of these samples are shown in Figure 5.13. Although detection limits and quantification capabilities cannot be determined from these data it is clear that all of the residues studied were detectable. One aspect not discussed in this work is the differentiation of compounds on a surface or in a mixture. Exploratory studies with acetaminophen, ibuprofen and several common pharmaceutical excipients displayed very similar trends to aspirin suggesting that discrimination would be difficult. The detection of surface residues in this work is essentially accomplished by the elimination of scatter arising from the underlying steel surface. The lack of chemical specificity, though, is not necessarily a weakness. If the method is sufficient for determining the presence of the target analyte(s) at pharmaceutically relevant concentrations the resulting data could be interpreted in a worst-case scenario approach similar to that used in TOC analysis for cleaning validation applications. This carries with it the advantage of simpler method development but in contrast to TOC does not require aqueous solubility and allows for *in-situ* analysis, which alleviates other issues such as surface adhesion and the uncertainties associated with analyte recovery from swabs. The results of this work are promising and certainly justify a more thorough investigation of this methodology, both experimental and theoretical, for cleaning validation.

Conclusion

This work demonstrated the potential of polarized light scattering measurements for pharmaceutical cleaning validation applications. The ability to detect residues of several APIs at pharmaceutically relevant concentrations as well as several other contaminants of interest on polished stainless steel surfaces with varying extents of wear was established. The prospect of quantifying surface residues was also established with the accurate determination of ethanol deposited aspirin residues on these surfaces. These results support future investigations into the use of this method for pharmaceutical cleaning validation. The relatively simple instrumentation used in this investigation suggests that the development of a rugged, portable instrument capable of *in-situ* surface analysis for cleaning validation applications is feasible.

Chapter Five Figures

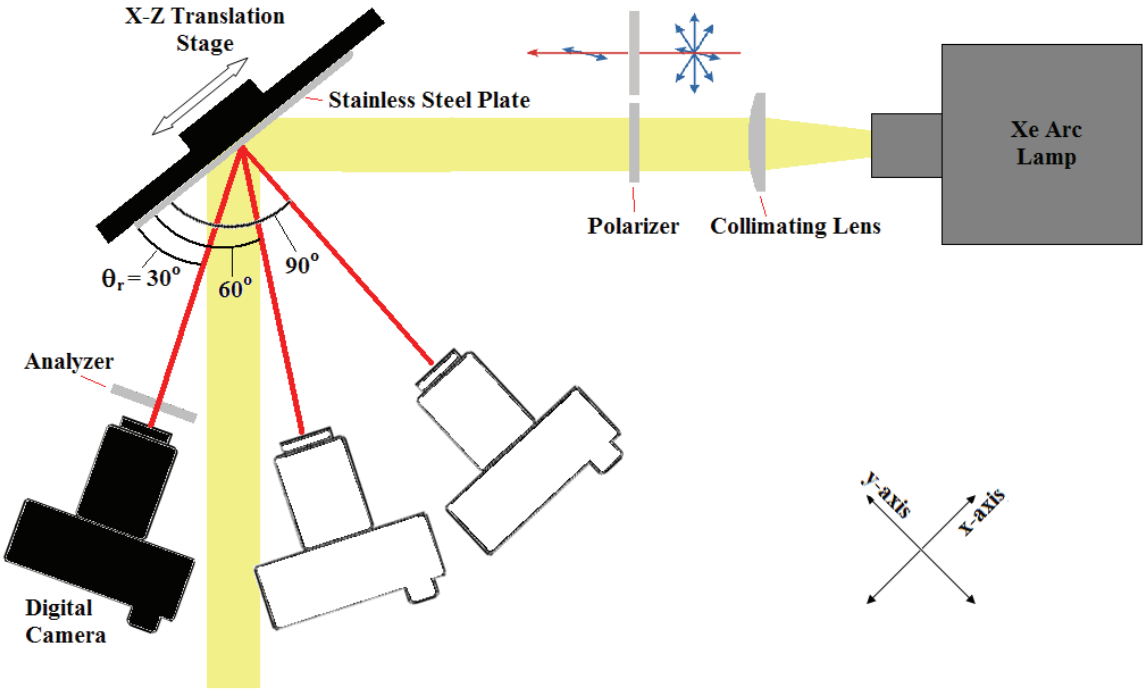

Figure 5.1: Schematic of the instrumentation used for polarized light scatter imaging.

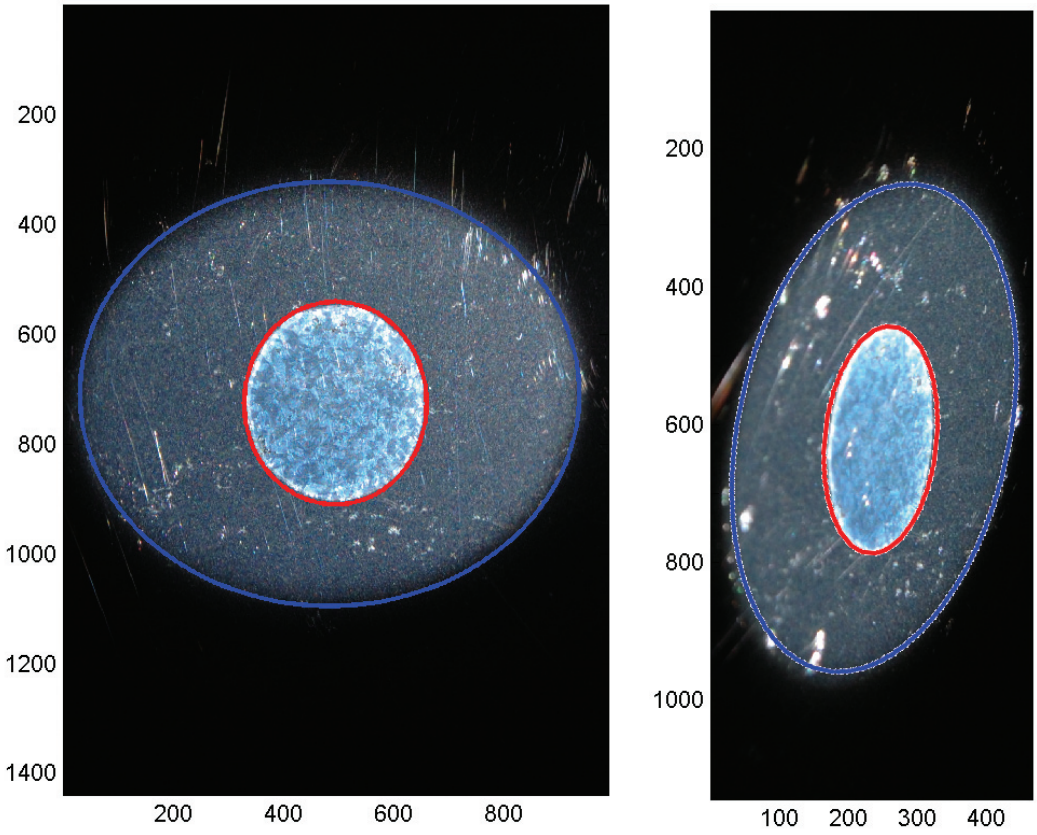

Figure 5.2: Example images from two different imaging angles (left panel: $\theta_r = 90^o$, $\Phi_r = 0^o$; right panel: $\theta_r = 30^o$, $\Phi_r = 30^o$). Scattered light intensity was estimated for steel and residue by isolating the image sections inside the red ellipses and between the red and blue ellipses.

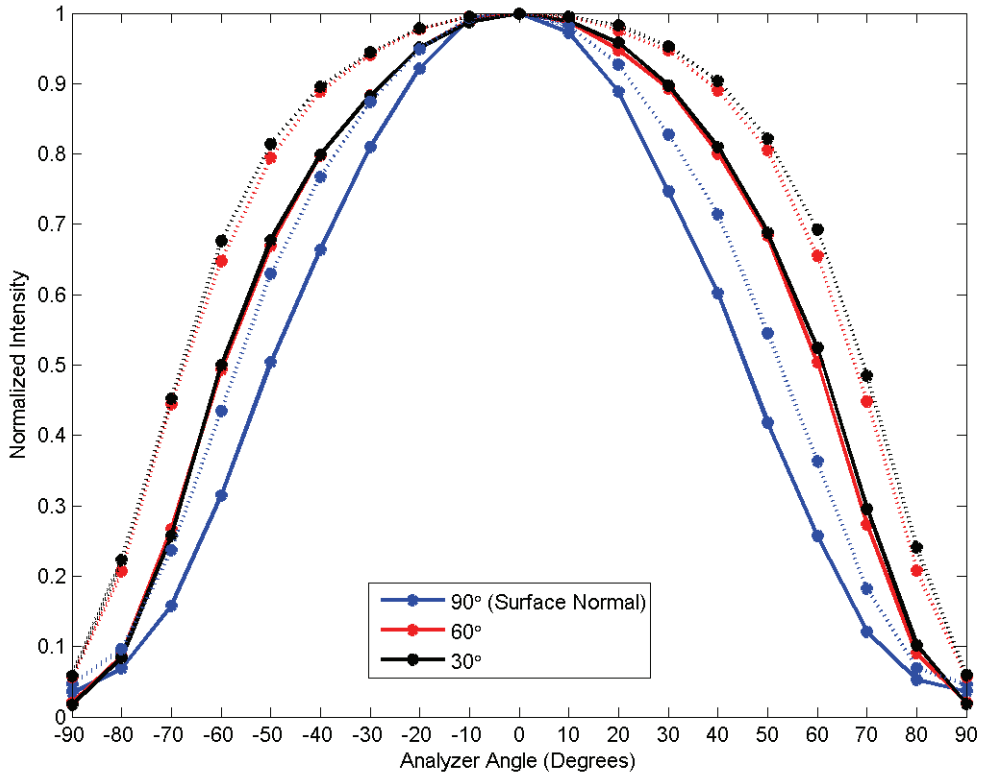

Figure 5.3: Normalized intensity of steel (solid lines) and aspirin residue (dashed lines) from imaging in the plane of incidence ($\Phi_r = 0$) at $\theta_r = 30$, 60 and 90º.

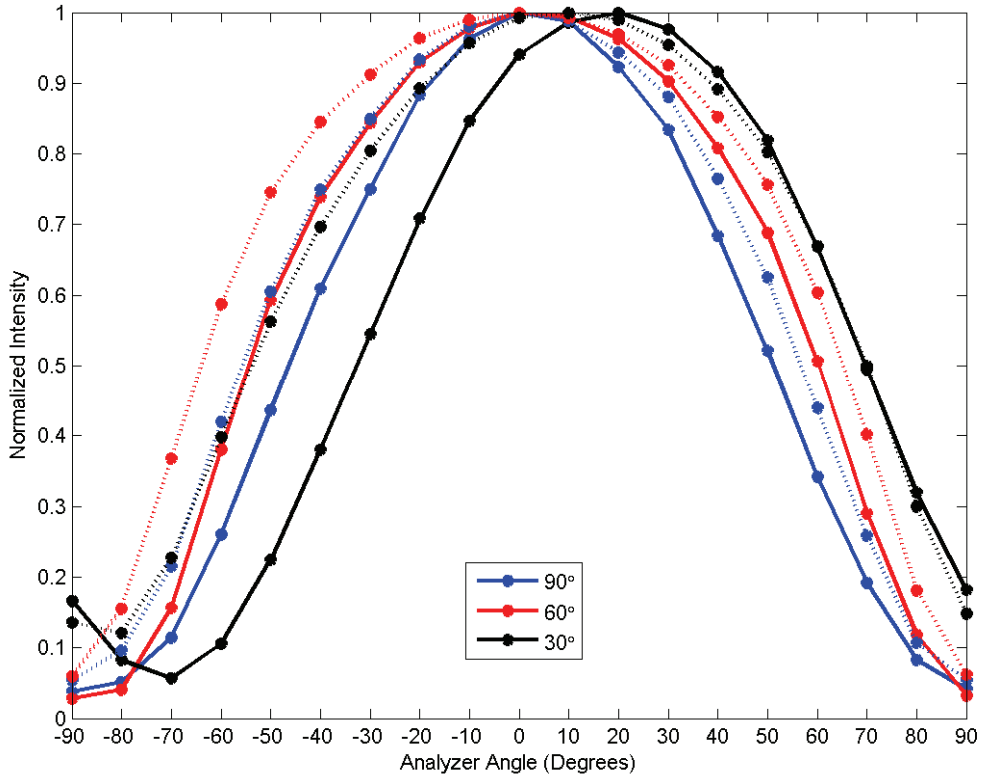

Figure 5.4: Normalized intensity of steel (solid lines) and aspirin residue (dashed lines) from imaging at Φ_r = 15 º for θ_r = 30, 60 and 90º.

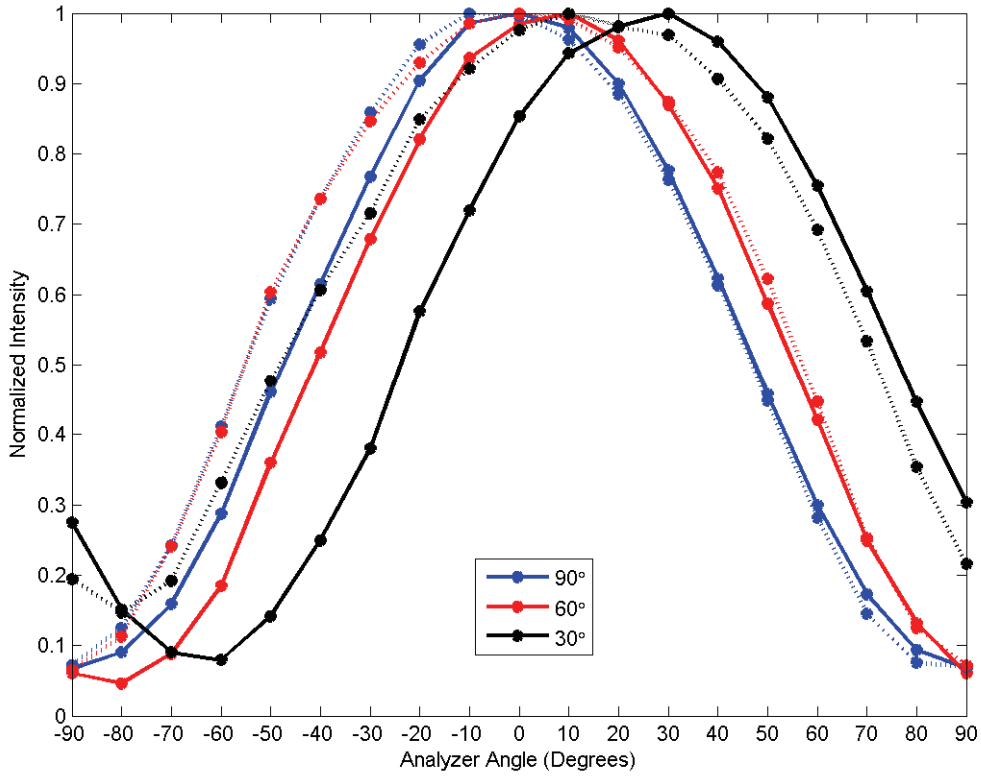

Figure 5.5:. Normalized intensity of steel (solid lines) and aspirin residue (dashed lines) from imaging at Φ_r = 30 º for θ_r = 30, 60 and 90º.

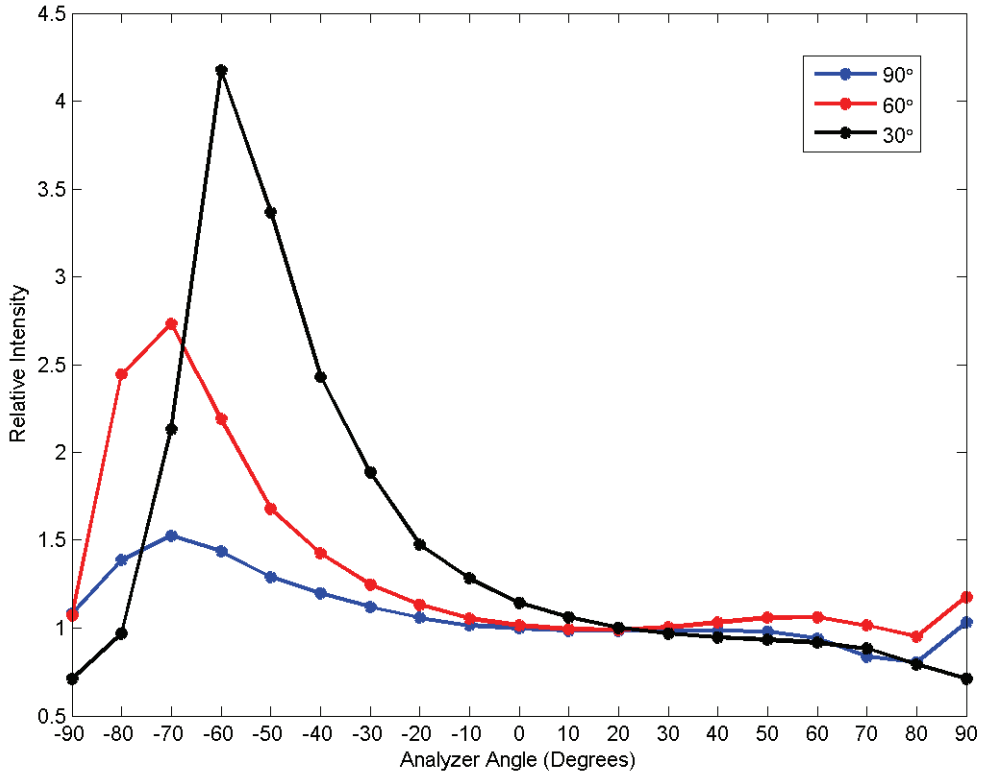

Figure 5.6: Ratio of the normalized intensities of steel and aspirin residue from Figure 5.5 (i.e., from imaging at Φ_r = 30 º for θ_r = 30, 60 and 90º).

