LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TH-623193

Statistical Data Analyses of Trace Chemical, Biochemical, and Physical Analytical Signatures

R. N. Udey

February 28, 2013

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

STATISTICAL DATA ANALYSES OF TRACE CHEMICAL, BIOCHEMICAL, AND PHYSICAL ANALYTICAL SIGNATURES

By

Ruth Norma Udey

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy

2013

ABSTRACT

STATISTICAL DATA ANALYSES OF TRACE CHEMICAL, BIOCHEMICAL, AND PHYSICAL ANALYTICAL SIGNATURES

By

Ruth Norma Udey

Analytical and bioanalytical chemistry measurement results are most meaningful when interpreted using rigorous statistical treatments of the data. The same data set may provide many dimensions of information depending on the questions asked through the applied statistical methods. Three principal projects illustrated the wealth of information gained through the application of statistical data analyses to diverse problems.

Firstly, novel aerosol test particles containing DNA barcodes were developed for the accurate assessment of aerosol transport and fate in populated locations. Aerosols are central to human and environmental health, and understanding the properties of these aerosols that are pervasive in our lives is essential. Test particles composed of FDA-approved saccharide food additives were generated using both a modified inkjet printer and a commercial spray dryer. Univariate statistical methods were used to evaluate generated particle size-distributions during production optimization. Non-coding DNA templates were incorporated into the particles as unique particle identifiers, which yielded customized test particles detectable using highly specific quantitative real-time polymerase chain reaction (QRT-PCR) assays. These safe, customizable, and specifically detected aerosol test particles will provide vital experimental feedback for evaluating aerosol dispersion and transport models. The project culminated with a successful demonstration of the aerosol test particles in an atmospheric release test.

Secondly, an original method for non-invasively analyzing the chemical profiles of latent fingerprint residues was developed in order to gain a new level of information from the most common type of forensic evidence. Passive solid-phase microextraction (SPME) headspace sampling collects both endogenous and exogenous volatile and semi-volatile compounds contained in the fingerprint residue while preserving the fingerprint for traditional analyses. The information-rich chemical profiles obtained from gas chromatography-mass spectrometry (GC-MS) analyses of the SPME samples were used to quantitatively compare fingerprint compounds both between subjects and over a time course of 30 days using multivariate statistical analyses.

Finally, endogenous metabolite profiles of cancer cells treated with anti-cancer agents were analyzed using gas chromatography- and high performance liquid chromatography- mass spectrometry (GC-MS and HPLC-MS), and the resulting complex data sets were interrogated using univariate and multivariate statistical analyses (*e.g.* ANOVA, PCA, PLS-DA, OPLS-DA). Possible modes of cytotoxicity of cisplatin and taxol, two commonly used cancer therapeutics, in breast and lung cancer cells were elucidated using statistical methods for data reduction in order to focus on the cellular biochemical processes most affected by drug treatment. Understanding how successful therapeutics interact with cells leads to design of novel anti-cancer agents that are more targeted and effective, minimizing dose-limiting side-effects and saving more lives.

Although the areas of study are diverse, the commonality is the generation of highly complex data tables that may be effectively analyzed and interpreted using statistical methods.

ACKNOWLEDGEMENTS

As my path through graduate school was unusual, there are many people that deserve gratitude for their help along the way. I am deeply grateful to Dan Jones, my research advisor at Michigan State University (MSU), for his valuable mentorship and support of my Master's degree work and moving out to California to do some of my doctoral work at Lawrence Livermore National Laboratory (LLNL). I truly appreciate all of the efforts to make my special arrangements successful. I am also indebted to George Farquar, my research supervisor at LLNL, for his support of my work and prudent advice. I thank my committee members, Gary Blanchard, Merlin Bruening, and Ruth Smith, for their guidance and sustained interest through the years. Other personnel at MSU that deserve my gratitude are Chrysoula Vasileiou and Babak Borhan, who collaborated with us on the cancer metabolomics work, as well as the Jones group and Forensic Chemistry group members for their suggestions and support. I also thank Christine Hara, Elizabeth Wheeler, Mary Rosa, Brian Baker, Maxim Shusteff, Cindy Thomas, and Beth Vitalis for research assistance and support at LLNL.

I am truly grateful to my supportive family: my parents, Dana and Joyce, my sister and brother-in-law, Amanda and Jonathan, all of my family members, and my sweetie, John. Thank you for being constant sources of encouragement and perspective.

Finally, I would like to acknowledge the Lawrence Livermore National Laboratory Lawrence Scholar Program and the Defense Threat Reduction Agency for supporting the aerosol test particles work and the Rapid Reaction Technology Office for supporting the

fingerprint project. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-TH-623193.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Motivations and Introduction

Modern instrumental analysis techniques generate torrents of data. At least 90% of all current analytical chemistry work is performed using instrumental methods as opposed to classical ("wet chemistry") analysis techniques (Miller and Miller 2005). This is because many modern instrumental methods are more sensitive, have a wider quantitative dynamic range, and offer capability to measure multiple analytes in a single analysis. In addition, modern analytical instruments are almost always interfaced with computers for sophisticated system control and the storage, treatment, and reporting of data. Thus it is common for thousands of data points to be collected during a single analysis. For example, methods combining chromatographic separations with spectroscopic or mass spectrometric detection can identify and quantify hundreds of components in a complex mixture within a few minutes. In addition, the analytical chemistry field is expanding towards characterization of the spatial distributions and time-dependence of abundances of multiple analytes, as well as automation of sample analyses enabling the processing of more samples, yielding results with high information content.

How do we deal with all of the data? The work presented in the following chapters illustrates the large amount of data generated by contemporary trace chemical and physical signature analyses and examines the statistical analyses required to interpret the data. One example is size-distribution data for sugar aerosol particles generated using varying parameters during the development of a novel aerosol test particle. After analyzing hundreds of thousands of particles from several different generated particle populations, how do we determine that

varying particle production parameters resulted in particles that are significantly different in size? Another example derives from analysis of all detectable endogenous and exogenous compounds in latent fingerprint residues deposited by different human subjects that were sampled at different times in order to discover subject habits and traits. Compound abundances could be altered by three variables: the subject, the analysis time point, and random analytical variance. How do we determine the magnitude of each variable's contribution to the differences in the results for each compound, and for each fingerprint chemical profile as a whole? A final example arises during analysis of all detectable endogenous metabolites in human lung and breast cancer cells both with and without treatment with the anticancer drugs cisplatin and taxol over the course of seven days to examine metabolic responses to cancer treatment. The cancer cell type, whether the cells were treated or not, what they were treated with, and the time post-dose that the cells were collected are all variables that could explain compound abundance differences in the analytical results. How do we determine which variables are the dominating ones for affecting metabolic response? Which metabolite abundances are varying significantly and are correlated with cell type and treatment variables? The following sections discuss statistical data analysis approaches for answering these questions and interpreting information-rich analytical chemistry analyses.

1.2 Statistical Approaches for Evaluating and Interpreting Analytical Data

The key question in experiments across many disciplines focuses on whether varying an experimental condition leads to differences in measurable outcomes. How do we evaluate analytical results for statistically significant differences? The simplest case involves comparing

two measurements to determine the effect of varying an independent experimental variable (*e.g.* fingerprint deposited by two different subjects, untreated and treated cancer cells). How do we determine if the result depends on the independent variable, or if the results are simply random? No measurement process is perfectly reproducible, and all analytical measurements are subject to fluctuations that are random in magnitude. The precision of a measurement is determined by repeating the measurement, and statistical significance can be derived from measures of the spread in measurement outcomes using the standard deviation, variance, and/or relative standard deviation (coefficient of variation) (Skoog *et al.* 1998). A major goal of statistical analyses is to separate and determine the magnitude of the sources of experimental variance (either random or controlled experimental variables) to establish the probability that differences in outcomes can be attributed to random chance. When this probability is sufficiently low, the investigator has greater confidence in assigning a relationship between experimental conditions and outcomes. Statistical approaches for evaluating both univariate (one variable) and multivariate (more than one variable) data for significant sample differences that facilitate experimental interpretation are discussed below.

1.2.1 Univariate Methods

Univariate statistical approaches for evaluating and interpreting analytical data are useful when there are more samples than analytes or properties being measured (*e.g.* mean particle size results for several different particle populations), and usually focus on comparisons of mean values for a single variable at a time (Trygg *et al.* 2006). The Student's *t*-test compares two experimental means and determines if they are significantly different using the standard

deviations and number of replicate measurements of the two experimental populations. Comparing the calculated *t* statistic to a table of critical *t* values assesses the level of confidence in the resulting decision that the experimental means are the same or different. When more than two sets of experimental means need to be compared (*e.g.* mean particle size results from six sugar particle populations generated using different sugar concentrations), analysis of variance (ANOVA) is a powerful statistical procedure that can separate and estimate the different sources of variation in a sample set. The ANOVA method is used to separate the variation caused by changing the experimental variable (between-sample variation; *e.g.* sugar concentration) from the variation due to random error (within-sample variation). The calculated *F* test statistic is the ratio between the between-sample variation and the within-sample variation, and the calculated *F* statistic is compared to a table of critical *F* values to assess the level of confidence in the experimental means being the same or different. Larger *F* values calculated from experimental results indicate higher confidence in the sample means differing due to the purposely-varied experimental variable and not random variation (Miller and Miller 2005).

In summary, the univariate Student's *t*-test and ANOVA analyses test whether altering a single experimental variable leads to statistically significant changes in analytical measurements. When multiple experimental parameters are varied or multiple analytes or properties are measured for each sample, multivariate statistical approaches are required to examine the results of changing each parameter, as well as the interactions between parameters or between analytes.

1.2.2 Multivariate Correlation Methods

Multivariate analytical data sets are more challenging to interpret compared to univariate data due to the interactions between the multiple variables. Testing whether altering one experimental variable significantly changes the results for many analytes or sample properties requires more sophisticated statistical methods. Pair-wise sample correlation analysis is one of the simplest approaches, where only two samples are compared using measurements of multiple analytes in each sample (*e.g.* all compounds detected in the latent fingerprint residues deposited by two different individuals). The goal of multivariate correlation analysis is to test how similar, or correlated, the two samples are based on all measured analyte signals (*e.g.* chromatogram peak heights resulting from gas chromatography-mass spectrometry (GC-MS) analyses of the latent fingerprint residues). The two most common pair-wise correlation coefficients are the Pearson product-moment correlation coefficient (PPMCC) and Spearman's rank correlation coefficient (SRCC). The coefficient calculations employ the same algorithm (covariance of the two samples divided by the product of their standard deviations), however, the inputs for PPMCC are the raw data (*e.g.* chromatographic peak heights), and the inputs for SRCC are the ranked data (most abundant peak height is given a rank of 1, the second most abundant peak height given a rank of 2, and so on in order of decreasing peak height). Calculating coefficients using ranked data (SRCC) does not assume that the data are normally distributed, and as a result reveals all monotonic correlations. Conversely, PPMCC calculations assume normally distributed raw data sets with linear correlations only, which is not always true of the data set being analyzed. Therefore, if the distribution of the experimental data is unknown, the SRCC is a robust metric for the

correlation between two samples using multivariate data (*i.e.* all linear correlations are monotonic, but not all monotonic correlations are linear) (Spearman 1904, Meier and Zund 2000, Miller and Miller 2005).

1.2.3 Multivariate Projection Methods

While multivariate correlation methods adequately describe relationships between two samples, simultaneous comparisons of multivariate experimental results from more than two samples require multivariate projection approaches to aid interpretation of complex data profiles. The goal of multivariate projection methods is to visualize similarities among many samples based on all measured analyte signals in a single presentation (*e.g.* metabolite abundances resulting from GC-MS analyses of breast and lung cancer cell extracts after treatment with either taxol or cisplatin) (Trygg *et al.* 2006, Trygg *et al.* 2007). Data containing complex sample chemical profiles collected using hyphenated analytical techniques (*e.g.* GC-MS and high-performance liquid chromatography-mass spectrometry (HPLC-MS)) require data processing steps to convert three-dimensional data for each sample (chromatogram retention time (RT), mass (*m/z*), and signal abundance data) into information that reflects abundances of individual analytes. The resulting multivariate data matrix is made up of the signal abundances (*e.g.* integrated peak areas) of all detected compound peaks, each identified using a unique RT-*m/z* pair, for each sample in the study. The multivariate data matrix is then normalized, scaled, and mean-centered as appropriate. Compound abundances within each sample may be normalized to the sum of all peak areas in that sample, which is a set value (*e.g.* 10000) for all samples to facilitate comparison. Scaling and mean-centering are performed by first calculating

the mean and standard deviation values for each variable (*e.g.* compound RT-*m/z* pair) in the sample data matrix. Mean-centering shifts the compound abundances in the data matrix so that the means are centered at the origin, which makes result visualization and interpretation more straightforward because comparisons are made to mean values for the population. Scaling is an important part of determining the weight, or importance, of each variable in fitting a projection model to the data, and ensures that a collective measure of similarity will be less likely to undervalue contributions by low abundance compounds with smaller standard deviations and that abundant compounds will not dominate sample comparisons. The two most common scaling methods are to unit variance and Pareto variance. Unit variance scales the variable according to the inverse of the variable's standard deviation, and Pareto variance scales the variable using the inverse of the square root of the variable's standard deviation. Pareto scaling lies between no scaling and unit variance scaling and divides each variable by the square root of its standard deviation, up-weighting features of medium abundance without magnifying baseline noise. In contrast, unit variance scaling adjusts all variables to yield equal variance and hence equal importance in fitting the projection model (Trygg *et al.* 2006, MKS Umetrics 2012).

When analytical measurements generate data sets with high dimensionality, as is the case with GC-MS and HPLC-MS analyses, recognition of the features that distinguish classes of samples is enabled by visualization tools that simplify data interpretation. Multivariate projection methods, including principal component analysis (PCA) and partial least squares (PLS), convert the multi-dimensional data set into a low-dimensional (typically two to five dimensions) model plane, which facilitates data visualization and interpretation (Figure 1.1).

Principal component analysis highlights covariance of different measured compound abundances and is unbiased by prior knowledge of sample class (*i.e.* unsupervised) (Pearson 1901, Wold *et al.* 1987, Trygg *et al.* 2006). It also allows reduction of data dimensionality to two dimensions that can be visualized on one plot. As described above, the multivariate data matrix resulting from GC-MS analyses is often composed of the integrated peak areas of all compounds (*i.e.* the variables) detected in all of the samples. A covariance matrix is then calculated, which examines how all variables (measured compounds) change with respect to one another, using Equation 1.1 for two variables, *x* and *y*:

$$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \qquad [1.1]$$

From the covariance values, linear combinations of variables that fluctuate similarly are calculated. These linear combinations are called eigenvectors, and are the principal components (PCs) of the analysis. Each eigenvector has an associated eigenvalue, which describes the amount of information contained in the eigenvector. Therefore, higher eigenvalues indicate more descriptive eigenvectors (PCs) for revealing the variance in the data set. The same number of PCs is calculated as the original number of variables in the data set. Each PC is calculated to be orthogonal to the preceding PC and describes the next greatest amount of variance in the data set. Principal component 1 is calculated to correlate the highest sources of variation in the data set, PC2 is orthogonal to PC1 and describes the next highest sources of variation, and so on. The result of PCA is that the largest and most informative

sources of variance in the data set can now be examined using a smaller number of variables (*i.e.* two or three PCs), instead of all of the original data variables. This also allows simple visualization of sample similarity as two PCs (a model plane), such as PC1 and PC2, may be plotted as the *x*- and *y*-axes of a single graph. For each individual sample, the loadings of the measured variables on each PC yield a score, and the *x,y* coordinates for each sample are displayed on this graph. This graph, called a scores plot, illustrates most of the variance in the samples projected into two dimensions. Samples with similar chemical composition will cluster together and away from chemically dissimilar samples on the scores plot, allowing conclusions to be drawn about how samples relate to one another. The scores plot may reveal groupings of chemically similar samples, sample relationships, and outliers (deviating samples). A corresponding plot called a loadings plot displays the variables that are responsible for the variance described by each PC. Variables that fluctuate similarly (*i.e.* have a high degree of covariance) will be positioned in the same area of the loadings plot. Again, PC1 is the *x*-axis and PC2 is the *y*-axis so that the position of a variable on the loadings plot corresponds to its contributions to sample positioning on the scores plot, and provides a powerful tool for understanding underlying patterns in the data (Miller and Miller 2005). In summary, PCA distinguishes samples and the variables responsible for sample differences, though clustering in PCA plots is usually dominated by signals from the most abundant analytes.

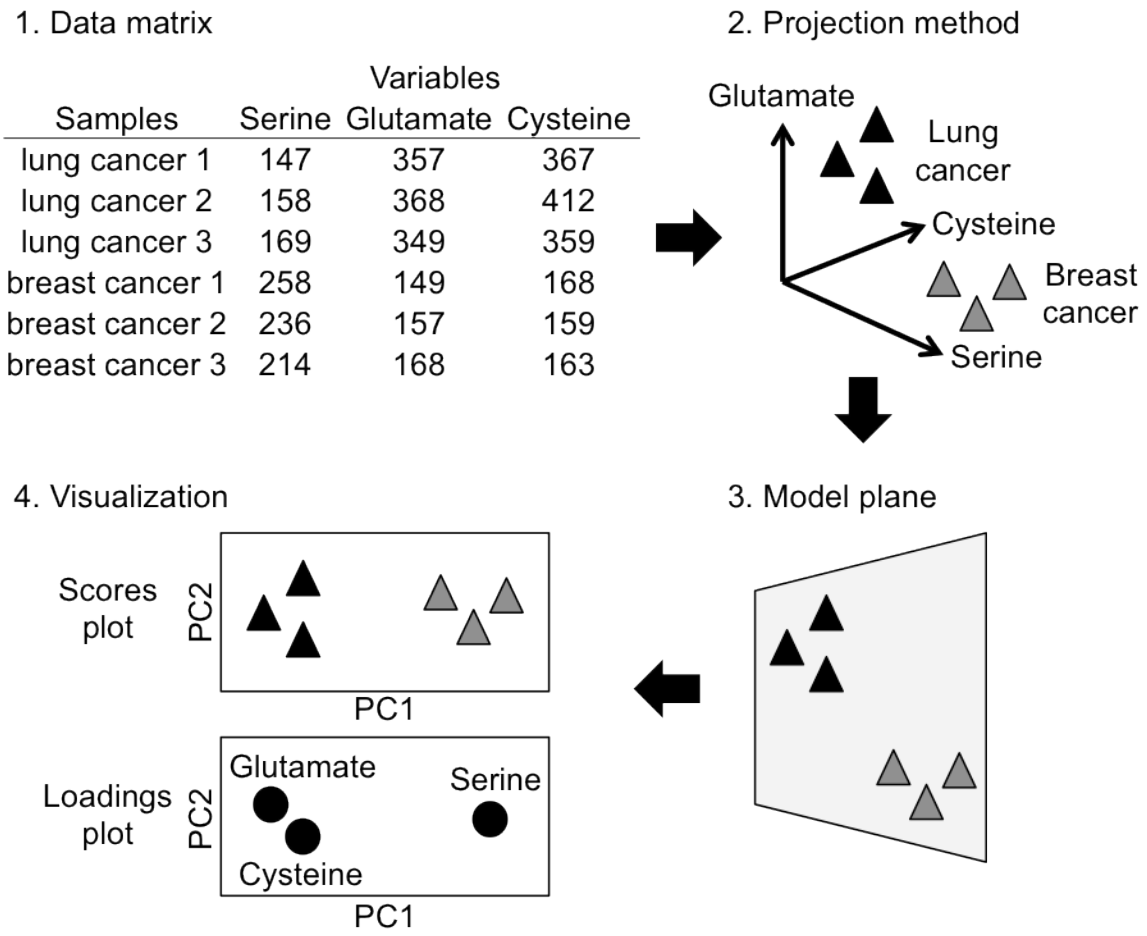

Figure 1.1. Overview of how multivariate projection methods work. (1) The data matrix containing values for each variable (*e.g.* metabolite abundances) measured in each sample observation (*e.g.* cancer cell samples) is projected into *n*-dimensional space (2), with an axis for each variable (three in the illustrated example) and one point for each sample (six in the example). Points (samples) positioned near one another are more similar than points that are far apart. (3) Projection methods such as PCA and PLS-DA find representative low-dimensional model planes that summarize the variation in the sample points for visualization (4). The scores plot gives a visual overview of sample relationships (*e.g.* groupings, outliers), and the corresponding loadings plot shows the contribution of the original variables to the sample positioning on the scores plot for interpretation.

The partial least squares (PLS) multivariate projection model relies on *a priori* knowledge of sample class (supervised) in order to extract information about measurements that most strongly discriminates the sample classes (Wold *et al.* 2001). This additional knowledge of sample class (*e.g.* cell type (breast cancer or lung cancer), dosing information

(treated or untreated)) is tabulated in a Y matrix, with the same number of samples as the data matrix, called the X matrix. The Y matrix can contain either quantitative information (*e.g.* serine concentration) or qualitative information to describe sample class (*e.g.* treated or untreated control). When the Y matrix is qualitative, each sample is assigned either a "1", indicating the sample is a member of that class, or a "0", indicating that it is not. For example, if a data set is made up of lung cancer cells and breast cancer cells, the Y matrix for lung cancer cells would have a "1" for all lung cancer samples and a "0" for all breast cancer samples. When a qualitative Y matrix is used to represent sample class, the method is called PLS discriminant analysis (PLS-DA) to distinguish it from analyses when a quantitative Y matrix is employed. Partial least squares is similar to PCA except there are two data matrices, X and Y. The PLS model focuses the model plane to describe the Y-related variation in X, *i.e.* variation correlated with different sample classes in the Y matrix, using a multivariate regression model. The results of PLS analyses are scores plots and loadings plots that display clusters of similar samples and the variables responsible for sample groupings, just like PCA models. Generally, the clustering of samples from the same class defined by the Y matrix is more clearly defined in a supervised PLS model compared to the unsupervised PCA model owing to the PLS model maximizing the sources of variation correlated with sample class. In summary, the PLS model is useful for finding correlations between the X and Y matrices in multivariate data sets with collinearity and more variables (analyte measurements) than samples, which is often the case. The variables highly correlated with sample class are distinguished from the background of weakly correlated signals, which is a powerful tool for experimental evaluation and interpretation (Wold *et al.* 2001, Trygg *et al.* 2006).

The interpretation of PLS models may be improved by using a slightly different model, the orthogonal partial least squares (OPLS) model. The X and Y matrices are the same as in PLS models (a qualitative Y matrix also makes it a discriminant analysis, OPLS-DA). The difference is that in OPLS the systematic variation in the X matrix that is not correlated (orthogonal) to the Y matrix is removed from the analysis. This reduces the complexity of the model while preserving the interpretation of the variables correlated with sample class. In summary, discriminant analysis approaches (PLS-DA and OPLS-DA) for analyzing multivariate data enhance contributions by measurements of low abundance signals compared to PCA. However, PLS-DA and OPLS-DA models are often confounded by the contributions of randomness in the measured values. When sufficiently large numbers of measurements are made, the chance of finding discrimination driven by randomness becomes substantial, and model validation must be performed to assess the probability that discrimination arises from random variance (Trygg and Wold 2002, Cloarec *et al.* 2005).

Finally, the PLS and OPLS methods can also be used to predict which class in the Y matrix an unknown sample belongs to. Prediction models develop a proposed mathematical relationship that is tested against a subset of samples to validate the results. The unknown sample could be a new measurement that needs to be evaluated in the context of the measurements already present in the data set, or it could be a subset of the original data set used for model validation and evaluation of its predictive ability. A predictive analysis begins with an X matrix (called the training set) of sample data and a Y matrix containing class information. A PLS or OPLS model is then fit to the training data, optimizing the model parameters to calculate the variance in the data set correlated with the Y variable. The

independent, unknown sample X matrix (called the prediction set) from the same data population as the training data set is then fit with the same PLS or OPLS model that was calculated for the training set, and the class in the Y matrix that the unknown sample belongs to is predicted. It is generally the case that the model does not fit the prediction set as well as it fits the training set because the prediction set is a smaller X matrix. However, PLS or OPLS models may have low predictive ability due to model overfitting. Overfitting arises when statistical models discriminate based on random noise in the data instead of the informative sample relationships superimposed on the noise. Overfitting is more likely to occur when the model is complex and the number of variables in the X matrix is larger than the number of samples (observations). Thus validating PLS and OPLS models by predicting the fit of the model to a prediction set is essential to ensure that the methods are describing informative sample variations and not the ever-present analytical variance. It is also important to note that the predictive ability of OPLS models may be less robust than PLS models owing to the removal of the information in the OPLS model orthogonal to the Y matrix. However, if the information uncorrelated with the Y matrix was the analytical variance, removal of the noise in the data set may yield an OPLS model with robust predictive ability that is valuable for interpreting differences between samples in the experiment (Wold *et al.* 2001, Trygg and Wold 2002, Cloarec *et al.* 2005, Trygg *et al.* 2006).

1.3 Overview of Projects

Chapter 2 discusses the development of a novel aerosol test particle containing DNA barcodes for the accurate assessment of aerosol transport and fate in populated locations.

Aerosols play a central role in comprehending the effects we have on our environment and the impacts of the environment on us. These aerosols both positively and negatively impact many facets of our lives, including climate (*e.g.* cloud formation), visibility (*e.g.* pollution hazes), and our health (*e.g.* pollen allergies, asthma inhalers). An understanding of the properties of aerosols is essential, as aerosol properties govern the generation, transport, and fate of airborne particulates, as well as efforts to control them. Test particles composed of FDA-approved saccharide food additives were generated using both a modified inkjet printer and a commercial spray dryer. Univariate statistical methods were used to evaluate particle size measurements to answer questions such as: Are two particle size-distributions different? How different? Does changing particle production parameters significantly change particle characteristics? Non-coding DNA templates were incorporated into the particles as unique particle identifiers, yielded customized test particles able to be specifically detected using quantitative real-time polymerase chain reaction (QRT-PCR) assays. These safe, customizable, and specifically detected aerosol test particles were generated with the same sizes as aerosols commonly observed in the environment on the gram-scale for studying large locations. No currently available simulant material meets all of these criteria for precisely studying atmospheric transport in populated environments. The project culminated with a successful demonstration of the aerosol test particles in an atmospheric release test.

Particle size is the principal parameter for characterizing the behavior of aerosols, as all properties of aerosols, including transport, are dependent on particle size to some extent. Aerodynamic diameter is used to describe aerosol particles instead of geometric diameter, as many aerosols have irregular shapes. Aerodynamic diameter is the diameter of a unit density

sphere that moves with the same velocity through a fluid as the aerosol particle. A particle's aerodynamic diameter is related to the Stokes' diameter, which assumes a particle is spherical, and its density by Equation 1.2:

$$d_{p,a} = d_{p,S}\sqrt{\rho_p} \quad [1.2]$$

where $d_{p,a}$ is the aerodynamic diameter of a particle, $d_{p,S}$ is the Stokes' diameter of the particle, and ρ_p is the density of the particle. The square of the aerodynamic diameter is also proportional to the terminal velocity of particles larger than 0.5 μm in diameter in still air (gravitational settling) according to Equation 1.3:

$$V_{TS} = \frac{\rho_0 d_{p,a}^2 g}{18\eta} \quad [1.3]$$

where V_{TS} is the terminal settling velocity of the particle, ρ_0 is the standard particle density (1000 kg/m^3), g is the acceleration due to gravity, and η is the viscosity of the gas (air) (Hinds 1999). Larger particles settle out of air more rapidly than smaller particles, which illustrates the importance of creating aerosol test particles that are similar in size to environmental aerosols. If the aerosol test particles are too large, they are deposited soon after release and are not

transported as far as the smaller environmental aerosols, yielding misleading simulations of transport of targeted aerosol types.

The motions of particles with diameters less than 0.5 μm are more complex compared to larger particles, as random Brownian motion is no longer negligible. The general case relating random particle diffusion to particle mobility is described by Equations 1.4 and 1.5:

$$D = \mu k_B T \quad [1.4]$$

where D is the diffusion constant, k_B is Boltzmann's constant, T is the absolute temperature, and μ is particle mobility, defined as the ratio of the particle's terminal drift velocity (v_d) to an applied force (F):

$$\mu = \frac{v_d}{F} \quad [1.5]$$

When an external force is applied to the particle, it dominates particle motion. However, when the particle is just drifting, diffusion due to random Brownian motion describes particle transport. A special case of Equation 1.4 is the Stokes-Einstein Equation (Equation 1.6), which relates the diffusion due to Brownian motion to particle size:

$$D = \frac{k_B T}{6\pi\eta r} \quad [1.6]$$

where D is the diffusion constant, k_B is Boltzmann's constant, T is the absolute temperature, η is the viscosity of the gas (air), and r is the Stokes radius of the particle (assumed spherical). The Stokes-Einstein equation predicts that smaller particles have larger diffusion constants compared to larger particles, and that diffusion must be considered when describing small particle transport (Dill and Bromberg 2003).

While predicting the transport of single particles using the relationships discussed above is possible, applying these relationships to a more realistic population of aerosol particles is challenging. Therefore there is a need for experimental feedback to inform and evaluate aerosol transport and dispersion models in order to accurately predict aerosol transport and fate in an environment. The novel aerosol test particles discussed in Chapter 2 were developed for this purpose, as they accurately mimic the size of environmental aerosols and are safe to use for atmospheric release tests in populated environments.

Chapter 3 presents the development of a novel method for non-invasively analyzing the chemical profiles of latent fingerprint residues in order to gain a new level of information from the most common type of forensic evidence. Passive solid-phase microextraction (SPME) headspace sampling collects both endogenous and exogenous volatile and semi-volatile compounds contained in the fingerprint residue while preserving the fingerprint for traditional analyses. The information-rich chemical profiles obtained from GC-MS analyses of the SPME samples were used to compare fingerprint compounds both between subjects and over a time course of 30 days. Multivariate correlation methods were used to answer the questions: Are two fingerprints different? How different? A multivariate projection method (PCA) was also

used to answer: Can multiple subjects be associated and discriminated using latent fingerprint residue chemical profiles? The changes in the fingerprint chemical profiles over 30 days are discussed as well.

Solid-phase microextraction, developed in 1989 by Pawliszyn and coworkers (Belardi and Pawliszyn 1989, Arthur and Pawliszyn 1990), revolutionized analytical sample preparation prior to chromatographic analysis. All common steps of sample preparation (extraction, enrichment, and introduction to the chromatograph) are rapidly achieved using a simple, inexpensive device that is reusable and does not require the use of solvents. Analytes are extracted and concentrated from a sample (either solid, liquid, or gas) onto the SPME fiber, which is an approximately 1 cm fused silica fiber coated with an organic polymer. Various polymer materials are used depending on the target compounds of the analysis. Analytes are absorbed or adsorbed into the fiber from the sample, and the quantity of analyte extracted by the fiber is proportional to the concentration in the sample. After extraction, the SPME fiber is transferred to the injection port of a gas chromatography or high-performance liquid chromatography system, where the analytes are desorbed from the fiber onto the column head and the separation and analysis take place (Wercinski and Pawliszyn 1999, Ulrich 2000).

The SPME fiber may extract compounds either directly by immersion into a sample solution or passively from the headspace above a sample. Headspace SPME sampling was chosen to extract compounds from latent fingerprints, as the SPME fiber does not directly contact the sample, leaving any friction ridge patterns unaltered for further analysis. Headspace sampling requires the analytes to partition between three phases: the solid or liquid sample (*e.g.* latent fingerprint residue), the gaseous phase above the sample, and the SPME fiber

polymer coating. Equation 1.7 gives the amount of an analyte that partitions into a SPME fiber during headspace sampling at equilibrium:

$$n_f = \frac{K^{fs} V_f n_0}{K^{fs} V_f + K^{hs} V_h + V_s} \quad [1.7]$$

where n_f is the amount of analyte extracted into the SPME fiber coating, K^{fs} is the distribution equilibrium constant of the analyte between the SPME fiber coating and the fingerprint sample, V_f is the volume of the SPME fiber coating, n_0 is the number of analyte molecules in the fingerprint residue prior to SPME, K^{hs} is the equilibrium constant of the analyte between the sample headspace and the fingerprint sample, V_h is the volume of the sample headspace, and V_s is the volume of the fingerprint residue sample. The kinetics of mass transport determine the headspace SPME sampling time required to achieve equilibrium and relates to the diffusion coefficients and concentrations of the analytes in the three phases (latent fingerprint residue, headspace, SPME fiber), the volumes of the three phases, and the partition coefficients (fingerprint to headspace, headspace to fiber) of the analytes. Extraction equilibrium is reached when analyte concentrations are homogeneous in the three system phases. The affinity of the analytes to the three phases, the temperature of the system, and whether the system is static or agitated affects sampling time as well. While the transfer of volatile analytes from the fingerprint sample to the headspace occurs quickly, transfer of semi-volatile compounds with

relatively larger molecular mass and/or higher affinity for the fingerprint sample ("like dissolves like") occurs more slowly and increases SPME sampling time. Increasing the temperature and/or agitating the system accelerates the diffusion of analytes from the sample to the headspace and the SPME fiber, however, these methods are not amenable to sampling evidentiary latent fingerprint residues (Zhang and Pawliszyn 1993, Wercinski and Pawliszyn 1999, Ulrich 2000). The work detailed in Chapter 3 applied passive headspace SPME sampling to latent fingerprint residues as a new approach for chemically profiling forensic evidence to yield suspect information that is currently inaccessible.

Chapter 4 discusses the combined use of GC-MS and HPLC-MS along with chemometric procedures to profile the patterns of endogenous metabolites in taxol- and cisplatin-treated human A549 lung cancer and MCF7 breast cancer cells. Such metabolite patterns revealed metabolic networks altered by drug treatment, which in turn serve as guides for further analyses and novel hypotheses regarding the biochemical mechanisms of taxol and cisplatin anticancer action in two different cell types. Despite recent advances in both diagnostic and therapeutic tools available, the mortality rate remains high and severe side effects are associated with common chemotherapy treatments. It is of paramount importance to continue the search for new chemotherapeutics with novel modes of cytotoxicity that offer higher and more selective potency and have fewer side effects. The generated multivariate metabolomics data sets were evaluated and interpreted using PLS-DA and OPLS-DA to discover: What are the treatment-distinguishing metabolites indicative of anticancer drug mode(s) of action? This work demonstrates a new way to elucidate the mode of cytotoxicity of potential cancer therapeutic agents to rapidly screen drug candidates for further development.

REFERENCES

1.4 References

Arthur, C. L., Pawliszyn, J. (1990). Solid phase microextraction with thermal desorption using fused silica optical fibers. Anal. Chem. 62:2145-2148.

Belardi, R. P., Pawliszyn, J. B. (1989). The application of chemically modified fused silica fibers in the extraction of organics from water matrix samples and their rapid transfer to capillary columns. Water Poll. Res. J. Can. 24:179-91.

Cloarec, O., Dumas, M. E., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., Nicholson, J. K., Holmes, E. (2005). Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ^{1}H NMR spectroscopic metabonomic studies. Anal. Chem. 77:517-526.

Dill, K. A., Bromberg, S. (2003). *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*. Taylor and Francis, New York, NY, pp. 327-328.

Hinds, W. C. (1999). *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles* (2nd Ed.). John Wiley and Sons, Inc., New York, NY, pp. 53-54.

Meier, P. C., Zund, R. E. (2000). Statistical Methods in Analytical Chemistry, in *Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications, Vol. 153*, J. D. Winefordner, ed, John Wiley & Sons, Inc., New York, NY, pp. 92-93.

Miller, J. N., Miller, J. C. (2005). *Statistics and Chemometrics for Analytical Chemistry* (5th Ed.). Pearson Education Limited, Harlow, UK, pp. 107-108, 110-111, 142, 167-169, 215-219.

MKS Umetrics (2012). *User Guide to SIMCA 13*. Malmo, Sweden, pp. 155-157, 569.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philos. Mag. 2:559-572.

Shurubor, Y. I., Matson, W. R., Willett, W. C., Hankinson, S. E., Kristal, B. S. (2007). Biological variability dominates and influences analytical variance in HPLC-ECD studies of the human plasma metabolome. BMC Clin. Pathol. 7:9-22.

Skoog, D. A., Holler, F. J., Nieman, T. A. (1998). *Principles of Instrumental Analysis* (5th Ed.). Brooks/Cole Thomson Learning, Pacific Grove, CA, pp. 99-113, A-1 – A-20.

Spearman, C. (1904). The proof and measurement of association between two things. Am. J. Psychol. 15:72-101.

Trygg, J., Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). J. Chemometrics 16:119-128.

Trygg, J., Gullberg, J., Johansson, A. I., Jonsson, P., Moritz, T. (2006). Chemometrics in Metabolomics – An Introduction, in *Biotechnology in Agriculture and Forestry, Vol. 57: Plant Metabolomics*, K. Saito, R. A. Dixon, L. Willmitzer, eds, Springer, New York, NY, pp. 117-128.

Trygg, J., Holmes, E., Lundstedt, T. (2007). Chemometrics in metabonomics. J. Proteome Res. 6:469-479.

Ulrich, S. (2000). Solid-phase microextraction in biomedical analysis. J. Chromatogr. A 902:167-194.

Wercinski, S. A. S, Pawliszyn, J. (1999). Solid Phase Microextraction Theory, in *Solid Phase Microextraction: A Practical Guide*, S. A. S. Wercinski, ed, Marcel Dekker, Inc., New York, NY, pp. 1-26.

Wold, S., Esbensen, K., Geladi, P. (1987). Principal component analysis. Chemometr. Intell. Lab. 2:37-52.

Wold, S., Sjostrom, M., Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. 58:109-130.

Zhang, Z., Pawliszyn, J. (1993). Headspace solid-phase microextraction. Anal. Chem. 65:1843-1852.

CHAPTER 2: AEROSOL TEST PARTICLES WITH DNA BARCODES

2.1 Motivations and Introduction

Airborne particles are present throughout our environment in many different forms, including dust, smoke, smog, and fog. Microscopic particles in the air may originate naturally (*e.g.* resuspended soil or sand, salt from ocean spray) or arise from man-made processes (*e.g.* smoke from power generation, smog from automobile exhaust, hairspray). These are all examples of aerosols, which are solid particles or liquid droplets suspended in a gas. These aerosols both positively and negatively impact many facets of our lives including climate, visibility, and our health and the health of all living things. Aerosols form clouds in the atmosphere that are vital components in the hydrologic cycle of water movement on Earth, while other atmospheric particles create hazes that adversely affect temperature and rainfall, which impact global climate. Airborne particles that enter the human body cause a range of adverse effects, including minor allergies to biological particles like pollen, diseases such as cancer caused by tobacco smoke, and life-threatening viral and bacterial infections. Conversely, aerosols may also be administered as therapeutic drugs (*e.g.* asthma inhalers), which is a rapidly growing portion of the pharmaceutical industry. Aerosols play a central role in understanding the effects we have on our environment and the impacts of the environment on us (Hinds 1999).