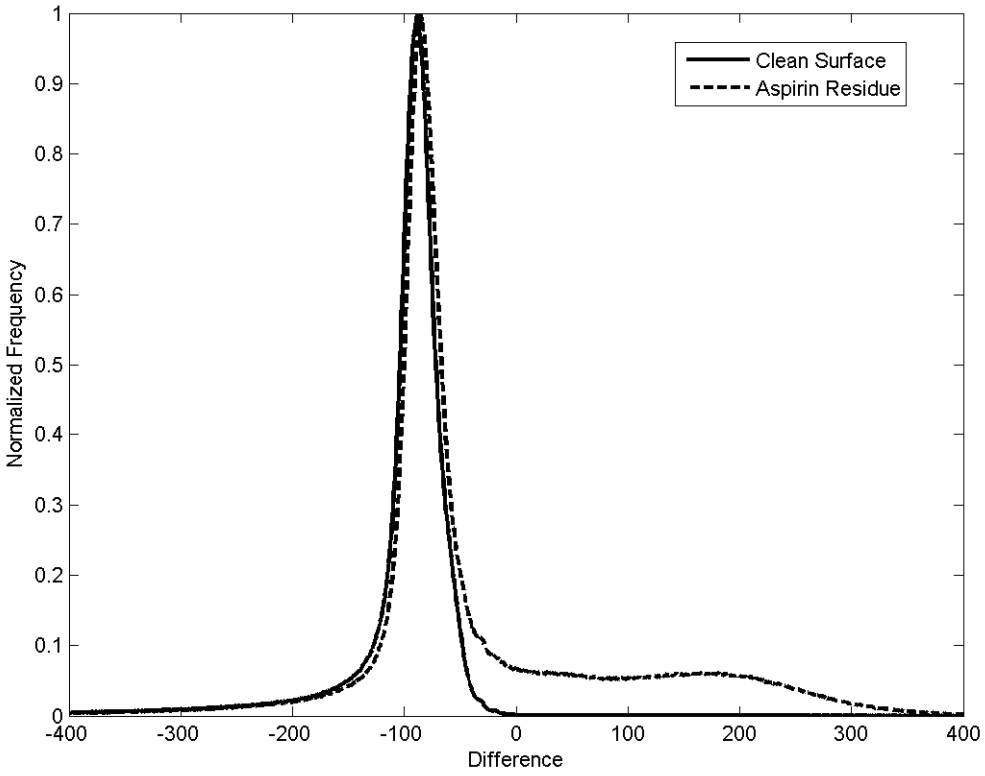

Figure 5.7: Normalized experimental pixel distributions from the difference images of clean steel (solid-line, n = 30) and aspirin contaminated surfaces (dashed-line, n = 36).

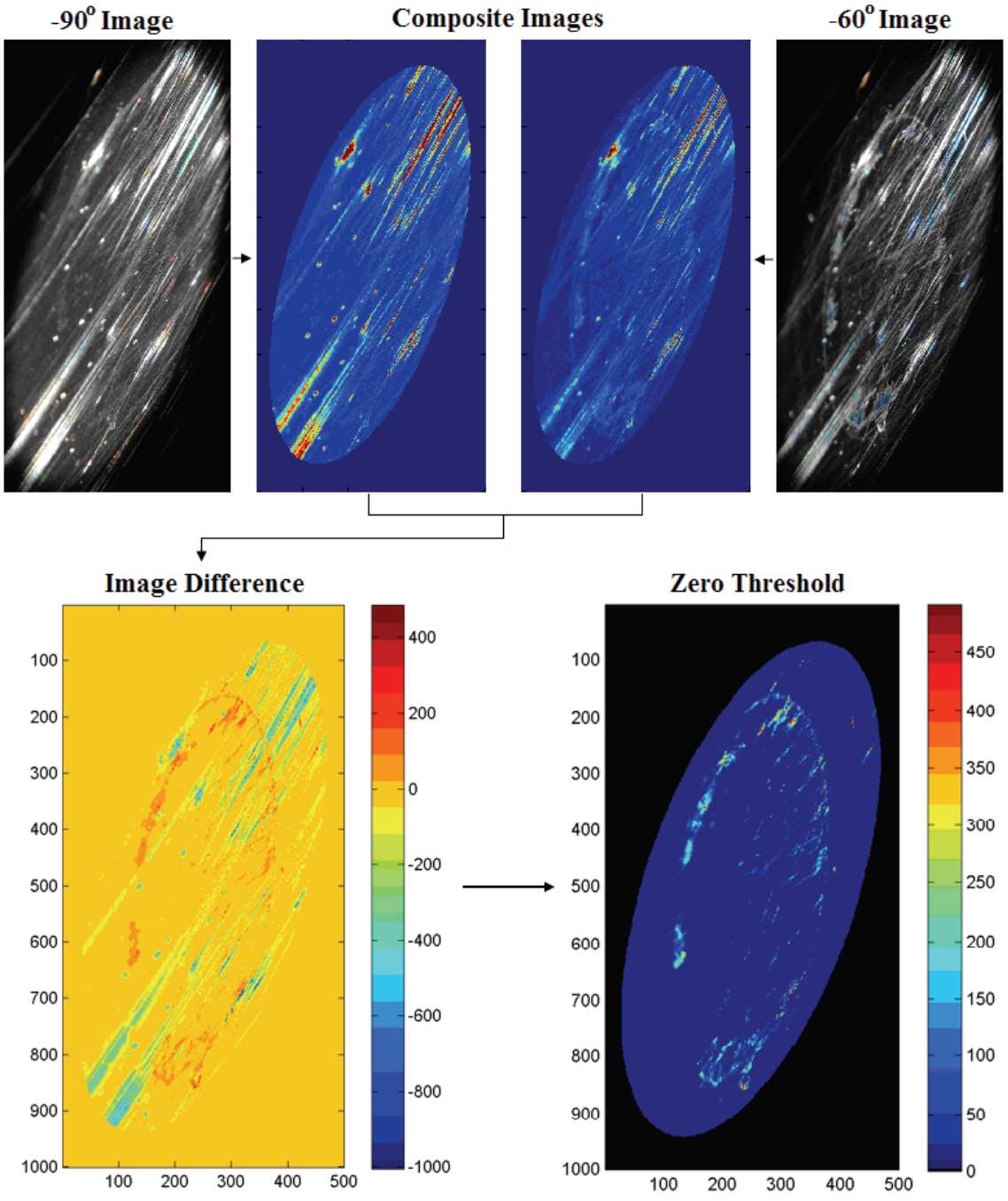

Figure 5.8:. Example of the image processing procedure using an image pair from a 5 μg/cm^2 aspirin residue sample. Note that the raw images (upper left and right corners) are RGB images at the longest exposure time.

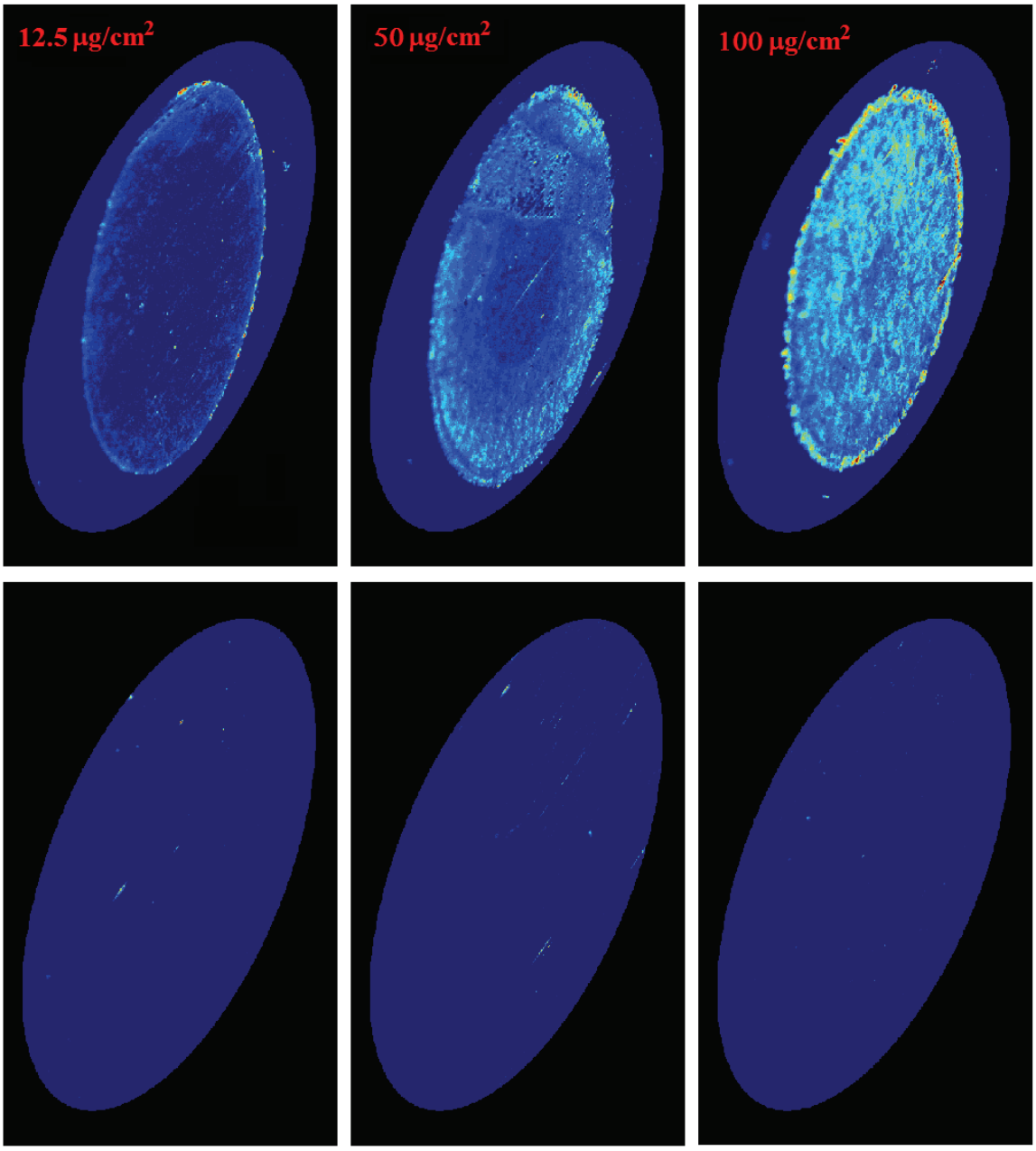

Figure 5.9: Examples of processed image pairs using the zero threshold for several aspirin contaminated (upper images) and clean surfaces (lower images). Surface concentrations of aspirin residue are denoted in the images.

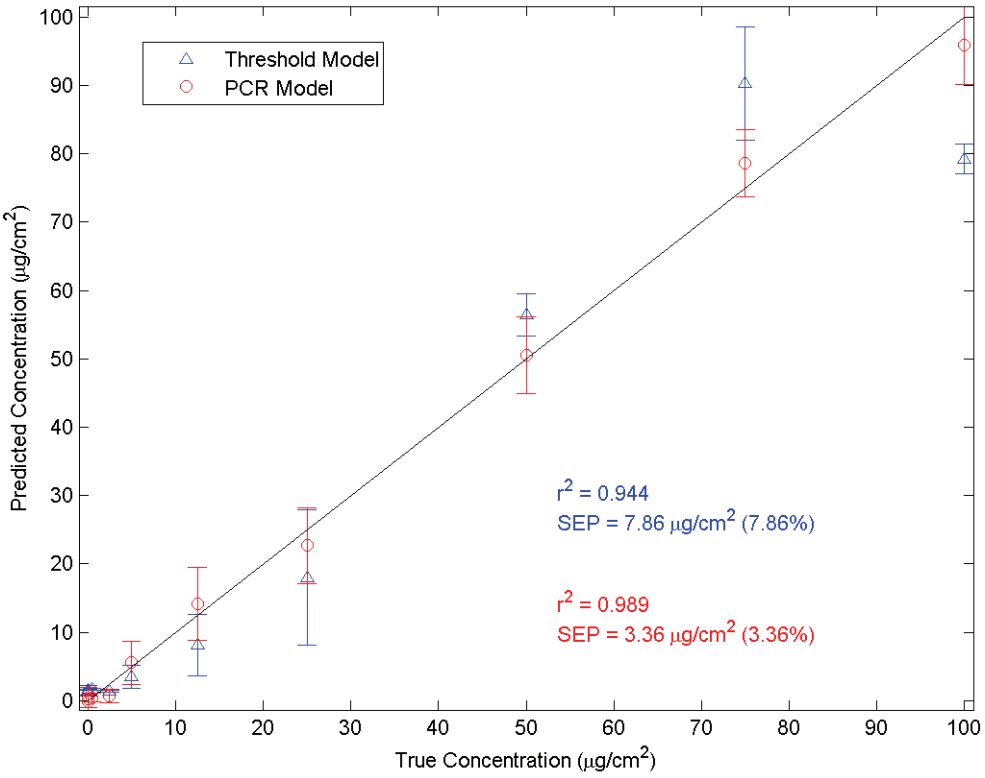

Figure 5.10: Cross-validated prediction results from the quantitative aspirin residue study for both calibration methods investigated: univariate threshold model (blue), PCR model (red). Error bars are plotted at ± 1 standard deviation.

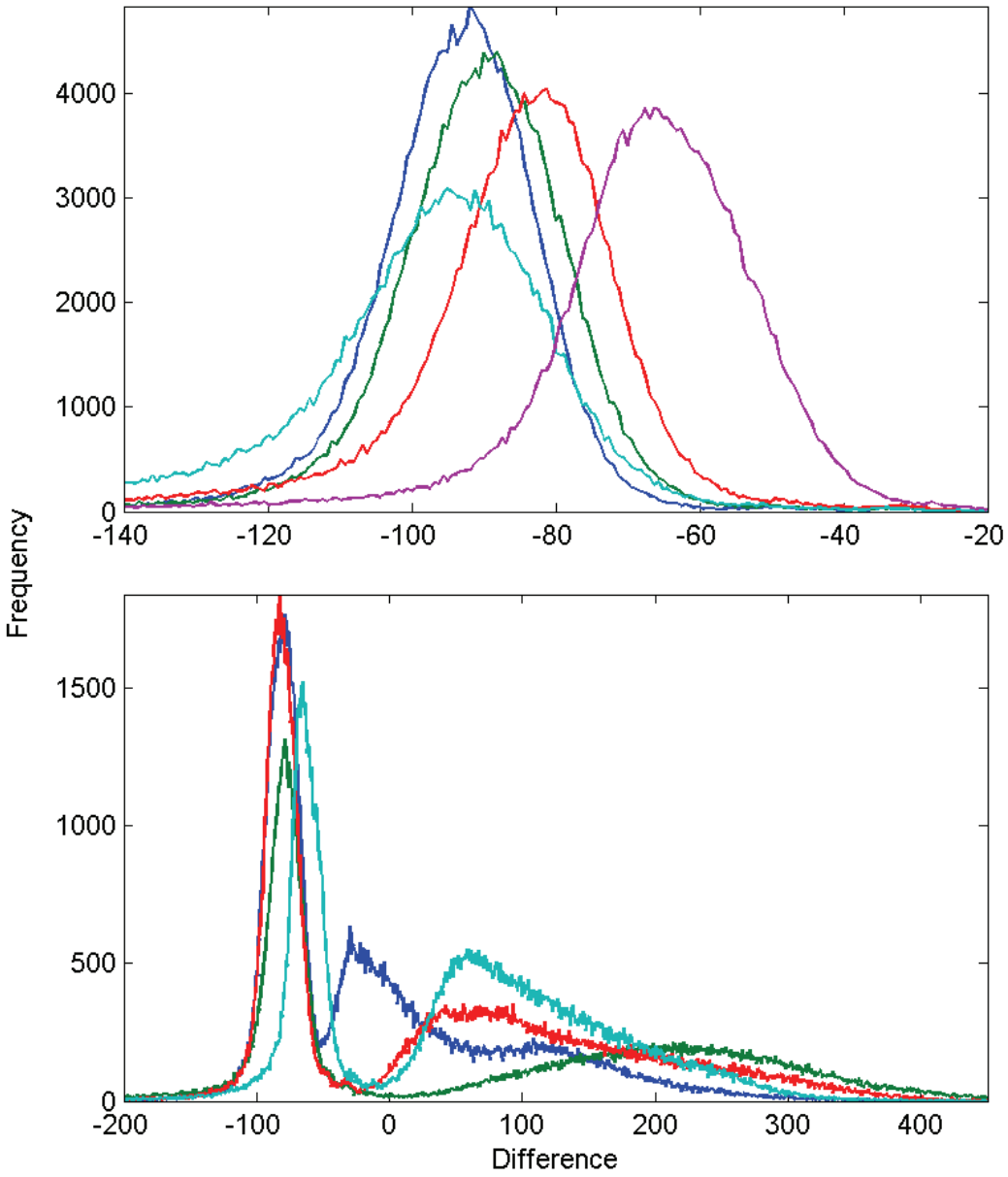

Figure 5.11: Examples of raw pixel distributions from individual clean (top) and aspirin contaminated (bottom) surfaces.

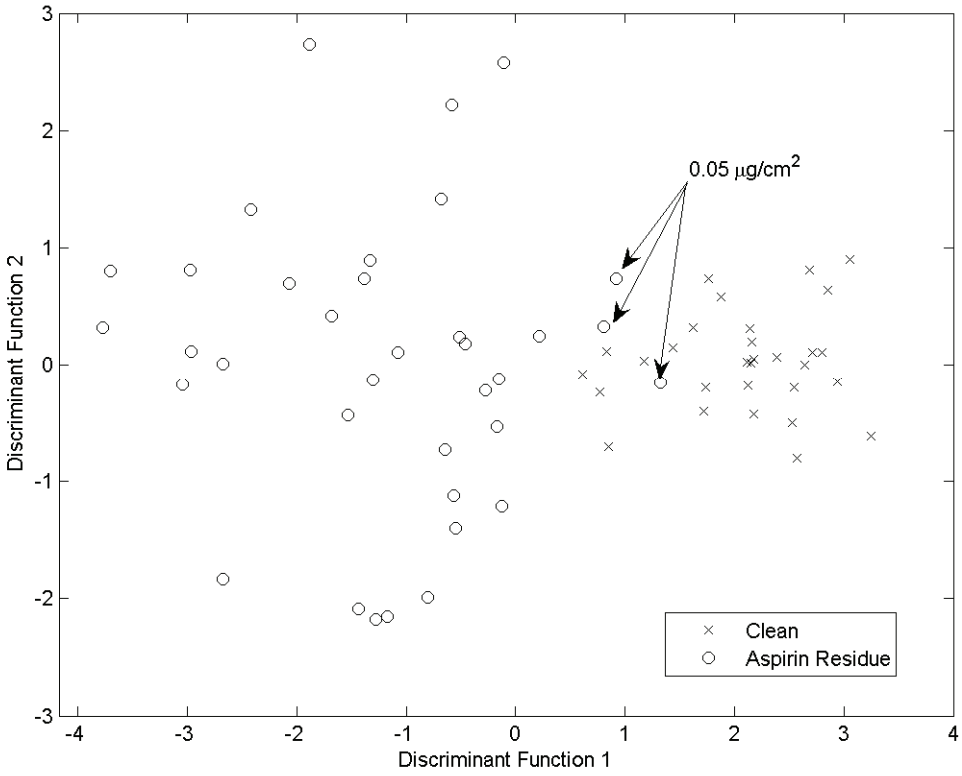

Figure 5.12: LDA results from aspirin quantification data for discrimination between clean and contaminated surfaces. Note that only the first discriminant function is significant and the second has been included for graphical purposes.

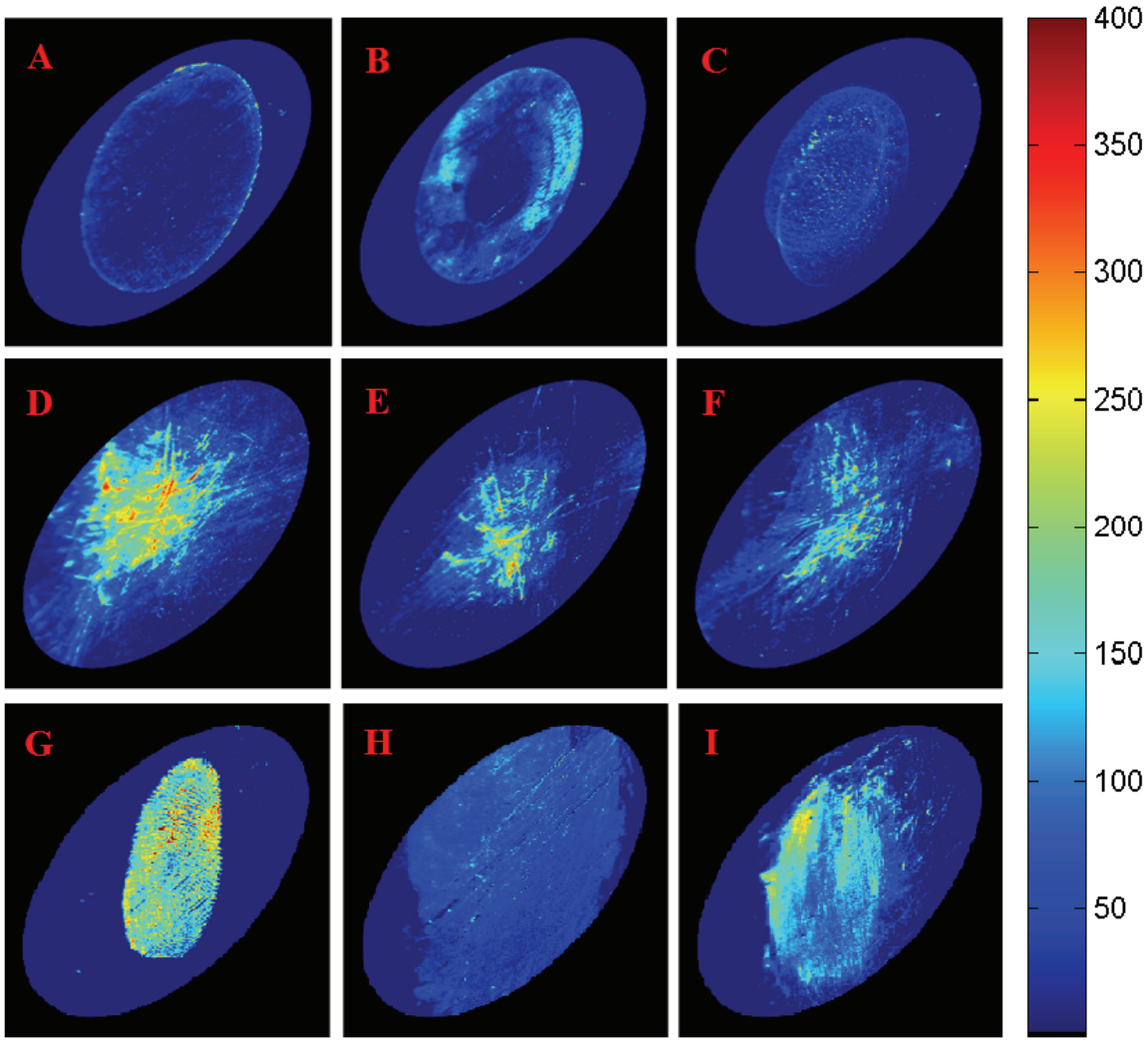

Figure 5.13: Examples of processed image pairs using the zero threshold for different surface residues: ethanol deposited residues of aspirin (A), acetaminophen (B) and ibuprofen (C) at surface concentrations of 12.5 µg/cm^2; dry smear deposited residues of aspirin(D), acetaminophen(E) and ibuprofen (F); fingerprint (G); industrial chain lubricant (H) and food grade lubricant (I). Note that surface concentrations are known for only the ethanol deposited samples.

Chapter Six – Near Infrared Spectroscopy and Molecular Factor Computing for Pharmaceutical Cleaning Validation

Introduction

The previous two chapters detailed research on the use of light-scatter and polarization phenomena for the detection of surface residues on glass and stainless steel. The applicability of a light-scattering method is limited if surface imperfections, such as abrasions. Examining the polarization of scattered light can permit the discrimination between contamination and surface imperfections. Although these methods demonstrated detection limits at pharmaceutically relevant concentrations they lack chemical selectivity. To address the selectivity issue, the present research investigates the use of NIRS for *in situ* cleaning validation applications. Near-infrared spectroscopy (NIRS) has been widely used in the pharmaceutical industry and is the most prevalent analytical method being explored for PAT applications.[180-183] However, very little work has been done to evaluate its capabilities for cleaning validation. One recent application has been reported that successfully employed NIRS to directly analyze swab samples of mineral oil residues without a solvent extraction step.[184] NIR can provide rapid, multicomponent chemical analysis with limited or no sample preparation. Our group has reported on the use of NIRS in a diverse range of pharmaceutical applications including dermal absorption of drugs[185, 186], identification of enteric coated tablets[142], testosterone quantification in thin-film composites[187], monitoring powder blend homogeneity[188], formaldehyde induced crosslinking in gelatin capsules[189, 190] and moisture uptake in intact pharmaceuticals.[161, 191] In this work NIRS was used in the qualitative differentiation of several active pharmaceuticals and excipients on stainless steel and glass surfaces. The first objective was to demonstrate that

traditional NIRS could be used to differentiate surface residues. Using a dispersive spectrometer, NIR reflectance spectra were collected from pure samples of different components deposited on stainless steel plates. The aim of this investigation was not method development but instead to establish that NIRS is capable of analyzing residues *in-situ*. For this methodology to be applicable to *in-situ* cleaning validation a rugged, portable instrument is necessary. Although portable full-spectrum NIR spectrometers exist, they sacrifice S/N and sensitivity that are necessary for this type of application. Although the eventual development of a suitable instrument similar to the mid-IR spectrometer mentioned previously is feasible, filter-based spectrometers also have potential to provide these capabilities. Filter-based instruments are inherently rugged and have already found wide use as process analytical tools in the pharmaceutical industry.[183, 192, 193]

The second objective of this research was to demonstrate that MFC based NIRS is applicable to pharmaceutical cleaning validation. A discussion of this methodology can be found in Chapter 1, and the potential application of it for evaluating the vulnerability of atherosclerotic plaques was presented in Chapter 3. In this work, MFC data, using transparent polymer sheets as molecular filters, were used to differentiate residues of active pharmaceuticals and excipients on glass surfaces to establish the potential of MFC for cleaning validation. In contrast to the traditional NIRS study, glass surfaces were used to allow collection of transmission measurements because a prototype MFC instrument for reflectance measurements is still in development.

Material and Methods

Traditional NIR Data Collection. Pure samples of the active pharmaceuticals naproxen, ibuprofen, aspirin (acetylsalicylic acid) and acetaminophen as well as the common pharmaceutical excipients lactose and magnesium stearate were obtained from Sigma-Aldrich (St. Louis, MO). NIR diffuse reflectance spectra of the pure samples are presented in Figure 6.1 and have been offset for comparison. Sample surfaces were prepared by depositing ethanol solutions of single components on 6 cm^2 polished stainless steel plates (Grade 316). The solubility in ethanol of the active pharmaceuticals is sufficient at the concentrations used, however, lactose and magnesium stearate are essentially insoluble. For these components, mixtures with ethanol were prepared and temporary suspensions were produced by agitation immediately prior to surface deposition. The specific aim of this study was qualitative identification of the residues so the surface concentrations were not determined during sample preparation. Furthermore, only the average surface concentration over plate surfaces was known and the uneven residue distribution made determination of the actual surface concentration in the spot illuminated of the spectrometer difficult. The range of average surface concentrations of the components for this study was approximately 0 – 1000 $\mu g/cm^2$. NIR reflectance spectra in the 1100-2500 nm wavelength range were collected from the surfaces using an Infraalyzer 500 spectrometer (Bran and Luebbe, Elmsford, NY). Ten samples of each component were analyzed in addition to 12 blanks resulting in a total of 72 spectra in the data set. The NIR reflectance spectra of the minimum and maximum concentration samples for the four active pharmaceuticals and a blank are presented in Figure 6.2 for comparison. Examples of the two excipients were omitted but are similar.

MFC Data Collection. A schematic of the instrumental setup used for MFC data collection is presented in Figure 6.3. Due to the additional complexity associated with a reflectance setup, this study was conducted using transmission spectroscopy and glass surfaces instead of stainless steel. The broadband output from the Infraalyzer 500, which included the broadband source, 1650 nm longpass filter, optical chopper and fiber optics, was used as the light source. In contrast to previous MFC investigations, a simplified approach to molecular filters was used which employed readily available transparent polymer sheets. Again, the purpose of this study was not method development but to demonstrate the applicability of MFC for cleaning validation purposes so optimal molecular filters were not selected from a NIR library as described in previous works. Clear polymer sheets with nominal thickness of 0.125-inches of the following were used: polycarbonate, polymethyl methacrylate (PMMA), polyethylene terephthalate glycol (PETG) and polyvinyl chloride (PVC) (K-Mac Plastics, Wyomin, MI). The transmission spectra of the filters were measured over the 1100-2500 nm wavelength range and are presented in Figure 6.4. The use of the 1650 nm longpass was due to the relatively minor absorption below 1600 nm for the filters coupled with fact that the majority of the distinguishing spectral features of the compounds under investigation are above 1600 nm. A subset of the components used in the traditional NIRS study was chosen for this study: aspirin, acetaminophen, lactose and magnesium stearate. Samples were prepared as described above except that 10 cm^2 borosilicate glass plates were used. Transmission measurements (0.5 s integration time) were recorded for each of 4 molecular filters as well as a filter-less measurement giving 5 total data points for each sample. Ten samples of each component and 12 blanks were analyzed resulting in a total of 52 samples in the MFC data set. Presented in Figure 6.5 are the raw MFC data if the left panel and the MFC data after standard normal variate (SNV) transformation and mean centering.

Data Analysis. All data analysis was performed using Matlab 7 (Mathworks, Inc., Natick, MA). Traditional NIRS data was preprocessed by approximating 2nd derivatives using cubic smoothing splines[15] then normalized to unit vector length. MFC data were preprocessed with the standard normal variate (SNV) transformation followed by normalization to unit vector length. Classification models were developed using principal components analysis (PCA) to reduce data dimensionality and/or remove collinearity between variables followed by linear discriminant analysis (LDA) to generate discriminant functions. A thorough treatment of these statistical techniques can be found in the literature.[6, 17, 24] The Boot-strap Error-adjust Single Sample Technique (BEST) metric was used for non-parametric estimation of class distributions in the LDA models. The BEST metric provides a multi-dimensional standard deviation (MSD) similar to Mahalanobis distance but does not assume that populations are normally distributed and instead uses a non-parametric boot-strap estimation of the population density. BEST MSDs from leave-one-out cross-validation routines were used to determine classification accuracy of the models. A thorough treatment of the BEST metric can be found in the literature[32, 142, 194]. The performance of classification models was assessed by examining several standard metrics: *Accuracy*, *Precision*, and *Recall*.

Results and Discussion

Traditional NIRS Data. Including the blank samples there are 7 classes present in the data set, which limits the potential number of discriminant functions from LDA to 6. Examination of the PCA results suggested that 8 PCs were sufficient to describe the data, accounting for 98.3% of the total variation. The results from LDA using 8 PCs showed 6 discriminant functions were statistically significant ($p < 0.00001$). Plots of the first 4 canonical variables (discriminant

functions) are presented in Figure 6.6. The separation between groups and relatively compact clusters is evident in the plots. The cross-validated prediction results from this data are presented in a confusion matrix in Table 6.1. Included in the table are several performance metrics (i.e., accuracy, precision and recall) based on the individual class results. Pooling the results of the complete data set gave the following global performance measures: 99.4% accuracy, 100% precision and 95.8% recall.