An understanding of the properties of these aerosols that are pervasive in our lives is essential. Aerosol properties govern the generation, transport, and fate of airborne particulates as well as efforts to control them. Particle size is the principal parameter for characterizing the behavior of aerosols, as all properties of aerosols are dependent on particle size to some

extent. Aerosols exist in a broad range of sizes, from 0.002 μm to more than 100 μm in diameter. It is common for an aerosol population to be rather polydisperse and contain particles spanning two orders of magnitude in size. Particles larger than 10 μm have limited stability in the atmosphere as they quickly fall out due to their large mass. Respirable particles range from 0.005-5 μm in diameter and their toxicity depends on both the physical and chemical properties of the aerosol, which are instrumental in evaluating airborne particulate hazards. Particle shape is also important for characterizing aerosol properties, but to a lesser extent than particle size, as typical variations in particle shape rarely produce more than a twofold change in any aerosol property. Bulk properties of aerosols, such as density and viscosity, differ imperceptibly from those of pure air as the particulate phase of an aerosol typically represents less than 0.0001% of its total mass and volume (even a dense combustion plume is still 99.999% pure air). While much is known about the properties of some aerosol populations, uncertainty is introduced when more realistic and complex aerosols are studied. More work is needed to understand the properties of an aerosol with multiple contributing sources and concentrations and how the aerosol is transported in a complex environment, such as an office building (Hinds 1999).

Aerosol transport monitoring is important for detecting the presence of airborne contaminants and tracking their fate in populated environments where they may adversely affect human health. The challenge is to selectively detect potentially hazardous aerosols within the background aerosol matrix of the location. Typical aerosol background is in the range of $10^{-5} – 10^{-3}$ g/m^3 mass concentration (the mass of particulate matter in a unit volume of aerosol) and is composed of both natural and urban aerosols. Most natural background aerosol

originates from direct emissions from desserts (soil dust), the oceans (sea salt), and vegetation (botanical debris) and varies by geographic location, altitude, and time of year. Urban aerosol is more complex and dominated by anthropogenic emissions with local concentrations varying greatly depending on proximity to sources and time of day (Hinds 1999). A building's heating, ventilation, and air conditioning (HVAC) air-handling system settings have substantial impacts on the aerosol transport throughout the building and also vary by season and time of day. Computational modeling studies are often used to evaluate the airflow in buildings under different HVAC system settings. Optimized evacuation routes are also determined by evaluating theoretical scenarios of fire or threat agent release at different points throughout the building. While computational models are a good starting point, there is a need for experimental feedback to inform and evaluate aerosol dispersion and transport models. The results of atmospheric release tests can provide insight into some effects of building planning, design, construction, and operation on the building's airflow and airborne contaminant transport performance (Underwood *et al.* 2007).

Materials not generally found in the building's aerosol background are used to study aerosol transport and fate to minimize background contamination and generate more accurate measurements. The most commonly introduced tracer material is sulfur hexafluoride (SF_6) gas, which is detected and quantified at different points in the building with high sensitivity using commercial fluorescence sensors (Underwood *et al.* 2007). However, SF_6 molecules are ~300 pm (~3 x 10^{-4} μm) in diameter, which is an order of magnitude smaller than the lower bound of the typical aerosol size range (0.002-100 μm in diameter). As particle size is the principal

characteristic governing aerosol transport, SF_6 does not accurately simulate the transport of aerosols in an environment because the molecules are too small. Another test material option is synthetic polymer microspheres with incorporated fluorophores. Particle transport and fate are monitored using fluorescence measurements that are tunable by incorporating different fluorophores into the plastic spheres. While these particles are the correct size for accurately simulating aerosol transport properties, they are unsafe for human inhalation exposure and therefore unsuitable for use in populated environments. Neither the SF_6 gas nor the fluorescent polymer microspheres are capable of generating accurate test data for evaluation of aerosol transport in areas where humans will be exposed, which are the areas of primary interest.

The goals for this project were to create a novel aerosol test particle for use in atmospheric release tests that accurately mimicked natural aerosols and had all of these desired characteristics:

- Physical properties that mimic a broad range of environmental aerosol populations, which are 1-5 μm in diameter (middle of the typical aerosol size range) and have generally spherical morphology
- Safe for human exposure, so they may be used to evaluate airflows in populated environments
- Specific detection of the aerosol test particles within the environmental aerosol matrix
- Cost-effective production of grams of aerosol test particles to effectively simulate aerosol transport over large areas and distances

- Customizable test particle production so that the material properties may be tuned to meet the needs of specific aerosol release tests

These new aerosol test particles were principally made of U.S. Food and Drug Administration (FDA)-approved saccharide food additives, which ensured the material was safe, biodegradable, and had a low burden to receive approval for atmospheric release tests (the particles met environmental safety and health guidelines for human exposure). Sensitive and selective detection of the particles was achieved by incorporating non-coding DNA templates into the particles and detecting them using highly specific quantitative real-time polymerase chain reaction (QRT-PCR) assays with no interference from the environmental background matrix. Several different DNA barcodes were available for incorporation, allowing several different types of aerosol test particles to be generated, which enabled simultaneous or sequential multiple release tests in the same environment without concern for background contamination from previous releases. Two different technologies, a modified commercial inkjet printer and a commercial spray dryer, were evaluated for particle production. The inkjet printer offered low-cost, rapid production of small quantities of particles for optimization studies, and the spray dryer generated grams of particles appropriate for atmospheric release tests. Both particle generation methods enabled tuning the particle properties, including size, size-distribution, DNA template barcode, and number of DNA barcodes per particle, to create customized aerosol test particles that were suitable for atmospheric release tests. These customizable aerosol test particles will provide vital experimental feedback for evaluating aerosol transport models that safeguard the occupants in an environment.

2.1.1 Historic and Current Methods for Producing Microparticles

Microdroplets and microparticles find use in many diverse scientific applications including investigation of solute atomization in flames for analytical spectroscopic analyses (Hieftje and Malmstadt 1968; Seymour and Boss 1983; Childers and Hieftje 1986), spray combustion and fuel injection (Sangiovanni and Labowsky 1982), agricultural spraying drift and deposition (Threadgill *et al*. 1974), evaporation (Yang *et al*. 1997) and other physical properties related to weather (Magarvey and Taylor 1956), and for assessing performance of particulate control devices and the effects of particulate air pollutants (Berglund and Liu 1973). More recently, aerosols have been tailored for use in drug delivery (Jain *et al*. 1998; Hauschild *et al*. 2005), Raman spectroscopy of levitated droplets (Trunk *et al*. 1994), maskless lithography for microelectromechanical system (MEMS) fabrication (Wang *et al*. 2004), and as particle standards containing trace levels of molecules for ion mobility spectroscopy (Fletcher *et al*. 2008) and trace levels of elements for X-ray fluorescence spectrometry and laser ablation inductively coupled plasma mass spectrometry (Fittschen *et al*. 2006; Fittschen *et al*. 2008).

All droplet generation systems share similar performance metrics, such as compatibility with liquid of interest, droplet size, frequency, and stability and reproducibility of operation. In the past, acceptable aerosol production was achieved using complex, custom-built systems designed for specific applications. While system components varied widely to meet the needs of diverse applications of aerosols, all systems contained a reservoir of the liquid to be aerosolized, an orifice or nozzle(s) for droplet release, and a link between the two, generally a tube or series of tubes. In general, aerosol generation systems are also capable of microparticle production by simply evaporating the solvent, creating solid particles composed of the solute

material. Most early droplet generation schemes employed the division of a regularly disturbed cylindrical jet as the mechanism. When mechanical disturbances are generated at a constant frequency and with sufficient amplitude on a liquid jet moving at a constant velocity, the jet will break up into droplets of equal size (Berglund and Liu 1973). These periodic vibrations originated from devices such as a speaker (Magarvey and Taylor 1956; Stricker and Sofer 1991), motor-driven plunger (Magarvey and Taylor 1956), magnetostrictive transducer (Sweet 1965; Sangiovanni and Labowsky 1982), or, most commonly, a piezoelectric transducer (Hieftje and Malmstadt 1968; Stemme and Larsson 1973; Threadgill *et al.* 1974; Buehner *et al*. 1977; Seymour and Boss 1983; Maehara *et al*. 1984; Childers and Hieftje 1986; Switzer 1991; Warnica *et al*. 1991; Trunk *et al*. 1994; Yang *et al*. 1997). These early studies all used custom-built aerosol generation systems that were unique and elaborate, requiring considerable time and effort to construct and patience and skill to operate. In addition, an attempt to duplicate or improve upon an aerosol generator described in the literature proved difficult as often details were oversimplified or not provided.

Currently there is a variety of commercially available aerosol generators designed to alleviate some of the operational drawbacks of using individual custom-built systems. The Vibrating Orifice Aerosol Generator (VOAG, TSI Inc., Shoreview, MN) uses a piezoelectric cylinder to precisely control the breakup of a liquid jet into monodisperse microdroplets. The diameters of the microdroplets are tunable by varying the liquid flow rate and the vibration frequency of the nozzle. However, droplets are created one at a time using a single nozzle, making the VOAG unsuitable for rapid, large-scale aerosol production (Berglund and Liu 1973). The Sono-Tek ultrasonic atomizer (Sono-Tek Corp., Milton, NY, www.sono-tek.com) uses the

piezoelectric transducer droplet production method as well, but produces a spray of polydisperse microdroplets with an atomizing surface at the end of the liquid feed tube instead of a nozzle. The resulting aerosol has a log-normal size-distribution, with the distribution maximum being tunable by varying the operating frequency. The resulting polydisperse microdroplets have the disadvantage of requiring sorting prior to use in some applications. Finally, commercial emulsion solvent extraction piezoelectric printing systems can print polymers such as poly(lactide-co-glycolide) (PLGA), which is a biocompatible polymer used for drug delivery and other environmental testing applications (Jain *et al.* 1998; Fletcher *et al*. 2008). The commercial systems (*e.g.* SphereJet, MicroFab Technologies, Inc., Plano, TX) are designed for oil/water emulsion solvent extraction microparticle production and are not capable of generating aerosols. As was the case with the VOAG, single monodisperse particle generation rates are not suitable for large-scale production. There are a range of other commercially available comprehensive aerosol and particle production and characterization systems based on inkjet printing technology that produce well-defined aerosols. However, most systems are tailored to a specific purpose and are unable to transition effectively to new applications.

2.1.2 Inkjet Printing

Reproducible, controlled-size aerosol production is typically achieved using commercial systems tailored for very specific applications. There is a need for a more accessible and universally applicable generation mechanism for controlled-size aerosols and microparticles using equipment already present in most laboratories. For this work we demonstrated the

successful use of a widely available commercial-off-the-shelf (COTS) inkjet printer to produce tunable and well-characterized aerosol size-distributions in a rapid, easily controlled manner.

A process for generating uniform microdroplets of ink for data recording applications was first developed by Sweet (1965). This first inkjet droplet production system, as well as the several modified systems by others that followed, operated continuously and was successfully applied to data recording and photocopying applications (Kamphoefner 1972). The next step for the inkjet printing effort was impulse (drop-on-demand) printing, where a droplet was only ejected when needed. This added functionality was more suited for printing characters than the previous continuous operation systems, and photocopying performance was improved as droplet frequency modulation enabled grayscale printing capability (Stemme and Larsson 1973; Carnahan and Hou 1977). The impulse inkjet printing systems were further developed into the early computer word processing output printers and formed the basis for the commercial inkjet printing industry (Buehner *et al*. 1977).

Commercial inkjet printer systems (*e.g.* Hewlett-Packard DeskJet series, Hewlett-Packard Company, Palo Alto, CA) are low-cost, typically less than $100 USD for the entire system (printer, cartridge, and operating software), are widely available, and are capable of rapid, large-scale, consistent size-distribution microdroplet production (Buskirk *et al*. 1988; Bohorquez *et al*. 1994; Shelley *et al*. 1997). Current COTS inkjet printers employ either a piezoelectric mechanism (discussed above) or a thermal bubble mechanism for microdroplet production. The basic components of a thermal inkjet printhead are resistors, ink channels, and exit nozzles. When the resistor is electrically heated, the nearby ink is vaporized to create a bubble, forcing an ink droplet out through the nozzle. The collapsing bubble and capillary force

then draw ink from the channel to refill the nozzle (Buskirk *et al*. 1988; Chen *et al*. 1998). High-throughput, reproducible size-distribution microdroplet generation is achievable as some inkjet cartridges contain approximately 300 identical nozzles with concurrent nozzle firing rates of around 8 kHz (Bohorquez *et al*. 1994; Shelley *et al*. 1997). Compared to custom-built and commercially available systems, minimal investment of time and money is needed to apply these widely available COTS inkjet printer systems to aerosol and microparticle production applications.

The U.S. Army's Edgewood Chemical Biological Center first demonstrated the potential for COTS inkjet-based aerosol generation from aqueous solutions (not conventional ink). The Inkjet Aerosol Generator (IJAG) coupled a 12-nozzle thermal inkjet printhead from a desktop inkjet printer to a heated desiccation tube to dry the aerosol into microparticles, and was controlled using custom-built electronics and software. The uniformity of aerosols produced over time was demonstrated, and particles with tunable sizes in the range of 1-10 μm were generated (Bottiger *et al*. 1998). The Digital Aerosol Generator (DAG) was the second generation of the IJAG, utilizing COTS industrial-sized piezoelectric inkjet array printheads. The printed droplets were rapidly dried using a flow of warm air in a custom-built droplet conditioning module, and the system was controlled using custom software and electronics (Dougherty *et al*. 2007). Sergeyev and Shaw (2006) also demonstrated the use of a COTS thermal inkjet cartridge operated using custom-built electronics to produce droplets in a constant size-distribution (17 μm ± 2 μm). The system was reproducible over several hours and the droplet generation rate was accurately controlled, however, no attempt was made to tune the size-distribution of the generated droplets. All of these custom-built microdroplet and

microparticle production systems incorporated COTS inkjet cartridges, leveraging decades of droplet production technology to produce aerosols with desirable properties. However, the other components of the aerosol generation systems (*e.g.* the control electronics and printhead enclosures) and the software to control the cartridges were custom-built, making it difficult to transition the technology in a cost-effective manner to other researchers and consumers.

This work employed widely available COTS inkjet printer systems, with the associated commercial control software, for comprehensive aerosol and microparticle generation. Simple modifications to the inkjet printers and cartridges were added, but with retained functionality of the commercial software to easily control the printer and aerosol generation. Tunable and well-characterized microparticle size-distributions were simply and reproducibly produced with diverse properties suitable for many applications using this commonly accessible system.

2.1.3 Spray Drying

The successful generation of novel aerosol test particles using the modified inkjet printer in test-scale batches led to production scale-up to grams of particles for atmospheric release tests using a commercial spray dryer. Spray drying is a common method in the pharmaceutical and food industries for producing a dry powder from a solution or slurry (BUCHI Corporation 2002, Arpagaus *et al.* 2010b). Commercially available systems employ an atomizer or spray nozzle to separate a liquid stream into a droplet spray. The droplets are quickly evaporated in a heated gas (usually air) stream parallel to the droplet spray direction, resulting in solid particles composed of the solutes in the spray dried solution. The particles are then collected from the moving gas stream using a cyclone collector. The particles and gas are drawn

into the top of the collector perpendicular to the cone-shaped cyclone, which initiates rotation of the airstream. The rotation becomes more rapid as the airstream travels down the cyclone until it reaches the bottom, where it then travels back up the center of the cyclone and out the top to be further filtered and exhausted or re-circulated in the instrument. Particles are deposited on the walls of the cyclone due to centrifugal force and fall down into the product collection vessel beneath the cyclone. The final product is a fine, amorphous or crystalline powder (Hinds 1999, BUCHI Corporation 2002). Particle size-distribution characteristics vary greatly, from hundreds of nanometers to tens of micrometers in diameter, depending on the droplet generation mechanism, gas and liquid flow rates, and solute concentrations in the spray dried solution (Arpagaus *et al.* 2010a). Due to the particle size tuning abilities of the spray dryer, it is capable of generating dispersible powders in the range of 1-5 μm in diameter that are desired by the pharmaceutical industry for delivery of inhalable drugs (Arpagaus *et al.* 2010b). Spray drying facilitates single step powder generation with particle size tunability, making it attractive for many industrial processes.

A common spray nozzle employed in spray dryers is the two-fluid nozzle, where one fluid is the liquid solution containing solutes to be dried into particles, and the other fluid is a spray gas, usually nitrogen, used to disperse the liquid into droplets. These nozzles are nebulizers, which are another commercially available aerosol production option. The aerosol production mechanism involves a rapidly flowing gas stream interacting with a relatively slowly moving liquid stream, where the turbulent gas flow breaks the liquid up into fine droplets of random size. More sophisticated nebulizers, such as those used for solution sample introduction in atomic spectroscopy, utilize impingement surfaces, such as beads or

ultrasonically vibrating plates, to break up the larger droplets into smaller droplets, resulting in a narrower size-distribution (Skoog *et al.* 1998, Rubinson and Rubinson 2000). The basic nebulizer in the spray dryer nozzle produces a log-normal size-distribution with most particles less than 10 μm in diameter (Hinds 1999, Miller and Miller 2005). Droplet generation by nebulization is also extremely rapid (milliliters of liquid dispersed into droplets per minute), making it attractive for generating grams of microparticles for aerosol release tests (Raabe 1976).

2.1.4 Microparticle Morphology Analyses

The principal characteristics of natural aerosols to mimic in these novel aerosol test particles were the size and shape of the microparticles. While aerosol populations can span several orders of magnitude in size, particles useful for simulating aerosol transport must have diameters less than 10 μm, as larger particles will have masses too great to be transported using normal air flows. Particle composition and morphology also play important roles in aerosol transport, as different aerosols (*e.g.* biological, soot, dust) have different aerodynamic properties in a moving air stream. The goals of this project included generation of spherical aerosol test particles 1-5 μm in diameter to simulate a broad range of natural aerosol populations during atmospheric tests. The sizes of the generated particles were evaluated using an Aerodynamic Particle Sizer (APS), which yields size-distribution information based on aerodynamic transport measurements in real time. This information allowed the particle generation process to be tuned to deliver the desired size by adjusting production parameters, and particle morphology was determined using Scanning Electron Microscopy (SEM). These

SEM analyses both confirmed the APS particle size results and added another dimension of information regarding the morphology of the microparticles and how well they mimic the shapes of natural aerosols.

2.1.4.1 Aerodynamic Particle Sizer

The sizes of generated aerosol test particles were characterized in order to determine if they accurately mimicked natural aerosols that are transported using typical airflow rates. Aerodynamic diameter is the physical diameter of a unit density sphere that settles through the air with a velocity equal to that of the particle being analyzed. It is the most meaningful aerosol size parameter because it allows prediction of a particle's behavior while airborne. Particles having the same airborne behavior due to physical properties such as size, shape, density, and composition have the same aerodynamic diameter. The APS detects particles from 0.37-20 μm using light scattering (determining whether a particle is in the measurement area or not only). The APS sizes particles in the range of 0.5-20 μm using a time-of-flight technique, with particle velocity providing a functional measure of aerodynamic diameter in real time. The time-of-flight approach provides a more accurate assessment of aerodynamic properties compared to light scattering as it accounts for particle shape and is unaffected by differences in refractive index. The optical detection system for this time-of-flight particle size analyzer is composed of two partially overlapping laser beams. As a particle moves through the measurement area in an accelerating sheath airflow, it scatters the light from the two beams and generates one signal with two crests. The presence of one or more crests during the measurement time window is used to count the particle, and the peak-to-peak time between the two crests serves as a

measure of particle velocity, which is related to the particle's aerodynamic diameter. Larger particles accelerate through the detection region more slowly compared to smaller particles due to increased mass, inertia, and aerodynamic drag forces. Particle velocity data is stored in 1024 time-of-flight bins. A particle size calibration by the manufacturer using polystyrene latex (PSL) spheres of known size is used to convert the time-of-flight measurements into aerodynamic particle diameters. The aerodynamic particle sizes are binned into 52 channels on a logarithmic scale, and the resulting size-distribution for a particle population is displayed as a histogram, with aerodynamic diameter bins on the *x*-axis and the number of particles detected in each size bin on the *y*-axis (Holm *et al.* 1997, Peters and Leith 2003, TSI Incorporated, 2004).

The APS was beneficial throughout this work as it gave real time aerodynamic size-distributions of generated microparticles. These rapid analyses provided feedback for tuning the microparticle generation parameters to yield desired size-distributions. The disadvantage of using an APS alone for aerosol characterization is that it is a non-selective measurement, meaning it measures all aerosols that enter into the instrument regardless of their origin. Background samples of the environmental aerosols were always collected before the aerosol test particles were introduced in order to get the most accurate results for the generated test particles. Atmospheric release tests may be monitored using APS instruments, however, more selective confirmatory testing is desired to separate the results from the aerosol test particles from the background aerosol population of the testing environment.

2.1.4.2 Scanning Electron Microscopy

The shapes and minute surface features of aerosols are informative for identification and influence the transport properties of the microparticles. Scanning Electron Microscopy (SEM) is commonly used to study the surface features and morphology of microorganisms and other objects that require resolution of features smaller than 200 nm (diffraction limit for the optical microscope). Electron microscopes use electrons instead of light beams and electromagnets instead of lenses to examine prepared samples under high vacuum. Samples are sputter coated with a thin layer of a heavy metal (*e.g.* gold) to make them conductive in order to minimize artifacts caused by charge buildup and thermal degradation. The finely focused electron beam travels in a raster pattern over the area of the sample being examined. Secondary electrons are generated by the interactions between the high-energy beam electrons (20 keV) and weakly bound conduction band electrons on the sample surface. The incident beam provides sufficient energy to eject conduction band electrons from the surface with < 50 eV of kinetic energy. These inelastic scattering interactions yield an exit beam of secondary electrons only slightly larger in diameter than the incident beam, with a typical resolution of 0.5 nm. Secondary electrons are most commonly detected using a scintillator-photomultiplier combination detector. The secondary electrons are attracted and accelerated toward a phosphor or scintillator, which is a doped glass or plastic target that emits visible photons upon impact. The photons are conducted by a light pipe to a photomultiplier tube outside of the high-vacuum region of the SEM, and the amplified signal is incorporated into an analog or digital image. These detected electrons are used to produce an image of the surface

features of the sample at magnifications from 10X to 100,000X (Flegler *et al*. 1993, Skoog *et al.* 1998, Madigan *et al.* 2000).

Aerosol test particle SEM analyses were used to confirm the sizes of the generated particles first measured using the APS and to examine the morphology of the individual particles. As SEM is capable of differentiating aerosol types (*e.g.* biological, soot, dust) and sometimes the individual class members based on morphology, it is a more selective characterization technique compared to data generated by the APS. However, SEM analyses require time-consuming sample preparation and greater expertise to carry out. The results obtained from SEM and APS analyses were complementary, and the combined information generated from these two analysis techniques fully characterized the sizes and shapes of generated aerosol test particles.

2.1.5 DNA Barcodes

Microparticle detection and characterization using APS and SEM are useful for particle counting and morphology information, however, some applications (including monitoring aerosol transport) require more specificity in determining the constituents of an aerosol population. As discussed above, common aerosol background levels are on the order of 1 mg/m^3 mass concentration (mass of particles in a unit volume of aerosol) and are composed of both natural and anthropogenic aerosols that vary greatly by location. Methods that specifically and selectively detect aerosols of interest within the background aerosol matrix of the environment are essential for accurate evaluation of aerosol transport and fate. One very specific method for characterizing the biological constituents of an aerosol is analyzing the DNA

of the microorganisms. A common method for detecting and quantifying DNA involves application of the polymerase chain reaction (PCR), which amplifies a target DNA sequence millions of times for detection. The PCR requires four principal reaction components: template DNA containing the target DNA sequence to be amplified and detected, synthetic oligonucleotide primers designed to anneal to the outside edges of the target DNA sequence to initiate replication, a heat stable DNA polymerase to synthesize copies of the target sequence, and deoxynucleoside triphosphates (dNTPs) as building blocks for synthesizing the target DNA copies. The PCR is carried out in a thermal cycler, which heats and cools the reaction tubes for the different stages of the amplification reaction. The first stage involves denaturation, where the double-stranded DNA is heated so that the hydrogen bonds break, yielding two single-stranded DNA molecules. The next stage is annealing, where the temperature is lowered so that the oligonucleotide primers may anneal to the complementary sequence on the DNA molecule. The final stage is elongation, where the temperature is held at the ideal level for optimal polymerase activity to replicate the target DNA sequence by starting at the primers and adding the appropriate dNTPs. This completes one thermal cycle, which has, in theory, doubled the number of target DNA sequences that were initially present in the reaction. The PCR then continues with another thermal cycle, up to approximately 40 cycles total, yielding millions of copies of the target DNA sequence. After the PCR is completed, the resulting DNA copies may be detected using agarose gel electrophoresis techniques (Sambrook and Russell 2001), though with limitations in quantitative precision and accuracy arising from the non-specific DNA staining protocols.

Quantitative real-time PCR (QRT-PCR) provides a valuable alternative to gel electrophoresis, and expands the capabilities of the PCR to include precise detection of the number of target DNA copies in the sample in real time throughout the PCR. This simultaneous amplification and quantification is achieved by adding a DNA-binding fluorophore to the PCR reaction mixture and employing fluorescence-detecting thermal cyclers. One popular approach to QRT-PCR is presented by TaqMan assays, which employ an oligonucleotide probe designed to anneal to an interior portion of the target DNA sequence. The probe is labeled with a fluorophore on one end and a quencher on the other end. When the probe is intact, fluorescence is quenched due to fluorescent resonance energy transfer (FRET). The probe anneals to the target DNA sequence during the annealing stage of the PCR, similar to the primers. During the elongation phase, the exonuclease activity of the polymerase cleaves the fluorophore from the probe. As the fluorophore is no longer being quenched, it fluoresces, which is detected by the thermal cycler. Since each target DNA replication releases a fluorophore, the fluorescence intensity of the sample increases with each thermal cycle in proportion to the amount of template DNA initially present in the sample. When a large number of template copies are present in the initial sample, only a few cycles of PCR amplification are necessary for the fluorescence in the solution to increase and cross a detection threshold. When there are fewer template copies in the initial sample, more thermal cycles of amplification are needed for the fluorescence to increase and cross the same threshold value. The relationship between the number of template copies in the sample and the number of PCR cycles it takes for the fluorescence to cross the threshold (called the C_T value) is the basis for the real-time quantitation of DNA. A calibration curve is constructed by

analyzing a series of ten-fold dilutions containing a known amount of template DNA. A plot of C_T vs. $\log_{10}$ of DNA copies yields a straight line, and the template concentration in experimental samples is determined by interpolation into the standard curve. Due to the low detection limits afforded by fluorescence detection, QRT-PCR is capable of quantifying a target DNA sequence over seven or eight orders of magnitude. The large dynamic range, low detection limits, high specificity, and the capacity to process many samples concurrently during a single PCR run make QRT-PCR a gold standard for detecting and quantitating DNA in a vast number of applications, including biological aerosol characterization (Holland *et al.* 1992, Heid *et al.* 1996).

As bioaerosols are present in most environmental aerosol populations they can be used as the release test material for aerosol transport studies and are especially advantageous when PCR-based assays are used for detection and quantification. Some currently used bioaerosol test particles are harmless bacterial spores such as *Bacillus subtilis* and *Bacillus thuringiensis*, which are commonly used as a simulant for *Bacillus anthracis*, the causal agent of anthrax (Madigan *et al.* 2000, Burton *et al*. 2005). These bacterial spores work well for particle counting release tests and PCR confirmation as they mimic typical aerosol size (bacterial spores are ~0.5-2 µm in diameter) and contain DNA for specific detection and quantification of the number of bioaerosols in a particular location. However, there is the possibility of inherent contamination in the background aerosol matrix due to natural *B. subtillis* reservoirs in the soil and *B. thuringiensis* use as a commercial biological insecticide (Madigan *et al.* 2000). Also, a simulant bacterial spore may only be used for one release test in a location because afterwards the environment is certainly contaminated with that spore. Background contamination compromises the accuracy of PCR-based measurements of targeted microbes, which is

undesirable. Another approach involves use of synthetic, non-virulent oligonucleotides or plasmids to simulate either single or multiple virulent organisms for PCR-based detection (Carrera and Sagripanti, 2009). The DNA sequences used for PCR detection of threat agents are reproduced *in vitro* to validate positive PCR results without exposing an environment or personnel to a pathogenic organism. While this approach is useful for PCR-based detection tests, it is not suitable for aerosol release tests as the physical characteristics of nucleotides and plasmids are not similar to most environmental aerosols. This limitation is resolved by incorporating the DNA sequences into a particle material so that both aerosol transport and PCR-based detection methods are challenged during an atmospheric release test.

Synthetic nucleotides were chosen as unique particle identifier molecules for these novel aerosol test particles. However, instead of using DNA sequences from virulent organisms, templates were chosen from the thermophilic bacterium *Thermotoga maritima*. *T. maritima* exhibits optimal growth at 80 °C and is found in terrestrial hot springs and deep ocean hydrothermal vents, so the potential for natural background contamination is low (Madigan *et al.* 2000). The genome for *T. maritima* is sequenced due to its potential use in biotechnology applications as a hyperthermophile with heat stable enzymes (Nelson *et al*. 1999). Four templates approximately 100 base pairs (bp) long were chosen from the *T. maritima* genome that were non-coding (do not code for a functional protein) and dissimilar from DNA sequences of other common biological aerosol microorganisms. Corresponding oligonucleotide primers and probes were also designed for highly specific QRT-PCR detection of the templates in experimental samples. The incorporation of DNA barcodes into the aerosol test particles enables PCR-based detection and quantification after atmospheric release tests. Multiple

barcodes are available for incorporation into different batches of the test particles to alleviate background contamination problems if several release tests are desired in the same area. The customizable DNA barcodes in the aerosol test particles, along with QRT-PCR detection, facilitate atmospheric release testing in most environments.

2.2 Materials and Methods

2.2.1 Inkjet Printers and Cartridges

Commercial-off-the-shelf inkjet printers and cartridges were purchased online from Hewlett-Packard (HP, http://www.shopping.hp.com). This work utilized HP Deskjet D1660 Printers and HP 60 Black and Tricolor ink cartridges. Hewlett-Packard printer systems use the thermal bubble mechanism of microdroplet production, which was better suited for this work compared to piezoelectric systems, as it is more robust to experimentation and modification due to the lack of moving mechanical parts in the printheads. The black ink cartridges contained 336 nozzles, 20 μm in diameter, arranged in two columns of staggered nozzles (Figure 2.1). Nozzles were spaced at 1/600 inch (600 dots per inch, or dpi, resolution), yielding a swath height of 0.56 inches. The nozzles have a reported firing rate of 8 kHz and generate 13.8 *p*L ink droplets, which equates to 30 μm in diameter (Bohorquez *et al*. 1994; Shelley *et al*. 1997).

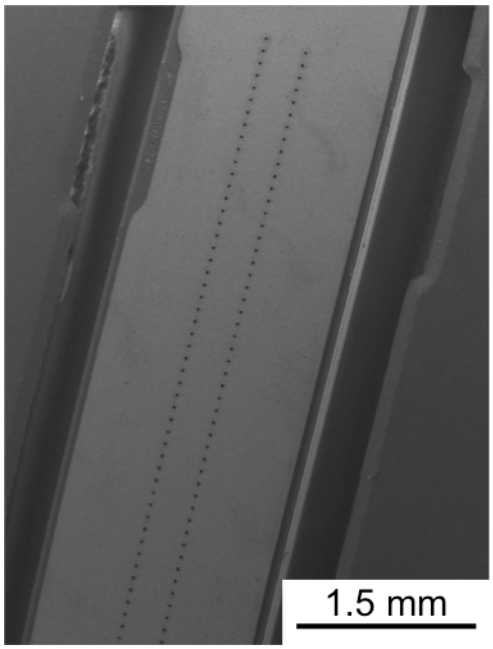

Figure 2.1. Scanning Electron Microscopy micrograph of the nozzle plate of a black inkjet cartridge (20X magnification).

Both the printers and cartridges were modified to enable printing a variety of solutions and collecting and analyzing the printed aerosol droplets. The top cover of the printer was removed to allow access to the cartridge carriage assembly inside (Figure 2.2a). The cartridge carriage was removed from a second printer, electrically connected to the carriage in the intact printer, and stationary mounted over a large plastic collection chamber. This configuration allowed the cartridge carriage within the intact printer to raster back and forth normally as if printing an image on the paper, but the printed droplets issued from the stationary mounted cartridge carriage for collection. This was necessary as the intact printer required cartridge position feedback for operation. The inkjet cartridges were prepared for experiments by removing the tops and flushing out the ink with distilled water. The solutions printed during experiments were always printed through clean black cartridges as they were more simply constructed compared to the tricolor cartridges. After the black cartridge was loaded with 500

μL of the solution to be printed, the top was replaced. The black cartridge was then placed in the stationary cartridge carriage over the initial collection chamber for production and collection of aerosol droplets.

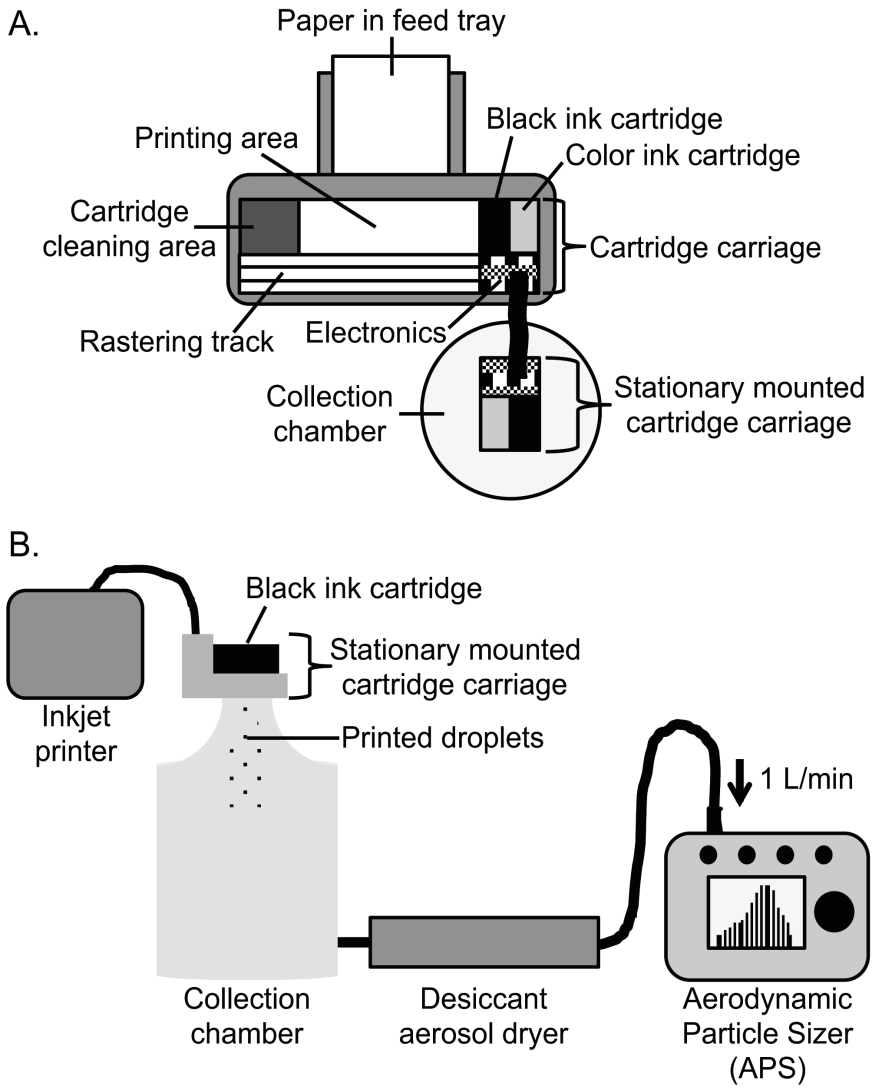

Figure 2.2. Schematic drawings of the experimental setup showing (A) the modified inkjet printer and stationary mounted cartridge carriage (top view) and (B) the droplet collection, dehydration, and microparticle characterization devices (side view).

Instead of striking paper to form an image, the droplets generated by the modified printer and stationary mounted black inkjet cartridge were collected in a five-gallon polycarbonate container and immediately drawn through a desiccant aerosol dryer (built in-house) to dehydrate them (Figure 2.2b). The desiccant aerosol dryer was constructed from 1.27 cm-thick anodized aluminum with a cover of 1.27 cm-thick glass (53.34 cm x 7.62 cm x 7.62 cm with the cover). A wire mesh tube 1.27 cm in diameter traversed the dryer lengthwise and had airtight couplings to the outside of the dryer for aerosol transport through the dryer. The dryer was filled with fresh regenerated desiccant daily (desiccant mesh size larger than the wire mesh tube to prevent the desiccant from entering the aerosol transport tube), covered with the glass with an o-ring for an airtight seal, and secured with 10 allen screws. As the generated microdroplets were drawn through the aerosol transport tube in the dryer, the surrounding desiccant dehydrated the droplets, resulting in solid microparticles consisting of the non-volatile dissolved substances from the printed aqueous solution. The microparticles (still suspended in the moving air stream) were then characterized using an Aerodynamic Particle Sizer (APS, TSI model 3321, Shoreview, MN) and Aerosol Instrument Manager Software (version 8.1, TSI). The APS drew air at a rate of 1 L/min (at room temperature, 20 °C) through the system and was the means for transporting the generated microdroplets from the initial collection chamber through the desiccant dryers to the APS to be analyzed. The APS counted the microparticles and determined their aerodynamic diameters simultaneously, yielding size-distribution results for the microparticles collected during each experiment.

The image file sent to the printer as the print job for all experiments was a solid black rectangle, 20.2 cm x 7.6 cm, created on a slide in Microsoft Office PowerPoint (Microsoft,

Redmond, WA). The slide was saved as a picture, and the picture was printed from the Preview application of a Mac OS X computer (Apple, Inc., Cupertino, CA). The HP Deskjet D1660 printer driver was downloaded from the HP website and installed on the same computer. When the black rectangle image was printed from Preview to the HP Deskjet D1660 printer, there were several customizable options available in the Paper Type/Quality portion of the Advanced print menu that were expected to influence the generated aerosol. The customizable options included: paper type, quality, color, photo fix, and grayscale mode. The most interesting paper type and print quality option combinations available in the commercial software were investigated (Table 2.1). The color option was always set to grayscale, as solution was only loaded into the black inkjet cartridge to be printed. The photo fix option was always "off" to simplify the results. The grayscale mode sub-options were high quality (both tricolor and black cartridges used) and black print cartridge only. The black print cartridge only option was employed at all times that it was available, however, for higher quality settings the high quality grayscale mode was the only option available.

Table 2.1. Paper type and print quality settings investigated for effect on printed aerosol characteristics (* indicates tested experimental condition, N/A indicates setting not available).