Erroneous results across the global cross-validation were limited to two false negatives, one each for aspirin and acetaminophen, which accounts for the relative disparity of the recall metric versus accuracy and precision. For these two samples, the cross-validated MSDs from their appropriate classes was 3.98 and 5.06 std. dev., while the closest distance to any other group was 15.01 and 18.44 std. dev., respectively. This indicates that the samples may be outliers or that the population variance is not adequately captured by the samples present and therefore the 3 std. dev. cutoff is not appropriate. Examination of the ibuprofen cluster in the lower panel of Figure 6.6 (canonical variables 3 and 4) suggests that the misclassified sample from that group may signify an outlier but more data would be required to confirm this.

The relative distances between component clusters were investigated by finding the average MSD from each sample to each group using the cross-validation data. This data is presented in Table 6.2. The table is arranged so that the elements in a particular row represent the average MSD of the samples in the row's class designation to each of classes listed in the columns. The diagonal elements represent the average intragroup MSDs. Note that the data in the table is not symmetric about the diagonal (for example the average MSD from blank samples to the naproxen group, 3.54, is markedly different from that of naproxen samples to the blank group, 32.29). Excluding the intragroup diagonal elements, examination of the table reveals that almost

all of the maximum row and minimum column elements correspond to the blank group. The large values in the blank column result from the relatively minor spectral variations among clean surface spectra, which results in a compact sample population compared to the "dirty" surface spectra, which cover a range of concentrations for each of the residues. The relatively small values in the blank row result from the fact that as the concentration of any residue diminishes the spectra are approaching that of the clean surface more so than the other components.

MFC Data. As in the traditional NIRS data set, the maximum number of possible discriminant functions was found to be statistically significant ($p < 0.01$), however the number is reduced to 4 since 2 components were removed in this data set. Plots of the 4 canonical variables are presented in Figure 6.7. The separation between groups is apparent in these plots, particularly in the first two canonical variables; however, it is not as significant in comparison with the traditional NIRS data. The same cross-validation procedure described above was performed on this data set and the results are presented in Tables 6.3 (confusion matrix) and 6.4 (average MSDs between groups). Pooling the prediction results of the MFC data set gave the following global performance measures: 98.1% accuracy, 100% precision and 90.4% recall. Erroneous results were once again limited to false negatives (5 in total), which accounts for the noticeably diminished recall rate compared to traditional NIRS. The average MSDs data in Table 6.4 demonstrate that the separation between groups as a whole is similar to NIRS (mean off diagonal elements in Tables 6.2 and 6.4 are 14.11 and 10.61, respectively), however, the trends in the data are somewhat different.

The efficacy of the MFC data for classification of the surface residues was better than expected. In contrast to previous work, the molecular filters used in this study were not specifically selected for the application. Examination of Figure 6.5 shows that the relative

intensities between filters for the raw MFC data (top panel) is virtually identical between samples. This is because, aside from a baseline offset, the data is dominated simply by the total intensity differences of the light transmitted by the filters. The effect of this is to waste the bit-range of the analog-to-digital converter. Despite these drawbacks, the results support a more thorough investigation of the MFC methodology for cleaning validation applications. Subsequent studies will address some of the instrumental limitations just described. In addition, the determination of limits of detection, linearity and analysis of mixtures are necessary.

Conclusion

This work demonstrated the potential use near-infrared spectroscopy in pharmaceutical cleaning validation applications. The ability to accurately discriminate between *in-situ* residues of several active pharmaceuticals and excipients at varied concentrations on polished stainless steel surfaces using NIRS was established. However, *in-situ* surface analysis in a realistic environment using this technique would require rugged, portable instrumentation with limited sacrifices in terms of performance. These requirements present a barrier to the use of traditional full-spectrum NIRS devices. To address this issue, this work also demonstrated that molecular factor computing data in the near-infrared spectral region, comprised of only 5 measurements, was sufficient to discriminate between surface residues on glass with good accuracy despite the lack of optimization for this application. Together, these results support future investigations into the use of these methods for pharmaceutical cleaning validation.

Chapter Six Tables

Table 6.1: Classification results of the traditional NIR reflectance data from cross-validated LDA models and the BEST metric.

True Class	Predicted Class						
	Clean	NAP	IBU	ASP	ACE	LAC	MAG
Clean	12	0	0	0	0	0	0
NAP	0	10	0	0	0	0	0
IBU	0	0	9	0	0	0	0
ASP	0	0	0	9	0	0	0
ACE	0	0	0	0	10	0	0
LAC	0	0	0	0	0	10	0
MAG	0	0	0	0	0	0	10
Accuracy	100%	100%	98.6%	98.6%	100%	100%	100%
Precision	100%	100%	100%	100%	100%	100%	100%
Recall	100%	100%	90%	90%	100%	100%	100%

*/ NAP denotes naproxen; IBU, ibuprofen; ASP, aspirin; ACE, acetaminophen; LAC, lactose; MAG, magnesium stearate.

Table 6.2: Average BEST MSDs between each class from cross-validated LDA models using traditional NIR reflectance data.*

Class	Clean	NAP	IBU	ASP	ACE	LAC	MAG
Clean	**1.52**	3.54	5.26	6.90	3.79	6.06	8.10
NAP	32.29	**1.25**	7.21	7.74	6.79	6.82	8.53
IBU	24.78	10.68	**1.41**	8.07	10.92	10.17	10.49
ASP	41.45	14.23	11.13	**1.85**	12.12	11.11	11.43
ACE	14.30	5.70	6.12	7.24	**1.55**	6.76	7.43
LAC	60.63	12.54	11.41	11.49	16.13	**1.21**	11.73
MAG	68.21	19.05	14.74	13.73	20.93	13.26	**1.38**

*/ NAP denotes naproxen; IBU, ibuprofen; ASP, aspirin; ACE, acetaminophen; LAC, lactose; MAG, magnesium stearate.

Table 6.3: Classification results of the MFC transmission data from cross-validated LDA models and the BEST metric.

True Class	Predicted Class				
	Clean	ASP	ACE	LAC	MAG
Clean	12	0	0	0	0
ASP	0	9	0	0	0
ACE	0	0	9	0	0
LAC	0	0	0	8	0
MAG	0	0	0	0	9
Accuracy	100%	98.1%	98.1%	96.2%	98.1%
Precision	100%	100%	100%	100%	100%
Recall	100%	90%	90%	80%	90%

*/ ASP denotes aspirin; ACE, acetaminophen; LAC, lactose; MAG, magnesium stearate.

Table 6.4: Average BEST MSDs between each class from cross-validated LDA models using MFC data.*

Class	Clean	ASP	ACE	LAC	MAG
Clean	**1.60**	12.70	8.21	8.48	15.60
ASP	48.32	**2.11**	6.15	9.76	23.13
ACE	38.80	5.74	**2.21**	5.14	11.59
LAC	15.09	9.87	5.84	**2.40**	10.99
MAG	7.67	12.32	5.32	4.62	**1.71**

*/ ASP denotes aspirin; ACE, acetaminophen; LAC, lactose; MAG, magnesium stearate.

Chapter Six Figures

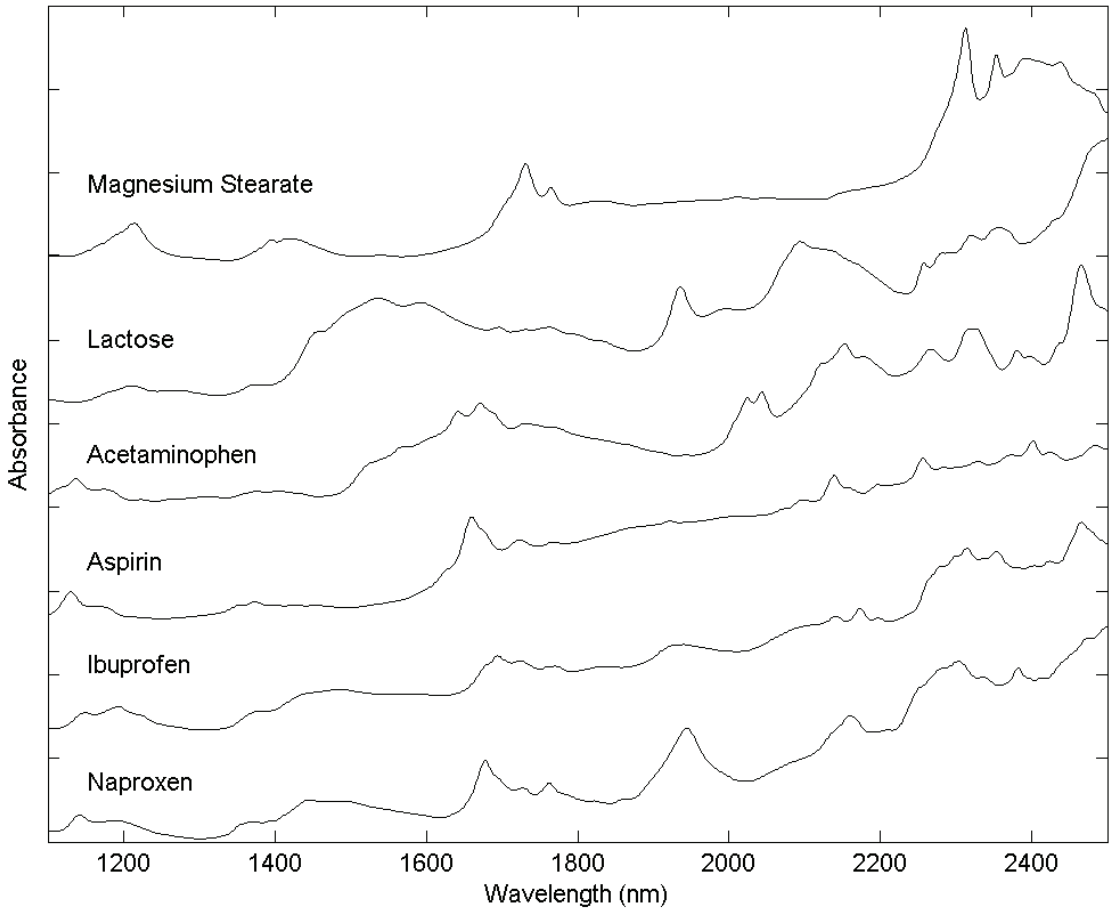

Figure 6.1: Diffuse reflectance near-IR spectra of pure samples of several APIs (naproxen, ibuprofen, aspirin, and acetaminophen) and excipients (lactose and magnesium stearate).

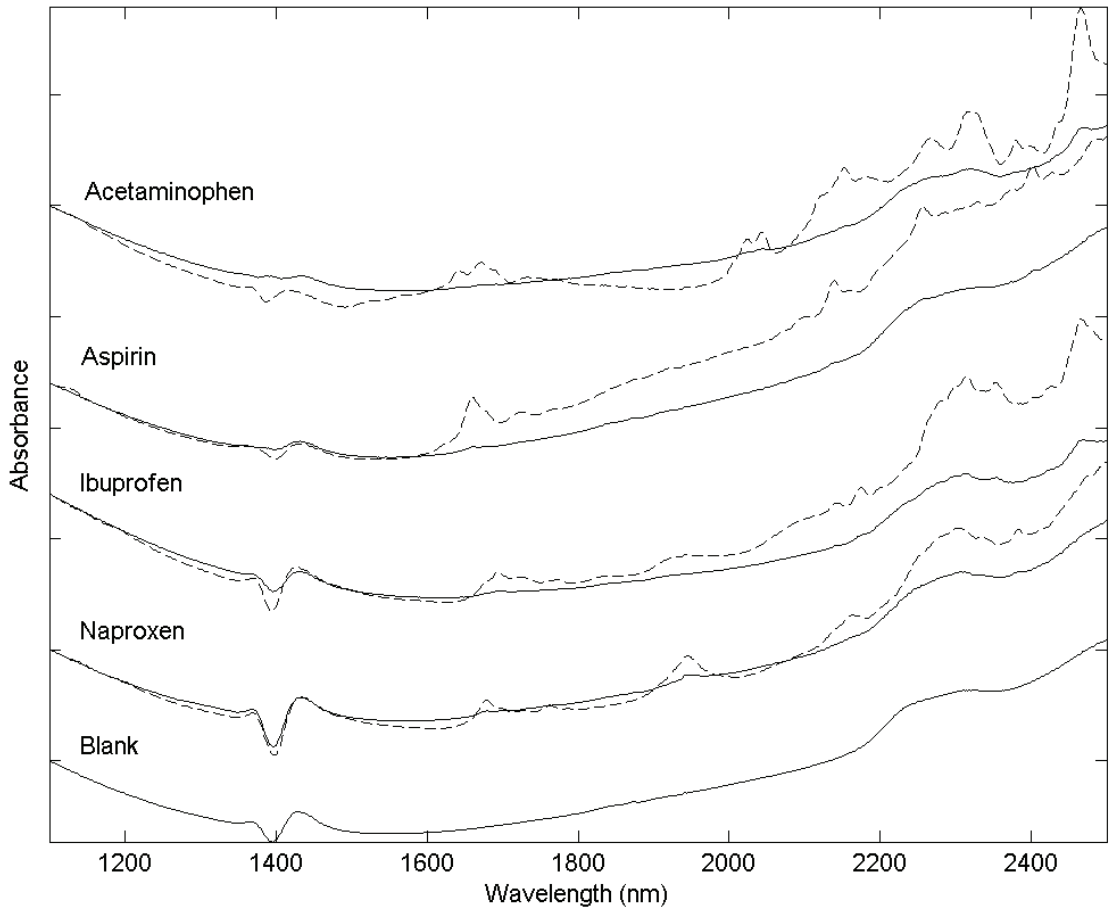

Figure 6.2: Examples of near-IR reflectance spectra from a clean polished stainless steel plate and plates contaminated with the APIs naproxen, ibuprofen, aspirin, and acetaminophen. Two spectra are shown for each API, which represent the minimum (solid line) and maximum (dashed line) concentration samples from the data set.

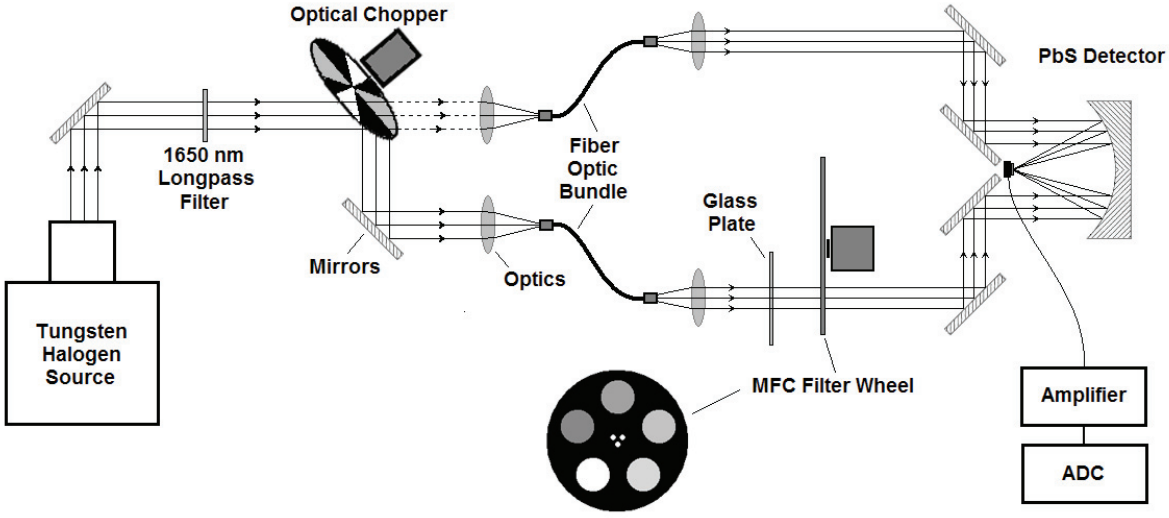

Figure 6.3: Instrument schematic of the molecular factor computing near-infrared spectrometer used in this work.

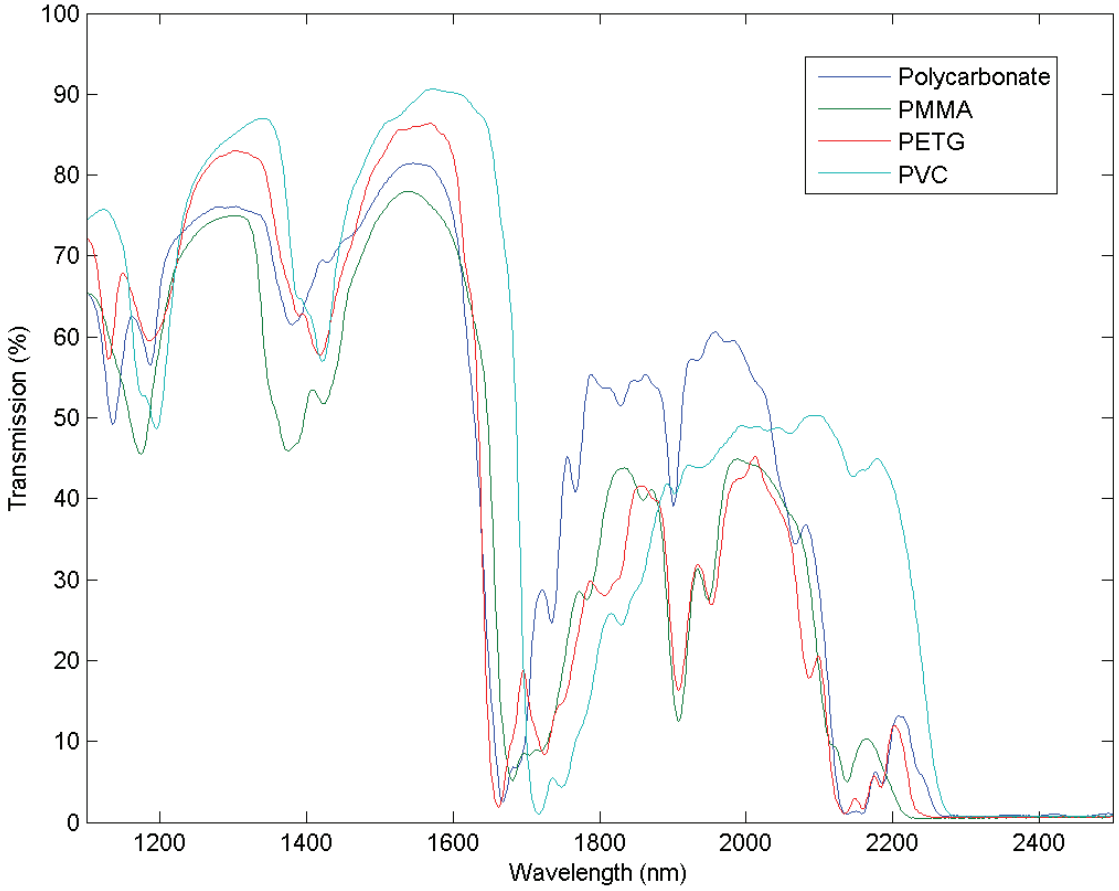

Figure 6.4: Transmission spectra of the polymers used as molecular filters in the work: polycarbonate, polymethyl methacrylate (PMMA), polyethylene terephthalate glycol (PETG), and polyvinyl chloride (PVC). Note that a 1650 nm longpass filter was used in the collection of MFC data.

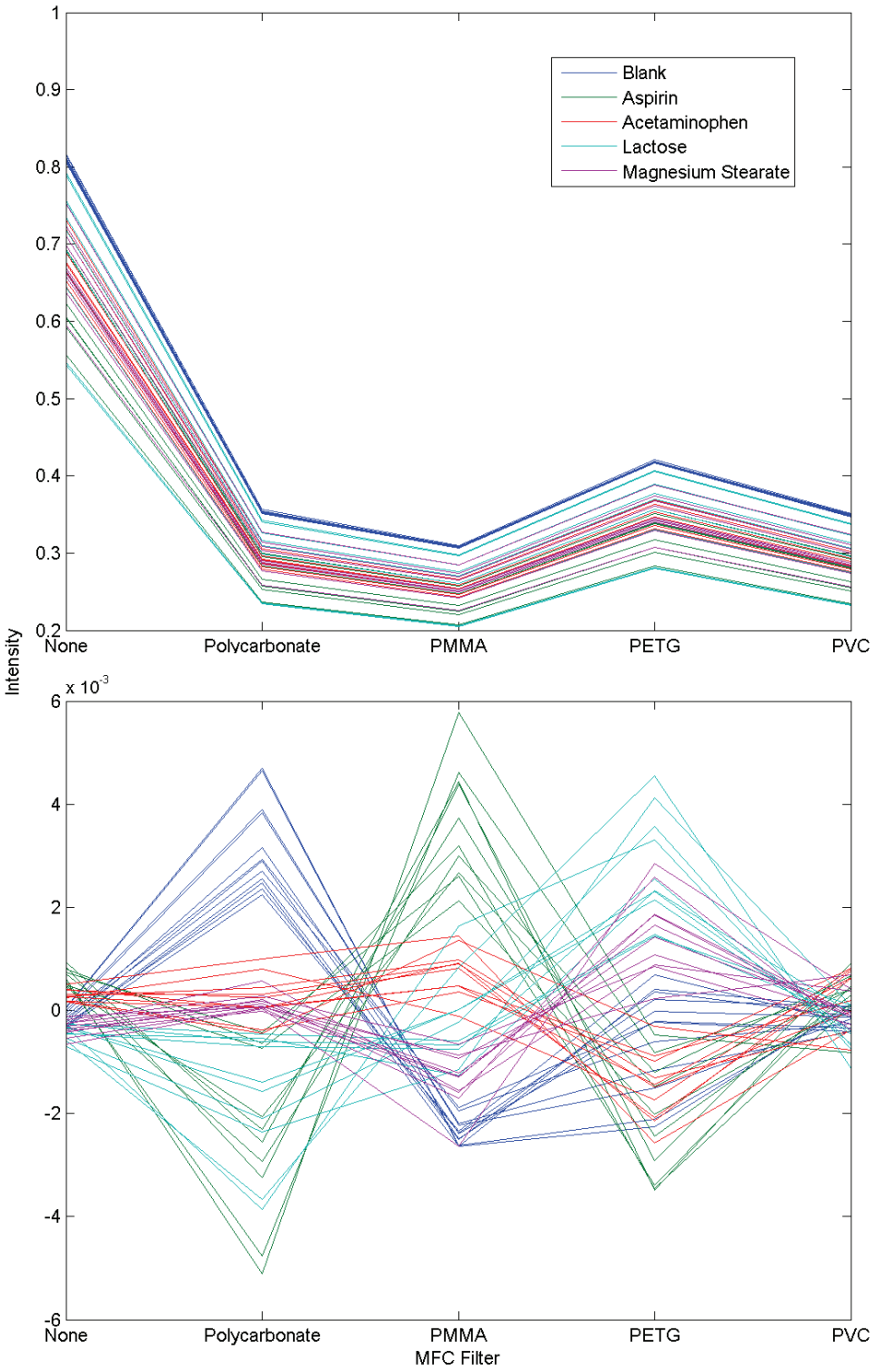

Figure 6.5: The raw MFC data set (top panel) and the data set after applying the standard normal variate transformation and mean centering. In contrast to the traditional near-IR data set, transmission measurements were collected for this data from surface residues on glass plates.

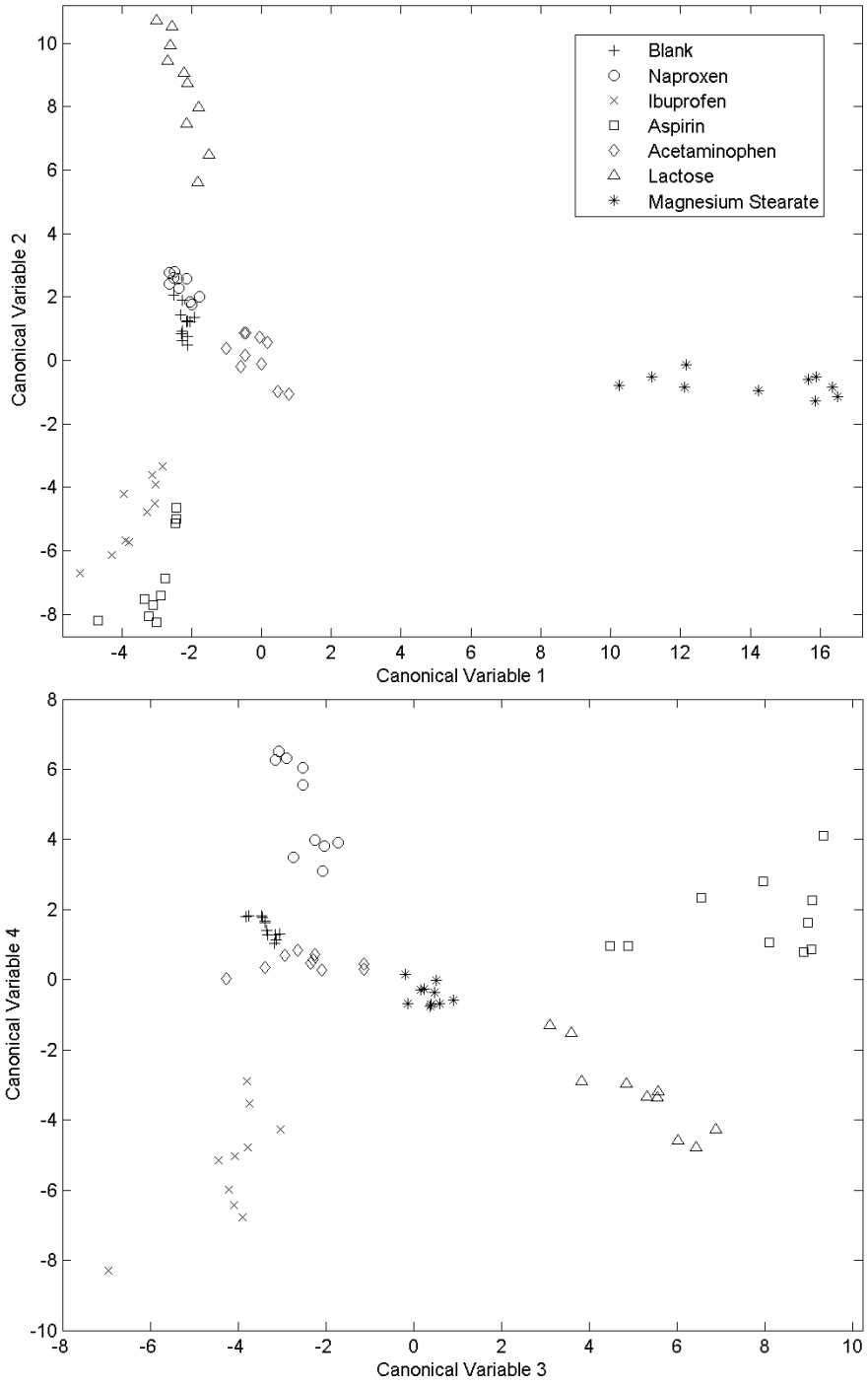

Figure 6.6: Graphs of the first four canonical variables from the traditional near-IR data set consisting of reflectance spectra from varying surface concentrations APIs and excipients on polished stainless steel plates. The legend in the top graph (CV 1 vs. CV 2) applies to the bottom graph (CV 3 vs. CV 4).

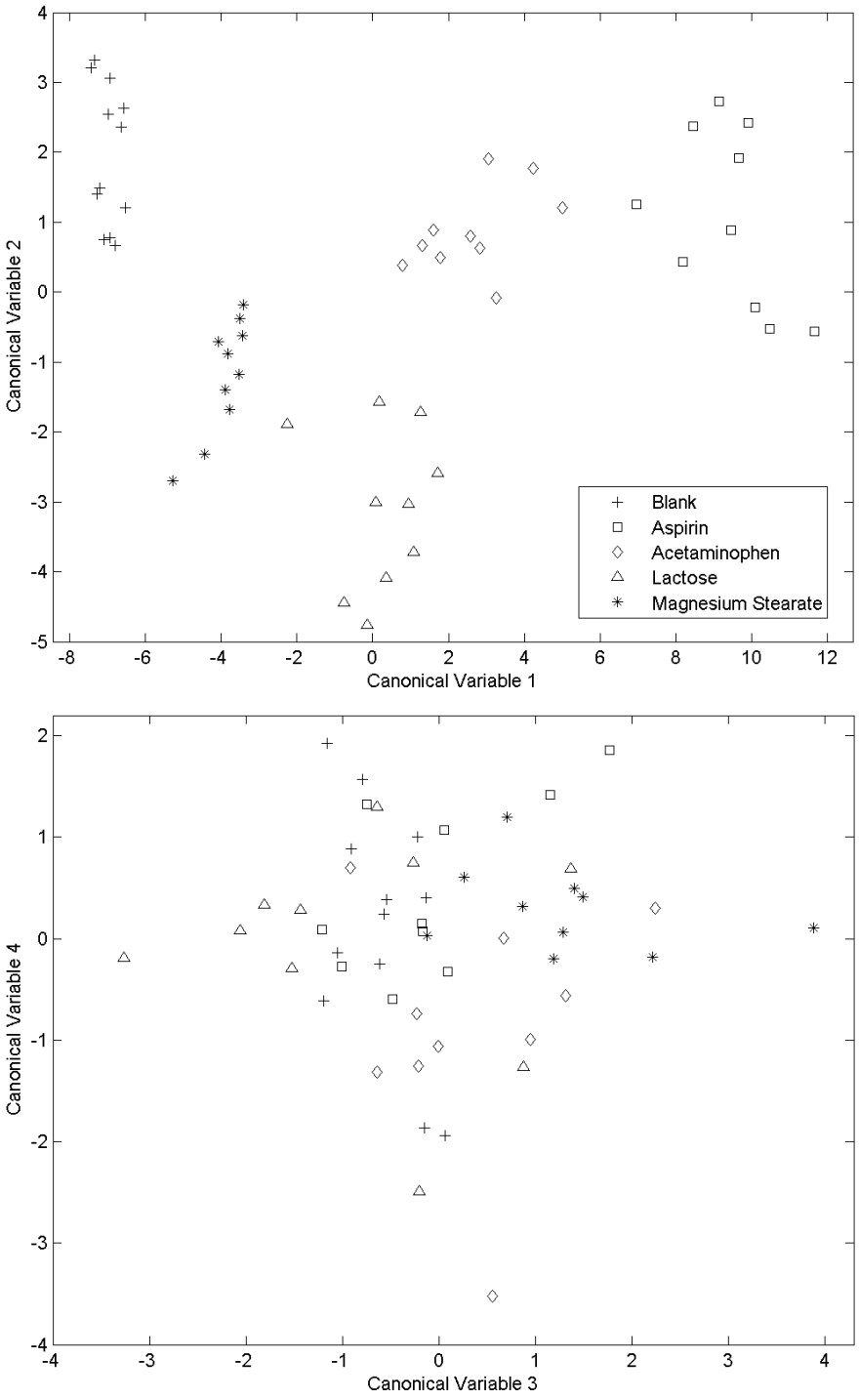

Figure 6.7: Graphs of the first four canonical variables from the transmission MFC data set of APIs and excipients on glass plates. The legend in the top graph (CV 1 vs. CV 2) applies to the bottom graph (CV 3 vs. CV 4).

Conclusion of Dissertation

The cumulative work contained in this dissertation was aimed at demonstrating that relatively simple instrumentation coupled with or motivated by chemometric data analysis techniques can be successfully applied to solving challenging analytical problems. The demand for such techniques can be driven by different motivations. For invasive biomedical spectroscopy and imaging applications, such as a coronary catheter, the effectiveness of traditional instrumental designs is greatly diminished. For this type of application, the development of novel instrumentation is critical for achieving the desired aims. To this end, the potential application of MFC in a NIR-based catheter system for diagnosing vascular diseases was demonstrated. Specifically, the differentiation of biological components with significance in abdominal aortic aneurysm and vulnerable atherosclerotic plaques was achieved through intervening red blood cell solutions.

The demand for similar techniques in pharmaceutical applications is driven by different motivations. Traditional instrumentation is often applicable to the problems at hand, but alternative techniques that are rapid, cost effective, and rugged are highly desirable. The validation of cleaning procedures represents one such area where conventional analytical techniques are effective but suffer greatly in terms of efficiency. Toward this goal light scatter and polarization imaging were investigated for the detection and quantification of surface residues on glass and stainless steel surfaces. The discrimination between different residues was not established with these techniques, but this is not necessarily a significant drawback. If further investigation shows that a wide variety of substances can be detected at pharmaceutically relevant concentrations, a worst-case interpretation of the results could yield a powerful cleaning validation tool. To address situations where residue discrimination is desired, the use of NIR

spectroscopy and MFC was demonstrated for the differentiation of pharmaceutical components on glass and stainless steel surfaces.