	Print Quality			
Paper Type	Fast Draft	Normal	Best	Maximum dpi
Plain	*	*	*	N/A
Photo	N/A	*	*	*
Inkjet	N/A	*	*	N/A
Transparency	N/A	*	N/A	N/A

The generated microparticles were composed of glucono-delta-lactone (GDL, Purac America, Inc., Lincolnshire, IL), a naturally-occurring food additive (Figure 2.3). Glucono-delta-

lactone was chosen as it is non-toxic and water soluble, making it amenable for use in the thermal inkjet cartridges that are manufactured to print water-based black ink (Hall *et al*. 1994). Solutions of varying GDL concentration (% w/v) in sterile water (Teknova, Inc., Hollister, CA) were tested in order to examine the tunability of the resulting microparticle size-distribution created using this inkjet printing method. Table 2.2 lists some physical properties, namely density and viscosity, of the tested GDL solutions, as well as the values measured for the HP 60 black ink that was in the cartridges. The density of each solution was measured by weighing 1.000 mL of the solution using an analytical balance in triplicate. The kinematic viscosity of each solution was determined using a Cannon-Fenske Opaque Viscometer (size 50, Cannon Instrument Company, State College, PA) in triplicate. Kinematic viscosity was converted to dynamic viscosity by multiplying by the density of the solution. Solution density and viscosity affect nozzle re-filling speed and droplet break-off time within the inkjet cartridge and are therefore expected to have an effect on expelled droplet size.

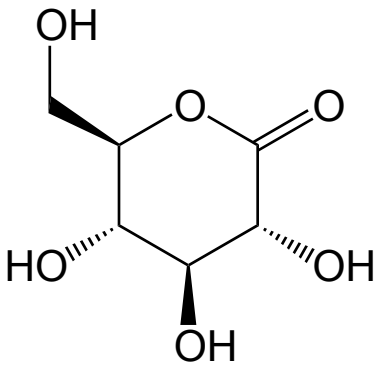

Figure 2.3. Structure of glucono-delta-lactone (GDL).

Table 2.2. Physical properties of aqueous GDL solutions used during experiments and HP 60 black ink for comparison (measured at 20 °C, *n* = 3, mean ± one standard deviation).

Solution	Density (g/mL)	Viscosity (cP)
1% GDL	1.005 ± 0.001	1.026 ± 0.002
2.5% GDL	1.009 ± 0.002	1.051 ± 0.008
5% GDL	1.018 ± 0.002	1.123 ± 0.005
7.5% GDL	1.029 ± 0.003	1.214 ± 0.004
10% GDL	1.035 ± 0.003	1.298 ± 0.001
12.5% GDL	1.048 ± 0.003	1.403 ± 0.007
15% GDL	1.055 ± 0.002	1.517 ± 0.004
17.5% GDL	1.067 ± 0.002	1.647 ± 0.007
20% GDL	1.078 ± 0.002	1.794 ± 0.006
Black Ink	1.043 ± 0.012	2.918 ± 0.003

2.2.2 Spray Dryer

A commercial spray dryer (Mini Spray Dryer B-290, BUCHI Corporation, Switzerland) was used to scale-up microparticle production to grams. Literature for the spray dryer indicated it was capable of producing highly dispersible, spherical particles 1 to 5 μm in diameter with narrow particle size-distributions, similar to the particles produced using the modified inkjet printer (Arpagaus *et al.* 2010b). However, it was discovered that GDL was not compatible with the spray drying process because it became sticky, causing product losses and agglomeration. This is a common problem encountered when spray drying low molecular weight sugars, as the glass transition temperature (T_g) is low compared to larger oligosaccharides. If the T_g is exceeded during the spray drying process, the particles change state from solid to glass and become sticky (Adhikari *et al.* 2005, Arpagaus *et al*. 2010a). Therefore, another naturally-occurring food additive, maltodextrin, was selected to be the bulk particle material when the spray dryer was used for generation (Figure 2.4). Organic Tapioca Maltodextrin DE 10 (Dextrose Equivalent 8-12, Ciranda, Hudson, WI) is an oligosaccharide with 10-15 glucose units and has a

T_g of 160 °C, which is sufficiently high to prevent stickiness during spray drying and caking during storage (Bhandari, no date).

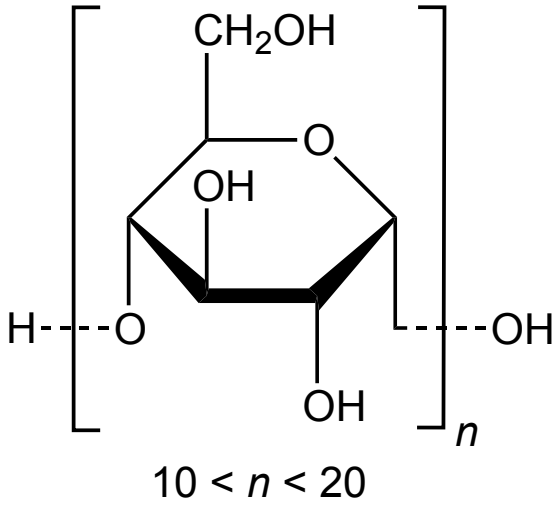

Figure 2.4. Structure of maltodextrin.

The spray dryer was operated as an open-cycle system, with room air as the drying gas (Figure 2.5). The air stream drawn into the spray dryer was electrically heated to the set inlet temperature before being drawn into the drying chamber. The aqueous maltodextrin solution (50-150 mL) was placed in a sterile beaker on a stir plate and was gently stirred throughout the spray drying process to ensure homogeneity. A peristaltic pump was used to draw the solution into the two-fluid spray nozzle, where it interacted with the ultra high purity compressed nitrogen (N_2) spray gas flow, controlled by a rotameter. The droplets were sprayed into the drying chamber in the same direction as the hot air (co-current flow), and the solvent quickly evaporated, resulting in solid microparticles. The microparticles were collected from the moving air stream using a high-performance cyclone, where inertial forces impacted the

particles on the walls of the cyclone and forced them down into the collector vial. Finally, the air stream passed through another filter before being exhausted through the aspirator to atmosphere. After all of the solution was spray dried, the system was allowed to cool before shutting down and removing the collector vial from the end of the cyclone containing the powder product. The free-flowing powder was transferred to an autoclaved scintillation vial (VWR International, Radnor, PA) for characterization and storage. The generated microparticles were re-aerosolized by shaking the vial for APS analyses and SEM sample preparation.

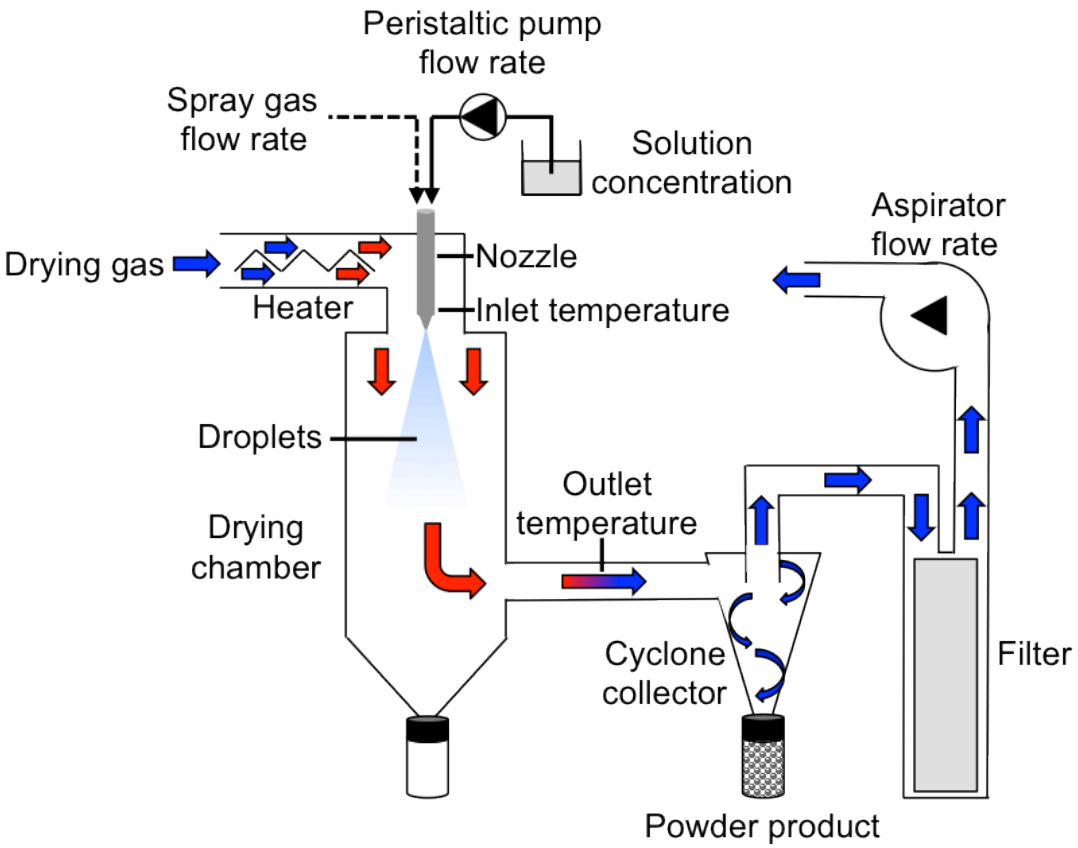

Figure 2.5. Schematic drawing of the spray dryer showing all gas and liquid flows, temperature measurement points, and product collection vessel. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

In order to optimize microparticle size and demonstrate particle size tunability, several different maltodextrin concentrations in aqueous solution and several different spray gas flow rates were tested, as those are the spray drying parameters that have the most effect on generated particle size. The tested maltodextrin concentrations were 1%, 3%, 5%, 10%, 12.5%, and 20% (w/v) in sterile water (Teknova). The tested spray gas flow rates were 30 mm (7.32 L/min), 40 mm (11.12 L/min), 45 mm (13.85 L/min), and 50 mm (17.53 L/min). The optimized operational settings for the various spray drying parameters used during experiments are shown in Table 2.3.

Table 2.3. Spray dryer operation settings for maltodextrin microparticle generation.

Operating Parameter	Setting
Solution concentration	3% (w/v)
Inlet temperature	190 °C
Aspirator flow rate	70% (475 L/min)
Peristaltic pump flow rate	30% (9 mL/min)
N_2 spray gas flow rate	45 mm (13.85 L/min)
Nozzle cleaner interval	3

2.2.3 Microparticle Characterization

2.2.3.1 Aerodynamic Particle Sizer

Microparticles generated using both the modified inkjet printer and the commercial spray dryer were characterized using the Aerodynamic Particle Sizer (APS). The GDL microparticles generated using the modified inkjet printer were analyzed directly after they were made, as the APS inlet airflow (1 L/min) was the means for transporting the microdroplets through the desiccant dryers and into the APS to be analyzed (Figure 2.2). Conversely, the APS was not directly integrated into the commercial spray dryer system. After the spray dried

maltodextrin powder was produced and transferred to a small vial for storage, the particles were re-aerosolized for APS analysis by shaking the vial with the tube leading to the APS inlet positioned over the vial. Both APS analysis methods yielded size-distribution data for characterizing the produced microparticles.

The microparticle APS size-distribution raw data collected during experiments were qualitatively compared to determine the effects of varying the generation parameters on the resulting microparticle size-distribution. The data sets were then treated more analytically by truncating them to only include the principal size-distribution data points (histogram size-bins with the number of particles detected in each bin) containing the majority of the collected microparticles and applying standard chromatographic peak processing and analysis methods (*e.g.* gas or liquid chromatography data analysis methods). The truncated size-distribution data from all experiments were imported into MATLAB (version R2011b, MathWorks, Inc., Natick, MA) and individually smoothed using the Savitzky-Golay algorithm with the best experimentally determined parameters for these data (data point span of seven, polynomial degree of three) (Savitzky and Golay 1964). The MATLAB Curve Fitting Toolbox was then used to fit a single Gaussian function (Equation 2.1) to the smoothed data from each experiment:

$$y = ae^{\left[-\left(\frac{x-b}{c}\right)^2\right]} \qquad [2.1]$$

where *a* is the peak amplitude, *b* is the peak maximum position, and *c* is the population standard deviation, which is a measure of the peak width. The resulting Gaussian equations that best fit the data were used to determine microparticle size-distribution maxima positions

and full width at half maximum (FWHM) values in terms of aerodynamic diameter. Statistically significant differences in the generated microparticle size-distribution characteristics resulting from changing experimental conditions were evaluated using Student's *t*-tests and one-way analysis of variance (ANOVA) procedures using Microsoft Office Excel (null hypotheses were that the experimental means were the same). The resolutions of relevant size-distributions were also calculated using the conventional chromatographic theory equation (Equation 2.2) using Excel:

$$R_S = \frac{2\left[(t_R)_B - (t_R)_A\right]}{W_A + W_B} \qquad [2.2]$$

As the microparticle size-distribution data were collected in terms of aerodynamic diameter on the x-axis and particle counts on the y-axis, the resolution was calculated using measurements in micrometers (μm). The mean size-distribution standard deviations from the Gaussian best-fit equations multiplied by four were used as the baseline peak width values, W, and the mean size-distribution maximum positions were used in a manner analogous to chromatographic retention times, t_R, in order to determine the resolution, R_s, of two adjacent microparticle size-distributions (Skoog *et al*. 1998; Rubinson and Rubinson 2000).

2.2.3.2 Scanning Electron Microscopy

Scanning Electron Microscopy (SEM) was used to confirm the size and examine the morphology of the GDL microparticles produced using the modified inkjet printer and the

maltodextrin microparticles generated using the spray dryer. Microparticles were collected on Teflon filter material (1.0 μm pore size, Zeflour supported PTFE, Pall Corp., Ann Arbor, MI) using a four-stage Impactor Stack with a 9 L/min airflow rate (California Measurements, Inc., Sierra Madre, CA). Samples were sputter coated with a thin layer of gold using a Hummer 6.2 Sputter Coater (Anatech, Union City, CA) and were viewed using a Hitachi S-800 field emission scanning electron microscope (Hitachi, South San Francisco, CA) and Quartz PCI imaging software (version 6.0, Quartz Imaging Corporation, Vancouver, BC, Canada). Images from SEM analysis were further analyzed using ImageJ software (National Institutes of Health, http://rsbweb.nih.gov/ij/) to measure particle sizes and count particles in the images.

2.2.4 DNA Barcodes and QRT-PCR Assays

Four ~100 bp DNA barcode templates from *T. maritima* were designed by a bioinformaticist to be dissimilar from amplicons used to detect other biological aerosol microorganisms and to have amplicons, primers, and probes that exhibited desirable characteristics for successful QRT-PCR assays (GC content of 50-60%, not likely to form dimers or secondary structures). The synthetic template oligonucleotides (Biosearch Technologies, Inc., Novato, CA) and the synthetic oligonucleotide primers and probes (Integrated DNA Technologies, Inc., Coralville, IA) were obtained from different manufacturers to avoid contamination and stored according to manufacturer's instructions. Calibration curves of ten-fold serial dilutions containing known DNA template concentrations (plotting the $\log_{10}$ of DNA copies vs. mean C_T result) were analyzed in order to quantify the amount of DNA in

experimental samples. Template solutions in Tris-EDTA buffer (TE buffer, 10 mM tris(hydroxymethyl)aminomethane (Tris), 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0, Teknova) were added to the aqueous solution of either GDL or maltodextrin (depending on the particle generation method employed) just prior to particle production. The DNA barcodes were detected in samples using real-time TaqMan PCR assays, where cleavage of fluorescent resonance energy transfer (FRET) quenched probes during amplification resulted in an increased fluorescence signal (Holland *et al.* 1992, Heid *et al.* 1996). The probes were synthesized with the fluorophore 6-carboxyfluorescein (6-FAM) on the 5' end and the quencher Black Hole Quencher 1 (BHQ-1) on the 3' end. The excitation maximum wavelength for 6-FAM is 494 nm and the emission maximum wavelength is 525 nm. The BHQ-1 excitation maximum wavelength is 534 nm, which overlaps with the emission spectrum for 6-FAM, enabling FRET quenching as BHQ-1 returns to the ground state through non-radiative decay. The heat stable DNA polymerase used for amplification was Platinum Taq (kits from Invitrogen, part of Life Technologies Co., Carlsbad, CA), which incorporated dNTPs (New England Biolabs, Inc., Ipswich, MA) to make copies of amplicon during thermal cycling. The thermal cyclers used to carry out the QRT-PCR were Smart Cyclers (Cepheid, Sunnyvale, CA) with integrated fluorescence detection optics. Smart Cycler optical channel 1 has a standard filter set optimized for the detection of free FAM fluorophores in solution after cleavage from the probe, with excitation from 450-495 nm and emission from 505-537 nm. The Smart Cycler software (version 2.0d) was used to program the thermal cycling parameters optimized for the QRT-PCR reaction (Table 2.4) and to analyze the results and export them into Microsoft Excel.

Table 2.4. Thermal cycling parameters used during QRT-PCR assays to quantify the amount of DNA barcode in samples.

Cycle Number	Number of Cycle Repeats	Temperature (°C)	Time (min:sec)
1	1	50	2:00
2	1	95	10:00
3	40	Step 1: 95	0:15
		Step 2: 55*	1:00

* Fluorescence measurement taken during this step

2.3 Inkjet Printer Production Method Results

2.3.1 Effect of Solute Concentration in Printed Solution on Microparticle Size-Distribution

The effect of varying the concentration of GDL in the printed solution on the resulting microparticle size-distribution was investigated by printing aqueous solutions ranging from 1-20% GDL in at least duplicate using the default inkjet printer software settings, namely plain paper and normal quality (1%, 2.5%, 5%, 10%, 12.5%, 15%, and 20% GDL $n = 3$; 7.5% GDL $n = 5$; 17.5% GDL $n = 2$). Aerosol generation took approximately 6.6 seconds using the default settings and was the same for all GDL solutions. The printed aerosol was drawn through desiccant dryers, and the resulting solid microparticles were characterized using an APS to yield both microparticle size and number collected. Figure 2.6 displays a representative subset of the mean microparticle APS size-distribution raw data results for the different GDL solutions. Very few particles (< 0.1%) were collected with aerodynamic diameters above 8 μm for all tested solutions. The majority of microparticles (> 80%) collected from all solutions were in the range of 1-5 μm in aerodynamic diameter. The APS results for the lower concentration GDL solutions displayed multimodal size-distributions, perhaps as a result of particle coalescence or abnormal inkjet cartridge operation. The solutions containing greater than 7.5% GDL created single size-distributions of microparticles as expected. This result may be due to the physical properties,

such as density and viscosity, of the GDL solutions that influence droplet production from the inkjet cartridges. The densities of the more concentrated GDL solutions (7.5-20% GDL, 1.029-1.078 g/mL) were more similar to the aqueous black ink density (1.043 g/mL) that the cartridges are designed to print compared to the lower concentration GDL solutions (1-5% GDL, 1.005-1.018 g/mL) (Table 2.2). Due to the undesirable multimodal operation of the inkjet cartridge when low concentration GDL solutions were printed, the data for the 1%, 2.5%, and 5% GDL solutions were not examined further. Particle size-distributions created by the modified inkjet printer approximated normal distributions, making them more similar to distributions generated using a nebulizer (log-normal size-distribution) and dissimilar from monodisperse droplets generated by the VOAG.

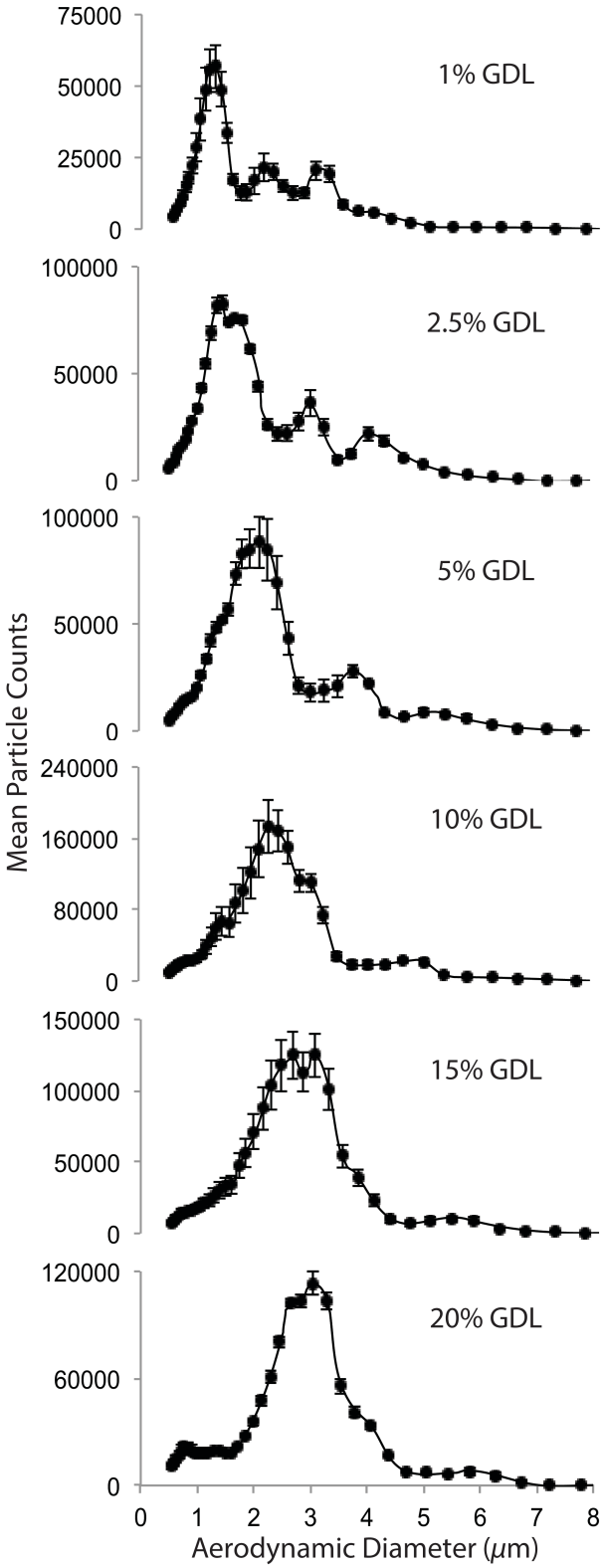

Figure 2.6. Mean APS size-distribution results for microparticles created by printing aqueous solutions with varying GDL concentrations demonstrate particle size tunability by varying solute concentration ($n \geq 2$, error bars are one standard deviation).

A trend in the microparticle size-distribution APS data indicated that as the GDL concentration in the printed solution increased, the size-distribution shifted towards larger microparticle aerodynamic diameters (Figure 2.6). This was the anticipated result, as the dry microparticles measured by the APS were composed of the solute in the solution and should display a dependence on solution concentration according to Equation 2.3:

$$D_p = C^{1/3} D_d \qquad [2.3]$$

where D_p is the dry particle diameter, C is the volumetric concentration of the nonvolatile solute in the volatile solvent, and D_d is the wet droplet diameter (Berglund and Liu 1973). A plot of the relationship between GDL concentration in the printed aqueous solution and the resulting mean microparticle aerodynamic diameter calculated from the APS results is displayed in Figure 2.7, along with theoretical particle diameters resulting from droplets 30 μm in diameter containing the same concentrations of GDL (Bohorquez *et al*. 1994; Shelley *et al*. 1997). The expected power relationship according to the cube root of the solute concentration is present in both the theoretical and experimental data sets, indicating consistent droplet diameter generation despite modifications to the inkjet printer and cartridge. However, the microparticles generated using the modified inkjet printer system and aqueous GDL solutions were smaller than predicted (note the separate y-axes for the experimental and theoretical results in Figure 2.7). The experimental droplet diameters were calculated using the tested GDL concentrations, mean particle diameters measured using the APS, and Equation 2.3, which yielded an average droplet diameter of approximately 5.3 μm issuing from the cartridge when

aqueous GDL solutions were printed. This difference between the reported 30 μm ink droplet diameter and the observed 5.3 μm GDL solution droplet diameters is due to the lower viscosities of the tested GDL aqueous solutions (1.026-1.794 cP, Table 2.2) compared to the black ink (2.918 cP). Higher viscosity solutions have longer break-off times of the exiting droplet from the solution remaining in the nozzle, resulting in more solution transferring to the droplet and therefore making the droplet larger compared to droplets of less viscous solutions (Chen *et al*. 1998).

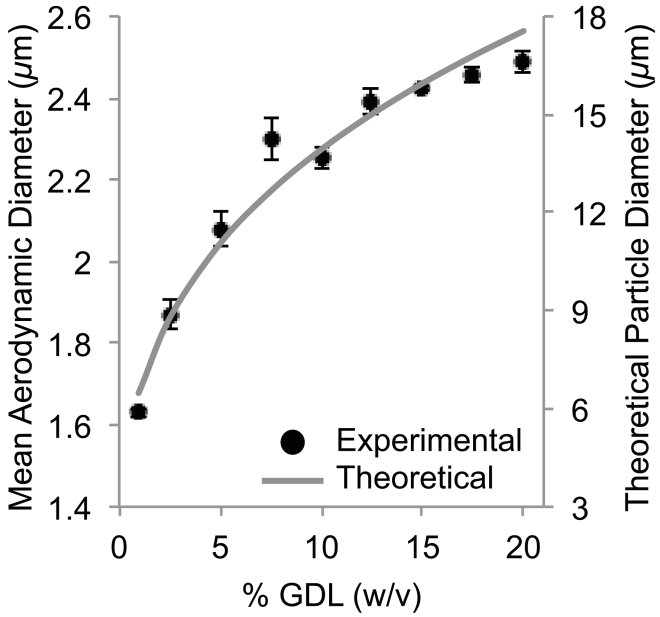

Figure 2.7. Mean microparticle aerodynamic diameters generated using various GDL concentrations in the printed aqueous solutions in comparison to theoretical particle diameters from droplets 30 μm in diameter produced using the same GDL solutions ($n \geq 2$, error bars are one standard deviation). Note separate *y*-axes for the experimental and theoretical results.

The raw APS size-distribution data from the microparticles produced using the different concentration GDL solutions were investigated further by applying the Savitzky-Golay smoothing algorithm and then fitting the smoothed data to Gaussian functions. The coefficients

of determination (R^2) of the Gaussian best-fit equations to the data points were all above 0.97 for this data set, indicating the Gaussian equations accurately represented the data (an R^2 of 1 indicates a perfect fit to the data) and that the distributions were normal (Miller and Miller 2005). The resulting best-fit Gaussian equations, shown in Figure 2.8, were used to determine mean microparticle size-distribution maximum positions and FWHM values for further analyses. The mean particle counts for all samples were normalized to unity for more straightforward comparison, which had no effect on the size-distribution positions or FWHM values. The microparticle size-distribution maxima and FWHM values in terms of aerodynamic diameter were compared to determine statistically significant differences and solution concentration effects on the resulting microparticles. Mean size-distribution maximum values ranged from 2.35 μm (7.5% GDL) to 2.93 μm (20% GDL) aerodynamic diameter, demonstrating a 22% relative change within the data set. One-way ANOVA determined that all GDL solution mean size-distribution maximum values differed significantly at greater than 99% confidence ($p < 0.01$), meaning that the microparticle size-distribution significantly shifts to higher aerodynamic diameters when the GDL concentration in the printed solution is increased. Student's *t*-test results from pairwise comparisons of the mean size-distribution maxima of adjacent concentration GDL solutions (*e.g.* comparing 7.5% with 10%, 10% with 12.5%) showed that almost all solutions produced microparticles that were significantly different in size at 99% confidence ($p < 0.01$) from the microparticles produced using other GDL solutions. The only exception was the 7.5% and 10% GDL size-distribution maxima comparison, which differed significantly at 86% confidence ($p = 0.137$). These size-distribution results demonstrate

microparticle size tunability using the commercial inkjet printer by simple adjustment of solute concentration in the printed solution.

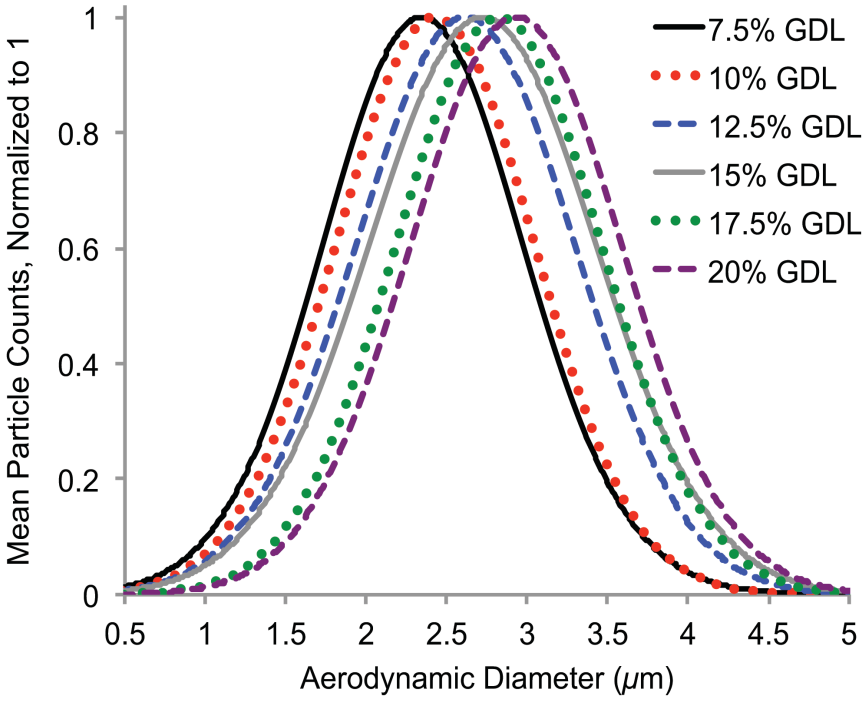

Figure 2.8. Mean microparticle size-distribution Gaussian fit results showing increasing distribution maximum positions with increasing GDL concentration and comparable size-distribution width for all GDL solutions ($n \geq 2$). Particle size distribution maxima are normalized to one to facilitate comparison.

Mean microparticle size-distribution FWHM values in terms of aerodynamic diameter from the best-fit Gaussian equations to the data varied from 2.05 μm wide (10% GDL) to 2.34 μm wide (15% GDL), which corresponds to a 13% relative change in the data set. One-way ANOVA of the microparticle size-distribution FWHM results for each tested GDL solution determined that the means were not significantly different ($p = 0.103$) when solution concentration was altered. There was no discernable relationship between the GDL concentration in the printed solution and size-distribution FWHM results. The microparticle

size-distribution resolutions (R_S) were calculated for all possible GDL solution comparisons using Equation 2.2, and the results are listed in Table 2.5. As in chromatographic peak analysis, a resolution value of 1.5 is considered complete separation, and a resolution of 1 indicates approximately 4% overlap between the two size-distributions (Skoog *et al*. 1998). The size-distributions in this data set were not appreciably resolved, with all comparisons showing resolution values less than 0.16. This result was expected due to the large size-distribution widths relative to the change in size-distribution maximum positions between microparticles produced using different GDL solutions. The FWHM results showed that the microparticle size-distribution widths remained relatively constant (13% relative change) despite significant changes in microparticle size due to varying GDL concentration in the printed solution. This suggested consistent printhead operation when different solutions were used for aerosol generation, which is a desirable characteristic of a system for use in generating tunable and diverse microdroplets and microparticles for a wide range of applications.

Table 2.5. Microparticle size-distribution resolutions from pair-wise comparisons of all tested GDL solutions (smoothed, Gaussian fit data used in calculations). All values were less than one, indicating considerable overlap among the different size-distributions.

	10% GDL	12.5% GDL	15% GDL	17.5% GDL	20% GDL
7.5% GDL	0.0224	0.0710	0.0981	0.1311	0.1590
10% GDL		0.0503	0.0781	0.1104	0.1389
12.5% GDL			0.0275	0.0556	0.0835
15% GDL				0.0263	0.0541
17.5% GDL					0.0296

2.3.2 Effect of Printer Software Settings on Microparticle Size-Distribution

Different print media require different ink droplet characteristics to produce the same degree of optical density and image resolution. Therefore, the inkjet printer is programmed with several different print modes selectable within the commercial software that yield droplets with the desirable characteristics for the specified print media and image quality (Bohorquez *et al*. 1994). The availability of different paper type and printing quality options suggested that the printed droplet characteristics would be altered depending on which settings were chosen. The paper type and print quality software settings investigated to determine if tunable microparticle size-distributions could be produced are listed in Table 2.1. The time and number of cartridge carriage passes used by the printer to execute the same print job using the different software settings was measured (Table 2.6) in order to characterize this novel aerosol generator. All printing times were less than four minutes, and most were less than a minute, demonstrating the high-throughput capability of this aerosol generator. As expected, the printing was executed most rapidly when the fast draft setting was employed, and the highest quality print modes took the most time, with higher numbers of cartridge passes and longer drying times between passes.

Table 2.6. The number of cartridge carriage passes and time expended by the modified inkjet printer to execute the same print job (solid black rectangle, 20.2 cm x 7.6 cm) to generate aerosols using various paper type and print quality software settings (n = 3, mean ± one standard deviation).

Printer Software Setting	Printing Time (sec)	Number of Cartridge Passes
Plain, Fast Draft	2.5 ± 0.5	6
Plain, Normal	6.6 ± 0.1	20
Plain, Best	26.8 ± 0.7	41
Photo, Normal	57.8 ± 1.0	83
Photo, Best	104.7 ± 0.4	123
Photo, Maximum dpi	223.1 ± 0.3	174
Inkjet, Normal	32.3 ± 0.2	62
Inkjet, Best	51.4 ± 0.3	62
Transparency, Normal	51.4 ± 0.4	62

The 15% GDL solution was printed with the different paper type and print quality settings in at least triplicate (photo paper, normal quality n = 5; photo paper, best quality n = 4; all other settings n = 3), and a summary of the APS size-distribution results for the generated microparticles is shown in Figure 2.9. The microparticles produced when the plain and photo paper settings were used displayed differences according to print quality setting. The plain paper results showed that the lowest quality setting (fast draft) produced the fewest number and smallest aerodynamic diameter microparticles, the moderate quality setting (normal) generated more and larger microparticles, and the highest quality setting (best) produced the most and largest microparticles. Similar results were observed when the photo paper setting was tested. When the print quality was increased from normal (lowest quality) to maximum dpi (highest quality), more and larger microparticles were produced. These results were expected, as the main differences between the programmed print modes for the inkjet printer are the number of cartridge carriage passes (Table 2.6) and the optical density of the black ink, where

higher quality print modes require more passes that deposit more and larger droplets compared to low quality print modes (Hall *et al*. 1994). The number of microparticles produced was much lower overall when the photo paper setting was used compared to the plain paper setting, which was expected as the glossy photo paper should require much less ink to achieve the same optical density levels as on more porous plain paper when printing with black ink.

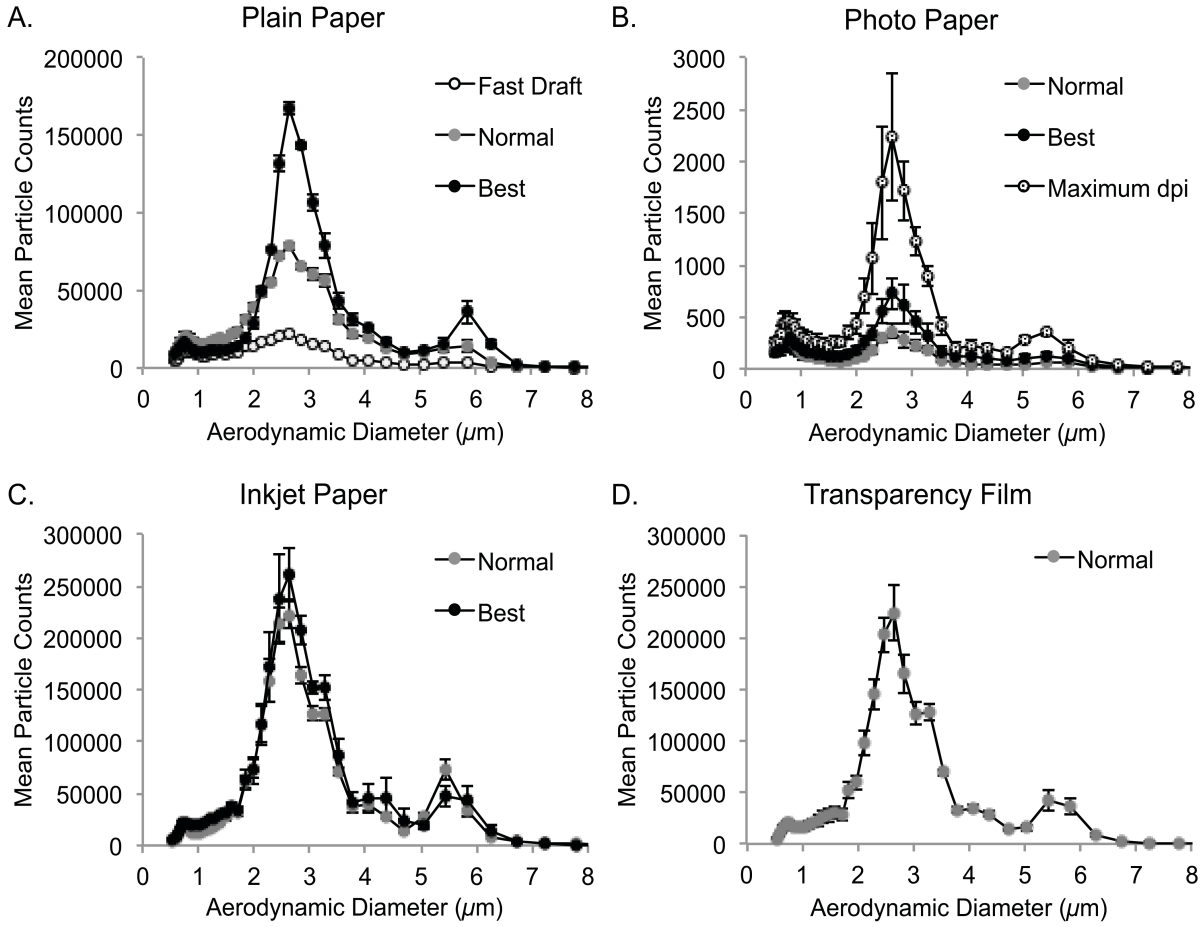

Figure 2.9. Mean size-distribution results from APS measurements of microparticles generated by printing 15% GDL solution using (A) plain paper, (B) photo paper, (C) inkjet paper, and (D) transparency film settings along with the associated print quality settings available in the commercial printer software ($n \geq 3$, error bars are one standard deviation).

The microparticle size-distribution results for the remaining two paper settings tested, inkjet paper and transparency film, were similar for both paper settings and both quality settings (Figure 2.9). The lack of difference in the produced microparticle characteristics when using the normal and best quality settings with inkjet paper was consistent with a common printing mechanism for the two conditions. These results agree with the timing data in Table 2.6, where the same number of cartridge passes were employed to generate droplets for the inkjet paper and transparency film, with the transparency film and best quality setting for the inkjet paper just taking longer with more drying time between passes. The number and aerodynamic diameter of the microparticles generated for inkjet paper and transparency film were both comparable to the results obtained when the plain paper, best quality setting was used, which was anticipated for the inkjet paper. It was expected that the size-distribution results for microdroplets printed onto transparency film would resemble the results for the photo paper setting. This was not observed, however, as the transparent media needs to be made opaque by the dense application of ink during printing, which requires a larger droplet number and size, similar to those generated for porous papers. In summary, the programmed print modes in the commercial software executed by the inkjet printer appeared to achieve the desired printed droplet features based on the time for aerosol production and the characteristics of the generated microparticles. The number of particles generated using the various software settings indicated that the photo paper option is unsuitable for high-throughput applications, while all other print media options yield high-throughput aerosol generation.