References

1. Hart, J. R.; Golumbic, C.; Norris, K. H., Determination of Moisture Content of Seeds by near-Infrared Spectrophotometry of Their Methanol Extracts. *Cereal Chemistry* **1962,** 39, (2), 94-99.
2. Norris, K. H.; Butler, W. L., Techniques for Obtaining Absorption Spectra on Intact Biological Samples. *Ire Transactions on Biomedical Electronics* **1961,** BME8, (3), 153-157.
3. Ciurczak, E. W.; Burns, D. A., *Handbook of Near-Infrared Analysis*. 2nd ed.; Marcel Dekker Inc.: New York, 2001; p 814.
4. Siesler, H. W.; Ozaki, Y.; Kawata, S.; Heise, H. M., *Near-Infrared Spectroscopy: Principles, Instruments, Applications*. Wiley-VCH: Weinheim, 2002; p 361.
5. Chalmers, J. M.; Griffiths, P. R., *Handbook on Vibrational Spectroscopy*. John Wiley $ Sons: Chichester, UK, 2001.
6. Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L., *Chemometrics: A Textbook*. Elsevier: Amsterdam, 1988; p 500.
7. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics - Part A*. Elsevier: Amsterdam, 1997; p 886.
8. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics - Part B*. Elsevier: Amsterdam, 1998; p 886.
9. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B., *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc.: New York, 1998.

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

11. Kowalski, B. R., *Chemometrics: Mathematics and Statistics in Chemistry*. Springer: New York, 2001.

12. Brereton, R. G., *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, Inc.: Chichester, UK, 2003.

13. Geladi, P.; MacDougall, D.; Martens, H., Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy* **1985,** 39, 491-500.

14. Isaksson, T.; Naes, T., The Effect of Multiplicative Scatter Correction and Linearity Improvement on NIR Spectroscop. *Applied Spectroscopy* **1988,** 42, 1273-1284.

15. Boor, C. d., *A Practical Guide To Splines*. 2nd ed.; Springer-Verlag: New York, 2001; p 358.

16. Dierckx, P., *Curve and Surface Fitting with Splines*. Oxford University Press: Oxford, England, 1993.

17. Jolliffe, I. T., *Principal Components Analysis*. 2nd ed.; Springer-Verlag: NY, 2002; p 502.

18. Boardman, A. E.; Hui, B. S.; Wold, H., The Partial Least-Squares Fix Point Method of Estimating Interdependent Systems with Latent-Variables. *Communications in Statistics Part a-Theory and Methods* **1981,** 10, (7), 613-639.

19. Gerlach, R. W.; Kowalski, B. R.; Wold, H. O. A., Partial Least-Squares Path Modeling with Latent-Variables. *Analytica Chimica Acta-Computer Techniques and Optimization* **1979,** 3, (4), 417-421.

20. Wold, H. O., Operative Aspects of Econometric and Sociological Models Current Developments of Fp (Fix-Point) Estimation and Nipals (Nonlinear Iterative Partial Least Squares) Modelling. *Economie Appliquee* **1973,** 26, (2-4), 385-421.

21. Wold, S.; Sjostrom, M.; Eriksson, L., PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001,** 58, (2), 109-130.

22. Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., The Collinearity Problem in Linear-Regression - the Partial Least-Squares (Pls) Approach to Generalized Inverses. *Siam Journal on Scientific and Statistical Computing* **1984,** 5, (3), 735-743.

23. Wold, S.; Martens, H.; Wold, H., The Multivariate Calibration-Problem in Chemistry Solved by the Pls Method. *Lecture Notes in Mathematics* **1983,** 973, 286-293.

24. Huberty, C. J., *Applied Discriminant Analysis*. John Wiley & Sons, Inc.: NY, 1994.

25. Mahalanobis, P. C., On the generalized distance in statistics. *Proceedings of the National Institute of Science of india* **1936,** 12, 49-55.

26. Efron, B., The bootstrap and modern statistics. *Journal of the American Statistical Association* **2000,** 95, (452), 1293-1296.

27. Efron, B., Better Bootstrap Confidence-Intervals. *Journal of the American Statistical Association* **1987,** 82, (397), 171-185.

28. Diaconis, P.; Efron, B., Computer-Intensive Methods in Statistics. *Scientific American* **1983,** 248, (5), 116-130.

29. Efron, B.; Gong, G., A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *American Statistician* **1983,** 37, (1), 36-48.

30. Efron, B., 1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife. *Annals of Statistics* **1979,** 7, (1), 1-26.

31. Lodder, R. A.; Selby, M.; Hieftje, G. A., Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. *Anal Chem* **1987,** 59, (15), 1921-1930.

32. Lodder, R. A.; Hieftje, G. A., Quantile BEAST Attacks the False-Sample Problem in Near-Infrared Reflectance Analysis. *Applied Spectroscopy* **1988,** 42, (8), 1351-1365.

33. Lodder, R. A.; Hieftje, G. A., Detection of Subpopulations in Near-Infrared Reflectance Analysis. *Applied Spectroscopy* **1988,** 42, (8), 1500-1512.

34. Zou, Y.; Xia, Y.; Jones, A. R.; Lodder, R. A., Making Your Best Case - Near-IR Spectral Identification of Soil. *Analytical Chemistry* **1993,** 65, (9), A434-A439.

35. Cassis, L. A.; Dai, B.; Urbas, A.; Lodder, R. A., In vivo applications of a molecular computing-based high-throughput NIR spectrometer. *Proc. SPIE-Int. Soc. Opt. Eng.* **2004,** 5329, (239-253).

36. Dai, B.; Urbas, A.; Douglas, C.; Lodder, R. A., Molecular Factor Computing for Predictive Spectroscopy. *Pharmaceutical Research* **2007,** In Press.

37. Dai, B. Simulations-Guided Design of Process Analytical Sensor using Molecular Factor Computing. University of Kentucky, 2007.

38. Damiani, P. C.; Escandar, G. M.; Olivieri, A. C.; Goicoechea, H. C., Multivariate calibration: A powerful tool in pharmaceutical analysis. *Current Pharmaceutical Analysis* **2005,** 1, (2), 145-154.

39. Kalivas, J. H., Multivariate calibration, an overview. *Analytical Letters* **2005,** 38, (14), 2259-2279.

40. Nelson, M. P.; Aust, J. F.; Dobrowolski, J. A.; Verly, P. G.; Myrick, M. L., Multivariate optical computation for predictive spectroscopy. *Analytical Chemistry* **1998,** 70, (1), 73-82.

41. Bialkowski, S. E., Species discrimination and quantitative estimation using incoherent linear optical signal processing of emission signals. *Analytical Chemistry* **1986,** 58, (12), 2561-2563.

42. Prakash, A. M. C.; Stellman, C. M.; Booksh, K. S., Optical regression: a method for improving quantitative precision of multivariate prediction with single channel spectrometers. *Chemometrics and Intelligent Laboratory Systems* **1999,** 46, (2), 265-274.

43. Soyemi, O.; Eastwood, D.; Zhang, L.; Li, H.; Karunamuni, J.; Gemperline, P.; Synowicki, R. A.; Myrick, M. L., Design and testing of a multivariate optical element: The first demonstration of multivariate optical computing for predictive spectroscopy. *Analytical Chemistry* **2001,** 73, (6), 1069-1079.

44. Myrick, M. L.; Soyemi, O. O.; Haibach, F.; Zhang, L.; Greer, A.; Li, H.; Priore, R.; Schiza, M. V.; Farr, J. R., Application of multivariate optical computing to near-infrared imaging. *Proceedings of SPIE-The International Society for Optical Engineering* **2002,** 4577, 148-157.

45. Myrick, M. L.; Soyemi, O. O.; Schiza, M. V.; Farr, J. R.; Haibach, F.; Greer, A.; Li, H.; Priore, R., Application of multivariate optical computing to simple near-infrared point measurements. *Proceedings of SPIE-The International Society for Optical Engineering* **2002,** 4574, 208-215.

46. Mitchell, M. E.; Sidawy, A. N., The pathophysiology of atherosclerosis. *Semin Vasc Surg* **1998,** 11, (3), 134-41.

47. Ross, R., Atherosclerosis--an inflammatory disease. *N Engl J Med* **1999,** 340, (2), 115-26.

48. Ross, R., Atherosclerosis is an inflammatory disease. *Am Heart J* **1999,** 138, (5 Pt 2), S419-20.

49. Hamilton, C. A., Low-density lipoprotein and oxidised low-density lipoprotein: their role in the development of atherosclerosis. *Pharmacol Ther* **1997,** 74, (1), 55-72.

50. Nordestgaard, B. G.; Nielsen, L. B., Atherosclerosis and arterial influx of lipoproteins. *Curr Opin Lipidol* **1994,** 5, (4), 252-7.

51. Berliner, J.; Leitinger, N.; Watson, A.; Huber, J.; Fogelman, A.; Navab, M., Oxidized lipids in atherogenesis: formation, destruction and action. *Thromb Haemost* **1997,** 78, (1), 195-9.

52. Navab, M.; Berliner, J. A.; Watson, A. D.; Hama, S. Y.; Territo, M. C.; Lusis, A. J.; Shih, D. M.; Van Lenten, B. J.; Frank, J. S.; Demer, L. L.; Edwards, P. A.; Fogelman, A. M., The Yin and Yang of oxidation in the development of the fatty streak. A review based on the 1994 George Lyman Duff Memorial Lecture. *Arterioscler Thromb Vasc Biol* **1996,** 16, (7), 831-42.

53. Navab, M.; Hama, S. Y.; Nguyen, T. B.; Fogelman, A. M., Monocyte adhesion and transmigration in atherosclerosis. *Coron Artery Dis* **1994,** 5, (3), 198-204.

54. Schwartz, C. J.; Valente, A. J.; Sprague, E. A., A modern view of atherogenesis. *Am J Cardiol* **1993,** 71, (6), 9B-14B.

55. DiCorleto, P. E., Cellular mechanisms of atherogenesis. *Am J Hypertens* **1993,** 6, (11 Pt 2), 314S-318S.

56. Stary, H. C., The sequence of cell and matrix changes in atherosclerotic lesions of coronary arteries in the first forty years of life. *Eur Heart J* **1990,** 11 Suppl E, 3-19.

57. Fuster, V.; Moreno, P. R.; Fayad, Z. A.; Corti, R.; Badimon, J. J., Atherothrombosis and High-Risk Plaque Part I: Evolving Concepts. *J Am Coll Cardiol* **2005,** 46, (6), 937-954.

58. Shah, P. K., Mechanisms of Plaque Vulnerability and Rupture. *J Am Coll Cardiol* **2003,** 41, (4), 15S-22S.

59. Virmani, R.; Burke, A. P.; Andrew, F.; D., K. F., Pathology of the Unstable Plaque. *Prog Cardiovasc Dis* **2002,** 44, (5), 349-356.

60. Virmani, R.; Kolodgie, F. D.; Burke, A. P.; Farb, A.; Schwartz, S. M., Lessons From Sudden Coronary Death: A Comprehensive Morphological Classification Scheme for Atherosclerotic Lesions. *Arterioscler Thromb Vasc Biol* **2000,** 20, (5), 1262-1275.

61. Virmani, R.; Burke, A.; Kolodgie, F.; al., e., Pathology of the thin-cap fibroatheroma: a type of vulnerable plaque. *J Interv Cardiol* **2003,** 16, (3).

62. Kolodgie, F.; Burke, A.; Farb, A.; al., e., The thin-cap fibroatheroma: a type of vulnerable plaque: the major precursor lesion to acute coronary syndromes. *Curr Opin Cardiol* **2001,** 16, (5), 285-292.

63. Yan, W. D.; Perk, M.; Chagpar, A.; Wen, Y.; Stratoff, S.; Schneider, W. J.; Jugdutt, B. I.; Tulip, J.; Lucas, A., Laser-Induced Fluorescence .3. Quantitative-Analysis of Atherosclerotic Plaque Content. *Lasers in Surgery and Medicine* **1995,** 16, (2), 164-178.

64. Oraevsky, A. A.; Jacques, S. L.; Pettit, G. H.; Sauerbrey, R. A.; Tittel, F. K.; Nguy, J. H.; Henry, P. D., Xecl Laser-Induced Fluorescence of Atherosclerotic Arteries - Spectral Similarities between Lipid-Rich Lesions and Peroxidized Lipoproteins. *Circulation Research* **1993,** 72, (1), 84-90.

65. Laifer, L. I.; Obrien, K. M.; Stetz, M. L.; Gindi, G. R.; Garrand, T. J.; Deckelbaum, L. I., Biochemical Basis for the Difference between Normal and Atherosclerotic Arterial Fluorescence. *Circulation* **1989,** 80, (6), 1893-1901.

66. Rokosova, B.; Rapp, J. H.; Porter, J. M.; Bentley, J. P., Composition and Metabolism of Symptomatic Distal Aortic Plaque. *Journal of Vascular Surgery* **1986,** 3, (4), 617-622.

67. Smith, E. B., Influence of Age and Atherosclerosis on Chemistry of Aortic Intima .2. Collagen and Mucopolysaccharides. *Journal of Atherosclerosis Research* **1965,** 5, (2), 241-248.

68. Noble, N. L.; Boucek, R. J.; Kao, K. Y. T., Biochemical Observations of Human Atheromatosis - Analysis of Aortic Intima. *Circulation* **1957,** 15, (3), 366-372.

69. Campa, J. S.; Greenhalgh, R. M.; Powell, J. T., Elastin degradation in abdominal aortic aneurysms. *Atherosclerosis* **1987,** 65, (1-2), 13-21.

70. Powell, J.; Greenhalgh, R. M., Cellular, enzymatic, and genetic factors in the pathogenesis of abdominal aortic aneurysms. *J Vasc Surg* **1989,** 9, (2), 297-304.

71. Rizzo, R. J.; McCarthy, W. J.; Dixit, S. N.; Lilly, M. P.; Shively, V. P.; Flinn, W. R.; Yao, J. S., Collagen types and matrix protein content in human abdominal aortic aneurysms. *J Vasc Surg* **1989,** 10, (4), 365-73.

72. Sakalihasan, N.; Heyeres, A.; Nusgens, B. V.; Limet, R.; Lapiere, C. M., Modifications of the extracellular matrix of aneurysmal abdominal aortas as a function of their size. *Eur J Vasc Surg* **1993,** 7, (6), 633-7.

73. Baxter, B. T.; Davis, V. A.; Minion, D. J.; Wang, Y. P.; Lynch, T. G.; McManus, B. M., Abdominal aortic aneurysms are associated with altered matrix proteins of the nonaneurysmal aortic segments. *J Vasc Surg* **1994,** 19, (5), 797-802; discussion 803.

74. White, J. V.; Mazzacco, S. L., Formation and growth of aortic aneurysms induced by adventitial elastolysis. *Ann N Y Acad Sci* **1996,** 800, 97-120.

75. Ghorpade, A.; Baxter, B. T., Biochemistry and molecular regulation of matrix macromolecules in abdominal aortic aneurysms. *Ann N Y Acad Sci* **1996,** 800, 138-50.

76. White, J. V.; Haas, K.; Phillips, S.; Comerota, A. J., Adventitial elastolysis is a primary event in aneurysm formation. *J Vasc Surg* **1993,** 17, (2), 371-80; discussion 380-1.

77. Minion, D. J.; Davis, V. A.; Nejezchleb, P. A.; Wang, Y.; McManus, B. M.; Baxter, B. T., Elastin is increased in abdominal aortic aneurysms. *J Surg Res* **1994,** 57, (4), 443-6.

78. Halloran, B. G.; Baxter, B. T., Pathogenesis of aneurysms. *Semin Vasc Surg* **1995,** 8, (2), 85-92.

79. Menashi, S.; Campa, J. S.; Greenhalgh, R. M.; Powell, J. T., Collagen in abdominal aortic aneurysm: typing, content, and degradation. *J Vasc Surg* **1987,** 6, (6), 578-82.

80. Glass, C. K.; Witztum, J. L., Atherosclerosis. the road ahead. *Cell* **2001,** 104, (4), 503-16.

81. Manning, M. W.; Cassis, L. A.; Huang, J.; Szilvassy, S. J.; Daugherty, A., Abdominal aortic aneurysms: fresh insights from a novel animal model of the disease. *Vasc Med* **2002,** 7, (1), 45-54.

82. Dempsey, R. J.; Davis, D. G.; Buice, R. G.; Lodder, R. A., Biological and medical applications of near-infrared spectroscopy. *Applied Spectroscopy* **1996,** 50, 18A-34A.

83. Spielman, A. J.; Zhang, G.; Yang, C. M.; D'Ambrosio, P.; Serizawa, S.; Nagata, M.; von Gizycki, H.; Alfano, R. R., Intracerebral hemodynamics probed by near infrared spectroscopy in the transition between wakefulness and sleep. *Brain Res* **2000,** 866, (1-2), 313-25.

84. Shaw, R. A.; Mansfield, J. R.; Kupriyanov, V. V.; Mantsch, H. H., In vivo optical/near-infrared spectroscopy and imaging of metalloproteins. *J Inorg Biochem* **2000,** 79, (1-4), 285-93.

85. McKinley, B. A.; Marvin, R. G.; Cocanour, C. S.; Moore, F. A., Tissue hemoglobin O2 saturation during resuscitation of traumatic shock monitored using near infrared spectrometry. *J Trauma* **2000,** 48, (4), 637-42.