The mean best-fit Gaussian equations to the smoothed data were used to determine microparticle size-distribution maximum positions and FWHM values for statistical evaluations of the effects of varying the commercial printer software settings. The Gaussian best-fit equations for this data set all had R^2 above 0.90, with most greater than 0.95, indicating the equations were accurately representing the data (Miller and Miller 2005). Mean size-distribution maximum values for this data set ranged from 2.57 μm (plain paper, fast draft quality) to 2.77 μm (plain paper, best quality), which represents an 8% relative change in the data set. One-way ANOVA showed that the mean size-distribution maxima values varied significantly at 99% confidence ($p < 0.01$) when the printer settings were varied. This result was due to small standard deviations within replicate analyses. Changing the printer settings had less effect on the resulting microparticle size in comparison to the first data set where the GDL concentration in printed solution was varied (22% relative change). When the experimental printer settings were ranked according to mean size-distribution maximum value, no meaningful relationship between the produced microparticle size and the printer settings was observed. Microparticle size-distribution maximum results were also compared within the individual paper types separately using one-way ANOVA. Significant differences were observed only for the plain paper type, where the mean size-distribution maxima produced using the different quality settings were significantly different at 99% confidence ($p < 0.01$). In summary, these results indicate that while the produced microparticle size-distribution maxima may shift slightly (8%) when different printer software settings are employed, the size is more appreciably changed by modifying the solute concentration in the printed solution (22%), as demonstrated in the previous section. Any undesirable change in the size of the microparticles

as a result of changing the printer software settings to tune the size-distribution width could easily be remedied by slight adjustment of the solute concentration.

Figure 2.10 shows the mean best-fit Gaussian equations to the smoothed APS size-distribution data of microparticles generated using varying printer software settings. The size-distribution maxima were normalized to unity and the size-distribution maximum positions are all aligned at zero for more straightforward comparison, as the size-distribution maxima remained comparatively constant (8% relative change) within this data set. Mean size-distribution FWHM results spanned from 1.47 μm wide (photo paper, maximum dpi quality) to 2.74 μm wide (plain paper, fast draft quality), displaying a 68% relative change within the data set. One-way ANOVA revealed significant differences in the mean FWHM values at 99% confidence ($p < 0.01$) for the microparticles generated using varying printer software settings. Changing the printing parameters had a much larger effect on the resulting microparticle size-distribution width compared to changing the GDL concentration in the printed solution (13% relative change). This result was reasonable, as aiming to generate images on different media with varying degrees of perceived print quality requires changing the way the printhead functions, while simply adjusting the solute concentration is not expected to have a large effect on the way the droplets are generated. When the printer software settings were ranked according to mean size-distribution FHWM values, a relationship was observed that was consistent with expectations based on printhead operation. The microparticles produced using the photo paper, maximum dpi quality setting had the narrowest size-distribution (smallest FWHM), and the widest size-distribution (largest FWHM) was produced using the plain paper, fast draft quality setting. Within each individual paper type, the higher quality settings

produced narrower size-distributions, which was true for all paper types. Finally, the resolutions of all microparticle size-distributions from this data set were calculated according to Equation 2.2. The resolution results listed in Table 2.7 show that the size-distributions were not appreciably resolved, with resolution values less than 0.06. This was expected, as the size-distribution maxima positions changed a relatively small amount within this data set, and some of the printer settings produced relatively wide size-distributions. In summary, the ability to tune the size-distribution width of a generated aerosol by varying the commercial printer software settings was demonstrated. More monodisperse microparticle populations were generated using the highest quality print settings, which is often desirable. However, some applications require a more polydisperse but still controlled aerosol size-distribution, which is also achievable using the lowest quality print settings available in the COTS inkjet printer software.

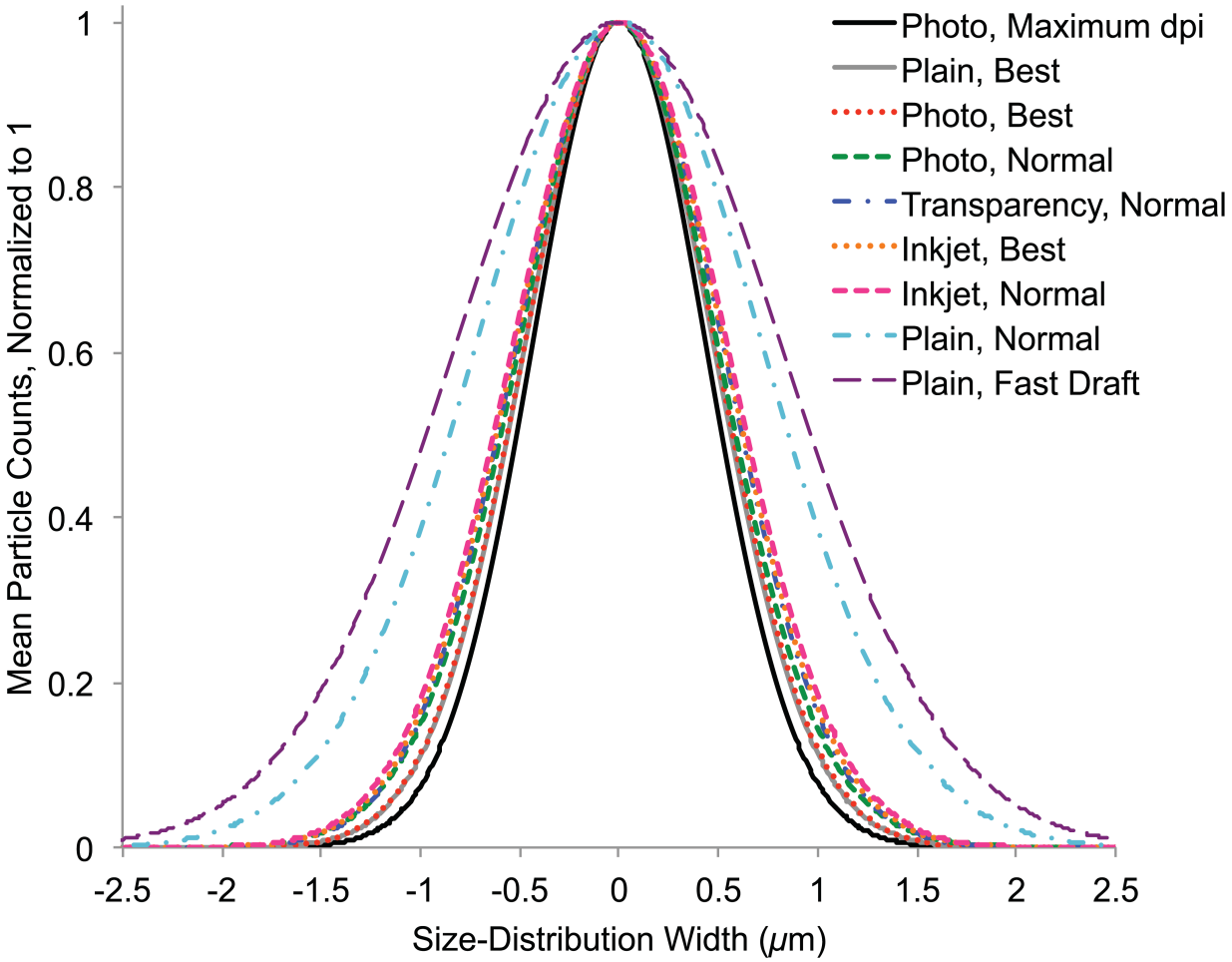

Figure 2.10. Mean microparticle size-distribution Gaussian fit results for 15% GDL solution printed using different paper type and print quality settings ($n \geq 3$), arranged in ascending order according to mean FWHM value. The mean particle counts for all samples are normalized to one, and all size-distribution maxima are aligned to zero for comparison. These results demonstrate the ability to tune the width of the generated microparticle size-distribution by altering the commercial printer software settings used during production.

Table 2.7. Microparticle size-distribution resolutions from pair-wise comparisons of all tested printer settings (smoothed, Gaussian fit data used in calculations). All values were less than one, indicating substantial overlap among the different size-distributions.

	Plain, Normal	Plain, Best	Photo, Normal	Photo, Best	Photo, Maximum dpi	Inkjet, Normal	Inkjet, Best	Transparency, Normal
Plain, Fast Draft	0.0374	0.0559	0.0469	0.0418	0.0392	0.0283	0.0390	0.0361
Plain, Normal		0.0124	0.0041	0.0028	0.0071	0.0152	0.0041	0.0074
Plain, Best			0.0099	0.0191	0.0251	0.0334	0.0199	0.0240
Photo, Normal				0.0085	0.0139	0.0229	0.0098	0.0137
Photo, Best					0.0053	0.0155	0.0018	0.0058
Photo, Maximum dpi						0.0111	0.0032	0.0010
Inkjet, Normal							0.0131	0.0094
Inkjet, Best								0.0038

The experiments discussed above demonstrated tunability of microparticle sizes produced using a commercial inkjet printer by varying the concentration of the solute in the printed solution, as well as the size-distribution width by changing the printer software settings. Changing these parameters usually altered the number of microparticles collected during each experiment and the duration of microdroplet generation as well, which are also parameters of interest when generating an aerosol or microparticle population for an application. The number of microparticles generated during an experiment is easily managed by printing different images from the computer. More microdroplets will be produced over a longer timespan when a larger image is printed, as the printer needs to cover a larger area on the print media. Large-scale, high-throughput production is also achievable by printing multiple pages covered almost completely by large black rectangles as the printed image. The same image was printed for all experiments during this study for experimental consistency, but changing the image printed is a straightforward method to tune the number of microparticles produced and the duration of production.

The microparticles generated during all experiments in this study were polydisperse to a limited extent (± ~0.7-1.4 μm, FWHM values ranging from 1.47-2.74 μm wide), which was likely due to the mechanical capabilities of the COTS inkjet cartridges as well as the software designed to operate them. Inkjet cartridges are manufactured to produce droplets that fall within a specified volume range, so some variation is expected among the approximately 300 nozzles in the cartridge. For example, the volume of the refill channel for each nozzle may vary slightly. Variability in microparticle size reportedly increases when continuous printing is employed, as was the case in the present study where the goal was to create a large and

representative microparticle population (Bohorquez *et al*. 1994). When droplets are ejected in close succession, differences in the meniscus position over the exit nozzle due to rapid refilling may cause slight volume differences in the microdroplets produced. The meniscus position varies due to the viscosity of the printed solution, which prevents the cartridge channels from completely refilling when droplets are printed in rapid succession. Also, as the printed solution is depleted, the backpressure in the cartridge decreases, yielding droplets with smaller volumes compared to the initially printed droplets, as the nozzles do not completely refill. Finally, as the temperature of the cartridge increases during continuous printing, the resulting droplet volumes also increase, an effect called thermal inkjet heating. The printhead is always heated to a designated temperature set-point customized for each printing mode before printing begins to correct for this effect, which is likely the source of the differing size-distribution widths when different printer software settings are used (Bohorquez *et al*. 1994).

Apart from the effects of the inkjet cartridges, software control algorithms also create the controlled polydispersity observed in the generated microparticles. The firmware- and hardware-based algorithms for generating images using the inkjet printer can be used to enhance edge smoothness and generate sharper-looking images by placing dots between the basic grid points or by changing the dot size (Bohorquez *et al*. 1994). As the image printed for this study's experiments was a black rectangle, the edges may have been smoothed in this manner, resulting in microparticles with different sizes. Also, the cartridge was directed to deliver overlapping swaths of solid black to achieve the printed image, so it is possible that the nozzles in the center of the printhead were programmed to produce more and larger droplets compared to the nozzles at the edges of the printhead so that the overlapping swath areas

would not become saturated. These design components of the inkjet cartridges and software control algorithms led to controlled and reproducible polydispersities in the generated microparticle populations. The degree of microparticle polydispersity observed in these experiments using the COTS inkjet printhead is acceptable for many scientific applications.

2.3.3 Scanning Electron Microscopy of Generated Microparticles

Scanning Electron Microscopy was used to confirm the APS size-distribution results by examining the morphology of the microparticles generated using the commercial inkjet printer. Microparticles created by printing 15% GDL solution using plain paper, normal quality (default) printer settings were collected on a filter using an impactor. Micrographs from SEM analysis are displayed in Figure 2.11, along with a size-distribution comparison between the SEM and APS data from Figure 2.6. The image on the left was obtained using relatively low magnification (150X) in order to observe the characteristics of the collected microparticle population. The large number of microparticles created and collected was expected and confirmed the high-throughput capability of this aerosol and microparticle generating technique. There was an observed uniform variability of size and shape in the microparticle population in the SEM micrographs, which demonstrated the consistency of the inkjet printer production method. There was also no evidence of particle aggregation (no observed fusion of two or more distinct particles into one non-spherical particle), indicating that the solvent is sufficiently evaporated before collection and analysis, resulting in dry microparticles composed of the solute material.

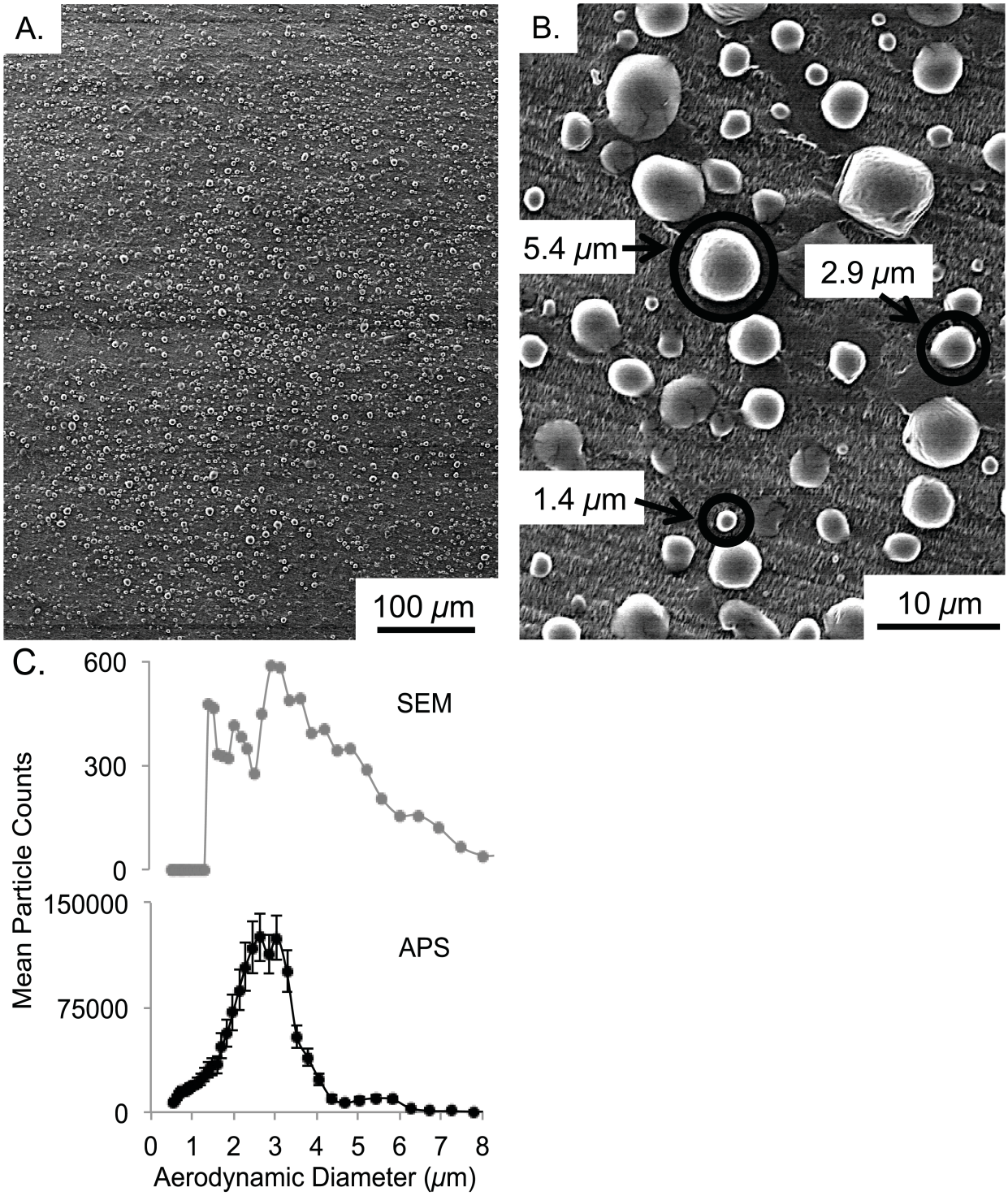

Figure 2.11. Micrographs taken during SEM analysis showing (A) the large number and representative characteristics of the GDL microparticle population produced using the inkjet printer at 150X and (B) the size and morphology of individual microparticles at 1500X, which agrees with the corresponding APS results (C).

The image on the right in Figure 2.11 was obtained using increased magnification SEM (1500X) to more precisely visualize the size and shape of the microparticles. The morphology of the microparticles is generally spherical as anticipated, with some displaying slightly oval or cubic features. Both the display features of the SEM Quartz PCI imaging software and ImageJ software were used to measure particle sizes and count particles in the images. The collected sample contained microparticles approximately 1-5.5 μm in diameter, with the majority in the range of 1.5-4 μm in diameter. The size-distribution generated using the SEM images generally agrees with the APS data, although three orders of magnitude less particles were analyzed using the SEM method, leading to an exaggeration of the contribution of particles with diameters greater than 5 μm. Also, the SEM samples were made using an impactor with a stage cutoff of 0.7 μm, so no particles smaller than that were collected for imaging as they passed to the next impactor stage. In summary, these SEM results indicated that spherical microparticles were generated by dehydrating microdroplets produced using a commercial inkjet printer, and that the microparticle size-distribution characteristics measured using the APS were accurate.

2.4 Spray Dryer Production Method Results

2.4.1 Effect of Solute Concentration in Spray Dried Solution on Microparticle Size-Distribution

The effects of varying the solute concentration and aerosol production parameters during spray drying on the resulting microparticle size-distributions were investigated in a manner analogous to the experiments described above for the modified inkjet printer. The oligosaccharide maltodextrin was used as the solute for spray dried microparticles as it produced a free-flowing, amorphous powder product with high yields, attributed to its high

glass transition temperature (T_g) that prevented sticking to the components of the spray dryer.

The effect of varying the maltodextrin concentration in the spray dried solution on the resulting microparticle size-distribution was investigated by spray drying aqueous solutions ranging from 1-20% maltodextrin and analyzing the generated powder using the APS. All other spray dryer operating parameters remained constant and are listed in Table 2.3. The spray dried powders were re-aerosolized into the inlet air flow of the APS by shaking the vial the powders were stored in by hand and summing the particle results for a minute, generating representative results for the bulk powder. A representative APS size-distribution result from the powder generated by spray drying the 10% maltodextrin solution is presented in Figure 2.12. The particle size-distribution appeared to be log-normal, displaying the characteristic tail on the larger diameter side of the distribution (Figure 2.12a). The same data were plotted on a logarithmic (base 10) size scale (Figure 2.12b), which yielded a symmetrical normal distribution as expected for distributions that are truly log-normal in shape (Miller and Miller 2005). While there is no fundamental theoretical reason for particle size data to approximate the log-normal distribution, it has been found to apply to most single-source aerosols, including those generated using nebulizers (Hinds 1999). This result is attributed to the spray dryer's nebulization droplet generation mechanism, which is different from the droplet generation mechanism in the modified inkjet printer cartridges described above. As the size-distributions generated by the spray dryer were not normal, it was not appropriate to fit them with Gaussian functions for more rigorous comparison as was done with the size-distributions generated by the inkjet printer. Comparisons of experimental results to theoretical particle diameters were

also not possible due to the random spray generated by the nebulizer in the spray dryer as opposed to the fixed nozzle diameter in the inkjet cartridges.

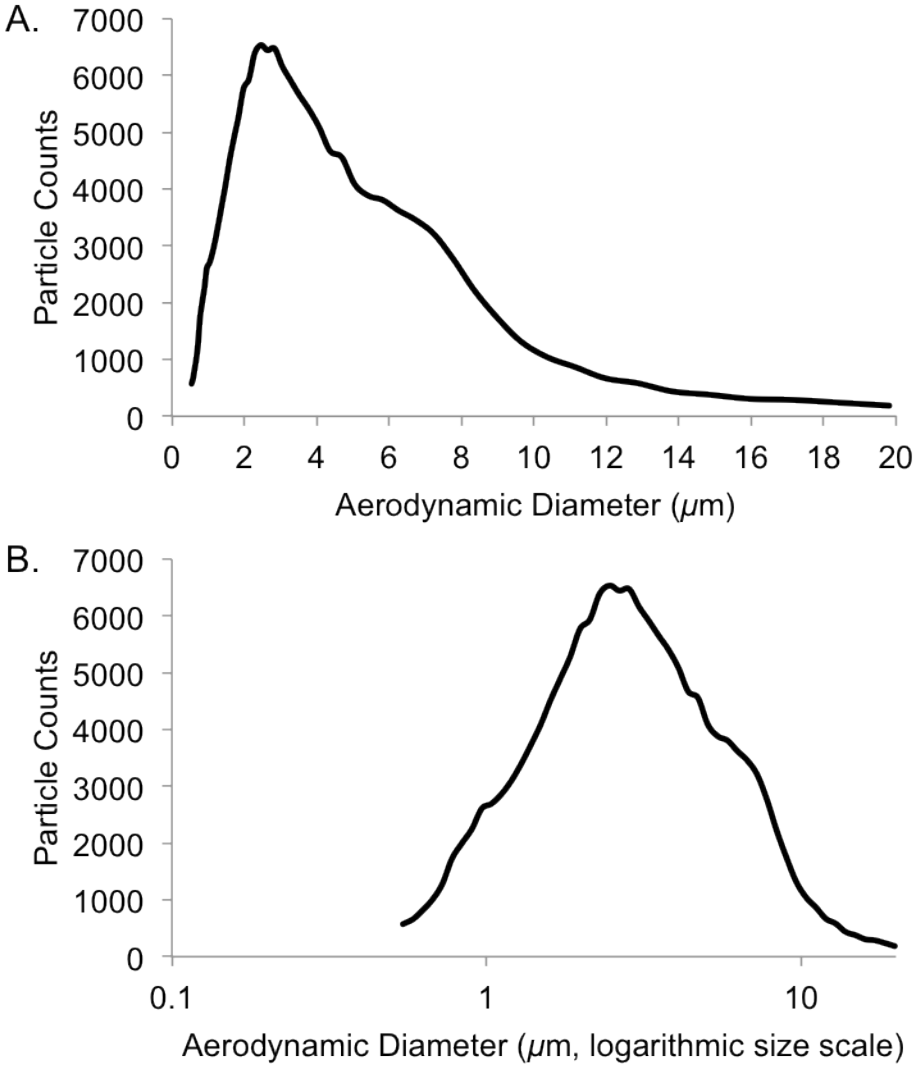

Figure 2.12. The size-distribution of microparticles produced using the spray dryer and 10% maltodextrin solution, analyzed using the APS. (A) Size-distribution data plotted on a linear scale displays the characteristic long tail at large particle sizes indicative of a log-normal distribution, which was verified by plotting the same data on a logarithmic scale (base 10), transforming the distribution to the symmetrical normal distribution (B).

Figure 2.13 displays APS size-distribution results from each powder generated from the different maltodextrin solutions. A relationship was observed between the maltodextrin concentration in the spray dried solution and the resulting size-distribution peak maximum positions, which shifted to larger particle aerodynamic diameters as the maltodextrin concentration in the spray dried solution was increased. This result was expected, as the higher concentration solutions have more solute material in them to form into microparticles when the water is evaporated. The size-distribution widths also increased as the maltodextrin concentration was increased, which is somewhat undesirable for the application of simulating environmental aerosols. The powder generated by spray drying the 1% maltodextrin solution (Figure 2.13) displayed the size-distribution most similar to those generated by the inkjet printer (Figure 2.6), with most of the microparticles in the range of 1-5 μm in aerodynamic diameter. However, the water was not completely evaporated from the generated powder product due to the high percentage of it in the spray dried solution, leaving the powder undesirably damp (observed during product transfer from the collector to the storage vial). A 3% maltodextrin solution was chosen as optimal for test particle generation as the powder product was completely dry and the size-distribution was still relatively small in diameter and narrow. As anticipated, the size of the microparticles generated by spray drying maltodextrin solutions was tunable according to solute concentration. This tunability led to the generation of a powder product with size-distribution characteristics similar to those generated using the modified inkjet printer and those of some environmental aerosols.

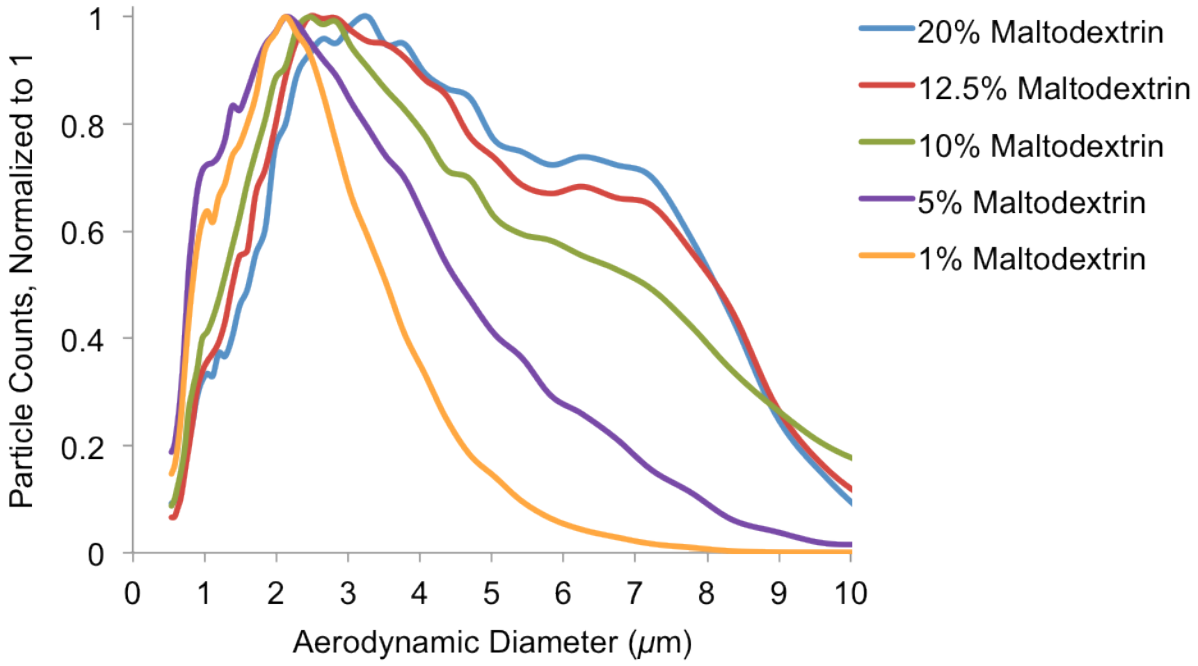

Figure 2.13. Size-distribution APS results of powders generated using the spray dryer and varying concentrations of maltodextrin in solution display particle size tunability according to solute concentration.

2.4.2 Effect of Spray Dryer Settings on Microparticle Size-Distribution

The effects of varying the spray gas flow rate during droplet generation on the resulting microparticle size-distribution was examined by generating powders using flow rates from 30-50 mm (rotameter values set on the spray dryer, which equate to 7.32-17.53 L/min) and analyzing them using the APS. The spray dried solution was 5% maltodextrin for all flow rates, and all other operating parameters remained constant and are listed in Table 2.3. The vials containing the generated powders were shaken by hand to re-aerosolize the microparticles into the inlet airflow of the APS for size-distribution analysis. Particle count and size results were summed over a minute to yield data representative of the microparticle population. Figure 2.14 shows APS size-distribution results for each maltodextrin powder generated using a different

spray gas flow rate in the nebulizer during droplet generation. The particle size-distributions were log-normal for all tested flow rates, which is characteristic for droplets produced by a nebulization mechanism. The slowest spray gas flow rate (7.32 L/min) generated the largest microparticles in the widest size-distribution. The microparticles had smaller diameters and narrower size-distributions as the spray gas flow rate was increased. This result was expected, as a higher flow rate of spray gas means there is more energy for fluid dispersion, which breaks up the liquid stream more vigorously into smaller droplets compared to a slower gas flow rate. While the highest tested flow rate (17.53 L/min) generated the smallest particles and narrowest size-distribution with most of the particles in the range of 1-5 μm in diameter as desired, the compressed nitrogen spray gas was being used too quickly. Therefore, the second highest flow rate of 13.85 L/min (45 mm) was selected as the optimal spray gas flow rate for aerosol test particle generation. The compromise of using the slightly lower flow rate to nebulize the less concentrated 3% maltodextrin solution generated microparticles principally 1-5 μm in diameter, accurately simulating environmental aerosols. The particle size tuning capabilities of the spray dryer facilitated generation of maltodextrin microparticles of the desired size, similar to the GDL microparticles generated using the modified inkjet printer, but in gram-scale quantities useful for atmospheric testing.

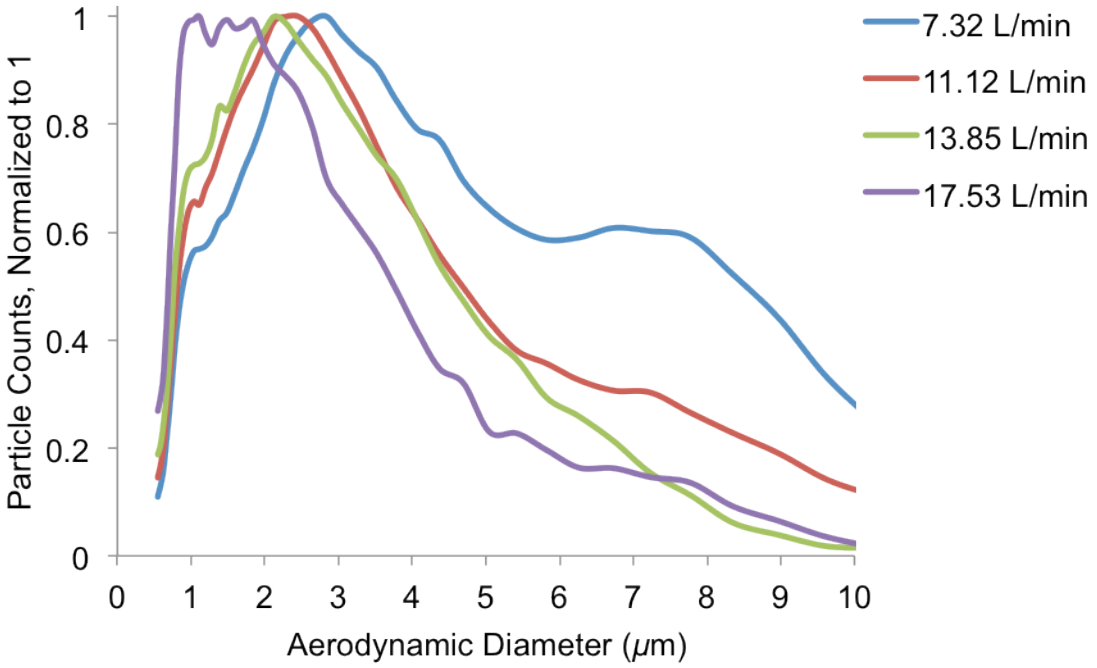

Figure 2.14. Microparticle size-distribution APS results generated using the spray dryer and varying spray gas flow rates demonstrates tunability of the particle size according to the spray drying parameters.

2.4.3 Scanning Electron Microscopy of Generated Microparticles

The morphology of maltodextrin microparticles generated using the spray dryer with optimal operating parameters (3% maltodextrin, 13.85 L/min spray gas flow rate, Table 2.3) was examined using SEM. The particles were deposited onto a filter using an impactor and sputter coated with a thin layer of gold to make them conductive for SEM analysis. Micrographs of the maltodextrin microparticles are displayed in Figure 2.15. The image on the left was acquired using relatively low magnification (250X) to observe the general microparticle population characteristics, and the image on the right was obtained using higher magnification (1200X) to examine individual particle morphology. The microparticle population generated using the spray dryer was more polydisperse compared to the population produced using the modified

inkjet printer. This result agrees with the APS results discussed above and is an effect of the nebulization droplet generation mechanism in the spray dryer, which generates a log-normal size-distribution that is wider compared to the normal size-distributions produced using the modified inkjet printer. There is no evidence of particle aggregation, indicating that the water is completely evaporated during spray drying as desired. The individual particles have precisely spherical morphology owing to the optimized droplet drying within the spray dryer. This contrasts with the results generated by the aerosol desiccant dryers used to dry the droplets generated by the inkjet printer, which yielded more varied particle morphologies that were generally spherical but with some oval and cubic features (Figure 2.11). The maltodextrin microparticles in Figure 2.15b are approximately 1-4.5 μm in diameter, and image analysis of several micrographs yielded the size-distribution shown in Figure 2.15c. Two orders of magnitude more particles were characterized when the same maltodextrin microparticle sample was analyzed using the APS, however, the size-distribution in the SEM image agrees with the APS results. Microparticle characterization using SEM confirmed particle sizes from 1-5 μm in diameter and revealed strictly spherical particle morphology, both of which are suitable for the aerosol test particles.

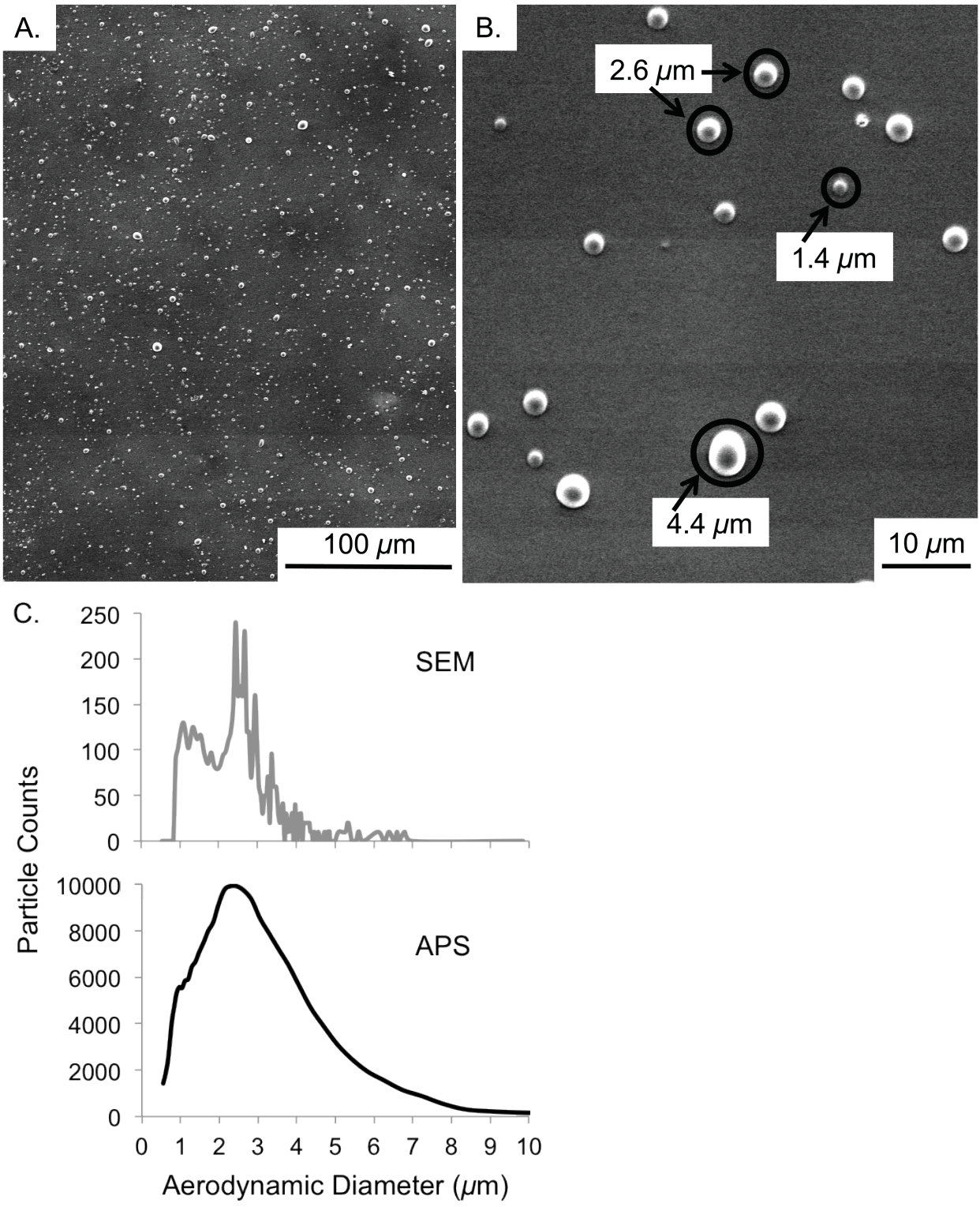

Figure 2.15. Micrographs from SEM analyses of maltodextrin microparticles generated using the spray dryer. (A) Displays the general characteristics of the microparticle population at 250X and (B) shows the size and morphology of individual microparticles at 1200X. A comparison of the microparticle size-distributions measured from the SEM image (A) and the APS is shown in (C).

2.5 DNA Quantification in Particles

The number of DNA barcodes contained in each particle is an important metric for evaluating how accurately the aerosol test particles simulate biological microorganisms. Microbes contain only one copy of DNA unless they are dividing, so only one or two copies of the DNA barcodes were desired per generated microparticle. Individual microbe populations, *e.g.* bacteria, are also relatively monodisperse (approximately 0.5-2 μm in diameter), whereas microparticle populations generated using the spray dryer were more polydisperse (approximately 1-10 μm in diameter) and in a log-normal size-distribution. Therefore, the goal was further refined to generating particle populations where the 1-2 μm particles had 1-2 DNA barcodes. As the number of DNA barcodes incorporated into each particle was assumed to be proportional to the volume of the particle (cubic dependence), the larger particles had up to two orders of magnitude more DNA barcodes, but the spray dryer produced fewer of them. Hence, the number of DNA barcodes per particle on average in an aerosol test particle population was determined first, and then the number of DNA barcodes within each particle size (particle sizes binned according to the APS size binning) was calculated to determine whether the particle population achieved the specified goal of 1-2 μm particles with 1-2 DNA barcodes.