86. Gabriely, I.; Wozniak, R.; Mevorach, M.; Kaplan, J.; Aharon, Y.; Shamoon, H., Transcutaneous glucose measurement using near-infrared spectroscopy during hypoglycemia. *Diabetes Care* **1999,** 22, (12), 2026-32.

87. Jaross, W.; Neumeister, V.; Lattke, P.; Schuh, D., Determination of cholesterol in atherosclerotic plaques using near infrared diffuse reflection spectroscopy. *Atherosclerosis* **1999,** 147, (2), 327-37.

88. Cassis, L. A.; Lodder, R. A., Near-IR imaging of atheromas in living arterial tissue. *Anal Chem* **1993,** 65, (9), 1247-56.

89. Dempsey, R. J.; Cassis, L. A.; Davis, D. G.; Lodder, R. A., Near-infrared imaging and spectroscopy in stroke research: lipoprotein distribution and disease. *Ann N Y Acad Sci* **1997,** 820, 149-69.

90. Moreno, P. R.; Lodder, R. A.; Purushothaman, K. R.; Charash, W. E.; O'Connor, W. N.; Muller, J. E., Detection of lipid pool, thin fibrous cap, and inflammatory cells in human aortic atherosclerotic plaques by near-infrared spectroscopy. *Circulation* **2002,** 105, (8), 923-7.

91. Daugherty, A.; Manning, M. W.; Cassis, L. A., Angiotensin II promotes atherosclerotic lesions and aneurysms in apolipoprotein E-deficient mice. *J Clin Invest* **2000,** 105, (11), 1605-12.

92. Bullivant, S., Freeze Etching and Fracturing. In *Some Biological Techniques in Electron Microscopy*, Parsons, D. F., Ed. Academic Press: New York, 1970; pp 123-146.

93. Sharifi, A. M.; Li, J. S.; Endemann, D.; Schiffrin, E. L., Effects of enalapril and amlodipine on small-artery structure and composition, and on endothelial dysfunction in spontaneously hypertensive rats. *J Hypertens* **1998,** 16, (4), 457-66.

94. Thomas, E. V.; Haaland, A. C., Comparison of multivariate calibration methods for quantitative spectral analysis. *Anal Chem* **1990,** 62, 1091-1099.

95. Beebe, K. R.; Kowalski, B. R., Introduction to multivariate calibration & analysis. *Anal Chem* **1987,** 59, 1007A-1017A.

96. Reid, J. C.; Wong, E. C., Data-Reduction and -Search System for Digital Absorbance Spectra. *Applied Spectroscopy* **1966,** 20, (5), 320-325.

97. Fayad, Z. A.; Fuster, V., Clinical imaging of the high-risk or vulnerable atherosclerotic plaque. *Circulation Research* **2001,** 89, (4), 305-316.

98. MacNeill, B. D.; Lowe, H. C.; Takano, M.; Fuster, V.; Jang, I. K., Intravascular modalities for detection of vulnerable plaque - Current status. *Arteriosclerosis Thrombosis and Vascular Biology* **2003,** 23, (8), 1333-1342.

99. Nemirovsky, D., Imaging of high-risk plaque. *Cardiology* **2003,** 100, (4), 160-175.

100. Pasterkamp, G.; Falk, E.; Woutman, H.; Borst, C., Techniques characterizing the coronary atherosclerotic plaque: Influence on clinical decision making. *Journal of the American College of Cardiology* **2000,** 36, (1), 13-21.

101. Pulido, M. A.; Angiolillo, D. J.; Costa, M. A., Imaging of atherosclerotic plaque. *International Journal of Cardiovascular Imaging* **2004,** 20, (6), 553-559.

102. Rudd, J. H. F.; Davies, J. R.; Weissberg, P. L., Imaging of atherosclerosis - Can we predict plaque rupture? *Trends in Cardiovascular Medicine* **2005,** 15, (1), 17-24.

103. Weissberg, P. L., Imaging of atherosclerosis: Can we predict plaque rupture? *Atherosclerosis Supplements* **2006,** 7, (3), 156-157.

104. Chen, J. W.; Wasserman, B. A., Vulnerable plaque imaging. *Neuroimaging Clinics of North America* **2005,** 15, (3), 609.

105. Nighoghossian, N.; Derex, L.; Douek, P., The vulnerable carotid artery plaque - Current imaging methods and new perspectives. *Stroke* **2005,** 36, (12), 2764-2772.

106. Bhatia, V.; Bhatia, R.; Dhindsa, S.; Dhindsa, M., Imaging of the vulnerable plaque: new modalities. *Southern Medical Journal* **2003,** 96, (11), 1142-1147.

107. Amirbekian, V.; Lipinski, M. J.; Briley-Saebo, K. C.; Amirbekian, S.; Aguinaldo, J. G. S.; Weinreb, D. B.; Vucic, E.; Frias, J. C.; Hyafil, F.; Mani, V.; Fisher, E. A.; Fayad, Z. A., Detecting and assessing macrophages in vivo to evaluate atherosclerosis noninvasively using molecular MRI. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, (3), 961-966.

108. Yuan, C.; Kerwin, W. S.; Yarnykh, V. L.; Cai, J. M.; Saam, T.; Chu, B. C.; Takaya, N.; Ferguson, M. S.; Underhill, H.; Xu, D. X.; Liu, F.; Hatsukami, T. S., MRI of atherosclerosis in clinical trials. *Nmr in Biomedicine* **2006,** 19, (6), 636-654.

109. Wilensky, R. L.; Song, H. K.; Ferrari, V. A., Role of magnetic resonance and intravascular magnetic resonance in the detection of vulnerable plaques. *Journal of the American College of Cardiology* **2006,** 47, (8), C48-C56.

110. Fayad, Z. A.; Fuster, V.; Fallon, J. T.; Jayasundera, T.; Worthley, S. G.; Helft, G.; Aguinaldo, J. G.; Badimon, J. J.; Sharma, S. K., Noninvasive in vivo human coronary artery lumen and wall imaging using black-blood magnetic resonance imaging. *Circulation* **2000,** 102, (5), 506-510.

111. Toussaint, J. F.; LaMuraglia, G. M.; Southern, J. F.; Fuster, V.; Kantor, H. L., Magnetic resonance images lipid, fibrous, calcified, hemorrhagic, and thrombotic components of human atherosclerosis in vivo. *Circulation* **1996,** 94, (5), 932-938.

112. Liu, X. S.; Xu, J. R., Accurate magnetic resonance imaging of atherosclerotic plaques: Change future strategies for the diagnosis and therapy of atherosclerotic disease. *Medical Hypotheses* **2007,** 68, (3), 656-660.

113. Nicholls, S. J.; Sipahi, I., Emerging role of intravascular ultrasound in the assessment of experimental anti-atherosclerotic therapies. *Current Medicinal Chemistry* **2006,** 13, (15), 1727-1734.

114. Nissen, S. E.; Yock, P., Intravascular ultrasound - Novel pathophysiological insights and current clinical applications. *Circulation* **2001,** 103, (4), 604-616.

115. Meyer, R. A., Intravascular ultrasound technological advances and clinical applications. *Progress in Pediatric Cardiology* **1997,** 7, (3), 141-153.

116. de Korte, C. L.; Woutman, H. A.; van der Steen, A. F.; Pasterkamp, G.; Cespedes, E. I., Vascular tissue characterisation with IVUS elastography. *Ultrasonics* **2000,** 38, (1-8), 387-90.

117. Yock, P. G.; Fitzgerald, P. J.; Linker, D. T.; Angelsen, B. A. J., Intravascular Ultrasound Guidance for Catheter-Based Coronary Interventions. *Journal of the American College of Cardiology* **1991,** 17, (6), B39-B45.

118. Tobis, J. M.; Mallery, J.; Mahon, D.; Lehmann, K.; Zalesky, P.; Griffith, J.; Gessert, J.; Moriuchi, M.; McRae, M.; Dwyer, M. L.; Greep, N.; Henry, W. L., Intravascular Ultrasound Imaging of Human Coronary-Arteries Invivo - Analysis of Tissue Characterizations with Comparison to Invitro Histological Specimens. *Circulation* **1991,** 83, (3), 913-926.

119. Sanghi, P. K.; Oh, J.; Kim, J.; Cilingiroglu, M.; Mancuso, J.; Milner, T. E.; Feldman, M. D., Molecular imaging of vulnerable plaque with optical coherence tomography. *American Journal of Cardiology* **2006,** 98, (8A), 220M-221M.

120. Prabhudesai, V.; Phelan, C.; Yang, Y.; Wang, R. K.; Cowling, M. G., The potential role of optical coherence tomography in the evaluation of vulnerable carotid atheromatous plaques: A pilot study. *Cardiovascular and Interventional Radiology* **2006,** 29, (6), 1039-1045.

121. Tearney, G. J.; Jang, I. K.; Bouma, B. E., Optical coherence tomography for imaging the vulnerable plaque. *Journal of Biomedical Optics* **2006,** 11, (2).

122. Fujimoto, J. G.; Pitris, C.; Boppart, S. A.; Brezinski, M. E., Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy. *Neoplasia* **2000,** 2, (1-2), 9-25.

123. ten Have, A. G.; Gijsen, F. J. H.; Wentzel, J. J.; Slager, C. J.; van der Steen, A. F. W., Temperature distribution in atherosclerotic coronary arteries: influence of plaque geometry and flow (a numerical study). *Physics in Medicine and Biology* **2004,** 49, (19), 4447-4462.

124. Madjid, M.; Naghavi, M.; Malik, B. A.; Litovsky, S.; Willerson, J. T.; Casscells, W., Thermal detection of vulnerable plaque. *American Journal of Cardiology* **2002,** 90, (10C), 36L-39L.

125. Stefanadis, C.; Diamantopoulos, L.; Vlachopoulos, C.; Tsiamis, E.; Dernellis, J.; Toutouzas, K.; Stefanadi, E.; Toutouzas, P., Thermal heterogeneity within human atherosclerotic coronary arteries detected in vivo - A new method of detection by application of a special thermography catheter. *Circulation* **1999,** 99, (15), 1965-1971.

126. Casscells, W.; Hathorn, B.; David, M.; Krabach, T.; Vaughn, W. K.; McAllister, H. A.; Bearman, G.; Willerson, J. T., Thermal detection of cellular infiltrates in living atherosclerotic plaques: Possible implications for plaque rupture and thrombosis. *Lancet* **1996,** 347, (9013), 1447-1449.

127. Zarrabi, A.; Gul, K.; Willerson, J. T.; Casscells, W.; Naghavi, M., Intravascular thermography: a novel approach for detection of vulnerable plaque. *Current Opinion in Cardiology* **2002,** 17, (6), 656-662.

128. Verheye, S.; Diamantopoulos, L.; Serruys, P. W.; Van Langenhove, G., Intravascular imaging of the vulnerable atherosclerotic plague: spotlight on temperature measurement. *Journal of Cardiovascular Risk* **2002,** 9, (5), 247-254.

129. van de Poll, S. W. E.; Romer, T. J.; Puppels, G. J.; van der Laarse, A., Raman spectroscopy of atherosclerosis. *Journal of Cardiovascular Risk* **2002,** 9, (5), 255-261.

130. Buschman, H. P.; Motz, J. T.; Deinum, G.; Romer, T. J.; Fitzmaurice, M.; Kramer, J. R.; van der Laarse, A.; Bruschke, A. V.; Feld, M. S., Diagnosis of human coronary atherosclerosis by morphology-based Raman spectroscopy. *Cardiovascular Pathology* **2001,** 10, (2), 59-68.

131. Motz, J. T.; Fitzmaurice, M.; Miller, A.; Gandhi, S. J.; Haka, A. S.; Galindo, L. H.; Dasari, R. R.; Kramer, J. R.; Feld, M. S., In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque. *Journal of Biomedical Optics* **2006,** 11, (2).

132. Scepanovic, O. R.; Fitzmaurice, M.; Gardecki, J. A.; Angheloiu, G. O.; Awasthi, S.; Motz, J. T.; Kramer, J. R.; Dasari, R. R.; Feld, M. S., Detection of morphological markers of vulnerable atherosclerotic plaque using multimodal spectroscopy. *Journal of Biomedical Optics* **2006,** 11, (2).

133. Arakawa, K.; Isoda, K.; Ito, T.; Nakajima, K.; Shibuya, T.; Ohsuzu, F., Fluorescence analysis of biochemical constituents identifies atherosclerotic plaque with a thin fibrous cap. *Arteriosclerosis Thrombosis and Vascular Biology* **2002,** 22, (6), 1002-1007.

134. Angheloiu, G. O.; Arendt, J. T.; Muller, M. G.; Haka, A. S.; Georgakoudi, I.; Motz, J. T.; Scepanovic, O. R.; Kuban, B. D.; Myles, J.; Miller, F.; Podrez, E. A.; Fitzmaurice, M.; Kramer, J. R.; Feld, M. S., Intrinsic fluorescence and diffuse reflectance spectroscopy identify superficial foam cells in coronary plaques prone to erosion. *Arteriosclerosis Thrombosis and Vascular Biology* **2006,** 26, (7), 1594-1600.

135. Pande, A. N.; Kohler, R. H.; Aikawa, E.; Weissleder, R.; Jaffer, F. A., Detection of macrophage activity in atherosclerosis in vivo using multichannel, high-resolution laser scanning fluorescence microscopy. *Journal of Biomedical Optics* **2006,** 11, (2).

136. Marcu, L.; Fang, Q. Y.; Jo, J. A.; Papaioannou, T.; Dorafshar, A.; Reil, T.; Qiao, J. H.; Baker, J. D.; Freischlag, J. A.; Fishbein, M. C., In vivo detection of macrophages in a rabbit atherosclerotic model by time-resolved laser-induced fluorescence spectroscopy. *Atherosclerosis* **2005,** 181, (2), 295-303.

137. Moreno, P. R.; Lodder, R. A.; Purushothaman, K. R.; Charash, W. E.; O'Connor, W. N.; Muller, J. E., Detection of lipid pool, thin fibrous cap, and inflammatory cells in human aortic atherosclerotic plaques by near-infrared spectroscopy. *Circulation* **2002,** 105, (8), 923-927.

138. Wang, J.; Geng, Y. J.; Guo, B.; Klima, T.; Lal, B. N.; Willerson, J. T.; Casscells, W., Near-infrared spectroscopic characterization of human advanced atherosclerotic plaques. *Journal of the American College of Cardiology* **2002,** 39, (8), 1305-1313.

139. Charash, W. E.; Lodder, R. A.; Moreno, P. R.; Purushothaman, K. R.; Swain, J. A.; O'Connor, W. N.; Muller, J. E., Detection of simulated vulnerable plaque using a novel near infrared spectroscopy catheter. *Journal of the American College of Cardiology* **2000,** 35, (2), 38A-38A.

140. Leardi, R., Genetic algorithm-PLS as a tool for wavelength selection in spectral data sets. *Data Handling in Science and Technology* **2003,** 23, 169-196.

141. Leardi, R., Application of genetic algorithm-PLS for feature selection in spectral data sets. *Journal of Chemometrics* **2000,** 14, (5-6), 643-655.

142. Medendorp, J.; Wyse, J.; Lodder, R. A., Rapid near-infrared qualification of microcrystalline cellulose and sodium caprate minitablets through intact enteric coated capsules. *PAT-The Journal of Process Analytical Technology* **2006,** 3, (3), 11-17.

143. Medendorp, J.; Buice, R. G.; Lodder, R. A., Acoustic-resonance spectrometry as a process analytical technology for the quantification of active pharmaceutical ingredient in semi-solids. *AAPS Pharmscitech* **2006,** 7, (3), 59.

144. Medendorp, J.; Lodder, R. A., Acoustic-resonance spectrometry as a process analytical technology for rapid and accurate tablet identification. *AAPS Pharmscitech* **2006,** 7, (1), 25.

145. ICH *Q7A Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients*; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use: August 2001, 2001.

146. PhRMA Quality Committee, B. P. W. G., PhRMA Guidelines for the Validation of Cleaning Procedures for Bulk Pharmaceutical Chemicals. *Pharm. Technol.* **1997,** 21, (9), 56-73.

147. FDA *Guide to Inspections of Validation of Cleaning Processes*; Food and Drug Administration: Washington, DC, July 1993, 1993.

148. Munson, J.; Stanfield, C. F.; Gujral, B., A review of process analytical technology (PAT) in the US pharmaceutical industry. *Current Pharmaceutical Analysis* **2006,** 2, (4), 405-414.

149. Hinz, D. C., Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative. *Analytical and Bioanalytical Chemistry* **2006,** 384, (5), 1036-1042.

150. Glover, C., Validation of the total organic carbon (TOC) swab sampling and test method. *PDA J Pharm Sci Technol* **2006,** 60, (5), 284-90.

151. Clark, K., How to develop and validate a total organic carbon method for cleaning applications. *Pda Journal of Pharmaceutical Science and Technology* **2001,** 55, (5), 290-294.

152. Jenkins, K. M.; Vanderwielen, A. J.; Armstrong, J. A.; Leonard, L. M.; Murphy, G. P.; Piros, N. A., Application of total organic carbon analysis to cleaning validation. *PDA J Pharm Sci Technol* **1996,** 50, (1), 6-15.

153. Carroll, J. J.; Le, T.; DeBono, R., An ultrafast alternative to HPLC for cleaning validation. *American Laboratory* **2004,** 36, (4), 32-35.

154. DeBono, R.; Carroll, J. J., Ion mobility spectrometry (IMS): An alternative to HPLC for cleaning validation. *Lc Gc North America* **2003**, 50.

155. Debono, R.; Stefanou, S.; Davis, M.; Walia, G., Using ion mobility spectrometry for cleaning verification in pharmaceutical manufacturing. *Pharmaceutical Technology North America* **2002,** 26, (4), 72-78.

156. Melling, P.; Thomson, M., Is this surface clean? Cleaning validation by direct spectroscopic surface analysis. *American Laboratory* **2006,** 38, (17), 16-21.

157. Hamilton, M. L.; Perston, B. B.; Harland, P. W.; Williamson, B. E.; Thomson, M. A.; Melling, P. J., Grazing-angle fiber-optic IRRAS for in situ cleaning validation. *Organic Process Research & Development* **2005,** 9, (3), 337-343.

158. Mehta, N. K.; Goenaga-Polo, J.; Hernandez-Rivera, S. P.; Hernandez, D.; Thomson, M. A.; Melling, P. J., Development of an in situ spectroscopic method for cleaning validation using mid-IR fiber optics. *Spectroscopy* **2003,** 18, (4), 14.

159. Perston, B. B.; Hamilton, M. L.; Williamson, B. E.; Harland, P. W.; Thomson, M. A.; Melling, P. J., Grazing-angle fiber-optic fourier transform infrared reflection-absorption spectroscopy for the in situ detection and quantification of two active pharmaceutical ingredients on glass. *Analytical Chemistry* **2007,** 79, (3), 1231-1236.

160. Quiesup, K., A Multifunctional Active Excitation Spectral Analyzer. *NASA TECH BRIEF* **2002,** 26, (1), i, 1-2, 1a-18a.

161. Malik, I.; Poonacha, M.; Moses, J.; Lodder, R. A., Multispectral Imaging of Tablets in Blister Packaging. *AAPS PharmSciTech* **2001,** 2, (2), 9.

162. Stover, J. C., *Optical Scattering: Measurement and Analysis*. SPIE Optical Engineering Press: Bellingham, WA, 1995.

163. Kim, J. H.; Ehrman, S. H.; Mulholland, G. W.; Germer, T. A., Polarized light scattering by dielectric and metallic spheres on oxidized silicon surfaces. *Applied Optics* **2004,** 43, (3), 585-591.

164. Kim, J. H.; Ehrman, S. H.; Mulholland, G. W.; Germer, T. A., Polarized light scattering by dielectric and metallic spheres on silicon wafers. *Applied Optics* **2002,** 41, (25), 5405-5412.

165. Germer, T. A., Polarized light scattering by microroughness and small defects in dielectric layers. *Journal of the Optical Society of America a-Optics Image Science and Vision* **2001,** 18, (6), 1279-1288.

166. Sung, L. P.; Mulholland, G. W.; Germer, T. A., Polarized light-scattering measurements of dielectric spheres upon a silicon surface. *Optics Letters* **1999,** 24, (13), 866-868.

167. Germer, T. A.; Asmail, C. C., Polarization of light scattered by microrough surfaces and subsurface defects. *Journal of the Optical Society of America a-Optics Image Science and Vision* **1999,** 16, (6), 1326-1332.

168. Germer, T. A., Angular dependence and polarization of out-of-plane optical scattering from particulate contamination, subsurface defects, and surface microroughness. *Applied Optics* **1997,** 36, (33), 8798-8805.

169. Germer, T. A.; Asmail, C. C.; Scheer, B. W., Polarization of out-of-plane scattering from microrough silicon. *Optics Letters* **1997,** 22, (17), 1284-1286.

170. Videen, G., Polarized light scattering from surface contaminants. *Optics Communications* **1997,** 143, (4-6), 173-178.

171. Videen, G.; Hsu, J. Y.; Bickel, W. S.; Wolfe, W. L., Polarized-Light Scattered from Rough Surfaces. *Journal of the Optical Society of America a-Optics Image Science and Vision* **1992,** 9, (7), 1111-1118.

172. Shkuratov, Y.; Bondarenko, S.; Ovcharenko, A.; Pieters, C.; Hiroi, T.; Volten, H.; Munoz, O.; Videen, G., Comparative studies of the reflectance and degree of linear polarization of particulate surfaces and independently scattering particles. *Journal of Quantitative Spectroscopy & Radiative Transfer* **2006,** 100, (1-3), 340-358.

173. Petrov, D.; Synelnyk, E.; Shkuratov, Y.; Videen, G., The T-matrix technique for calculations of scattering properties of ensembles of randomly oriented particles with different size. *Journal of Quantitative Spectroscopy & Radiative Transfer* **2006,** 102, (1), 85-110.

174. Grynko, Y.; Shkuratov, Y.; Videen, G., Polarization of near-forward-scattered light from particulate substrates illuminated at near-grazing angles. *Journal of Quantitative Spectroscopy & Radiative Transfer* **2006,** 101, (3), 522-526.

175. Zubko, E.; Shkuratov, Y.; Hart, M.; Eversole, J.; Videen, G., Backscatter of agglomerate particles. *Journal of Quantitative Spectroscopy & Radiative Transfer* **2004,** 88, (1-3), 163-171.

176. Shkuratov, Y.; Ovcharenko, A.; Zubko, E.; Volten, H.; Munoz, O.; Videen, G., The negative polarization of light scattered from particulate surfaces and of independently

scattering particles. *Journal of Quantitative Spectroscopy & Radiative Transfer* **2004,** 88, (1-3), 267-284.

177. Holler, S.; Auger, J. C.; Stout, B.; Pan, Y.; Bottiger, J. R.; Chang, R. K.; Videen, G., Observations and calculations of light scattering from clusters of spheres. *Applied Optics* **2000,** 39, (36), 6873-6887.

178. Germer, T. A.; Rinder, T.; Rothe, H. In *Polarized light scattering measurements of polished and etched steel surfaces*, Proc SPIE, 2000; Gu, Z.-H.; Maradudin, A. A., Eds. 2000; pp 148-155.

179. Kotz, S.; Nadarajah, *Extreme Value Distributions: Theory and Applications*. Imperial College Press: London, 2001.

180. Blanco, M.; Alcala, M., Use of near-infrared spectroscopy for off-line measurements in the pharmaceutical industry. In *Process Analytical Technology*, Bakeev, K., Ed. Blackwell Publishing Ltd.: Oxford, UK, 2005; pp 362-391.

181. Ciurczak, E. W.; Drennen, J. K., III, *Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy.* Marcel Dekker, Inc.: New York, N.Y., 2002.

182. Ciurczak, E. W.; Drennen, J. K., Pharmaceutical applications of near-infrared spectroscopy. *Practical Spectroscopy* **2001,** 25, 349-366.

183. Skibsted, E., Near infrared spectroscopy: the workhorse in the PAT toolbox. *Spectroscopy Europe* **2006,** 18, (5), 14-17.

184. Grout, B. In *Cleaning validation by direct swab NIR analysis.*, 38th Middle Atlantic Regional Meeting of the American Chemical Society, Hershey, PA, 2006; American Chemical Society: Hershey, PA, 2006.

185. Medendorp, J. P.; Paudel, K. S.; Lodder, R. A.; Stinchcomb, A. L., Near infrared spectrometry for the quantification of human dermal absorption of econazole nitrate and estradiol. *Pharmaceutical Research* **2007,** 24, (1), 186-193.

186. Medendorp, J.; Yedluri, J.; Hammell, D. C.; Ji, T.; Lodder, R. A.; Stinchcomb, A. L., Near-infrared spectrometry for the quantification of dermal absorption of econazole nitrate and 4-cyanophenol. *Pharmaceutical Research* **2006,** 23, (4), 835-843.

187. Fountain, W.; Dumstorf, K.; Lowell, A. E.; Lodder, R. A.; Mumper, R. J., Near-infrared spectroscopy for the determination of testosterone in thin-film composites. *Journal of Pharmaceutical and Biomedical Analysis* **2003,** 33, (2), 181-189.

188. El-Hagrasy, A. S.; Morris, H. R.; D'Amico, F.; Lodder, R. A.; Drennen, J. K., 3rd, Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity. *J Pharm Sci* **2001,** 90, (9), 1298-307.

189. Gold, T. B.; Buice, R. G., Jr.; Lodder, R. A.; Digenis, G. A., Detection of formaldehyde-induced crosslinking in soft elastic gelatin capsules using near-infrared spectrophotometry. *Pharm Dev Technol* **1998,** 3, (2), 209-14.

190. Gold, T. B.; Buice, R. G., Jr.; Lodder, R. A.; Digenis, G. A., Determination of extent of formaldehyde-induced crosslinking in hard gelatin capsules by near-infrared spectrophotometry. *Pharm Res* **1997,** 14, (8), 1046-50.

191. Buice, R. G., Jr.; Gold, T. B.; Lodder, R. A.; Digenis, G. A., Determination of moisture in intact gelatin capsules by near-infrared spectrophotometry. *Pharm Res* **1995,** 12, (1), 161-3.

192. Cogdill, R. P.; Anderson, C. A.; Drennen, J. K., Using NIR spectroscopy as an integrated PAT tool. *Spectroscopy* **2004,** 19, (12), 104-109.

193. Bakeev, K. A.; Ciurczak, E. W., Near-infrared spectroscopy as a process analytical tool - Part 1: Laboratory applications. *Spectroscopy* **2003,** 18, (11), 32-35.

194. Lowell, A. E.; Ho, K.-S.; Lodder, R. A., Remote Hyperspectral Imaging of Endolithic Biofilms Using a Robotic Probe. *Contact in Context* **2002,** 1, (1), 1-10.

Vita

AARON A. URBAS
Born: November 6, 1976
Cincinnati, Ohio

EDUCATION

Ohio University, Bachelor of Specialized Studies in Chemometrics, June 2001

PROFESSIONAL

Research Assistant, University of Kentucky Department of Chemistry, 2001-2007
Teaching Assistant, University of Kentucky Department of Chemistry, 2001-2002

HONORS

2004-2006 NIH Cardiovascular Training Grant Fellowship
2003-2004 Kentucky Opportunity Fellowship
2002-2004 IGERT-NSF Fellowship
2001-2002 Franklin R. Tuttle Fellowship

PUBLICATIONS

"Anisotropic Polarized Light Scatter Imaging for Pharmaceutical Cleaning Validation." Urbas AA, Lodder RA. 2007, Submitted to *AAPS PharmSciTech*.

"Molecular Factor Computing for Predictive Spectroscopy." Dai B, Urbas AA, Douglas CC, Lodder RA. 2007, *Pharmaceutical Research. In Press*.

"Centrifugal Microfluidics with Integrated Sensing Microdome Optodes for Multi-Ion Detection." Watts AS, Urbas AA, Moschou E, Gavalas VG, Madou MJ, Bachas LG. 2006. Submitted to *Analytical Chemistry*.

"Understanding the Kinetics and Reaction Progression in Microdome Optodes using Confocal Microscopy." Watts AS, Urbas AA, Chopra N, Gavalas VG, Bachas LG. 2006. In Preparation.

"Molecular Factor Computing Near-Infrared Spectroscopy for Differentiating Cholesterol, Collagen and Elastin through Red Blood Cell Solutions." Urbas AA, Shamsaie A, Dai B, Lodder RA. 2006, Submitted to *Analytical Chemistry*.

"Decyl Methacrylate-Based Microspot Optodes." Watts AS, Urbas AA, Finley T, Gavalas VG, Bachas LG. *Analytical Chemistry,* 2006, 78(2), pp 524-529.

"Real-time broad-band measurement of cholesterol, collagen, and elastin using a novel rotary switch spectrometer." Claps R, Guynn R, Serafin W, Virojanapa J, Urbas A, Lodder R. *Proceedings of SPIE,* 2006, 6078. pp 60782G/1-60782G/18.

"Hyperspectral integrated computational imaging." Cassis LA, Urbas AA, Lodder RA. *Analytical and Bioanalytical Chemistry,* 2005, 382(4), pp 868-872.

"Magnetoelastic transducers for monitoring coagulation, clot inhibition, and fibrinolysis." Puckett LG, Lewis JK, Urbas AA, et al. *Biosensors & Bioelectronics*, 2005, 20(9), pp 1737-1743.

"^{26}Al-containing acidic and basic sodium aluminum phosphate preparation and use in studies of oral aluminum bioavailability from foods utilizing ^{26}Al as an aluminum tracer." Yokel RA, Urbas AA, Lodder RA, Selegue JP, Florence RL. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms,* 2005, 229(3-4), pp 471-478.

"In vivo applications of a molecular computing-based high-throughput NIR spectrometer." Cassis LA, Dai B, Urbas AA, Lodder RA. *Proceedings of SPIE,* 2004, 5329, pp 239-253.

"Near-Infrared Spectrometry of Abdominal Aortic Aneurysm in the Apo $E^{-/-}$ Mouse." Urbas AA, Manning MW, Daugherty A, Cassis LA, Lodder RA. *Analytical Chemistry*, 2003, 75, pp 3318-3323.

"*In situ* spectroscopic cleaning validation." Urbas, AA, Lodder, RA. *NIR News,* 2003, 14(2), pp 22-24.

"Two-Dimensional Wavelet Compression of Ion Mobility Spectra." Urbas, AA, Harrington, PB. *Analytica Chimica Acta*, 2001, 446, pp 393-412.

"Evaluation of Neural Network Models with Generalized Sensitivity Analysis." Harrington PB, Urbas AA, Wan, C. *Analytical Chemistry*, 2000, 72, pp 5004-5013.

"Two-dimensional Correlation Analysis." Harrington PB, Urbas AA, Tandler PJ. *Chemometrics and Intelligent Laboratory Systems*, 2000, 50, pp 149-174.

"Temperature-Constrained Cascade Correlation Network and Its Application to Bacteria Identification." Zhang Z, Urbas AA, Harrington PB, Voorhees KJ, Rees J. *Chemical Journal of Chinese Universities*, 23:4 (2002) 570-572.

PRESENTATIONS

"Quantification of BSA on Glass by Laser Scatter for Cleaning Validation." Aaron A. Urbas, Robert A. Lodder. PITTCON, March 10, 2004, Chicago, IL. Oral Presentation.

"Magnetohydrodynamic Currents in Blood Coagulation." Aaron A. Urbas, Robert A. Lodder, Gerald Thompson, Karthik Venkatachalam, Erika Wanner, PITTCON, March 11, 2004, Chicago, IL. Oral Presentation.

"Near-Infrared Spectrometry of Abdominal Aortic Aneurysm in the $ApoE^{-/-}$ Mouse." Aaron A. Urbas, Robert A. Lodder, Michael Manning, Alan Daugherty, Lisa A. Cassis. PITTCON, March 12, 2004, Chicago, IL. Oral Presentation.

"Reduction of Manufacturing Problems Using Near-Infrared Spectrometry (NIRS) as a Process Analytical Technology." Aaron A. Urbas, Robert A. Lodder. Institute of Validation Technology - Process Analytical Technology (PAT) Conference, October 23, 2002, Gaithersburg, MD. Oral Presentation.

"Near-Infrared Spectrometry of Abdominal Aortic Aneurysm in the Apo E-/- Mouse." Aaron Urbas, Tracy Henriques, Debra Rateri, Alan Daugherty, Lisa Cassis, Robert Lodder. Gill Heart Institute Cardiovascular Research Day, October 24, 2003, Lexington, KY. Poster Presentation.

"Synchrotron Infrared Microspectroscopic Analysis of Human Thin-Cap Fibroatheromas." Aaron Urbas, David Wetzel, Ginell Post, Robert Lodder. Gill Heart Institute Cardiovascular Research Day, October 25, 2004, Lexington, KY. Poster Presentation.

"Molecular Factor Computing for Differentiating Cholesterol, Elastin and Collagen Samples Through Blood." Aaron Urbas, Bin Dai, Ali Shamsaie, Robert Lodder. Gill Heart Institute Cardiovascular Research Day, October 7, 2005, Lexington, KY. Poster Presentation.