In order to determine the average number of DNA barcodes per microparticle, both the number of particles present and the number of DNA barcodes present in a particle population must be known. A maltodextrin powder was generated using the spray dryer, with template DNA added to the spray dried solution so that DNA barcodes were incorporated into the generated microparticles. A small representative particle population (on the order of 10^5

particles) was removed from the bulk powder and individually optically counted (proprietary method) to quantify the number of microparticles. After the total was obtained, the particle population was washed into solution and the number of DNA barcodes present in the sample was quantified using QRT-PCR. It was previously verified that the QRT-PCR assay was not adversely affected by the presence of maltodextrin in the reaction (up to 5% w/v) by testing maltodextrin concentrations ranging from 0-5% (0%, 0.1%, 1%, 2%, 3%, 4%, and 5% tested) with two different DNA barcode concentrations (9.6×10^3 and 9.6×10^5 copies) in the QRT-PCR reaction ($n = 3$). One-way ANOVA of the C_T results at 95% confidence indicated that the means did not significantly differ for both the samples containing 9.6×10^3 DNA barcodes ($p = 0.192$) and the samples containing 9.6×10^5 DNA barcodes ($p = 0.062$). Dividing the number of DNA barcodes quantified by the QRT-PCR assay by the counted number of microparticles gave the average number of DNA barcodes per microparticle. The same bulk powder was also characterized using the APS to determine the size-distribution of the particle population. The number of DNA barcodes in each particle size was then determined mathematically. It was assumed that aerodynamic diameter measured by the APS could be treated as the diameter of a sphere (good assumption from the SEM results (Figure 2.15) that the particles are spheres), from which the volume of a single particle of each size was calculated. The particle size-distribution results from the APS were then adjusted so that the total number of particles in the distribution was the same as the number of particles in the population that was quantified. The adjusted number of particles in each size bin was then multiplied by the particle volume to determine the total particle volume in each size bin. The total particle population volume was

then determined by summing all bins, and the volume fraction of the total in each size bin was calculated. The volume fraction for each size bin was then multiplied by the total number of DNA barcodes quantified in the particle population, yielding the total number of DNA barcodes in the size bin. This assumed that the DNA was homogeneously distributed throughout the particle volume, which is a good assumption as the spray dried maltodextrin solution containing the DNA templates was stirred throughout the spray drying process. Finally, the number of DNA barcodes in the size bin was divided by the number of particles in the size bin to yield the number of DNA barcodes in each particle of that size. If the particles did not have enough DNA barcodes, more template DNA was added to the initial spray dried solution. Likewise, if the particles had an unrealistically high number of incorporated DNA barcodes, less DNA was added to the spray dried solution. In this manner the optimal amount of template DNA to add to the spray dried solution was empirically determined to yield an aerosol test particle population where 1-2 μm particles had 1-2 DNA barcodes per particle.

A representative DNA barcodes per microparticle quantitation result for a spray dried microparticle population is illustrated in Figures 2.16 and 2.17. The powder was completely analyzed twice (n = 2), with PCR samples for three sample dilution levels in triplicate (n = 9 for each analysis, yielding n = 18 total) to ensure there was no observed PCR reaction inhibition (no inhibition observed). The histogram of the microparticle size-distribution measured using the APS is displayed in Figure 2.16, with the particle sizes binned according to aerodynamic diameter on a logarithmic scale on the x-axis and the number of particles in each size bin on the y-axis. Note that the logarithmic-scaled x-axis causes the log-normal particle size-distribution generated by the spray dryer to appear normal. The same aerodynamic diameter size bins used

by the APS software were used in the calculations to determine the number of DNA barcodes per particle for each particle size bin. The numbers of particles in the characterized microparticle sub-populations were 308164 and 410992, and the number of DNA barcodes present in samples was quantified using QRT-PCR. The final result was 14.6 ± 3.5 DNA barcodes per particle on average, meaning that the particle size was not taken into account in this calculation (n = 18, mean ± one standard deviation). This value was used to calculate the number of DNA barcodes incorporated into each particle size, which is a more useful metric for evaluating the similarity of the aerosol test particles to microorganisms. The results of these calculations, described above, are shown in Figure 2.17. Figure 2.17a shows the number of DNA barcodes per particle for all of the size bins measured by the APS. The relationship is cubic owing to the dependence of the number of DNA barcodes on the particle volume, resulting in a three order of magnitude increase in the number of DNA barcodes per particle for particles increasing one order of magnitude in aerodynamic diameter. Figure 2.17b focuses on the most interesting portion of the graph displaying the number of DNA barcodes incorporated into particles less than 2.5 μm in aerodynamic diameter, the size of microorganisms. From these results it was determined that particles in the 1.037 μm aerodynamic diameter bin contain 0.53 ± 0.16 DNA barcodes per particle (mean ± one standard deviation; approximately one particle out of two has a DNA barcode). Particles in the 1.286 μm size bin had 1.01 ± 0.30 DNA barcodes in each particle, and particles in the 1.981 μm size bin had 3.68 ± 1.08 DNA barcodes per particle. Of the entire particle population, 87% ± 3% (mean ± one standard deviation) of the particles contained at least one DNA barcode and 13% ± 3% of the particles did not contain a DNA barcode (some of the smallest particles did not have a DNA template molecule

incorporated). These aerosol test particles achieved the goal of 1-2 μm particles containing 1-2 DNA barcodes and therefore accurately mimic properties displayed by typical biological aerosols. The tunability of the spray dryer particle production process was instrumental in accomplishing the desired number of DNA barcodes per particle. It should be noted that similar experiments were successfully carried out with GDL microparticles generated using the modified inkjet printer as well, with tunability of the number of DNA barcodes per particle demonstrated (data not shown). These novel aerosol test particle generation methods enabled abundant flexibility to tune both particle size as well as the number of DNA barcodes incorporated into the particles by straightforward adjustments to production parameters. Microparticles accurately simulating biological aerosols both in terms of size and amount of DNA available for PCR detection were created on the gram scale for use in atmospheric release tests. The DNA barcodes themselves are also tunable by adding different templates to allow multiple releases in the same environment, either simultaneously or sequentially, which may be accurately characterized by PCR with minimal chance of errors due to background contamination.

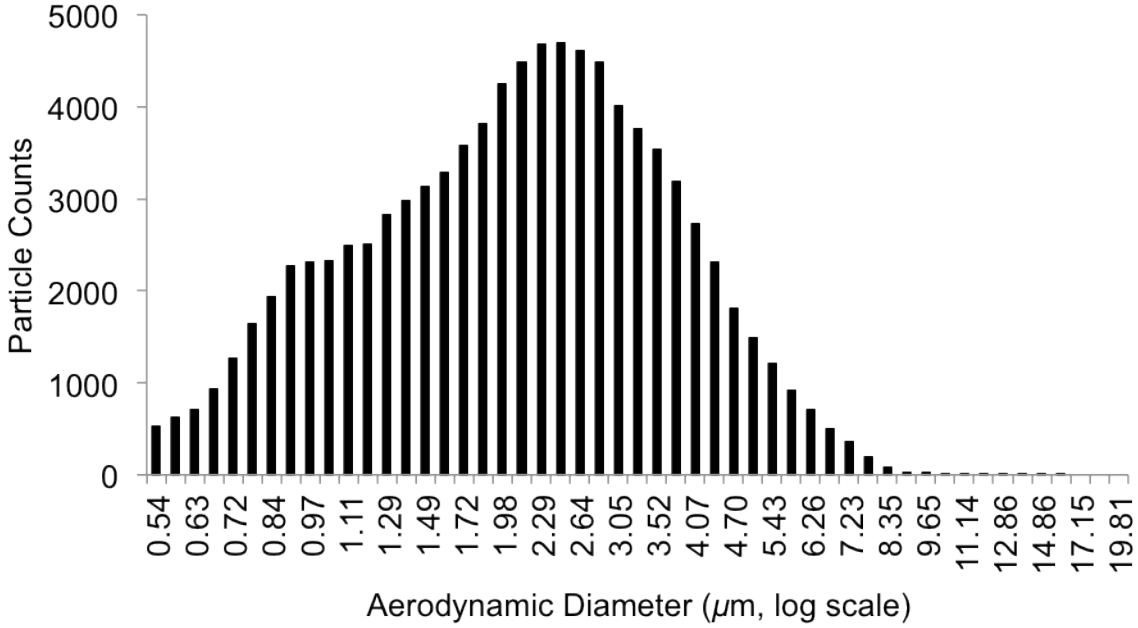

Figure 2.16. The microparticle APS size-distribution for quantifying the number of DNA barcodes per particle in each particle size bin. The aerodynamic diameter bins are on a logarithmic scale dictated by the APS instrument software.

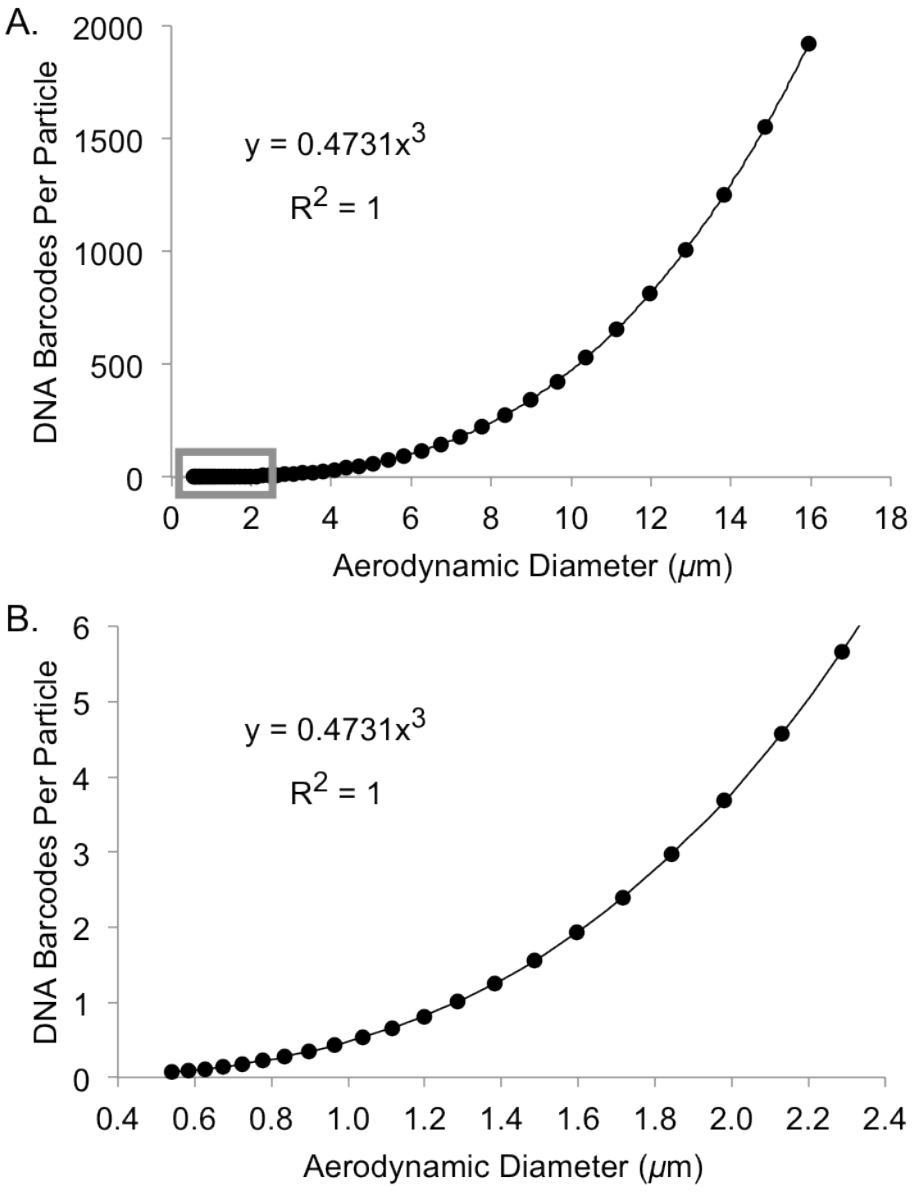

Figure 2.17. Calculated number of DNA barcodes per particle as a function of particle aerodynamic diameter for (A) the entire range of particle sizes measured by the APS and (B) only the particles smaller than 2.5 μm (decreased scale to focus on the region outlined by the square in (A)).

2.6 Atmospheric Release Test

A small-scale atmospheric release test was conducted using the aerosol test particles described in the previous section (spray dried maltodextrin containing 14.6 DNA barcodes per particle on average) to demonstrate their utility for this application. The particles were aerosolized using a pesticide sprayer (Figure 2.18) in unused laboratory space towards a Dry Filter Unit (DFU) aerosol collector (Lockheed Martin, Bethesda, MD) positioned 6.7 meters away. The DFU draws 800 L/min of room air through two adjacent polyester felt filters (1.0 μm pore size, Lockheed Martin) which trap the aerosols for downstream analyses including QRT-PCR for DNA content. Prior to the release, the room background air was sampled for 15 minutes. After the DFU filters were changed, 0.74 g of aerosol test particles was released using the pesticide sprayer over approximately 3 seconds in the general direction of the DFU. There were no obstructions between the release point and the DFU, and the normal room airflows were operational. Particle fallout directly in front of the release point was not visually observed, indicating that the majority of the aerosol test particles were small enough to remain airborne and travel throughout the room. The room air was sampled using the DFU for 15 minutes post-release. At the conclusion of the test, one DFU filter from the room background and one DFU filter from the aerosol release were analyzed using QRT-PCR for the DNA barcodes incorporated into the particles (the other DFU filter was archived in the freezer for follow-up testing if necessary). A blank DFU filter was also processed as a negative control. The DFU filters were placed in 50 mL conical tubes (Becton, Dickinson and Company, Franklin Lakes, NJ) containing 10 mL of PBS buffer (Phosphate Buffered Saline, pH 7.4, Amresco, Solon, OH) with 0.1% Triton X-100 (Acros Organics, part of Thermo Fisher Scientific, Pittsburgh, PA) and shaken vigorously

by hand for 2 minutes. The conical tubes were then vortexed for 30 seconds and a 1 mL aliquot was removed for QRT-PCR analysis. The aliquot was serially diluted by factors of two with QRT-PCR-grade water (Teknova) up to 1:10 to check for PCR reaction inhibition by any room contaminants drawn into the filters. The undiluted, 1:2, 1:4, and 1:10 samples were analyzed by QRT-PCR in triplicate along with a calibration curve for results interpretation.

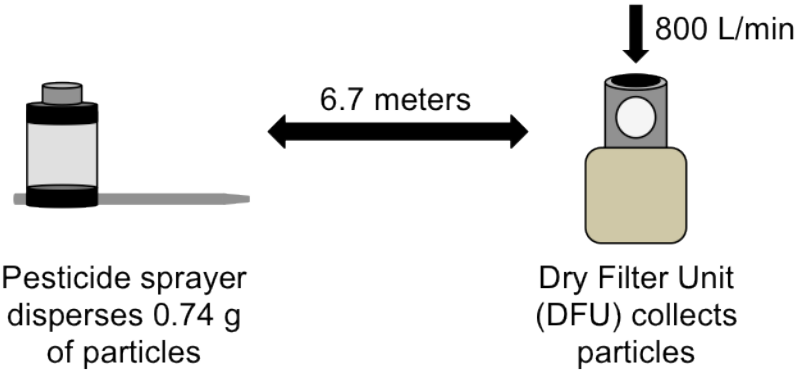

Figure 2.18. Schematic drawing of the equipment used to perform the atmospheric release test.

The QRT-PCR results for the blank DFU filter were negative, indicating no template contamination in the PCR reaction. The aerosol release DFU filter was positive for the DNA barcodes. The undiluted sample showed some PCR inhibition and was excluded from the results. Factoring in the dilution factors and the 14.6 DNA barcodes per particle on average, 6.30×10^7 ± 3% relative standard deviation (mean ± RSD, *n* = 9) aerosol test particles were collected on the DFU filter (100% recovery of the DNA barcodes from the filter was previously experimentally verified (data not shown)). The room background DFU filter collected prior to the aerosol release was weakly positive for the DNA barcodes as well, indicating the presence of 9.92×10^3 ± 53% RSD (*n* = 9) aerosol test particles. The magnitude of background

contamination was 0.02% of the result after the release, which is trivial. The contamination likely originated from preparing for the aerosol release in the room adjacent to the testing area while the room background sample was being collected. A door separated the two laboratory areas, but the airflow of the building could have carried the test particles from the staging area to the testing area, which is an interesting result in itself. This preliminary atmospheric release test demonstrated that the novel aerosol test particles travel in the airflows of an indoor environment as expected. The carefully tuned particle size and number DNA barcodes per particle enable the particles to travel in manner similar to natural aerosols and be detected and quantified using selective QRT-PCR assays. The aerosol test particles show great promise for large-scale atmospheric release testing to evaluate airflows in buildings and study aerosol drift and dispersion in populated environments.

2.7 Conclusions and Future Directions

The developed aerosol test particles exhibited all of the target characteristics outlined at the beginning of the project for accurate assessment of environmental aerosol transport and fate in populated locations. Food additives approved by the FDA for human exposure were suitable as bulk materials for the test particles using either the modified inkjet printer or the commercial spray dryer for particle generation. The non-coding DNA templates added as unique particle identifiers were also safe for human contact and yielded customized test particles able to be specifically detected using QRT-PCR assays. These safe, customizable, and specifically detected aerosol test particles were generated with the same sizes as aerosols commonly observed in the environment on the gram-scale for studying large locations. No

currently available simulant material meets all of these criteria for precisely studying atmospheric transport in populated environments. The project culminated with a successful demonstration of the aerosol test particles in an atmospheric release test, which was their desired end use.

The sizes and shapes of the generated particles were true to the general physical properties of aerosols observed in the environment. The principal particle sizes (1-10 μm) that were created overlap with the sizes of many environmental aerosols of interest, including dust, fog, mist, smog, spray, cloud droplets, cement dust, coal dust, flour, coal fly ash, machining fluids, tobacco smoke, paint spray, bacteria, and fungal spores (Hinds 1999). Simple modifications to the spray dryer parameters (*e.g.* higher concentration of solute in spray dried solution, lower spray gas flow rate) would enable production of particles larger (up to 25 μm) than those that were generated for this work if larger environmental aerosol simulants were required (*e.g.* pollen is 10-100 μm). Generating sub-micrometer particles useful for simulating aerosols including smoke, metal fumes, and viruses is more of a challenge. The cyclone collector dictates the lower limit of particle sizes able to be generated by the spray dryer used in this work as the cyclone collection efficiency markedly decreases for particles less than 1 μm in diameter. There is a different spray dryer model available (the BUCHI Nano Spray Dryer B-90) that employs an electrostatic particle collector for particles 0.3-5 μm in diameter that could be used to generate smaller test particles. However, this is at the expense of not collecting the particles larger than 5 μm and working with smaller quantities (less than 1 g of product per run) than what was possible using the spray dryer with the cyclone collector. It is possible that sub-micrometer particles could be generated using the modified inkjet printer, however, a great

deal more work would need to be dedicated to the formulation of the printed solution. Theory dictates that smaller particles will be produced if either the droplets are smaller or there is less solute material in the printed solution. The inkjet cartridges operated multi-modally when the lowest concentration solutions were printed, indicating that they were not able to function correctly when the printed solution had low viscosity. This could potentially be remedied by incorporating highly viscous additives that are safe for human exposure. The cartridge nozzle sizes are fixed, so altering droplet size would be challenging. An impingement surface could possibly be employed that would break up the droplets further prior to desolvation. In addition, a different particle analysis instrument would need to be employed to analyze the created particles smaller than 0.5 μm, such as the TSI Scanning Mobility Particle Sizer Spectrometer 3936 which generates size-distribution results for particles 2.5 nm – 1 μm in diameter. Generating sub-micrometer particles by building on the methods of this research has the potential for creating aerosol test particles suitable for assessing the transport of hazardous aerosols such as smoke and viruses, which has great potential to improve the safety and health of people in the studied environments.

The test particle size-distributions observed during this work (~2-10 μm wide) are actually less polydisperse than most aerosols in the environment, where aerosols that span two or three orders of magnitude in size are common. The narrower size-distributions that were made were more easily characterized by the APS and SEM compared to polydisperse environmental aerosols. The APS only characterized particles from 0.5-20 μm in diameter, so any particles outside of this range would have been inadvertently excluded from size-distribution results. It was important to verify that there were no particles in the generated

populations that were outside of the APS size range using a secondary method, which was SEM. The SEM micrographs did not show any particles smaller than 0.5 μm or larger than 20 μm in the examined samples, so the APS size-distribution results were accurate for the generated aerosol test particles.

The primary particle sizes of 1-10 μm generated throughout this work also permitted realistic DNA barcode incorporation of 1-2 DNA molecules per 1-2 μm particles, accurately mimicking bioaerosols. Both the normal distributions generated using the modified inkjet printer and the log-normal distributions created using the spray dryer for particle production had their peak maxima between 1-3 μm, allowing most of the particles (> 80%) to contain DNA barcodes for detection after release tests. A more cost-effective source for the DNA barcode oligonucleotides than commercial synthesis would permit the addition of more DNA template to the droplet generation solution if desired. Some possibilities are to use plasmids during commercial generation or to simply amplify the templates using PCR and add the purified amplicon as the DNA barcode in the particles. Finally, developing a multiplexed PCR assay so that multiple DNA barcodes may be detected simultaneously in samples will greatly accelerate sample processing after synchronized release tests using different DNA barcodes in particles released from different locations. The optical detection systems in the Smart Cycler thermal cyclers used in this work can monitor four fluorescence emission channels simultaneously during multiplexed assays. The requirements for successful multiplexed PCR assays are (1) no fluorophore emission spectral overlap and (2) no primer cross-amplification. The four optical channels in the Smart Cyclers are designed to detect certain classes of common fluorophores, so the four probes need to be synthesized with one fluorophore picked from each list of

detectable fluorophores in each optical channel. The fidelity of the primers and probes designed for each DNA barcode also needs to be assessed to check for non-specific amplification of the other templates in the reaction. The primers and probes should only anneal to their designed target template to yield accurate quantitation of the amount of each template in the sample.

One of the principal advantages of using these novel aerosol test particles to study aerosol transport is that introducing unique DNA barcodes to an environment enables sensitive and specific detection of the released particles. A final note about the 100 bp DNA barcodes is that even this relatively short template allows the creation of 10^{60} possible unique barcodes (4 different DNA bases, ATGC, in 100 different positions in the template, $4^{100} = 1.6 \times 10^{60}$). Even if some templates do not have appropriate characteristics for successful PCR amplification (GC content of 50-60%, not likely to form dimers or secondary structures) or are similar to sequences detected in the background aerosol population, there is still a nearly unlimited supply of DNA barcodes. Customized particles could be used to study aerosol transport and fate in the same location over and over, with accurate results for every experiment, due to the novel aerosol test particle generation methods developed during this project.

REFERENCES

2.8 References

Adhikari, B., Howes, T., Lecomte, D., Bhandari, B. R. (2005). A glass transition temperature approach for the prediction of the surface stickiness of a drying droplet during spray drying. Powder Technol. 149:168-179.

Arpagaus, C., Schafroth, N., Meuri, M. (2010a). Laboratory scale spray drying of lactose: a review. Best@Buchi Information Bulletin Number 57. www.buchi.com.

Arpagaus, C., Schafroth, N., Meuri, M. (2010b). Laboratory scale spray drying of inhalable drugs: a review. Best@Buchi Information Bulletin Number 59. www.buchi.com.

Berglund, R. N., Liu, B. Y. H. (1973). Generation of monodisperse aerosol standards. Environ. Sci. Technol. 7:147-153.

Bhandari, B. Stickiness during spray drying. http://www.uq.edu.au/lcafs/documents/kmutt.ppt. Accessed 12/05/12.

Bohorquez, J. H., Canfield, B. P., Courian, K. J., Drogo, F., Hall, C. A. E., Holstun, C. L., Scandalis, A. R., Shepard, M. E. (1994). Laser-comparable inkjet text printing. Hewlett-Packard J. 45:9-17.

Bottiger, J. R., Deluca, P. J., Stuebing, E. W., VanReenen, D. R. (1998). An ink jet aerosol generator. J. Aerosol Sci. 29:S965-S966.

BUCHI Corporation (2002). Training papers: spray drying. www.buchi.com.

Buehner, W. L., Hill, J. D., Williams, T. H., Woods, J. W. (1977). Application of ink jet technology to a word processing output printer. IBM J. Res. Dev. 21:2-9.

Burton, N. C., Adhikari, A., Grinshpun, S. A., Hornung, R., Reponen, T. (2005). The effect of filter material on bioaerosol collection of *Bacillus subtilis* spores used as a *Bacillus anthracis* simulant. J. Environ. Monit. 7:475-480.

Buskirk, W. A., Hackleman, D. E., Hall, S. T., Kanarek, P. H., Low, R. N., Trueba, K. E., Van de Poll, R. R. (1988). Development of a high-resolution thermal inkjet printhead. Hewlett-Packard J. 39:55-61.

Carnahan, R. D., Hou, S. L. (1977). Ink jet technology. IEEE T. Ind. Appl. 13:95-104.

Carrera, M., Sagripanti, J.-L. (2009). Artificial plasmid engineered to simulate multiple biological threat agents. Appl. Microbiol. Biotechnol. 81:1129-1139.

Chen, P. H., Chen, W. C., Ding, P. P., Chang, S. H. (1998). Droplet formation of a thermal sideshooter inkjet printhead. Int. J. Heat Fluid Fl. 19:382-390.

Childers, A. G., Hieftje, G. M. (1986). An improved uniform-size-droplet generator. Appl. Spectrosc. 40:688-691.

Dougherty, G. M., Hadley, D. R., O'Connor, P. R., Bottiger, J. R. (2007). Engineered Aerosol Production for Laboratory Scale Chemical/Biological Test and Evaluation. Available at https://e-reports-ext.llnl.gov/pdf/347374.pdf.

Fittschen, U. E. A., Hauschild, S., Amberger, M. A., Lammel, G., Streli, C., Forster, S., Wobrauschek, P., Jokubonis, C., Pepponi, G., Falkenberg, G., Broekaert, J. A. C. (2006). A new technique for the deposition of standard solutions in total reflection X-ray fluorescence spectrometry (TXRF) using pico-droplets generated by inkjet printers and its applicability for aerosol analysis with SR-TXRF. Spectrochim. Acta B 61:1098-1104.

Fittschen, U. E. A., Bings, N. H., Hauschild, S., Forster, S., Kiera, A. F., Karavani, E., Fromsdorf, A., Thiele, J., Falkenberg, G. (2008). Characteristics of picoliter droplet dried residues as standards for direct analysis techniques. Anal. Chem. 80:1967-1977.

Flegler, S. L., Heckman, J. W., Klomparens, K. L. (1993). *Scanning and Transmission Electron Microscopy: An Introduction*. Oxford University Press, New York, NY, pp. 1-10, 65-76.

Fletcher, R. A., Brazin, J. A., Staymates, M. E., Benner, B. A., Gillen, J. G. (2008). Fabrication of polymer microsphere particle standards containing trace explosives using an oil/water emulsion solvent extraction piezoelectric printing process. Talanta 76:949-955.

Hall, C. A. E., Scandalis, A. R., Broder, D. W., Moore, S. I., Movaghar, R., Rhoads, W. W., Schwiebert, W. H. (1994). Inkjet printer print quality enhancement techniques. Hewlett-Packard J. 45:35-40.

Hauschild, S., Lipprandt, U., Rumplecker, A., Borchert, U., Rank, A., Schubert, R., Forster, S. (2005). Direct preparation and loading of lipid and polymer vesicles using inkjets. Small 1:1177-1180.

Heid, C. A., Stevens, J., Livak, K. J., Williams, P.M. (1996). Real time quantitative PCR. Genome Res. 6:986-994.

Hieftje, G. M., Malmstadt, H. V. (1968). A unique system for studying flame spectrometric processes. Anal. Chem. 40:1860-1867.

Hinds, W. C. (1999). *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles* (2nd Ed.). John Wiley and Sons, Inc., New York, NY, pp. 1-14, 90-91, 252-253, 304-307, 394-396.

Holland, P. M., Abramson, R. D., Watson, R., Will, S., Saiki, R. K., Gelfand, D. H. (1992). Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. Clin. Chem. 38:462-463.

Holm, R. L., Caldow, R., Hairston, P. P., Quant, F. R., Sem, G. J. (1997). An enhanced time-of-flight spectrometer that measures aerodynamic size plus light-scattering intensity. J. Aerosol Sci. 28:S11-S12.

Jain, R., Shah, N. H., Malick, A. W., Rhodes, C. T. (1998). Controlled drug delivery by biodegradable poly(ester) devices: different preparative approaches. Drug Dev. Ind. Pharm. 24:703-727.

Kamphoefner, F. J. (1972). Ink jet printing. IEEE T. Electron Dev. 19:584-593.

Madigan, M. T., Martinko, J. M., Parker, J. (2000). *Brock Biology of Microorganisms* (9th Ed.). Prentice Hall, Upper Saddle River, NJ, pp. 55-56, 333, 508, 543, 916.

Maehara, N., Nakane, S., Yamamoto, K., Iga, K. (1984). A pinhole-plate ultrasonic atomizer. Ultrasonics 22:259-260.

Magarvey, R. H., Taylor, B. W. (1956). Apparatus for the production of large water drops. Rev. Sci. Instrum. 27:944-947.

Miller, J. N., Miller, J. C. (2005). *Statistics and Chemometrics for Analytical Chemistry* (5th Ed.). Pearson Education Limited, Harlow, UK, pp. 24-25, 54-61, 142-144.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., Fraser, C. M. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 399:323-329.

Peters, T. M., Leith, D. (2003). Concentration measurement and counting efficiency of the aerodynamic particle sizer 3321. J. Aerosol Sci. 34:627-634.

Raabe, O. G. (1976). The Generation of Aerosols of Fine Particles, in *Fine Particles: Aerosol Generation, Measurement, Sampling, and Analysis*, B. Y. H. Liu, ed, Academic Press, New York, NY, pp. 61-63.

Rubinson, K. A., Rubinson, J. F. (2000). *Contemporary Instrumental Analysis.* Prentice Hall, Upper Saddle River, NJ, pp. 373-374, 617.

Sambrook, J., Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual* (3rd Ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 8.4-8.24, 8.86-8.95.

Sangiovanni, J. J., Labowsky, M. (1982). Burning times of linear fuel droplet arrays: a comparison of experiment and theory. Combust. Flame 47:15-30.

Savitzky, A., Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627-1639.

Sergeyev, A. V., Shaw, R. A. (2006). An inexpensive uniform-size aerosol generator. Meas. Sci. Technol. 17:N41-N44.

Seymour, R. J., Boss, C. B. (1983). Design modification for a uniform droplet generator system. Appl. Spectrosc. 37:375-379.

Shelley, D. J., Majewski, J. T., Thackray, M. R., McWilliams, J. L. (1997). A lower-cost inkjet printer based on a new printing performance architecture. Hewlett-Packard J. 48:6-11.

Skoog, D. A., Holler, F. J., Nieman, T. A. (1998). *Principles of Instrumental Analysis* (5th Ed.). Brooks/Cole Thomson Learning, Pacific Grove, CA, pp. 201-202, 549-553, 688.

Stemme, E., Larsson, S. G. (1973). The piezoelectric capillary injector- a new hydrodynamic method for dot pattern generation. IEEE T. Electron Dev. 20:14-19.

Stricker, J., Sofer, D. (1991). Monosize droplet stream generator. Rev. Sci. Instrum. 62:3047-3050.

Sweet, R. G. (1965). High frequency recording with electrostatically deflected ink jets. Rev. Sci. Instrum. 36:131-136.

Switzer, G. L. (1991). A versatile system for stable generation of uniform droplets. Rev. Sci. Instrum. 62:2765-2771.

Threadgill, E. D., Williamson, R. E., Miles, G. E. (1974). Development of controlled-size droplet generators. T. ASAE 17:837-840.

Trunk, M., Lankers, M., Popp, J., Kiefer, W. (1994). Simple and inexpensive design for a uniform-size droplet generator. Appl. Spectrosc. 48:1291-1293.

TSI Incorporated (2004). Model 3321 aerodynamic particle sizer spectrometer. P/N 1930087 Rev. C. www.tsi.com.

Underwood, D. M., Herron, D. L., Croisant, W. J. (2007). Whole-building dispersion of tracer gas after internal release in an administrative/classroom building. ASHRAE Tran. 113:PT 2.

Wang, Y., Bokor, J., Lee, A. (2004). Maskless lithography using drop-on-demand inkjet printing method. P. Soc. Photo-Opt. Ins. 5374:628-636.

Warnica, W. D., Van Reenen, M., Renksizbulut, M., Strong, A. B. (1991). A piezoelectric droplet generator for use in wind tunnels. Rev. Sci. Instrum. 62:3037-3046.

Yang, J. C., Chien, W., King, M., Grosshandler, W. L. (1997). A simple piezoelectric droplet generator. Exp. Fluids 23:445-447.

CHAPTER 3: CHEMICAL PROFILING OF LATENT FINGERPRINT RESIDUES USING SOLID-PHASE MICROEXTRACTION WITH GAS CHROMATOGRAPHY-MASS SPECTROMETRY ANALYSIS

3.1 Motivations and Introduction

Objects and locations connected with nearly every type of crime are routinely examined for latent fingerprint evidence. However, fingerprints are often smudged or overlapping, and it is not possible using current methods to determine how long a fingerprint has been on a surface. Such information would aid in establishing a crime's timeline and whether the fingerprint is relevant to the investigation at hand. There is a need for a portable, non-destructive method to gain information from latent fingerprint residue when the friction ridge detail is obscured, yet remains compatible with traditional forensic analyses. Here we demonstrate chemical profiling of latent fingerprint residues using solid-phase microextraction (SPME) headspace sampling coupled to gas chromatography-mass spectrometry (GC-MS) detection in order to determine subject traits and prior activities, as well as to determine the time window since the fingerprint was deposited, while leaving the fingerprint intact for traditional forensic examinations.

3.1.1 Fingerprint Compounds

Fingerprints are among the oldest and most important evidence categories in forensic science (Gaensslen and Young 2003). They remain one of the most commonly gathered types of forensic evidence today and are considered a fundamental tool for identifying people with criminal histories in nearly every police agency worldwide. The friction ridge skin on the fingertips has pores through which eccrine sweat glands secrete their contents, and typically

sebaceous gland secretions from the face and scalp are present as well due to touching of the face and hair. While human sweat is approximately 99% water, it also contains both inorganic salts (from eccrine glands on the palms) and organic compounds (from sebaceous glands on the face and scalp) including proteins, amino acids, nucleic acids, lipids, sugars, vitamins, and organic acids (Mong *et al.* 1999, Ramotowski 2001, Bernier *et al*. 2000). While the composition of human sweat is well understood, latent fingerprint residues are more complex due to the presence of exogenous contaminants (*e.g.* personal care products, cosmetics, food residues, etc.) (Mong *et al.* 1999; Ramotowski 2001). An almost unlimited variety of substances from the environment, unique for every person's recent exposure history, can be retained on the friction ridge skin and be deposited into latent fingerprint residue when the fingers touch a surface. Thus, a number of common endogenous constituents of most latent print residues are based on the composition of sweat, however, many individualizing compounds and materials can be present in a latent fingerprint as well (Mong *et al.* 1999, Gaensslen and Young 2003).

While the value of using the unique friction ridge patterns in the fingerprint for human identification has been recognized for centuries (Gaensslen and Young 2003), the chemical profile of the residue is just beginning to gain recognition for providing additional information about an individual and/or the individual's prior activities. For example, the chemical composition of children's latent fingerprint residue is markedly different compared to that of adults as the lipid content in sweat increases after puberty, allowing determination of the age-range of the subject (Noble 1995, Buchanan *et al*. 1996, Antoine *et al*. 2010). Differences between individuals' endogenous fingerprint compound levels have been observed, and it is thought that these slight inter-personal variations in sebaceous fatty acid mixture yield unique

individualizing scents, which is the basis for canine tracking (Nicolaides 1974, Knowles 1978, Ramotowski 2001). As a result, several groups have demonstrated first steps toward using human scent profiles as a biometric (Mong *et al*. 1999; Zhang *et al*. 2005, Curran *et al*. 2007, Curran *et al.* 2010). Exogenous compounds from an individual's environment are also present in fingerprints, and the combination of exogenous and endogenous compounds detectable in a latent fingerprint may give clues about personal traits, such as age, habits, activities, gender, and disease state (Buchanan *et al*. 1996). The secretions and dead cells in a human fingerprint also contain evidence of ingested substances and their metabolites (*e.g.* pharmaceuticals, illegal drugs), demonstrating a non-invasive method for sampling the human body that is of interest to law enforcement and forensic personnel (Johnson and Maibach 1971, Naitoh *et al*. 2000). Amphetamines and their metabolites are excreted in human sweat (Vree *et al*. 1972), and nicotine has been detected in latent fingerprint residue (Buchanan *et al*. 1996). Antibody tags have been used to detect individuals' prior cigarette and marijuana smoking, as well as methadone, heroin, and cocaine use in latent fingerprint residue as well (Leggett *et al*. 2007, Hazarika *et al.* 2008, Hazarika *et al*. 2009, Hazarika *et al*. 2010). While useful for proof that these substances may be detected in latent fingerprint residue, the disadvantage in these approaches lies in their inability to profile many substances simultaneously in a non-targeted manner. The changes in response over time were not examined as well. The possibility of associating and discriminating individuals based on latent fingerprint residue chemical profiles has also not been previously examined. These studies show a need for evaluating the full chemical profiles of latent fingerprint residue for homeland security purposes.

3.1.2 Solid-Phase Microextraction

Solid-phase microextraction (SPME) provides a powerful approach to collect and concentrate volatile and semi-volatile organic compounds for subsequent analysis in a variety of applications (Mills and Walker 2000, Martin *et al.* 2010). The collected chemicals are then desorbed in the heated inlet of a gas chromatograph (GC), where a temperature program separates the chemicals for detection using mass spectrometry (MS), which yields structural information for chemical identification. The combination of SPME headspace sampling with GC-MS analysis has not, to the authors' knowledge, been previously applied to latent fingerprint residue analysis, and is expected to yield information-rich chemical profiles of the volatile components contained in latent fingerprint residues. In addition, SPME headspace sampling is non-invasive, leaving the latent fingerprint intact for further processing with traditional forensic methods such as powder dusting, photography, and even DNA profiling (Schulz and Reichert 2002). Chemical profiling using SPME-GC-MS also has potential to provide useful information from smudged or partial latent fingerprints where individual identification using friction ridge pattern analysis is not possible (Asano *et al*. 2002).

3.1.3 Fingerprint Changes Over Time

Fingerprints degrade over time due to exposure to light and heat, and the outgassing of volatile components, most notably water (Wargacki *et al*. 2008, De Paoli *et al*. 2010). Mong and coworkers (1999) observed substantial oxidative degradation of squalene and other unsaturated compounds such as wax esters and fatty acids over time compared to their saturated analogs. It has also been determined that the decrease in abundance of volatile fatty

acids and other lipid compounds over time occurs at different rates in latent fingerprint residues from children and adults, and different temperature and humidity storage conditions affect degradation rates as well (Noble 1995, Buchanan *et al*. 1996, Antoine *et al*. 2010). Weyermann and coworkers (2011) observed similar relationships, and normalized chemical abundance to squalene and/or cholesterol in an effort to reduce the large degree of variability, both inter- and intra-subject, observed in the collected latent residues. In addition to outgassing, oxidative and bacterial degradation may take place in the deposited fingerprint material over time as well, generating degradation products (Ramotowski 2001). A better understanding of these alterations in latent fingerprint residues over time would aid the development of improved visualization reagents and potentially lead to methods for determining the age of the fingerprint. Knowing the time-window since the fingerprint was deposited would help to determine the crime's timeline and to assess the relevance of the fingerprint evidence to the investigation.

This study fully examined the chemical profiles present in a latent fingerprint residue using non-destructive SPME headspace sampling and GC-MS analysis. To the authors' knowledge, this study is the first to apply SPME headspace sampling to latent fingerprint residues. Multivariate statistical analyses associated and discriminated five subjects based on their fingerprint residue chemical profiles up to three days post-deposition. A second latent fingerprint sample collected from one of the subjects five months after the original study was also strongly associated with the initial sample using principal component analysis (PCA). The changes in the chemical profiles over 30 days also show promise for dating the fingerprints.

3.2 Materials and Methods

3.2.1 Fingerprint Deposition for Headspace Sampling

Five adult subjects (3 males and 2 females) donated fingerprints in accordance with an approved Lawrence Livermore National Laboratory Institutional Review Board human subjects protocol. The subject's activities prior to sample donation were not controlled. All subjects donated five fingerprints, one from each finger of the right hand, by pressing each finger individually on a different area of the glass surface inside of a wide-mouth septa jar (short bottle style, 60 mL capacity, Thermo Fisher Scientific, I-Chem Brand, Rockwood, TN) for five seconds (fingerprints not overlapping). Septa jars containing fingerprint samples, as well as one blank jar with no fingerprints as a control, were stored at room temperature (~21 °C) and humidity (~50% relative humidity) under continuous fluorescent lighting for the duration of the study. The septa jar lids were sealed during SPME sampling time periods (fingerprint residues sampled through the septum in the lid) and the lids were removed during degradation time periods to expose the latent fingerprint residue samples to the ambient atmosphere. Five months after the initial study, one subject donated a second latent fingerprint residue sample in the same manner as the first. The fingerprint sample was stored, sampled, and analyzed using identical procedures to determine whether the second fingerprint sample would be associated with the first sample using data analyses.

3.2.2 Headspace SPME Sampling

Volatile and semi-volatile compounds inside the septa jars were passively sampled using SPME Portable Field Samplers with 65 μm polydimethylsiloxane/divinylbenzene (PDMS/DVB)

fibers (Supelco, St. Louis, MO). The SPME fibers were thermally conditioned in the GC injection port per manufacturer's instructions prior to sample collection, and it was verified that the method left the fibers free of carry-over from previous samples (data not shown). The PDMS/DVB fiber was chosen for its ability to efficiently collect a broad range of compounds in air, from volatile (general-purpose PDMS) to semi-volatile and larger volatiles (DVB), for analysis. The SPME fibers were inserted through the septa in the lids of the closed septa jars and were exposed to the headspace inside the septa jars for 16 h at room temperature. At the end of the sampling time period the fibers were retracted and subsequently analyzed using GC-MS. In most cases, the SPME fibers were analyzed directly after sample collection. However, if analysis was delayed, the SPME fibers were stored individually in sealed containers at -20 °C until GC-MS analysis. The SPME sampling time was not optimized for this work, and 16 h was chosen after a review of the literature, for convenience, and to ensure the fiber was saturated with volatile and semi-volatile chemical sample by diffusion at room temperature (Curran *et al.* 2007). An optimized SPME sampling method for latent fingerprint residues is the subject of ongoing research.

As SPME headspace sampling is non-destructive, the same latent fingerprint residues (and control sample with no fingerprints) were sampled and analyzed over 30 days to assess chemical profile changes over time. The first SPME sample (Day 0) was collected directly after fingerprint residue collection in the septa jars without exposure to ambient conditions. After the 16 h SPME sample collection was completed, the lids of the septa jars were removed and the latent fingerprint residues were exposed to ambient conditions for 4 h to simulate evidentiary sample exposure and degradation. After the degradation time period was over, the

lids to the septa jars were replaced and the next SPME sampling time point was collected. During times when neither SPME sampling nor degradation exposure was taking place, the septa jars remained sealed. The length of time that the latent fingerprint residue samples were exposed to ambient conditions between SPME samplings increased as the degradation study progressed, as chemical profile changes were occurring more gradually compared to the more rapid changes observed during the first few days of the study. Table 3.1 lists a summary of SPME sampling time points and the cumulative time that the latent fingerprint residue samples were exposed to ambient conditions during the course of the 30-day study.

Table 3.1. Latent fingerprint residue SPME headspace sampling times during the 30-day study.

Day	Cumulative Time Post-Deposition (h)	Cumulative Exposure Time to Ambient Conditions (h)
0	0	0
1	24	4
2	48	8
3	72	12
5	120	40
7	168	72
10	240	128
13	312	184
17	408	264
21	504	344
24	576	406
30	720	528

3.2.3 GC-MS Analysis of Fingerprint Compounds

Volatile and semi-volatile chemicals adsorbed onto the SPME fiber were directly analyzed using an Agilent 6890 GC with 5973 MSD system and ChemStation software (Agilent Technologies, Inc., Santa Clara, CA) with manual injection and thermal desorption. The capillary

column was 0.25 mm x 30 m coated with 0.25 μm DB5-MS (5% phenyl methyl siloxane, Agilent Technologies), and the following conditions were employed: helium flow rate of 1.2 mL/min, inlet at 300 °C, column at 50 °C for 1 min, followed by a 10 °C/min ramp to 300 °C, held for 2 min, transfer line at 300 °C, 70 eV electron ionization, full scan acquisition from *m/z* 40 to 550, and autotuning and mass calibration performed using perfluorotributylamine every day prior to sample analyses. ChemStation GC-MS total ion chromatogram datafiles were exported as comma-separated values (.csv) files for analyses in Microsoft Office Excel (Redmond, WA). ChemStation datafiles were also converted to NetCDF format files and then to Waters MassLynx file format (Milford, MA) for statistical analyses. Compounds were tentatively identified by comparing the mass spectra with the NIST 2008 Standard Reference Database of mass spectra (RMatch > 750, NIST MS Search 2.0, National Institute of Standards and Technology, Gaithersburg, MD) as well as mass spectra published in the literature.

3.2.4 Chemical Profile Data Analyses

The latent fingerprint residue chemical profiles obtained from headspace SPME-GC-MS analyses contained numerous compounds at varying abundances indicative of both subject and time point, yielding a multivariate data set. This multivariate data set was examined using two different statistical data analysis strategies: pairwise sample comparison using non-parametric Spearman rank correlation coefficient analyses, and simultaneous comparison of all samples using principal component analysis (PCA). The Spearman correlation coefficients were used to evaluate the similarity between two samples, while PCA examined the differences among all samples and provided graphical visualization tools that highlighted those differences to

facilitate interpretation. Therefore, the two techniques are complementary (Miller and Miller 2005).

Spearman rank correlation coefficients were calculated between all five subjects' latent fingerprint residue samples and the blank sample collected on day 1 post-deposition. Each sample total ion chromatogram resulting from GC-MS analysis was exported into Excel, where the measured abundances (peak heights) at each retention time were numerically ranked in descending order within the sample (*i.e.*, the most abundant retention time given a rank of 1, the second most abundant retention time given a rank of 2, and so on). Abundant siloxane contaminant peaks from the septa in the septa jars were excluded. The rank given to each retention time was then compared between the two samples being investigated in order to yield the correlation coefficient describing the similarity of the samples. All possible combinations of the six samples (five containing latent fingerprint residue and one blank) were investigated, yielding 15 correlation coefficients describing the ability to associate and discriminate the samples using the raw SPME-GC-MS data.

The SPME-GC-MS data sets were then pre-processed using Waters MarkerLynx software (version 4.1) prior to additional univariate and multivariate statistical analyses. Again, the siloxane contaminant peaks present in all samples were excluded from further analysis. Chromatograms from each subject at each sampling time were pre-processed using peak detection, integration, and retention time alignment with noise reduction. This process generated a list of peak areas labeled using both retention time (RT) and *m/z*, known as an RT-*m/z* pair or marker. Markers not detected or with peak areas below the set cutoff value in a sample were assigned a value of zero in the matrix. The resulting peak area matrix was

exported into Excel and was used to examine abundance changes for individual fingerprint compounds over the time course of the experiment (30 days). This data matrix was not normalized, as chemical components generally decreased in abundance over time.

In order to compare the chemical profiles from all subjects using multivariate statistical procedures, the aligned peak area matrix was further processed using the MarkerLynx software. The peak areas within each sample were normalized to the sum of the peak areas in that sample, which was arbitrarily set to 10000 for all samples. The normalized data matrix was then exported into SIMCA (software version 13, Umetrics, San Jose, CA) for further pre-treatment and multivariate statistical analyses. The data matrix exported from MarkerLynx was mean-centered and scaled to Pareto variance to mitigate the high loading of abundant compounds in the mathematical models. Finally, principal component analysis (PCA) was used to assess the relationships between fingerprint samples donated by different subjects and blank samples analyzed shortly after deposition (0-3 days post-deposition). Samples with similar chemical composition were clustered together in the PCA scores plot, while dissimilar samples were spatially distant. The association and discrimination of samples visualized in the PCA scores plot were interpreted using the associated loadings plot of the individual chemical components varying the most in the data set (Miller and Miller 2005).

3.3 Latent Fingerprint Residue Chemical Profiles from Five Subjects

Chemical profiles of latent fingerprint residue samples containing fingerprints from five volunteers were passively collected using headspace SPME and analyzed using GC-MS over a 30-day time course. As headspace SPME sampling is non-destructive, the same five fingerprint

samples (as well as a blank control) were re-analyzed throughout the study, eliminating intra-subject fingerprint deposition variability effects that have plagued previous studies (*e.g.* Mong *et al*. 1999). However, as subject activities prior to fingerprint sample donation were not controlled as a part of this study, intra-subject differences were expected in the resulting volatile and semi-volatile chemical profiles. Total ion chromatograms from GC-MS analysis of the SPME samples collected one day after the latent fingerprint residue samples were deposited are presented in Figure 3.1. Chemical profiles from each subject, as well as the blank jar profile of the ambient environment and instrumental background, are plotted on the same abundance scale for comparison. Numerous volatile and semi-volatile chemical species present in latent fingerprint residues were successfully collected using SPME and desorbed and analyzed using GC-MS. The number of peaks above background per sample ranged between approximately 350 for Subject 36 to approximately 500 for Subject 16. Tentative compound identifications by comparing the mass spectra to the NIST Mass Spectral Library are compiled in Table 3.2.

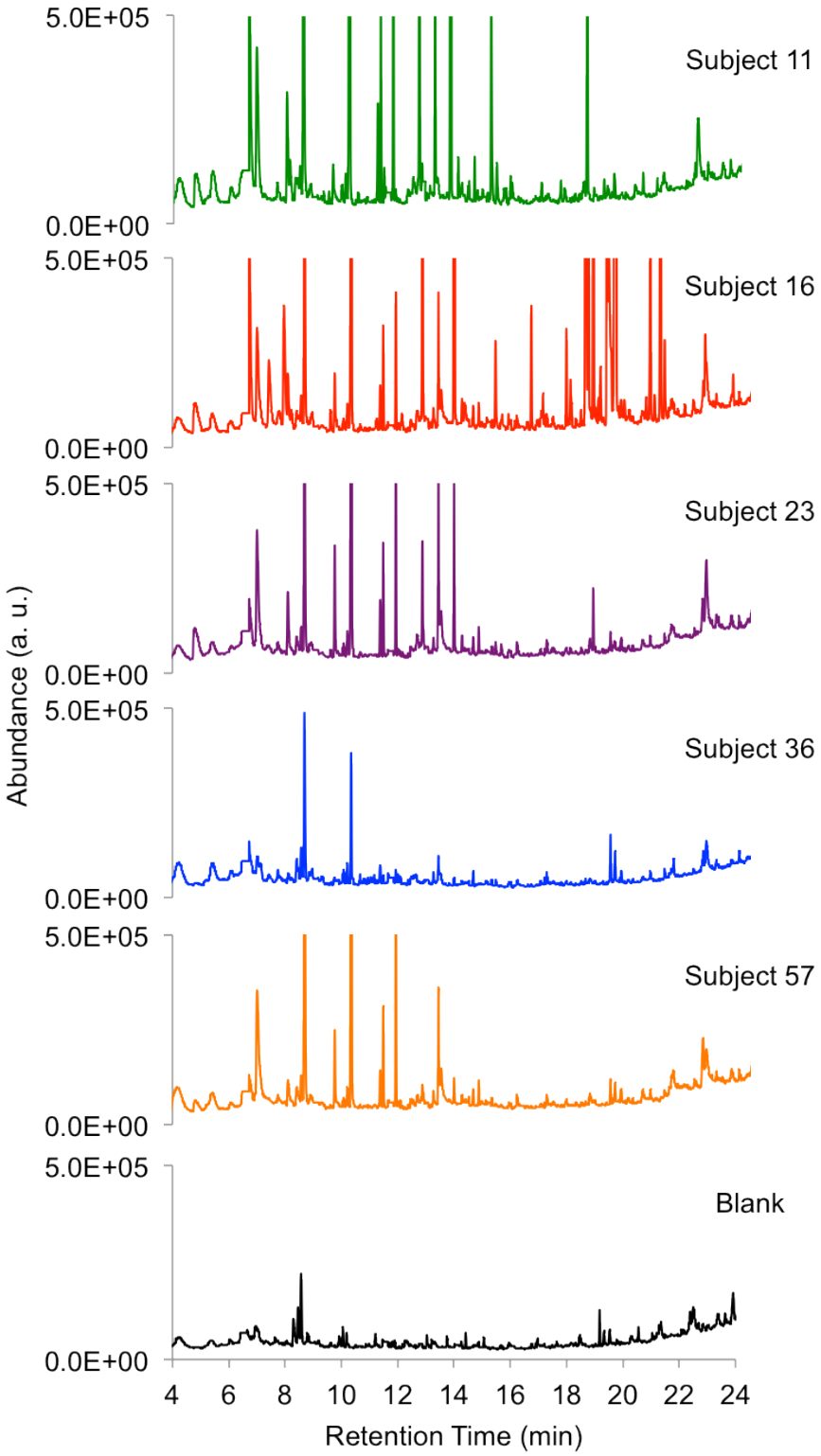

Figure 3.1. Total ion chromatograms from SPME-GC-MS analyses of latent fingerprint residues donated by five subjects sampled one day after deposition (plotted on the same abundance scale for comparison). The siloxane peaks resulting from the SPME fiber coating and septa jar septum were removed.

Also evident from Figure 3.1 are both similarities and differences in the SPME-GC-MS chemical profiles from the five subjects' latent fingerprints. Chromatogram abundances varied over two orders of magnitude, from approximately 500,000 a.u. in Subject 36's profile to approximately 25,000,000 a.u. in Subject 16's chemical profile. This was likely due to inter-subject variability in the amount of latent fingerprint residue deposited, partially caused by uncontrolled activity prior to fingerprint sample donation. It was observed that Subject 36 washed her hands just prior to donating fingerprints, so it follows that Subject 36's chemical profile contained fewer compounds and lower abundance compounds compared to the other subjects that did not wash their hands right before fingerprint sample donation. Despite prior activity variation, some compounds are present in all subject chemical profiles, such as nonanal (8.57 min) and decanal (10.18 min). Compounds detected in all subjects' latent fingerprint residues are most likely endogenous compounds from the human body. Varying abundances of endogenous compounds were deposited by each subject, which reflects inter-personal variability in both the amount of fingerprint residue deposited and also the ratios of the different compounds within the residue, consistent with previous observations by other groups (Nicolaides 1974, Mong *et al*. 1999, Bernier *et al*. 2000). Confirming the presence of these compounds in a questioned residue may be useful for determining that the residue is indeed a latent fingerprint and should be examined more closely. Other compounds are only present in one subject's fingerprints, such as those eluting between 18 and 20 minutes in Subject 16's chemical profile. Individualizing compounds are useful for subject habit and prior activity information. Certain chemicals may indicate subject age, gender, habits, and/or prior activities,

providing more information from the latent fingerprint residue evidence than is accessible using current non-destructive analysis methods.

Table 3.2. Annotated fingerprint compounds with corresponding GC retention times in SPME-GC-MS analyses of latent fingerprint residues from five subjects, with the number of subjects whose fingerprint chemical profiles contained the compounds indicated as well.

Peak	Compound	Chemical Class	RT (min)	Number of Subjects
1	Butanone	Aliphatic ketone	4.75	4
2	6-Methyl-5-hepten-2-one	Branched aliphatic ketone	6.67	5
3	Octanal	Aliphatic aldehyde	6.93	4
4	2-Ethylhexanol	Branched aliphatic alcohol	7.35	1
5	2,5-Dimethyl-2,5-hexanediol	Alcohol	7.84	1
6	2-Nonen-1-ol	Aliphatic alcohol	7.99	4
7	Unidentified compound		8.36	4
8	Nonanal	Aliphatic aldehyde	8.57	5
9	2-Nonenal	Aliphatic aldehyde	9.47	4
10	2-Decen-1-ol	Aliphatic alcohol	9.62	4
11	Menthol	Monoterpenoid alcohol	9.84	1
12	Decanal	Aliphatic aldehyde	10.18	5
13	Phenoxyethanol	Aromatic alcohol	10.49	5
14	4-Methoxybenzaldehyde	Aromatic aldehyde	11.06	3
15	2-Undecen-1-ol	Aliphatic alcohol	11.3	5
16	Hydroxycitronellal	Monoterpenoid aldehyde	11.45	2
17	Undecanal	Aliphatic aldehyde	11.73	5
18	Limonene diepoxide or unidentified isomer	Other	12.66	4
19	Dodecanal	Aliphatic aldehyde	13.21	5
20	6,10-Dimethyl-5,9-undecadien-2-one (Geranylacetone)	Branched aliphatic ketone	13.75	5
21	Methylparaben	Aromatic ester	14.07	4
22	Dodecanol	Aliphatic alcohol	14.14	2
23	Butylated hydroxytoluene	Phenolic antioxidant	14.54	1
24	Tridecanal	Aliphatic aldehyde	14.6	4
25	Phenoxyethyl isobutyrate	Aromatic ester	14.7	4
26	Lilial	Aldehyde	14.92	3

Table 3.2. (cont'd)

Peak	Compound	Chemical Class	RT (min)	Number of Subjects
27	9-(3,3-Dimethyloxiran-2-yl)-2,7-dimethylnona-2,6-dien-1-ol	Alcohol	15.19	4
28	δ-Undecalactone	Aliphatic lactone	15.48	1
29	Tetradecanal	Aliphatic aldehyde	15.93	4
30	Methyl dihydrojasmonate or unidentified isomer	Aliphatic ester	16.43	1
31	Vanillin isobutyrate	Phenolic ester	16.59	1
32	Methyl dihydrojasmonate or unidentified isomer	Aliphatic ester	16.79	1
33	Hexyl salicylate	Phenolic ester	16.83	2
34	Patchouli alcohol	Alcohol	16.97	5
35	Hexyl cinnamaldehyde	Aromatic aldehyde	17.64	4
36	Unidentified compound		17.73	1
37	Lilac alcohol	Alcohol	17.78	3
38	Cubenol	Alcohol	18	1
39	Octyl salicylate	Phenolic ester	18.34	1
40	Isopropyl myristate	Fatty acid ester	18.42	2
41	5,9,13-Trimethyl-4,8,12-tetradecatrienal	Terpenoid aldehyde	18.58	4
42	2-Hydroxycyclopentadecanone	Macrocyclic keto-alcohol	18.75	1
43	Musk 36A (CAS Number 88-29-9)	Ketone	18.83	2
44	Homosalate or unidentified isomer	Ester	19.06	1
45	Palmityl alcohol	Aliphatic alcohol	19.15	1
46	Homosalate or unidentified isomer	Ester	19.33	1
47	Musk ambrette	Ketone	19.83	1
48	Isopropyl palmitate	Fatty acid ester	20.56	1
49	Musk T (CAS Number 105-95-3)	Ketone	20.7	1
50	Oxybenzone	Ketone	20.9	1
51	N-Acetylserotonin	Indole amide	21.93	1
52	Octinoxate or unidentified isomer	Ester	22.04	1
53	Methyl podocarpa-8(14),9(11),12-trien-15-oate	Ester	22.32	1
54	Octinoxate or unidentified isomer	Ester	23.53	1

Volatile and semi-volatile compounds from several chemical groups, including ketones, aldehydes, alcohols, carboxylic acids, and esters, composed latent fingerprint residue chemical

profiles detected using SPME-GC-MS (Table 3.2). The abundant peaks present in all subject samples were aldehydes, with chain lengths of C_8-C_{14} detected as well as some unsaturated and aromatic compounds. Nonanal, decanal, undecanal, and dodecanal were present in all five subject's profiles above background and have been previously reported as constituents of human emanations (Bernier *et al.* 2000). Aldehydes are formed endogenously by peroxidation of lipids that are abundant on the surface of human skin (O'Brien *et al.* 2005). Some more complex aldehydes, such as hydroxycitronellal, lilial, and hexyl cinnamaldehyde are exogenous scent compounds, and 5,9,13-trimethyl-4,8,12-tetradecatrienal is potentially derived from the synthesis or degradation of squalene, the endogenous biosynthesic precursor to cholesterol and steroids (Nicolaides 1974).

The majority of detected carboxylic acid derivatives were the more volatile esters including isopropyl myristate and isopropyl palmitate. Free fatty acids are generated endogenously and also result from bacterial hydrolysis of triglycerides on the skin (Puhvel *et al.* 1975). Hexyl salicylate, vanillin isobutyrate, and methyl dihydrojasmonate were exogenous personal care product scent additives detected. The abundant peaks unique to Subject 16's chemical profile that eluted between 18 and 20 minutes in Figure 1 were identified as homosalate and octyl salicylate, which are sunscreens. These compounds illustrate the potential of SPME-GC-MS latent fingerprint residue chemical profiling to reveal information about a subject's habits and prior activities.

Detected alcohols included several short-chain, branched, and unsaturated compounds, some with combinations of functional groups. Identified fatty alcohols included C_{12} and C_{16}

saturated alcohols as well as C_9-C_{11} fatty alcohols with unsaturations in the 2-position. As C_{12}-C_{16} alcohols are common constituents of skin lotions, these were likely exogenous compounds. Exogenous scent alcohols commonly added to personal care products, including menthol, lilac alcohol, and patchouli alcohol, were present, as well as another possible precursor/degradation product of squalene, putatively assigned as 9-(3,3-dimethyloxiran-2-yl)-2,7-dimethylnona-2,6-dien-1-ol.

A few ketones were observed, including butanone and 6-methyl-5-hepten-2-one, which have been reported previously (Bernier *et al.* 2000). Geranylacetone, another likely squalene oxidation product, was tentatively identified. Another sunscreen compound, oxybenzone, as well as several musk scent additives and peach lactone were exogenous personal care product additives detected as well from subject latent fingerprint residues using SPME-GC-MS.

The most probable origins (endogenous or exogenous) of the compounds detected in the latent fingerprint residues should be interpreted with care, as some endogenous compounds are also contained in exogenous sources, causing additive effects in compound abundance that are difficult to discriminate (Bernier *et al.* 2000). As the diet and prior activities of subjects in this study were uncontrolled, latent fingerprint residue chemical components of both endogenous and exogenous origin were expected. Future work with more controlled sample collection will aid the interpretation of sources of compounds detectable in latent fingerprint residues using SPME sampling followed by GC-MS analysis.

In summary, these results show that non-destructive SPME sampling collects many endogenous and exogenous volatile compounds that yield information-rich chemical profiles.

These latent fingerprint residue chemical profiles provide a new level of information from the most common type of forensic evidence while leaving the fingerprint undisturbed for traditional forensic processing and interpretation. The methods used in this study simulate discovering latent fingerprint residues using oblique lighting or other techniques that do not modify the latent fingerprint to enhance visualization. Future work will examine whether common fingerprint residue visualization methods, including powder dusting and cyanoacrylate fuming, interfere with subsequent SPME headspace sampling.

3.4 Subject Discrimination Using Spearman Rank Correlation Analysis of Fingerprint Chemical Profiles

The SPME-GC-MS chemical profiles from the blank sample and all five subjects' latent fingerprint residue samples were compared pair-wise to determine if they could be differentiated using the numerous compounds detectable in fingerprints. Pair-wise comparisons would be used to associate a suspect's fingerprint residue to an evidentiary fingerprint from a crime scene. The two most common pair-wise correlation coefficients are the Pearson product-moment correlation coefficient (PPMCC) and Spearman's rank correlation coefficient (SRCC). The coefficient calculations are the same (covariance of the two samples divided by the product of their standard deviations), however, the input for PPMCC is the raw data (peak heights in this case), and the input for SRCC is the ranked data (most abundant peak height is given a rank of 1, the second most abundant peak height given a rank of 2, and so on in order of decreasing peak height). Calculating coefficients using ranked data (SRCC) does not assume that the data are normally distributed, and as a result reveals all correlations that are

monotonic and useful for data interpretation. Conversely, PPMCC calculations are limited to normally distributed data sets and only consider linear correlations, which are too restrictive for exploratory analyses. Therefore, the raw GC-MS total ion chromatograms from latent fingerprint residue samples collected on day one post-deposition (truncated to the most variable region eluting between 4 and 24 minutes as shown in Figure 3.1) were compared using SRCC analyses (Miller and Miller 2005, Curran *et al.* 2007, Curran *et al*. 2010). Chromatogram peak heights in each latent fingerprint sample at each measured retention time were ranked numerically in order of abundance within the sample, and then the rank given to each retention time was compared between the two samples being analyzed in order to yield the correlation coefficient. Correlation coefficients may vary between -1 and 1, with values close to 1 indicating a strong positive correlation, values close to 0 indicating minimal correlation, and values close to -1 indicating a strong negative correlation (Miller and Miller 2005). All possible pair-wise comparisons were examined, yielding 15 correlation coefficients from this data set that are listed in Table 3.3. The correlation coefficients ranged from 0.521 to 0.913, indicating positive linear correlations that increase in significance as the coefficients approach 1 (perfect linear correlation). Correlation coefficient thresholds of 0.8 and 0.9 were tested for associating and discriminating samples in this small data set (*i.e.,* if the correlation coefficient is larger than 0.9, the samples cannot be differentiated at the 0.9 correlation threshold level) (Curran *et al.* 2007).

Table 3.3. Summary of Spearman's rank correlation coefficients comparing SPME-GC-MS chemical profiles of the blank and subjects' latent fingerprint residues on day one post-deposition. **Bold** indicates samples not distinguished at a correlation threshold of 0.9. *Italics* indicate samples not distinguished at a correlation threshold of 0.8.

	Subject 11	Subject 16	Subject 23	Subject 36	Subject 57
Blank	0.685	0.603	0.775	**0.906**	*0.832*
Subject 11		0.762	*0.899*	0.691	*0.839*
Subject 16			*0.809*	0.521	0.780
Subject 23				0.741	**0.913**
Subject 36					0.765

The blank sample was compared with all five subjects' latent fingerprint residue samples and was successfully differentiated from Subjects 11, 16, and 23 (correlation coefficients less than 0.8 in Table 3.3). The blank and Subject 36 could not be differentiated at the 0.9 correlation threshold, and the blank and Subject 57 could not be differentiated at the 0.8 correlation threshold. As observed from Figure 3.1, Subject 36 deposited a small amount of fingerprint material due to washing her hands prior to sampling, so it follows that the chemical profile is more correlated with the blank sample and successfully differentiated from all other latent fingerprint residue samples. Considering only the correlations between the different subjects (excluding comparisons with the blank sample), 90% of the pairs were distinguished at a correlation threshold of 0.9, and 70% of the pairs were distinguished at the 0.8 correlation threshold. Subject 16 was successfully differentiated from the blank and all subjects except for Subject 23 at the 0.8 correlation threshold, owing to the presence of individualizing compounds in the chromatogram. Subjects 11, 23, and 57 could not be differentiated from two of the other subjects at the 0.8 correlation threshold owing to the presence of the same endogenous fingerprint compounds present in each sample, but they could be differentiated from some subjects due to differing abundance ratios and the absence of some compounds. In summary,

these results illustrate the potential utility of comprehensive analysis of all compounds present in an individual's latent fingerprint residue chemical profile for subject association and discrimination.

3.5 Subject Association and Discrimination Using PCA of Fingerprint Chemical Profiles

The informative headspace SPME-GC-MS volatile chemical profiles from the five subjects' latent fingerprint residues were compared to one another and the blank jar sample using principal component analysis (PCA). The GC-MS chemical profiles from all subjects and the blank at all time points were combined into a single data matrix for preprocessing (peak integration, alignment, normalization, and scaling) followed by PCA to visualize sample association and discrimination based on analysis of the entire chemical profile (Miller and Miller 2005). The resulting PCA scores plot is presented in Figure 3.2, where each data point on the plot represents a sample at a certain time point. This PCA analysis only included samples analyzed on days 0, 1, 2, and 3 after the latent fingerprint residue samples were deposited, as later time points exhibited an appreciable level of degradation and resulting loss of chemical signals above background. The *x*-axis represents principal component 1 (PC1) and is derived from a combination of chemical signals in the data set that vary in a similar manner and explain the largest contributions to the variation within the entire data set. In this analysis, 21% of the total variance in the data set is explained by PC1. Similarly, the *y*-axis represents PC2, which describes a second combination of chemical signals in the data set that vary similarly but are unrelated to the signals in PC1 and describe the second-largest contributions to the variation within the data set. The fingerprint chemical profile variance described by PC2 in this analysis

was 15%, so the scores plot visualizes how all of the subject samples relate to one another with 36% of the total variance in the data set described by a single plot (Figure 3.2). Samples that are positioned close together on the scores plot are chemically similar, while samples that are spatially distant are chemically different. The individual chemical signals varying the most in the data set are displayed in a corresponding loadings plot that facilitates the interpretation of the scores plot (not shown). Subsets of highly loading compounds (chemical signals varying the most in the data set) were identified, and representative examples are presented in Figure 3.3.

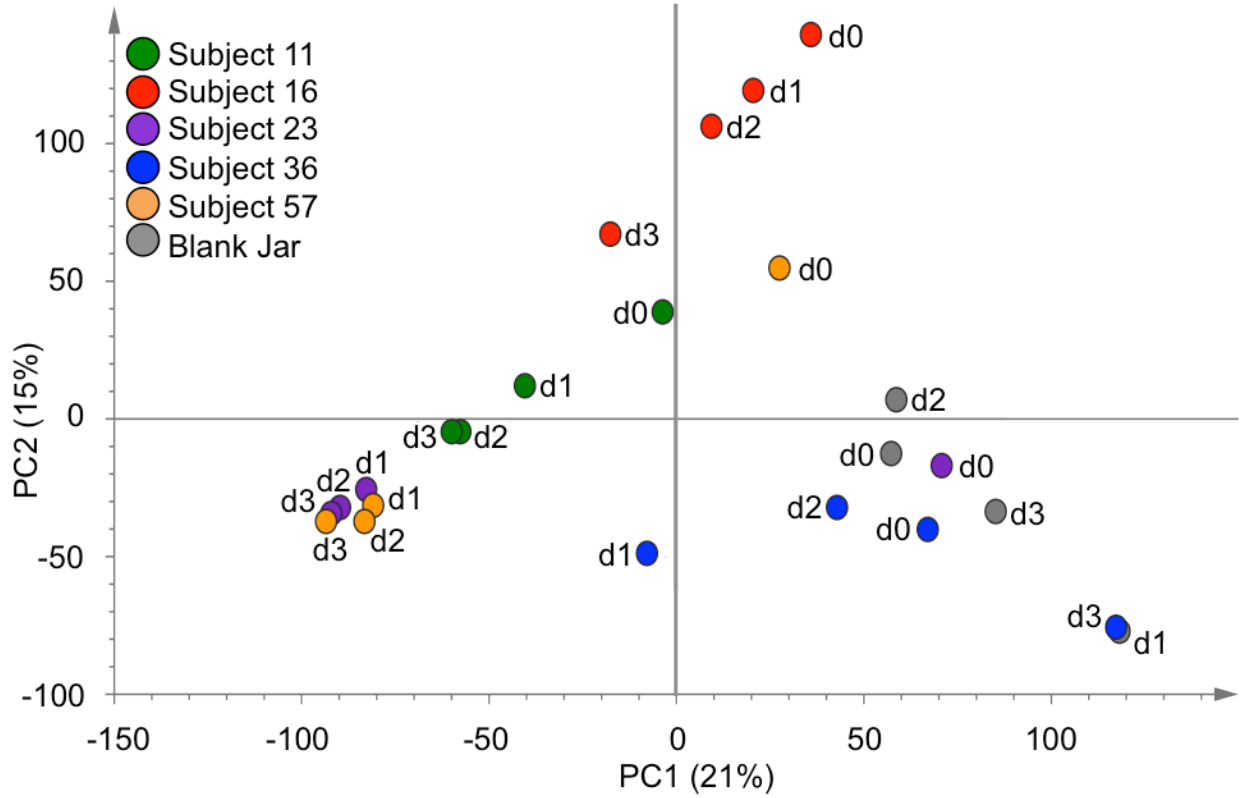

Figure 3.2. Principal component analysis scores plot of all subject and blank headspace SPME-GC-MS chemical profiles analyzed 0-3 days post-deposition. Samples positioned near one another are chemically similar, while samples located far apart are chemically different.

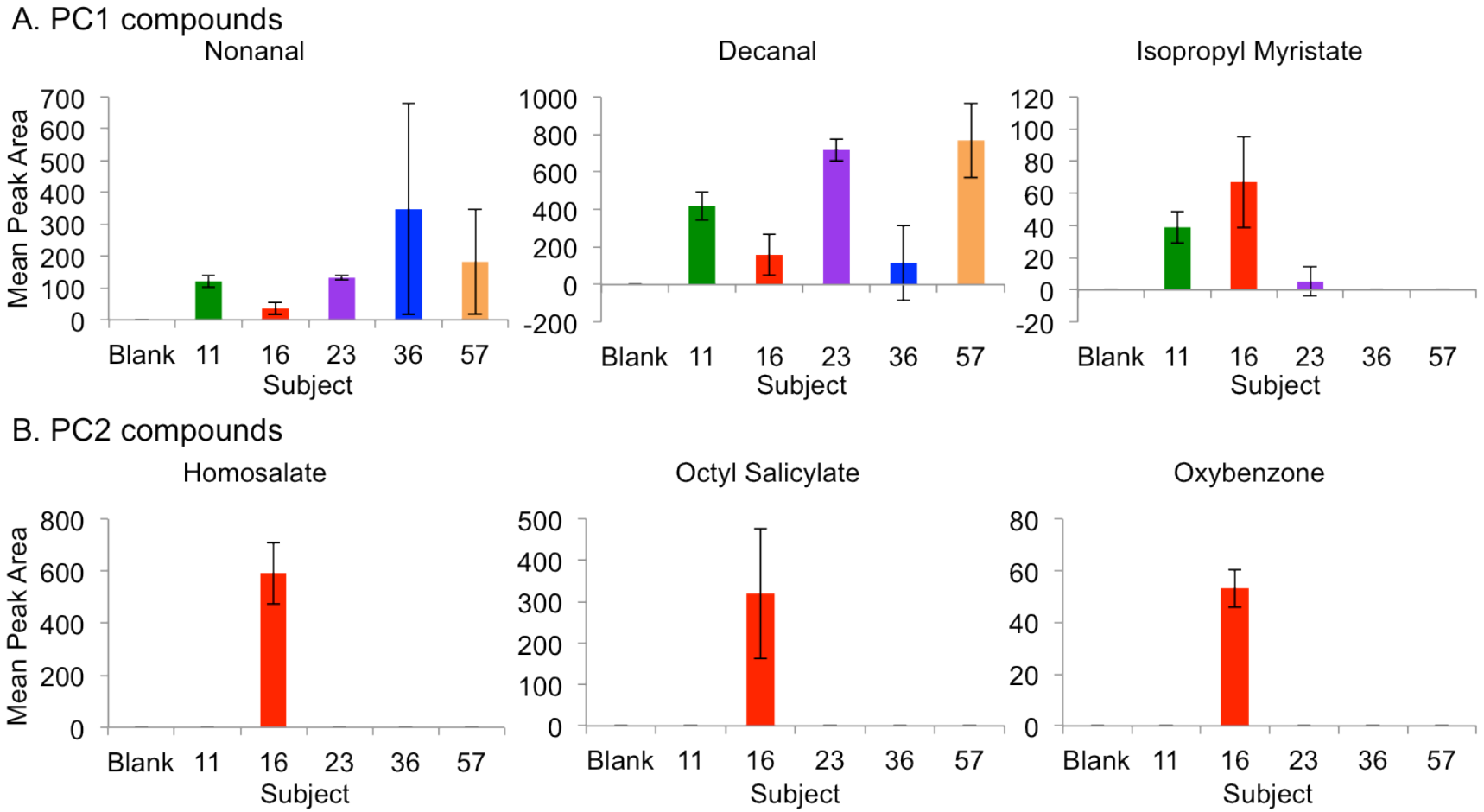

Figure 3.3. Compound abundances contributing to sample positions in the PCA scores plot (Figure 3.2). (A) Nonanal, decanal, and isopropyl myristate levels make substantial contributions to a sample's position on the PC1 *x*-axis, and (B) homosalate, octyl salicylate, and oxybenzone abundances make important contributions to where a sample is positioned on the PC2 *y*-axis (mean peak areas from days 1-3 (n = 3), error bars are one standard deviation).

All samples from the blank control and Subject 36 are clustered together in the lower right quadrant of the PCA scores plot (Figure 3.2), along with the sample from Subject 23's latent fingerprint residue analyzed on day zero. Subject 36 washed her hands just prior to donating latent fingerprints (subject behavior not controlled as a part of this study) and had the lowest abundance chemical profile components detected (Figure 3.1), so it follows that Subject 36 and the blank samples without any fingerprints would be associated together and discriminated from all of the rest of the subject samples containing volatile fingerprint components. The presence of the Subject 23 day zero sample in the cluster of blank samples was unexpected, and further inspection of the scores plot shows that all subject day zero samples containing fingerprint material are positioned to the right (more positive on the PC1 *x*-axis) of the later time point samples. Inspection of the raw data chemical profiles revealed that profiles collected and analyzed on day one generally had more peaks containing fingerprint compounds and greater peak area abundances compared to profiles on day zero, the day the fingerprints were deposited. This is principally due to the loss of the water in the fingerprint over the first approximately 24-48 hours after deposition, after which the fingerprint becomes "brittle" and the compounds that were initially held in the fingerprint by interactions with water may more easily escape into the headspace to be sampled by SPME. These results suggest that each sample's position on the PC1 x-axis describes the relative amount of fingerprint compounds present in the samples, which are similar for all subjects, *i.e.* endogenous compounds. When chemical signals contributing the most to PC1 (*x*-axis) were analyzed, several compounds were identified, including nonanal, decanal, and isopropyl myristate (Figure 3.3a). Nonanal and decanal are present in all five subjects' fingerprint

chemical profiles, and isopropyl myristate was detected in three subjects. These identified endogenous compounds are indicative of the presence of fingerprint material in an unknown residue.

All samples from Subjects 11 and 16 were positioned in distinct clusters, most notably Subject 16's samples located most positively on the PC2 *y*-axis (Figure 3.2). The samples from Subjects 23 and 57 collected 1-3 days post-deposition were all positioned closely together. One striking feature of the profiles shows that Subjects 11, 23, and 57 group closely and are all males, while Subjects 16 and 36 are females and their samples form distinct clusters separate from the males and each other. These results suggest that a sample's positioning on the PC2 *y*-axis reflects the presence of exogenous compounds that serve as indicators of certain individuals. Relative levels of these exogenous compounds from the surface of the skin decrease in the first few days post-deposition, as the sample positions decrease on the PC2 *y*-axis as the time-since-deposition increases. When chemical signals contributing the most to PC2 (*y*-axis) were investigated, the most prominent identified compounds were homosalate, octyl salicylate, and oxybenzone (Figure 3.3b). These three compounds are sunscreens and were only present in the latent fingerprint residues donated by Subject 16, who applied face lotion containing sunscreen approximately six hours prior to donating her fingerprint sample. This explains the grouping of Subject 16 samples relative to other subject samples. The presence of individualizing compounds from cosmetics or other personal care items in fingerprint samples donated by females has been previously observed (Mong *et al.* 1999). These results illustrate the usefulness of chemically profiling latent fingerprint residues using headspace SPME-GC-MS

in order to provide chemical information about an individual's habits and recent history that is inaccessible through friction ridge analysis alone.

Due to the presence of unique chemical compounds in the latent fingerprint residue of Subject 16 that permitted discrimination from all other subjects, we took the study a step further by asking Subject 16 to donate another latent fingerprint residue in an identical manner five months after the original study had been conducted. The second fingerprint residue was sampled and chemical profiles generated using the same protocol as the initial study, and the resulting GC-MS chemical profiles from both data sets were processed and analyzed using PCA to see if both sets of fingerprints from Subject 16 would be associated with each other. A comparison of the GC-MS chromatogram results is presented in Figure 3.4a, and Figure 3.4b shows the PCA scores plot with the samples from Subject 16's latent fingerprints donated five months after the initial study incorporated into the original data set. The chromatogram from the sample donated five months after the initial sample contains all of the principal peaks observed in the original results (same as in Figure 3.1). The chemical profiles sampled from fingerprints donated five months apart contained many of the same compounds, with some peak abundance variations in relation to one another observed. The cluster of compounds eluting from 18-20 minutes, identified as sunscreens that distinguished Subject 16's original sample from the fingerprints donated from the other individuals, was also present in the chemical profiles of fingerprints collected five months later. This suggests that the Subject 16 maintained a daily routine of applying sunscreen, which remained on the fingers or was transferred from touching other parts of the body and was detected in the latent fingerprint residue using headspace SPME-GC-MS. The presence of the sunscreen compounds in the

fingerprint chemical profiles enabled the samples from the original study to be associated with the samples collected five months later from the same individual in the PCA scores plot (Figure 3.4b). Both sets of chemical profiles collected from Subject 16 five months apart were clustered together and away from both blank samples and fingerprint samples without sunscreens detected in the chemical profiles. These results illustrate the value of knowing more information from a latent fingerprint than just friction ridges, such as unique compounds indicative of an individual's habits and prior activities, as more points of similarity increase the case for association.

Comparisons of chemical profiles collected from latent fingerprint residues using non-destructive headspace SPME-GC-MS were successful in distinguishing fingerprint samples from blanks and other subject's fingerprint residues. Fingerprints were discriminated from blanks by the presence of endogenous chemicals such as aldehydes and fatty acid esters, and fingerprint samples were discriminated from fingerprints donated by other individuals in some cases by the presence of unique exogenous chemical components. These latent fingerprint residue samples were associated and discriminated at up to three days post-deposition, indicating a substantial portion of the fingerprint material is still present for volatile and semi-volatile chemical profiling using SPME-GC-MS when some traditional friction ridge analysis techniques would fail due to the absence of water from the residues.

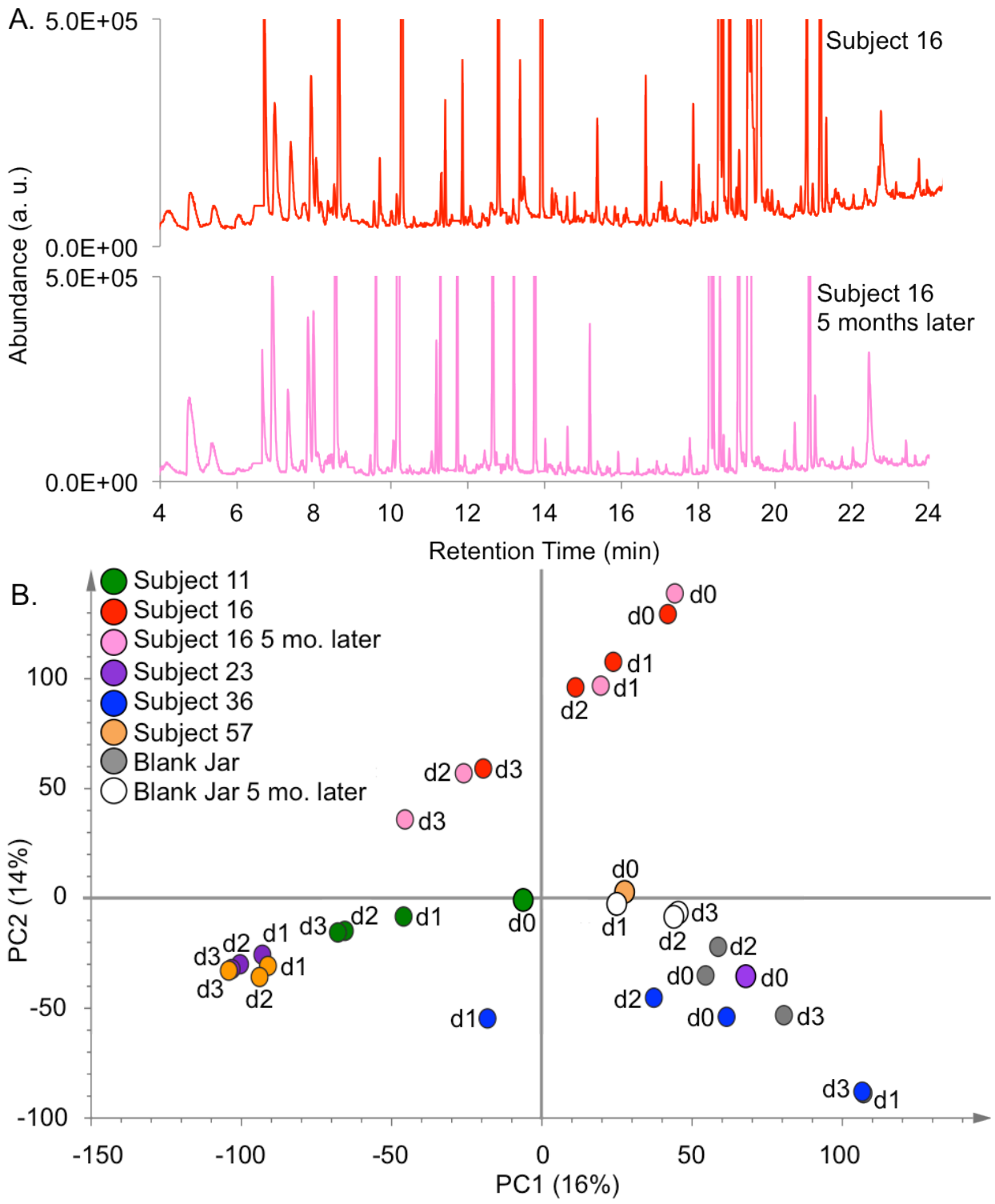

Figure 3.4. Association of latent fingerprint residue chemical profiles donated by one individual, Subject 16, five months apart and sampled 0-3 days after deposition. (A) Total ion chromatogram comparison of SPME-GC-MS chemical profiles sampled one day after fingerprints were deposited. (B) PCA scores plot including Subject 16 fingerprint samples from five months later with the original fingerprint data set sampled daily 0-3 days post-deposition.

3.6 Latent Fingerprint Chemical Changes Over Time

The latent fingerprint residue samples donated by five subjects were analyzed over the course of 30 days to assess chemical profiles changes over time and if those changes could be used to inform criminal investigations. As the outgassing fingerprint compounds were collected passively using SPME, the same latent fingerprint residue samples were re-analyzed at each point in the time course, minimizing intra-personal variation in the results. As expected, the GC-MS chromatograms displayed an appreciable decrease in the abundance and number of compounds detectable above background at later time points. This was anticipated, as fresh latent fingerprint samples have a higher probability of containing volatile compounds, which proceed to dissipate and degrade as the samples age (Curran *et al.* 2007). Similar observations were also made by Mong and coworkers (1999), who noted that the most significant chemical losses occurred within the first week after fingerprint deposition. However, some fingerprint compounds were still detectable in some of the samples after 30 days of storage at ambient conditions. The same data matrix that was used to perform PCA was used to examine how the individual compound peak areas varied over the time course of the experiment. Peak area changes over time were then further examined using linear (0 order), exponential (first-order), and power law decay curve fitting, as well as second- and third-order reaction analyses (Atkins 1994). The resulting coefficients of determination (R^2) for the least-squares fits to the experimental data were finally evaluated to determine the function that most accurately described the decreasing levels of compounds for interpretation.

The time-dependent abundances of decanal, an endogenous fingerprint compound that was highlighted by PCA of the fingerprint results at the early time points (Figure 3.3), are shown

for four of the subjects' samples in Figure 3.5. Decanal was only detectable on days zero and one in Subject 36's latent fingerprint residue sample, preventing examination of abundance changes over time for that sample. The results from Subjects 11, 16, 23, and 57 all show a low abundance of decanal on day zero, which is likely due to the water-containing matrix of the fingerprint residue still present soon after deposition. Moisture may also play important roles in determining the kinetics of lipid oxidation in latent fingerprint residue, including *via* the action of microbial enzymes. An increased abundance of decanal is sampled from the gas above the fingerprint residues on day one post-deposition in all subjects' latent fingerprint residue. The level of decanal in the latent fingerprint residues donated by Subjects 23 and 57 decayed steadily after day one until it was no longer detectable above background on day 17. However, a different temporal relationship was observed for decanal in fingerprint samples from Subjects 11 and 16, where the abundance of decanal continued to increase after day one until day seven for Subject 11 and day five for Subject 16. After decanal levels peaked several days post-deposition, the decanal signal gradually decreased, but was still detectable in the day 30 sample from both subjects. The different abundance kinetics displayed by Subjects 11 and 16 compared to Subjects 23 and 57 is likely related to interpersonal variation in rates of lipid oxidation, as well as the amount of latent fingerprint residue material deposited by each subject. Levels of antioxidant compounds, including vitamins C and E, in each subject's latent fingerprint residue may delay lipid oxidation to varying extents. This may explain the increases in decanal abundance observed for two of the subjects over the first few days of the study. Once the antioxidants are consumed, the levels of decanal increase due to active enzymes in the absence of antioxidants. The level of decanal detected in Subject 11's sample was the largest of the four

subjects, with a maximum peak area of 58712 at seven days post-deposition. Subject 16 had a maximum decanal peak area of 39281 on day five, which was less than Subject 11 and peaked sooner. Subjects 23 and 57 had the smallest abundances, with relative peak areas of 21244 and 15925, respectively, peaking on day one post-deposition. More latent fingerprint residue time course analyses from more subjects are needed to more fully characterize these findings, but it appears that the amount of fingerprint residue initially deposited, which is difficult to control, has a substantial effect on the amounts of endogenous compounds outgassed from the latent fingerprints for SPME-GC-MS detection.

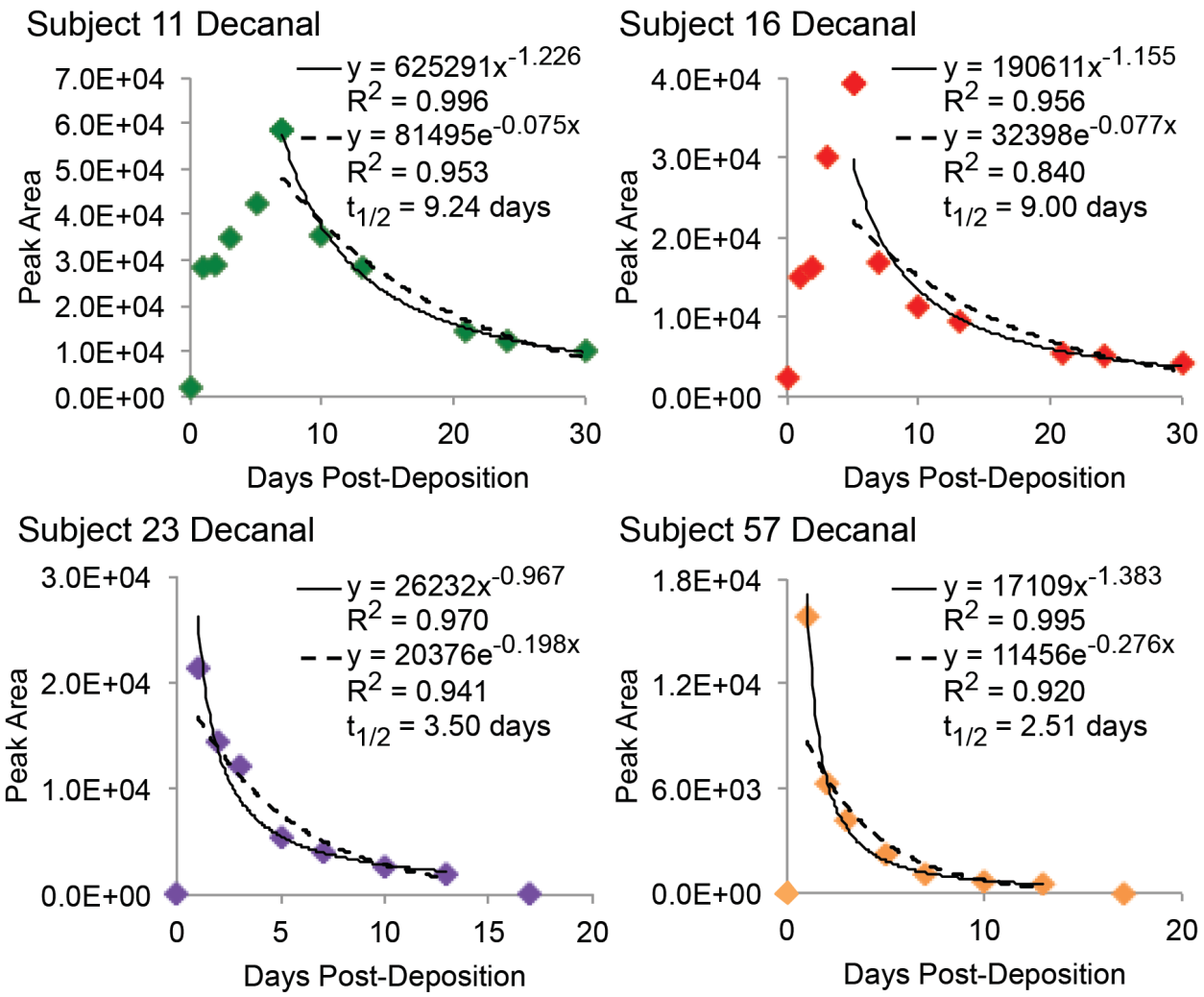

Figure 3.5. Decanal abundance over time in latent fingerprint residue samples from Subjects 11, 16, 23, and 57. Power (solid line) and exponential (dashed line) fits to the data are shown, as well as the calculated half-life ($t_{1/2}$) of decanal in each subject's latent fingerprint residue.

The decay portion of the decanal abundance plot from each subject was best described by a power law dependence (Figure 3.5), as the power function fits to the data resulted in the highest R^2 values. Exponential functions (first-order decay kinetics) are also shown for comparison, however, the R^2 values indicate that decanal decay is not as highly correlated with an exponential function compared to the power function. This was generally true for the other endogenous fingerprint compounds examined, including other aldehydes, which were present

in the fingerprint chemical profiles for several days so that compound abundances over time could be examined. The power law mathematical relationship is described by the function $y = cx^a$ where c is a constant and a is the scaling exponent of the law (Clauset *et al.* 2009, Stumpf and Porter 2012). The most interesting feature of power laws is they are scale invariant, meaning the suspected differences in the amount of fingerprint residue deposited by each subject has no effect on the shape of the decay curve as they're simply scaled versions of the same decay process. The scaling exponent is considered to be characteristic of a process, so from the four scaling exponent values shown in the plots in Figure 3.5, the scaling exponents for the decay of decanal vary by approximately 35% in this preliminary study of four subjects. This is very promising, as scaling exponents for the degradation of endogenous fingerprint compounds could lead to determining the age of an evidentiary fingerprint. However, power laws need to be interpreted with care, as they routinely follow the trends in complex biological processes, which may be decomposed to show trends of many individual overlapping components that do not follow power law dependence (Stumpf and Porter 2012). More work needs to be done to understand the mechanisms underlying fingerprint chemical degradation in order to statistically validate the power law models for latent fingerprint residue compounds (Clauset *et al*. 2009, Stumpf and Porter 2012). The endogenous fingerprint chemical degradations may also be displaying second-order decay kinetics, as the half-lives increased as the time post-deposition increased, and second-order reaction plots generally had the second highest R^2 values for fitting this small data set.

Compound abundance changes over time for the exogenous sunscreen compounds homosalate, octyl salicylate, and oxybenzone detected in Subject 16's latent fingerprint residue chemical profile were also examined (Figure 3.6). All three chemicals were present at maximum abundance on day zero, after which levels steadily decreased until they were no longer detectable above baseline on day five for octyl salicylate and day seven for homosalate and oxybenzone. The same decay curve fitting methods applied to the endogenous compounds described above were applied to these exogenous sunscreen compounds, and exponential curves (first-order decay kinetics) yielded the highest R^2 values for the least-squares fits to the data (Figure 3.6). The exponential decay equations were used to calculate the chemical half-lives of the compounds (Atkins 1994), which were just under one day for homosalate and octyl salicylate (0.86 and 0.85 days, respectively) and 2.23 days for oxybenzone. Oxybenzone has a higher boiling point temperature (224-227 °C) compared to homosalate and octyl salicylate (161-165 °C and 189 °C, respectively), resulting in a longer half-life compared to the other two (Budavari *et al.* 1996).

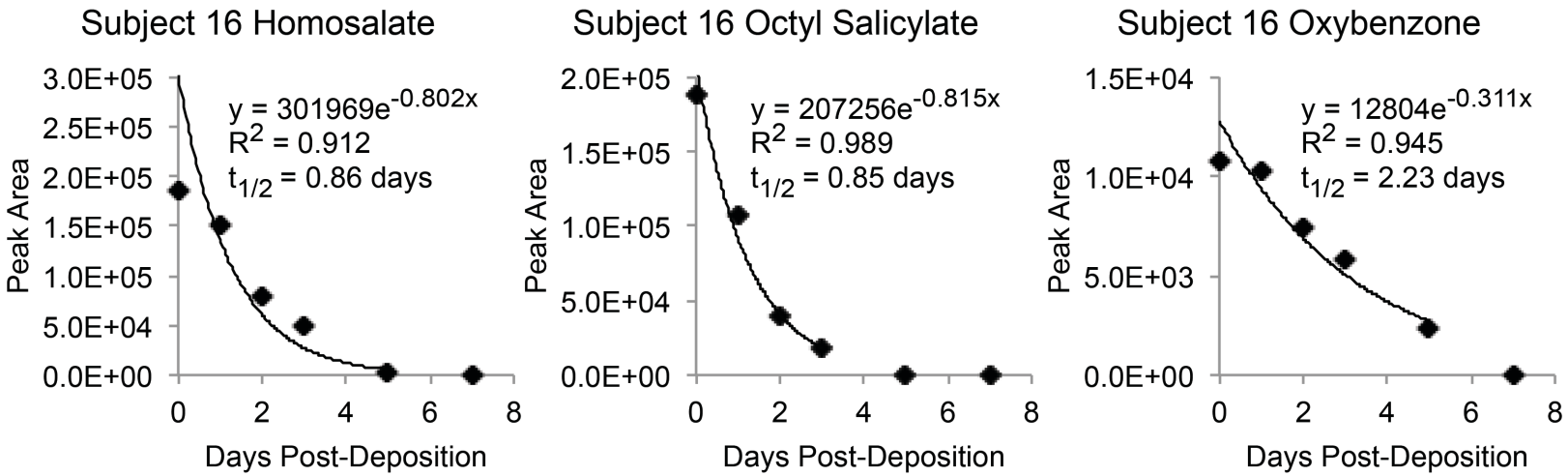

Figure 3.6. Decay of the exogenous compounds homosalate, octyl salicylate, and oxybenzone over time in Subject 16's latent fingerprint residue sample. Exponential curve fits to the data are shown, as well as the calculated half-lives ($t_{1/2}$) of the three compounds.

The observed exponential decays of exogenous sunscreen compounds in the latent fingerprint residues (Figure 3.6) are similar to the power law decays of the endogenous compounds (Figure 3.5) in that the half-lives are independent of the initial concentrations of the compounds (Atkins 1994). These scale invariant mathematical relationships are promising, as inherent heterogeneity in the amount of fingerprint residue deposited with each surface contact is the principal reason quantitative comparisons between fingerprint samples are difficult (Mong *et al.* 1999). It makes sense that exogenous compounds from the skin's surface would decay with a straightforward exponential dependence, as they are most concentrated at the time of deposition and disperse over time. The presence of exogenous compounds from the donor's environment is useful for determining prior activities and habits. Endogenous latent fingerprint residue compounds have far more complex mechanisms underlying their observed levels over time, including both generation and degradation through human or microbial enzymatic processes in addition to outgassing. The fact that endogenous compounds generally remain present in the latent fingerprint residues for longer periods of time compared to the exogenous compounds makes them more promising for universal fingerprint dating. The approximately 35% variation in scaling exponents between the four subjects in this small proof-of-principle study demonstrates the potential for determining the time since a fingerprint was deposited regardless of who deposited it. A larger study of many more subjects and latent fingerprint residue samples is the next step for examining the utility of endogenous chemical decay rates for determining the time-window since a fingerprint was deposited. Such information would greatly aid criminal investigations by helping to determine a crime's timeline and if fingerprint evidence is germane to the investigation at hand.

3.7 Conclusions and Future Directions

Objects and locations connected with nearly every type of crime are routinely examined for latent fingerprint evidence. A novel method for non-invasively analyzing the chemical profiles of latent fingerprint residues has been developed in order to gain a new level of information from the most common type of forensic evidence. Passive SPME headspace sampling collects both endogenous and exogenous volatile and semi-volatile compounds contained in the fingerprint residue while preserving the fingerprint for traditional analyses. There is no additional sample preparation or cleanup following SPME sampling, making this fingerprint sampling method field deployable and usable by relatively unskilled personnel.

The information-rich chemical profiles obtained from GC-MS analyses of the SPME samples were used to quantitatively compare fingerprint compounds both between subjects and over a time course of 30 days. Subjects were associated and discriminated from each other and blank control samples using both pairwise Spearman rank correlation coefficient analyses and principal component analysis of all subjects simultaneously. Both endogenous and exogenous compounds were detectable and responsible for sample differentiation. Endogenous compounds identify a residue as a latent fingerprint and were in some cases detectable 30 days post-deposition. The decay rates of endogenous compounds over time were comparable in different subjects and show promise for universal fingerprint dating methods. Knowledge of the time since a fingerprint was deposited would greatly aid criminal investigators in establishing a crime timeline and if fingerprint evidence is relevant to the investigation at hand. The exogenous compounds that were detected were individualizing and degraded over a shorter period of time compared to the endogenous compounds. Rare compounds may serve to link

latent fingerprints and the individuals that deposited them to specific locations or activities (*e.g.* explosives on the hands of a bomb maker). Other compounds may infer certain subject traits or habits, such as a cigarette smoker or drug user, all of which is additional information from a latent fingerprint residue that is currently inaccessible using traditional fingerprint analysis methods.

In the future, studies will be expanded to include SPME-GC-MS chemical profiling of latent fingerprint residues deposited on porous surfaces, as well as degradation studies under varying storage conditions (*e.g.* light, temperature, humidity) to determine effects on decay rate.

REFERENCES

3.8 References

Atkins, P. W. (1994). *Physical Chemistry* (5th Ed.). W. H. Freeman and Company, New York, NY, pp. 861-893.

Antoine, K. M., Mortazavi, S., Miller, A. D., Miller, L. M. (2010). Chemical differences are observed in children's versus adults' latent fingerprints as a function of time. J. Forensic Sci. 55:513-518.

Asano, K. G., Bayne, C. K., Horsman, K. M., Buchanan, M. V. (2002). Chemical composition of fingerprints for gender determination. J. Forensic Sci. 47:805-807.

Bernier, U. R., Booth, M. M., Yost, R. A. (1999). Analysis of human skin emanations by gas chromatography/mass spectrometry. 1. Thermal desorption of attractants for the yellow fever mosquito (*Aedes aegypti*) from handled glass beads. Anal. Chem. 71:1-7.

Bernier, U. R., Kline, D. L., Barnard, D. R., Schreck, C. E., Yost, R. A. (2000). Analysis of human skin emanations by gas chromatography/mass spectrometry. 2. Identification of volatile compounds that are candidate attractants for the yellow fever mosquito (*Aedes aegypti*). Anal. Chem. 72:747-756.

Buchanan, M. V., Asano, K., Bohanon, A. (1996). Chemical characterization of fingerprints from adults and children. Proc. SPIE 2941:89-95.

Budavari, S., O'Neil, M. J., Smith, A., Heckelman, P. E., Kinneary, J. F. (Eds.). (1996). *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals* (12th ed.). Merck & Co., Inc., Whitehouse Station, NJ, p. 811.

Clauset, A., Shalizi, C. R., Newman, M. E. J. (2009). Power-law distributions in empirical data. SIAM Rev. 51, 661-703.

Curran, A. M., Ramirez, C. F., Schoon, A. A., Furton, K. G. (2007). The frequency of occurrence and discriminatory power of compounds found in human scent across a population determined by SPME-GC/MS. J. Chromatogr. B 846:86-97.

Curran, A. M., Prada, P. A., Furton, K. G. (2010). The differentiation of the volatile organic signatures of individuals through SPME-GC/MS of characteristic human scent compounds. J. Forensic Sci. 55:50-57.

De Paoli, G., Lewis, S. A., Schuette, E. L., Lewis, L. A., Connatser, R. M., Farkas, T. (2010). Photo- and thermal-degradation studies of select eccrine fingerprint constituents. J. Forensic Sci. 55:962-969.

Gaensslen, R. E., Young, K. R. (2003). Fingerprints, in *Forensic Science: An Introduction to Scientific and Investigative Techniques*, S. H. James, J. J. Nordby, eds, CRC Press, Boca Raton, FL, pp. 277-296.

Hazarika, P., Jickells, S. M., Wolff, K., Russell, D. A. (2008). Imaging of latent fingerprints through the detection of drugs and metabolites. Angew. Chem. Int. Ed. 47:10167-10170.

Hazarika, P., Jickells, S. M., Russell, D. A. (2009). Rapid detection of drug metabolites in latent fingermarks. Analyst 134:93-96.

Hazarika, P., Jickells, S. M., Wolff, K., Russell, D. A. (2010). Multiplexed detection of metabolites of narcotic drugs from a single latent fingermark. Anal. Chem. 82:9150-9154.

Johnson, H. L., Maibach, H. I. (1971). Drug excretion in human eccrine sweat. J. Invest. Dermatol. 56:182-188.

Leggett, R., Lee-Smith, E. E., Jickells, S. M., Russell, D. A. (2007). "Intelligent" fingerprinting: simultaneous identification of drug metabolites and individuals by using antibody-functionalized nanoparticles. Angew. Chem. Int. Ed. 46:4100-4103.

Martin, A. N., Farquar, G. R., Jones, A. D., Frank, M. (2010). Human breath analysis: methods for sample collection and reduction of localized background effects. Anal. Bioanal. Chem. 396:739-750.

Miller, J. N., Miller, J. C. (2005). *Statistics and Chemometrics for Analytical Chemistry* (5th Ed.). Pearson Education Limited, Harlow, UK, pp. 110-111, 142, 167-169, 215-219.

Mills, G. A., Walker, V. (2000). Headspace solid-phase microextraction procedures for gas chromatographic analysis of biological fluids and materials. J. Chromatogr. A 902:267-287.

Mong, G. M., Petersen, C. E., Clauss, T. R. W. (1999). Advanced fingerprint analysis project final report- fingerprint constituents. Pacific Northwest National Laboratory Report PNNL-13019.

Naitoh, K., Inai, Y., Hirabayashi, T. (2000). Direct temperature-controlled trapping system and its use for the gas chromatographic determination of organic vapor released from human skin. Anal. Chem. 72:2797-2801.

Noble, D. (1995). Vanished into thin air: the search for children's fingerprints. Anal. Chem. 67:435A-438A.

O'Brien, P. J., Siraki, A. G., Shangari N. (2005). Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. Crit. Rev. Toxicol. 35:609-62.

Puhvel, S. M., Reisner, R. M., Sakamoto, M. (1975). Analysis of lipid composition of isolated human sebaceous gland homogenates after incubation with cutaneous bacteria- thin-layer chromatography. J. Invest. Dermatol. 64:406-411.

Ramotowski, R. S. (2001). Composition of Latent Print Residue, in *Advances in Fingerprint Technology* (2nd Ed.), H. C. Lee, R. E. Gaensslen, eds, CRC Press, Boca Raton, FL, pp. 63-104.

Schulz, M. M., Reichert, W. (2002). Archived or directly swabbed latent fingerprints as a DNA source for STR typing. Forensic Sci. Int. 127:128-130.

Stumpf, M. P. H., Porter, M. A. (2012). Critical truths about power laws. Science 335:665-666.

Vree, T. B., Muskens, A. T. J., Van Rossum, J. M. (1972). Excretion of amphetamines in human sweat. Arch. Int. Pharmacod. T. 199:311-317.

Wargacki, S. P., Lewis, L. A., Dadmun, M. D. (2008). Enhancing the quality of aged latent fingerprints developed by superglue fuming: loss and replenishment of initiator. J. Forensic Sci. 53:1138-1144.

Weyermann, C., Roux, C., Champod, C. (2011). Initial results on the composition of fingerprints and its evolution as a function of time by GC/MS analysis. J. Forensic Sci. 56:102-108.

Zhang, Z.-M., Cai, J.-J., Ruan, G.-H., Li, G.-K. (2005). The study of fingerprint characteristics of the emanations from human arm skin using the original sampling system by SPME-GC/MS. J. Chromatogr. B 822:244-252.

CHAPTER 4: ASSESSMENT OF *IN VITRO* TOXICITY OF CHEMOTHERAPEUTIC AGENTS IN CANCER CELLS USING A METABOLOMIC APPROACH

4.1 Motivations and Introduction

4.1.1 Treatment of Cancers with Chemotherapeutic Agents

Cancer is a collection of diseases that are characterized by uncontrolled growth of abnormal cells. Approximately 577,190 Americans were expected to die of cancer in 2012, more than 1,500 people per day. Cancer accounts for nearly one of every four deaths in the U.S. and is the second most common cause of death, exceeded only by heart disease. Lung cancer and breast cancer are two of the most prevalent cancers in America (surpassed only by prostate cancer), with estimates of 229,060 new cases of breast cancer and 226,160 new cases of lung cancer diagnosed in 2012 (American Cancer Society 2012). The development and progression of cancer are influenced by the combined effects of genetic factors (*e.g.* inherited mutations, hormones, immune conditions) and environmental factors (*e.g.* exposures to tobacco, mutagenic and tumor-promoting chemicals and radiation, infectious agents). Most cancer treatments rely on selective removal or destruction of tumors with surgery, radiation, and/or a variety of pharmaceutical therapies including chemotherapy, hormone therapy, biological therapy, and targeted therapy (American Cancer Society 2012).

Anticancer drugs are a vital component of cancer treatment, however, severe side effects limit the dose that patients receive. There is an urgent need for more targeted and selective chemotherapeutics, and evaluation of new anticancer drugs is a critical step in improving mortality rates for the disease. Despite recent advances in both diagnostic and therapeutic tools available, the mortality rate remains high and severe side effects are associated with common chemotherapy treatments. It is of paramount importance to continue

the search for new chemotherapeutics with novel modes of cytotoxicity that offer higher and more selective potency and have fewer side effects. *In vitro* screening of candidate chemotherapeutic agents for their potency in altering metabolic phenotypes has potential to provide multidimensional measures that reflect how these compounds act on a wide variety of cellular processes.

4.1.2 The Role of Metabolism in Cancer Cell Proliferation

The rapid proliferation of cancer cells requires efficient biosynthesis of essential biomolecules, particularly those that conduct biochemical energy to the processes of cell growth and division. The transformation of normal cells to tumor cells involves a shift in metabolic traits that allow them to proliferate with less dependence on extracellular signaling mechanisms, yet many questions remain regarding the complex networks that regulate metabolic processes in tumors (DeBerardinis 2008). Even less is known about metabolic processes in cancer cells during chemotherapeutic treatments designed to either kill cancer cells or arrest their growth and proliferation.

Chemotherapeutic treatments of a variety of cancers are accomplished through use of an assortment of alkylating agents that cross-link DNA and prevent cell replication. One of the more successful alkylating agents is cisplatin (Figure 4.1), which was developed by Barnett Rosenberg at Michigan State University in the 1960s (Rosenberg *et al.* 1965, Rosenberg *et al.* 1969). Cisplatin acts both by alkylation of DNA and through imparting stress to the endoplasmic reticulum (ER), and both of these events can trigger programmed cell death known as apoptosis (Rabik and Dolan 2007). The effectiveness of cisplatin in cancer treatment is compromised by

side effects including toxicity in noncancerous tissues, as well as development of cisplatin-resistant cancers. Another of the most successful chemotherapeutic agents is the plant natural product paclitaxel, also known by its trade name Taxol, which does not act through covalent modification of DNA. Instead, its potency is based on interruption of mitosis of cancer cells by stabilizing microtubules via non-covalent binding (Wani *et al.* 1971, Bayet-Robert *et al*. 2010a).

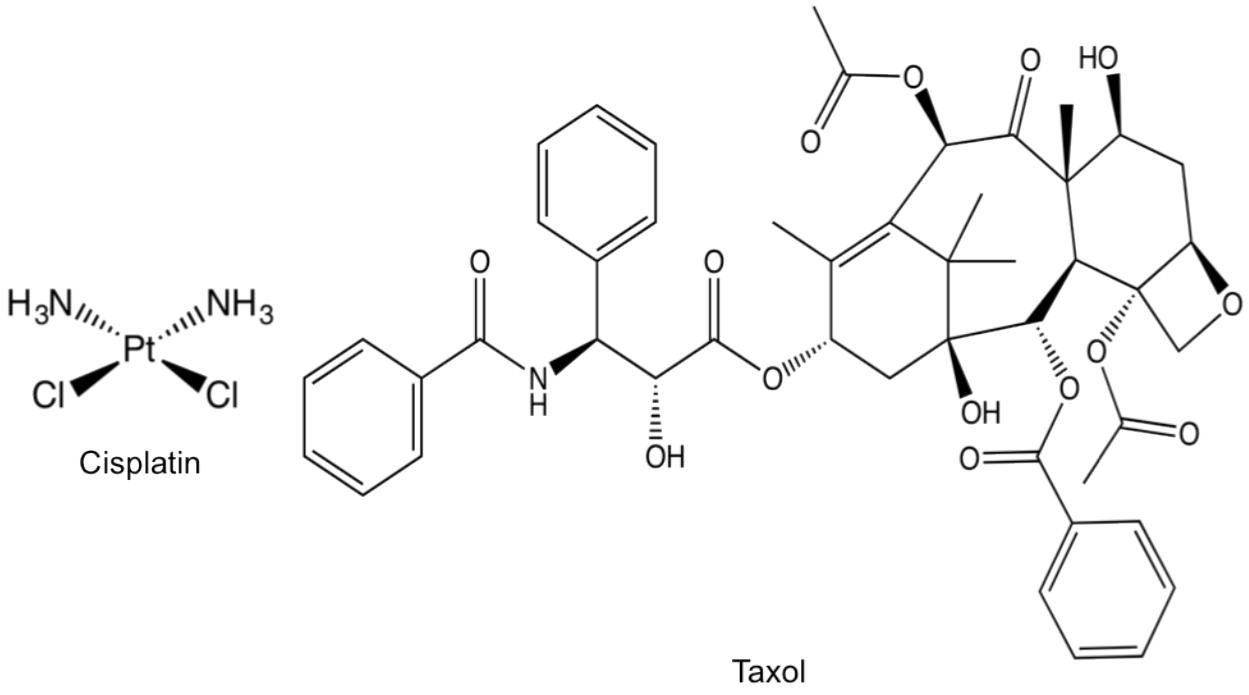

Figure 4.1. Chemical structures of the cancer chemotherapeutic agents cisplatin and taxol.

Since many chemotherapeutic agents disrupt normal cellular processes involved in cell division, they provide many opportunities for modification of fluxes through the various cellular metabolic processes. Recognition of so-called non-target mechanisms offer the potential for new understanding of how anticancer agents disrupt the generation of biochemical energy in

cancer cells, and it is hoped that this knowledge will help guide future development of more effective cancer chemotherapies.

4.1.3 Cellular Metabolism

Central cellular metabolism is composed of a few main pathways of interest: glycolysis, the pentose phosphate pathway (PPP), the tricarboxylic acid (TCA) cycle, anaplerotic reactions that replenish pools of key metabolic intermediates and thus allow intermediates to be withdrawn to support other biosynthetic reactions, and the biosynthesis of non-essential amino acids, fatty acids, and phospholipids (Griffin and Shockcor 2004, Yang *et al.* 2007, Yang *et al.* 2008, Oakman *et al.* 2011). These biochemical pathways form the central backbone of metabolism, providing energy, cofactors, and building blocks for cell growth and replication. Glycolysis and the PPP are the pathways for glucose catabolism and generation of biochemical energy that drives many cellular processes, and these enzymatic reactions take place in the cytosol. In contrast, the reactions of the TCA cycle occur in the mitochondrial matrix where respiration takes place. The PPP supplies NADPH, needed for reducing intermediates in fatty acid and other biosynthetic pathways, and ribose-5-phosphate, the precursor of nucleotides (Becker *et al.* 2006, KEGG database http://www.genome.jp/kegg/).

The TCA cycle transfers electrons to the respiratory chain in the mitochondria of healthy cells supplied with adequate oxygen. Hypoxic cells suppress metabolism through the TCA cycle due to the lack of adequate oxygen to accept electrons at the end of the ATP-producing respiratory chain. Cancer cells also characteristically rely on glycolysis rather than the TCA cycle/respiratory chain for ATP production regardless of oxygen level (called the Warburg

effect, discussed below). It is anticipated that fluxes through TCA cycle reactions are low in cancer cells, which has potential to decrease cellular levels of TCA cycle intermediates to below conventional detection limits. The anaplerotic reactions supply intermediates to the TCA cycle, supplementing the levels produced through the conversion of acetyl-CoA entering the TCA cycle from glycolysis. The most important reaction is the conversion of pyruvate to oxaloacetate by pyruvate carboxylase. Aspartate can also be converted to oxaloacetate. Glutamate is readily converted to α-ketoglutarate, and the β-oxidation of fatty acids yields succinyl-CoA, another intermediate in the TCA cycle. Anaplerosis attempts to compensate for the lack of TCA cycle activity characteristic of cancer cells (Becker *et al.* 2006, KEGG database http://www.genome.jp/kegg/).

The non-essential amino acids are synthesized from glycolysis and TCA cycle intermediates. Glycine and cysteine are formed from serine, which is derived from 3-phosphoglycerate, a glycolytic intermediate. Alanine is synthesized from pyruvate, the end product of glycolysis. Glutamine and proline are synthesized from glutamate, which is generated from α-ketoglutarate, a TCA cycle intermediate. Aspartate is the precursor for asparagine and is formed from the TCA cycle intermediate oxaloacetate. Fatty acid biosynthesis starts with acetyl-Co-A (the starting point for the TCA cycle as well), which is synthesized in the mitochondria, moves into the cytosol with citrate cleavage, and is used to generate malonly-CoA. Palmitic acid (C16:0) is formed from acetyl-CoA and malonyl-CoA, and can be elongated to form stearic acid (C18:0). Desaturation of palmitic and stearic acids yields palmitoleic (C16:1) and oleic (C18:1) acids. Palmitic, stearic, and oleic acids are the main fatty acids in human cells. These fatty acids are incorporated into glycerolipids and glycerophospholipids. The backbone of

glycerophospholipids, glycerol phosphate, is derived from a glycolytic intermediate. The other components are acetyl-CoA, two fatty acids, and a headgroup (choline, inositol, serine, glycerol, or ethanolamine) that classifies the type of glycerophospholipid (Becker *et al.* 2006, KEGG database http://www.genome.jp/kegg/).

4.1.4 The Warburg Effect

Normal cells with access to adequate oxygen create energy through mitochondrial oxidation of pyruvate through the TCA cycle, yielding 36 molecules of ATP along with H_2O and CO_2 per glucose molecule metabolized. Glycolysis provides an anaerobic alternative energy production mechanism, producing two molecules of ATP and lactate for every molecule of glucose metabolized. When there is not enough oxygen, the pyruvate at the end of glycolysis remains in the cytoplasm (instead of being transported into the mitochondria) and is converted to lactate, which is then exported from the cell. The persistence of aerobic glycolysis is a trait of most cancers, first reported by Warburg and is known as the Warburg effect (Warburg *et al.* 1927, Warburg 1956). The high density of most malignant cells creates an anaerobic environment that requires up-regulated glycolysis for energy generation. However, persistent glycolysis has been observed in some cancer types despite the presence of adequate oxygen levels.

4.1.5 Metabolomics and Mass Spectrometry

One of the most informative methods for studying the global metabolism of a cell (or other biological system) under a given set of conditions is metabolomics. Metabolomics is defined as a comprehensive and quantitative analysis of all of the metabolites of the biological

system under study (Fiehn 2001). Metabolites are the intermediates and products of metabolism, which are mostly small molecules less than 1000 Da (Asiago *et al.* 2010). Metabolites result from the interaction of the system's genome with its environment, so they are not simply the end-products of gene expression, but part of the regulatory system for the cell, tissue, or organism. The human genome contains approximately 40,000 genes encoding up to one million proteins. The up- or down-regulation of metabolic enzymes results in the synthesis or degradation of small molecule metabolites. At any given time in a cell thousands of energy transformation processes are occurring, and these conversions are collectively referred to as metabolism. In addition to anabolism and catabolism, cellular processes such as absorption, distribution, and detoxification of endogenous and exogenous compounds are all reflected in the metabolome as well (Claudino *et al.* 2007). The number of known endogenous metabolites is still increasing, with approximately 3000 identified so far (Di Leo *et al.* 2007). The metabolites are most closely related to the observed phenotype in the central dogma of biology (genes in DNA transcribed into RNA, RNA translated into proteins, enzymes regulate levels of metabolites), so they reflect the actual physiological conditions of the system (cell, tissue, or organism) at the time of sampling. Disease states and drug treatments can alter the metabolic phenotype of a cell, tissue, or human, making metabolic profiling relevant to clinical medicine and the pharmaceutical industry (Yang *et al.* 2007, Yang *et al.* 2008).

Metabolite profiling first appeared in the scientific literature in the 1950s, but was relatively slow to develop into a distinct research field (Di Leo *et al.* 2007). Early work involved analyzing the metabolites of exogenous pharmaceutical products (*e.g.* synthetic estrogens) (Williams *et al.* 1975). The field gradually shifted to profiling classes of endogenous compounds

of interest, such as catecholamines, serotonin, and melanin precursors (Muskiet *et al.* 1981) and prostanoids (Robinson *et al.* 1984). Researchers in the past only profiled certain compound classes due to limitations in analytical detection capabilities. Technological advancements in the 1960s and 1970s enabled quantitative (as opposed to qualitative) measurements of metabolite profiles. Horning and Horning (1971) and Pauling and coworkers (1971) pioneered the use of GC-MS for quantitative metabolite profiling in human urine and tissue extracts. Around the same time, Hoult and coworkers (1974) demonstrated NMR detection of metabolites in human tissue. Today, analytical instrumentation has advanced to allow detection of greater numbers of endogenous and exogenous metabolites at biologically relevant concentrations, and the field of metabolomics applications has expanded as well. Metabolite profiling aims to rapidly analyze a large number of compounds in a non-targeted manner. Often the goal is not quantitation but relative comparisons of metabolite levels that characterize a given sample. Not all metabolites must be identified and quantified for a metabolite profiling experiment to be successful. Rather, the detection of a specific metabolite profile that is characteristic of a sample type is useful for differentiating that sample from others (Claudino *et al.* 2007).

The analytical technologies that enable metabolomic analyses include separation techniques and detection methods. Gas Chromatography (GC) and High Performance Liquid Chromatography (HPLC) are the most commonly used separation techniques. Gas chromatography generates very high chromatographic resolution but requires chemical derivatization of polar metabolites to make them volatile for gas-phase separation. Even with derivatization, some large, polar, or chemically reactive metabolites cannot be analyzed using GC. Compared to GC, HPLC has lower chromatographic resolution, but it is capable of

separating a much wider range of analytes without derivatization. The most common detection method coupled to both GC and HPLC separations is mass spectrometry (MS), which is used to identify and quantify metabolites. Mass spectrometry is both sensitive and can be very specific by detecting the masses of analyte ions. The analyte ionization process in GC-MS creates mass spectral fragmentation patterns that enable analyte identification by comparison to mass spectral libraries of known compounds, which is highly advantageous. It is also possible to use MS as a stand-alone metabolite detection method with direct sample infusion and MS information to identify metabolites.

Another common metabolite detection method is nuclear magnetic resonance (NMR), which does not rely on analyte separation prior to analysis and is non-destructive for downstream analyses. A broad range of small molecule types may be quantitated simultaneously with minimal sample preparation, high analytical reproducibility, and molecular structure information for identification. Intact material, such as cell pellets or biopsies, may be analyzed using NMR without extraction of metabolites, which is useful for clinical studies. However, metabolite analyses are generally done on extracted materials, which are amenable to high-throughput analyses using MS. The sensitivity of MS is a few orders of magnitude better than NMR, allowing detection of more metabolites at biologically relevant concentrations. However, MS is less specific than NMR for metabolite identification, as isomers cannot be distinguished using MS alone (Bayet-Robert *et al.* 2010b). A further advantage of MS analysis is automated data analysis options, which generate a larger number of resolvable metabolites and aid in the identification of unknown metabolites (Claudino *et al.* 2007). Therefore, MS is the more useful analytical platform for global metabolite profiling, but it should be noted that the

majority of pharmacometabolomic studies published to date have gained useful knowledge using NMR analysis techniques as well.

4.1.6 Chemometric Procedures

The major challenge of metabolomics is relating the vast amounts of measured biochemical data to systems biology (Bayet-Robert *et al.* 2010b). All metabolomics studies yield complex multivariate data sets that require data filtering using chemometric and bioinformatics methods and visualization software for interpretation. The aim of these procedures is to elucidate biochemical "fingerprints" that are characteristic of the phenotype or indicative of changes to the phenotype as a result of altering the sample conditions. The ultimate goal is to identify the metabolites causing the characteristic biochemical fingerprint, which yield information about biochemical pathways involved and are potential biomarkers that define the phenotype in a biological or clinical context (Griffin and Shockcor 2004, Claudino *et al.* 2007, Di Leo *et al.* 2007).

Data reduction and chemometric approaches enable efficient data mining and extraction of useful information from large, multivariate metabolomics data sets. Multivariate pattern recognition methods can classify samples based on the identification of inherent patterns in the measured metabolite profiles, and these patterns enable visualization of sample relationships in the data set. There are two general approaches for multivariate pattern recognition: unsupervised and supervised methods. Both approaches may be applied to the same data set in order to extract different information. Unsupervised methods examine the inherent sample relationships without knowledge of sample classification. Examples of

unsupervised methods include principal component analysis (PCA) and clustering methods. Supervised algorithms use class information to maximize the separation between different sample classes. Examples of supervised methods include partial least squares discriminant analysis (PLS-DA) and artificial neural networks. Supervised methods are more useful for analyzing metabolite profile patterns associated with different sample conditions (*e.g.* normal or cancerous cells, metabolic response to treatment with different drugs) (Claudino *et al.* 2007, Trygg *et al.* 2007, Oakman *et al.* 2011).

A typical gas chromatography-mass spectrometry (GC-MS) or high-performance liquid chromatography-mass spectrometry (HPLC-MS) metabolomics analysis generates thousands of data points, of which only a small subset might be needed to distinguish different anticancer drug treatments. Extracting the most meaningful features of these data is thus key to generating useful new knowledge with mechanistic or explanatory power (Goodacre *et al.* 2004). Supervised multivariate statistical analysis methods are employed to process mass spectrometry data collected during metabolomics experiments of differentially treated cancer cells. This work utilized both PLS-DA and orthogonal partial least squares discriminant analyses (OPLS-DA) to detect metabolite signals that differentiated samples. The goal of PLS-DA is to find the metabolites (RT-*m/z* pairs; X variables) that describe the greatest variations in the spectra while at the same time have maximal correlation with the class assignment (input Y variable). The metabolite signals that display similar variation and correlation patterns in the data set are then linearly combined to form principal components. The score plot, limited to the most significant principal components, gives a visual image of sample variations from a global point of view. The corresponding loadings plot allows the evaluation of the contribution that each

metabolite makes to the total information of the metabolome (Boccard *et al.* 2007). Supervised orthogonal partial least squares discriminant analyses (OPLS-DA) are complementary to PLS-DA and are used to highlight interclass variance and provide associated confidence intervals (Trygg *et al.* 2007). These statistical data analysis methods help to determine how endogenous metabolism is affected by anticancer drug treatment by reducing the dimensionality of complex data sets to facilitate recognition of outlier metabolites that differ in cytotoxic modes of action. While the multivariate statistical analyses identify important molecules by mass, these compounds still need to be identified. Mass spectra collected using GC-MS can be compared to the National Institute of Standards and Technology (NIST) library of known compound mass spectra in order to identify metabolites of interest in cancer cell samples. Metabolite identification points to specific mechanisms of altered cellular biochemistry that result from treatments with anticancer agents.

4.1.7 Application of Metabolomics to Studying Cancer

There is a growing interest in investigating the biochemical mechanisms of action of drugs and identifying targeted enzymes and metabolic pathways using metabolomic analyses (Chung and Griffiths 2008, Bayet-Robert *et al.* 2010c). Cancer phenotypes are especially suited for study using metabolomics. While RNA and protein regulation and expression levels play a definitive role in cancer initiation and progression, malignant cells undergo a range of changes in metabolism as well (Yang *et al.* 2007). The most studied and well-known example of metabolic alteration as a result of cancer is the Warburg effect, which is the shift to aerobic glycolysis from normal cellular respiration in cancer cells (Warburg *et al.* 1927). The

metabolome of a cancer cell is also likely to show changes in response to anticancer drug treatment (Chung and Griffiths 2008). These metabolism changes can indicate if the drug is working and the mechanism of action. Cellular toxicity of anticancer agents can also be studied using metabolomics, with particular focus on rapid screening of potential chemotherapeutic agents for lead compound selection (Kim *et al.* 2010).

Previous studies have primarily focused on three areas of cancer research with clinical applications. The first is early detection and diagnosis of cancer (including stage, receptor status, and metastasis) and recurrence using metabolomics techniques to differentiate normal and cancerous cells and tissues (Ting *et al.* 1996, Yang *et al.* 2007, Yang *et al.* 2008, Frickenschmidt *et al.* 2008, Asiago *et al.* 2010). The second is detection and prediction of drug toxicity, with application to screening drug candidates for organ toxicity prior to clinical trials (Lindon *et al.* 2005). The third area of prior research is monitoring and predicting response to anticancer drug treatment and prognosis, with the field moving towards individualized medicine based on an individual's unique metabolism (Beloueche-Babari *et al.* 2005, Bathen *et al.* 2007, Chung and Griffiths 2008).

The present work focuses on the application of metabolomics to elucidating the cytotoxic mechanisms of action of both known and potential anticancer agents. Evaluating the response of cancer cells to anticancer therapeutics by differential metabolite profiling gives insights into drug cytotoxicity and tumor cell adaptive mechanisms (Bayet-Robert *et al.* 2010b). As global metabolomics is not presumptive on the involvement of any subset of metabolism, it is an optimal tool for discovering active metabolic pathways and biomarkers. In addition,

tracking metabolite profiles has the potential to uncover vital enzymatic steps that could be targeted in the drug discovery process (Yang *et al.* 2008).

Here we introduce the combined use of GC-MS and HPLC-MS along with chemometric procedures to profile the patterns of various metabolites in taxol- and cisplatin-treated human lung adenocarcinoma A549 and human breast adenocarcinoma MCF7 cells. Such metabolite patterns revealed metabolic networks altered by drug treatment, which in turn serve as guides for further analyses and novel hypotheses regarding the biochemical mechanisms of taxol and cisplatin anticancer action in two different cell types.

4.2 Materials and Methods

4.2.1. Cell Culture and Anticancer Drug Treatments

A full factorial experimental design was used to analyze the effects of and the interactions between three experimental factors: cancer cell type, anticancer drug treatment, and time post-dose. Human lung carcinoma A549 cells (American Type Culture Collection CCL-185, Manassas, VA) were grown in Professor Babak Borhan's laboratory at Michigan State University using F-12K Nutrient Mixture Kaighn's Modification media (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Sigma-Aldrich, St. Louis, MO), 61.4 μg/mL penicillin G (Sigma-Aldrich), 100 μg/mL streptomycin (Sigma-Aldrich), 292 μg/mL L-glutamine (Sigma-Aldrich), and 0.01 M HEPES buffer (pH 7.2) at 37 $^{\circ}$C and 5% CO_2 atmosphere. Human breast adenocarcinoma MCF7 cells (American Type Culture Collection HTB-22) were grown in Eagle Modified Minimum Essential Media (Invitrogen) supplemented with 10% fetal bovine serum, 61.4 μg/mL penicillin G, 100 μg/mL streptomycin, 292 μg/mL L-glutamine, 0.01 mg/mL bovine

insulin (Sigma-Aldrich), and 0.01 M HEPES buffer (pH 7.2) at 37 °C and 5% CO_2 atmosphere. Both types of cells were grown to log-phase, trypsinized, and resuspended in the corresponding media at a final concentration of approximately 6 x 10^5 cells/mL. Aliquots of these cultures (2 mL) were placed in 6-well plates, and the cells were incubated for 10 hours at 37 °C and 5% CO_2 atmosphere to allow for proper attachment to the plate prior to anticancer drug addition.

The A549 and MCF7 cell cultures were treated with two different concentrations of each anticancer drug, one high concentration and one low concentration, based on experimentally determined LD_{50} values. The high concentration dose of taxol (paclitaxel, Sigma-Aldrich) was 5 μg/mL (5.86 μM) and the low concentration dose was 0.1 μg/mL (0.12 μM). The high concentration dose of cisplatin (Sigma-Aldrich) was 5 μg/mL (16.61 μM) and the low concentration dose was 0.1 μg/mL (0.33 μM). Cells were exposed to 5.86 μM taxol, 0.12 μM taxol, 16.61 μM cisplatin, 0.33 μM cisplatin, or the vehicle (dimethyl sulfoxide (DMSO)) in triplicate and were incubated at 37 °C and 5% CO_2 atmosphere for the desired period of time. At various time points post-dose (14 hours, 2 days, 4 days, and 7 days for A549 cells; 1 day, 3 days, 5 days, and 7 days for MCF7 cells), the cells from the 2 mL cultures were washed with PBS to remove detached and dead cells, trypsinized, harvested by centrifugation at 3000 x *g* for 30 seconds at room temperature, and washed with PBS buffer to remove medium components. The final cell pellet was flash-frozen in liquid N_2 and stored at -80 °C until extraction.

4.2.2. Cell Extraction

In order to extract metabolites with a wide range of physical properties, 1 mL ice-cold acetonitrile:isopropanol:water (3:3:2 v/v/v; HPLC-grade 2-propanol and acetonitrile from Merck, Darmstadt, Germany; Milli-Q water, Milli-Q Academic, Millipore, Billerica, MA) extraction solvent was added to the frozen cell pellet (Taylor *et al.* 2010). A 100 μL aliquot of 1 mM adonitol (ribitol, Sigma-Aldrich) in Milli-Q water internal standard solution was also added, and the cell pellet was extracted for 30 minutes at 0 $^{\circ}$C. Finally, the cell pellet was vortexed, centrifuged at 3000 x *g* for 30 seconds at room temperature (20 $^{\circ}$C), and 100 μL aliquots of supernatant were removed for GC-MS and HPLC-MS analyses.

4.2.3. Metabolite Analyses

4.2.3.1. GC-MS Analysis of Metabolites

Prior to GC-MS analysis, the polar metabolites were derivatized by removing a 100 μL aliquot of cell extract solution and evaporating to dryness using a SpeedVac. Aldehyde and ketone groups were methoximated by addition of 50 μL of 10 mg/mL methoxylamine hydrochloride in pyridine (Sigma-Aldrich) for 12-16 hours at room temperature (20 $^{\circ}$C). Acidic groups were then trimethylsilylated with 50 μL of MSTFA (*N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide, Sigma-Aldrich) for 1.5 hours at 60 $^{\circ}$C. The derivatized cell extract was then transferred to a 2 mL autosampler vial containing a 100 μL low volume insert and directly analyzed using an Agilent 6890 GC with 5973 MSD system and ChemStation software (Agilent Technologies, Inc., Santa Clara, CA) with splitless injection of 1.0 μL using an

autosampler. The capillary column was 0.25 mm i.d. x 30 m coated with 0.25 μm HP-5MS (5% phenyl methyl siloxane, Agilent Technologies), and the following conditions were employed: helium flow rate of 1.2 mL/min, injector at 300 $^{\circ}$C, column at 80 $^{\circ}$C for 2 min, followed by a 30 $^{\circ}$C/min ramp to 140 $^{\circ}$C, then 15 $^{\circ}$C/min to 310 $^{\circ}$C held for 5 min (20.33 min total), transfer line at 280 $^{\circ}$C, 70 eV electron ionization, full scan acquisition from *m/z* 50 to 700 at a rate of 2.28 scans/second, and mass calibration performed using perfluorotributylamine. ChemStation GC-MS datafiles were converted to NetCDF format files and then to Waters MassLynx format using the Waters DataBridge program (Milford, MA) for peak detection and integration and for statistical analyses. Metabolites of interest were identified using the ChemStation software by comparing the mass spectrum with the NIST 2008 mass spectral database (National Institute of Standards and Technology, Gaithersburg, MD) as well as mass spectra published in the literature.

4.2.3.2. HPLC-MS Analysis of Metabolites

Lipids in cell extracts were prepared for HPLC-MS analysis by removing a 100 μL aliquot of cell extract solution to an autosampler vial and adding 10 μL of 100 mM ammonium acetate and 10 μL of 0.12 mM propylparaben as an internal standard. Samples were analyzed using a 3.5 μm particle 4.6 x 75 mm Waters Symmetry C18 Reversed Phase Column held at 35 $^{\circ}$C in a Shimadzu column oven. Solvent gradients were delivered by ternary Prominence LC-20AD HPLC pumps (Shimadzu, Columbia, MD) and a Shimadzu SIL-5000 LC-autosampler coupled to a Waters LCT Premier time-of-flight mass spectrometer. Non-targeted lipid screening was

performed using a 12-minute HPLC gradient using a flow of 0.4 mL/min. Solvent A (10 mM ammonium formate), solvent B (2-propanol), and solvent C (acetone) were programmed using linear gradients as follows: initial conditions 99% A/1% B for 0.5 minutes, linear to 10% A/90% B at 2 minutes, then linear to 100% B at 6 minutes, holding at 100% B for 2 minutes, then 100% C at 8 minutes, holding at 100% C for 2 minutes, then back to initial conditions at 10 minutes for the remaining 2 minutes. Eluted molecules were ionized for MS analysis using electrospray ionization in negative mode with a capillary voltage of -3 kV. Multiplexed non-selective collision-induced dissociation was used to generate quasi-simultaneous spectra with varying degrees of fragmentation by maintaining the cone voltage at 25 V and switching Aperture 1 voltages among 15, 30, 45, 60 and 75 V with a scan acquisition time of 0.2 seconds at each voltage (Gu *et al*. 2010). Mass analysis was performed using V-mode ion optics at a mass resolution (full width half-maximum) of approximately 5000, and full-mass scan data were collected in centroid mode over *m/z* 100-1500 using the instrument's dynamic range extension feature. Data were acquired and analyzed using Waters MassLynx software (version 4.1).

4.2.4. Data Pre-Treatment and Chemometric Procedures

Waters MarkerLynx software (version 4.1) was used to pre-process both the GC-MS and HPLC-MS data sets separately prior to multivariate statistical analyses. Internal standard and contaminant retention time windows were excluded due to their presence in every sample. Each sample chromatogram was processed using chromatographic peak detection, peak integration, and retention time alignment, with thresholds set to eliminate low-level signals. The resulting peak areas were organized using a combination of retention time (RT) and *m/z*

ratio, known as an RT-*m/z* pair or marker. Markers not detected or with signal intensities below the threshold value in a sample were assigned a value of zero in the matrix. The detected peak areas within each sample were then normalized to the sum of all peak areas in that sample, which was set to 10000 for all samples. The normalized data matrix was then exported to SIMCA-P (software version 11.5, Umetrics, San Jose, CA) for further pre-treatment and multivariate statistical analyses. The data matrix exported from MarkerLynx was mean-centered and scaled to Pareto variance prior to supervised chemometric procedures. Partial Least Squares or Projection on Latent Structures-Discriminant Analyses (PLS-DA) and Orthogonal Partial Least Squares or Projection on Latent Structures-Discriminant Analyses (OPLS-DA) were carried out using appropriate Y variables indicating which treatment group each cancer cell sample belonged to, with the X variables being the markers (peak area for each RT-*m/z* pair) in the data matrix. The PLS-DA models were used to visualize how several treatment groups were similar to one another, and OPLS-DA was used to compare only two treatment groups per model. The OPLS-DA models were validated using the Leave-One-Out Cross-Validation method. Finally, Student's *t*-test and one-way analysis of variance (ANOVA) comparisons were made between metabolite levels of interest using Microsoft Excel (Redmond, WA). Statistical tests were two-tailed and differences were considered statistically significant for $p < 0.05$. Biologically relevant metabolic pathways involving the identified metabolites were determined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) online database (http://www.genome.jp/kegg/).

4.3 GC-MS and HPLC-MS Analyses of Untreated Lung Cancer and Breast Cancer Cells

The metabolites present in untreated A549 lung cancer and MCF7 breast cancer cell extracts were analyzed using both GC-MS and HPLC-MS to investigate differences between the two cell lines and for comparison after treatment with anticancer drugs. The cell extracts were derivatized (methoximated and trimethylsilylated) before GC-MS analysis so that the less volatile polar metabolites would be included in the results. Both the GC-MS and HPLC-MS analyses were conducted using non-targeted methods to detect as many of the metabolites present in the cell extracts as possible. Example total ion chromatogram results from GC-MS analyses of the untreated A549 lung cancer and MCF7 breast cancer cell extracts collected on day 0 of the study are displayed in Figure 4.2. The information-rich chromatograms were comprised of approximately 300-500 peaks corresponding to detected compounds. The GC-MS analyses were also rapid, with metabolite profiles collected during a sample analysis time of only 20.33 minutes. Approximately 20 of the abundant peaks visible in Figure 4.2 are present in both cell types, however, their levels were 10-60% higher in the lung cancer cell extract compared to the breast cancer results. Differences in metabolite abundances between the two cell types were expected, as lung tissue and breast tissue have very different functions in the human body. The A549 lung cancer cells are derived from alveolar epithelial cells which have normal functions of substance transport relevant to respiration, and are expected to exhibit greater metabolic activity compared to breast cells, which are engaged in lipid metabolism and hormone response, and are positive for estrogen receptor. The GC-MS results supported these anticipated metabolic differences, as more peaks (~500 compared to ~300) and peaks with greater abundances were observed in the lung cancer cell extracts compared to the breast

cancer cell extracts. These baseline metabolomic analyses of the A459 lung cancer and MCF7 breast cancer cells at the beginning of the study were important for assessing significant metabolism changes in response to anticancer drug treatment and changes over the time course of the study.

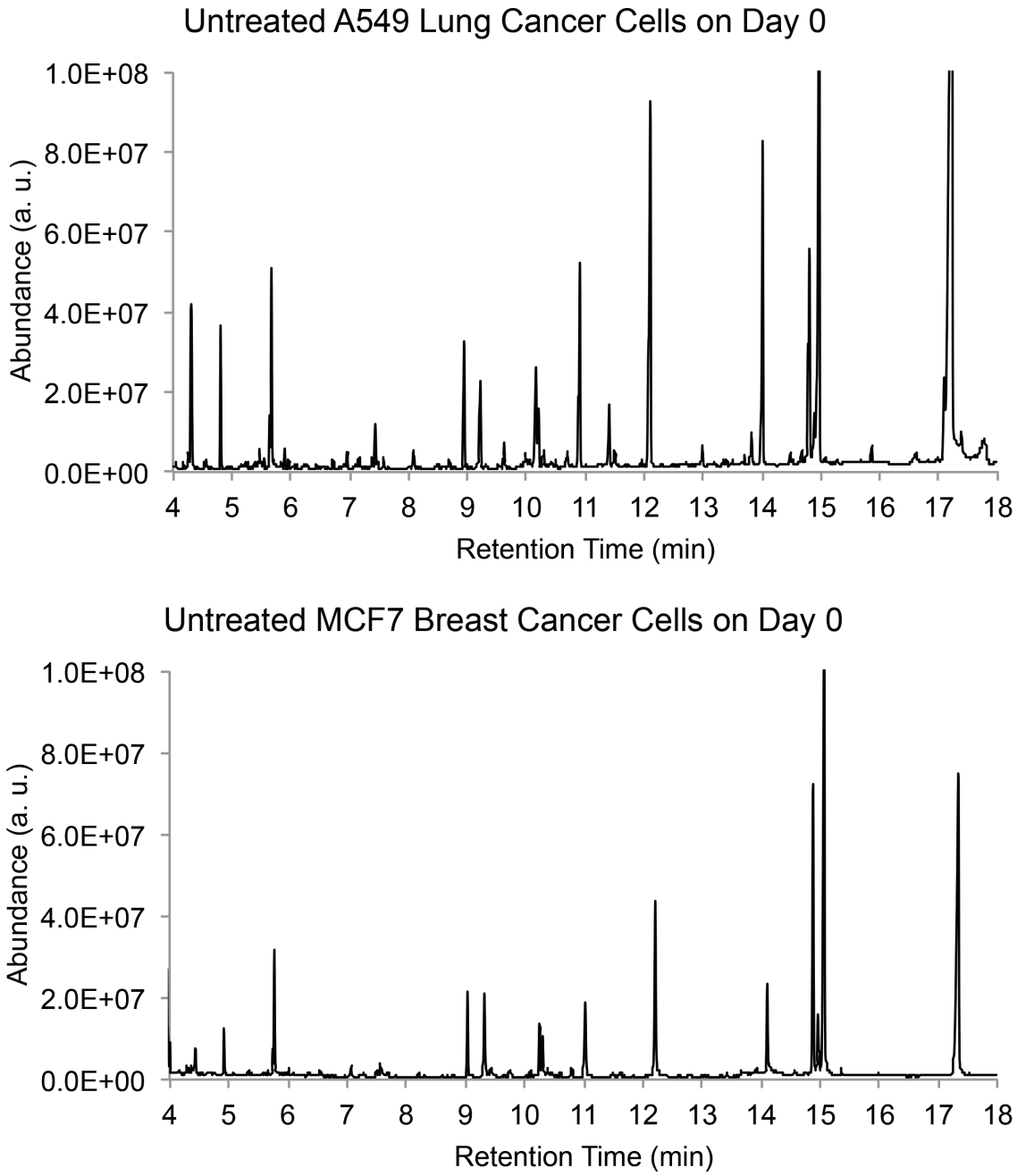

Figure 4.2. Example total ion chromatograms from GC-MS analyses of untreated A549 lung cancer and MCF7 breast cancer cells on day 0 of the study. The chromatograms are plotted on the same abundance (*y*-axis) scale for comparison.

Example total ion chromatogram results from HPLC-MS analyses of the untreated A549 lung cancer and MCF7 breast cancer cell extracts collected on day 0 of the study are displayed

in Figure 4.3. The HPLC-MS data were generated using multiplexed non-selective collision-induced dissociation to generate quasi-simultaneous spectra with varying degrees of fragmentation to aid in metabolite identification. The reversed phase HPLC separation and electrospray ionization in negative mode yielded non-targeted lipid screening results rapidly in a 12-minute sample analysis. While the chromatographic peaks from HPLC-MS analysis are not resolved as well as the GC-MS results, analyte separation and identification based on the mass spectral information was still achieved. Since electrospray ionization is a gentle process when low Aperture 1 potentials are employed, the mass spectra are simpler, usually lacking fragment ions. As a result, overlapping LC-MS chromatographic peaks are more readily resolved using ion masses than can be achieved with GC-MS, where spectra have a wide assortment of fragment ions. Similar to the GC-MS results discussed above, some of the detected metabolites (*e.g.* those eluting from 5.1-5.4 minutes) are approximately 50% more abundant in the A549 lung cancer cells compared to the MCF7 breast cancer cells (Figure 4.3). Further data analyses identifying metabolite peaks and comparing HPLC-MS results between the untreated A549 lung cancer and MCF7 breast cancer cells and cells treated with anticancer agents over time post-dosing are discussed below.

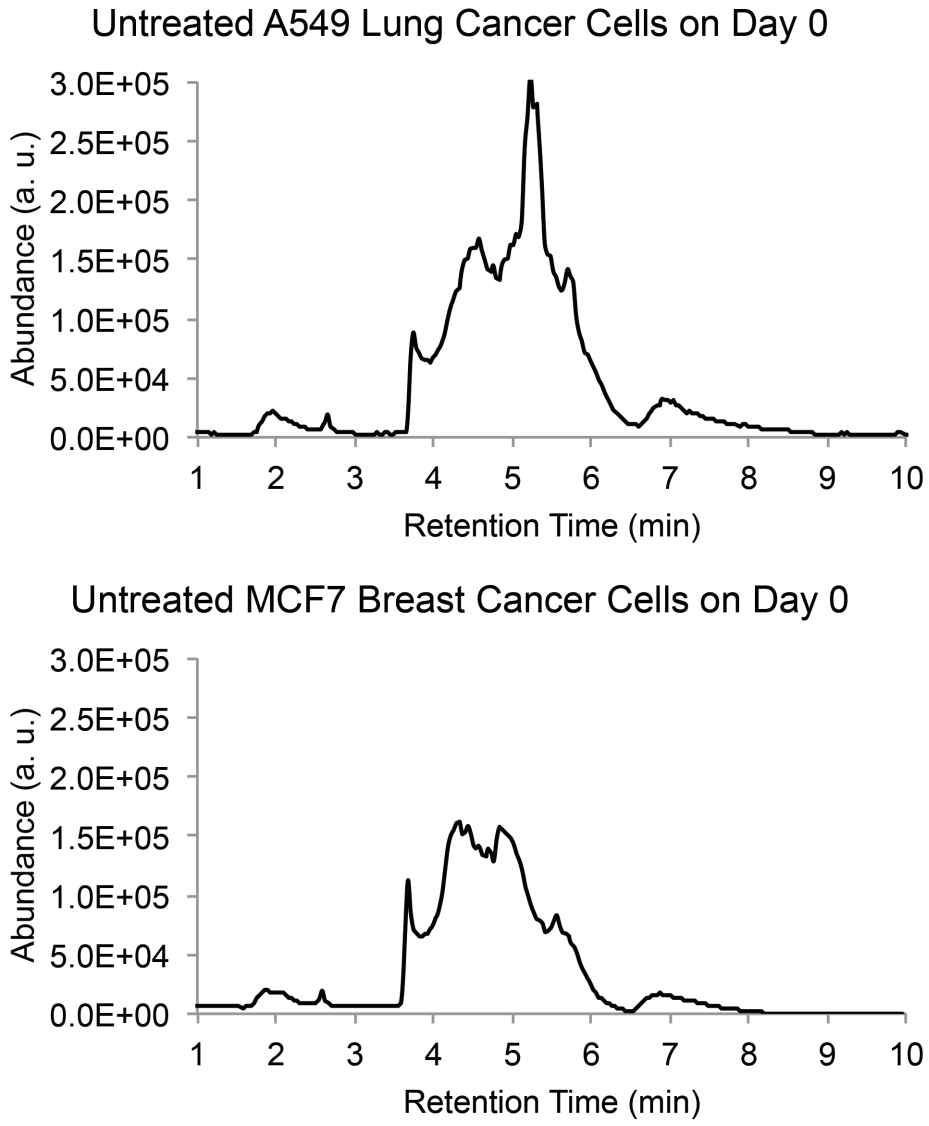

Figure 4.3. Example total ion chromatograms from HPLC-MS analyses of untreated A549 lung cancer and MCF7 breast cancer cells on day 0 of the study. The total ion abundances measured quasi-simultaneously at the five Aperture 1 voltages were summed and plotted on the first voltage's retention time scale for straightforward chromatogram visualization and comparison. The chromatograms are plotted on the same abundance (*y*-axis) scale for comparison as well.

These baseline metabolomic analyses of the untreated A459 lung cancer and MCF7 breast cancer cells at the outset of the study were essential for assessing significant metabolism changes in response to treatment with anticancer drugs over the time course of the study.

The similarities and differences between untreated lung and breast cancer cell metabolite profiles shown in Figures 4.2 and 4.3 visualize only one dimension of metabolomic information collected, the chromatography information. Including the underlying second dimension of information, the mass spectra, in the assessment of sample similarity yields a dramatic increase in information content that can be used to assess differences in metabolite profiles more deeply. Defining metabolite peaks using both retention time and ions observed in the mass spectra generates metabolite abundance values less subject to interference by co-eluting substances. Due to the complexity of the GC-MS and HPLC-MS data, multivariate statistical data analyses are required to assess significant differences between samples and relationships between multiple metabolite levels. All of the information (peak areas organized by retention time and ion *m/z* values) contained in the GC-MS and HPLC-MS metabolomics data was used to differentiate metabolic response of A549 lung cancer and MCF7 breast cancer cells to treatment with two anticancer drugs in order to gain knowledge about how these compounds affect metabolic phenotypes.

4.4 Anticancer Drug-Induced Changes in Lung Cancer and Breast Cancer Metabolite Profiles

4.4.1 Partial Least Squares-Discriminant Analyses

The A549 lung cancer and MCF7 breast cancer cell lines were dosed with the anticancer drugs taxol and cisplatin and cells were collected at time points up to seven days post-dose to assess changes in cellular metabolism due to the dose of the therapeutic agent. Both high and low doses of taxol and cisplatin were administered to the cells for comparison as well. After all cell pellets were collected, they were extracted, derivatized (GC-MS analyses only), and

analyzed using GC-MS and HPLC-MS. The vast amount of information contained in the GC-MS and HPLC-MS metabolomic results for the differentially treated cancer cell samples was then evaluated using partial least squares-discriminant analysis (PLS-DA). The PLS-DA method is a supervised multivariate pattern recognition model, meaning that a Y variable is added to the data matrix that informs the model of which class a sample belongs to (*e.g.* was the sample treated with the high dose of taxol or not). The goal of PLS-DA is to find the metabolites (RT-*m/z* pairs; X variables) whose abundances contribute most to discriminating samples based on class assignment (input Y variable). The metabolite signals that display similar variation and correlation patterns in the data set are then linearly combined to form principal components (PCs). The first principal component describes the largest and most correlated sources of variation, and the degree of variation and correlation decreases with the subsequent principal components. The first two principal components are then plotted as the *x*- and *y*-axes of a single chart (scores plot) to visualize sample relationships. Samples that are clustered together on the scores plot have similar metabolite profiles, while samples that are spatially distant have substantial differences in metabolite levels. A corresponding PLS-DA loadings plot is also generated which displays the metabolites responsible for the positioning of the samples on the scores plot. In this manner, the metabolite levels that are discriminating between treatment groups are recognized so that they may be identified and used to interpret the effects of anticancer drug treatment on the metabolism of cancer cells.

The PLS-DA models used to evaluate the metabolism changes in A549 lung cancer and MCF7 breast cancer cells due to treatment with taxol and cisplatin were performed on only one cancer cell type (breast or lung cancer) using metabolomics data from one analytical method

(GC-MS or HPLC-MS) per model to decrease complexity in results interpretation. A Y variable indicating whether samples were treated with a drug (either taxol or cisplatin) or was an untreated control was used in all PLS-DA models. This allowed comparisons between all three treatments (taxol-treated, cisplatin-treated, and untreated control) to be visualized and assessed in a single mathematical model. Results from A549 lung cancer and MCF7 breast cancer cells treated with the low doses of taxol and cisplatin did not display appreciable metabolite profile differences from the untreated cells over the course of the study (data not shown) and were excluded from further analyses. This result was not surprising since the low doses were substantially lower than the LD_{50} values. The cells treated with the high doses of taxol and cisplatin and collected over the course of seven days displayed differential metabolite profiles, both from untreated cells and between the two different treatments. The PLS-DA scores plots illustrating the clustering of similar samples together and away from dissimilar samples in the four data subsets (lung cancer GC-MS data, breast cancer GC-MS data, lung cancer HPLC-MS data, and breast cancer HPLC-MS data) are displayed in Figures 4.4 – 4.7, respectively.

The PLS-DA scores plot for the A549 lung cancer samples treated with the high dose of taxol or cisplatin as well as the untreated control samples collected at various time points up to seven days post-dose and analyzed using GC-MS is shown in Figure 4.4. Principal component 1 (*x*-axis) accounted for 45% of the variation in the data set, and PC2 (*y*-axis) explained 33% of the data set variation, meaning that 78% of the variance in this sample set was visualized in one plot. The cisplatin-treated, taxol-treated, and untreated control sample groups all clustered together and away from differently treated samples, indicating substantial treatment-

dependent metabolite profile changes. Anticancer drug treatment is therefore a more dominant factor than the time post-dose for creating significant metabolome variations. This PLS-DA scores plot indicated treatment-dependent changes were induced in A549 lung cancer cells that were detectable using GC-MS metabolomics techniques. Identifying the metabolites that contribute to the principal components from the corresponding loadings plot (not shown) will likely reveal metabolic networks altered by taxol- and cisplatin-treatment in A549 lung cancer cells.

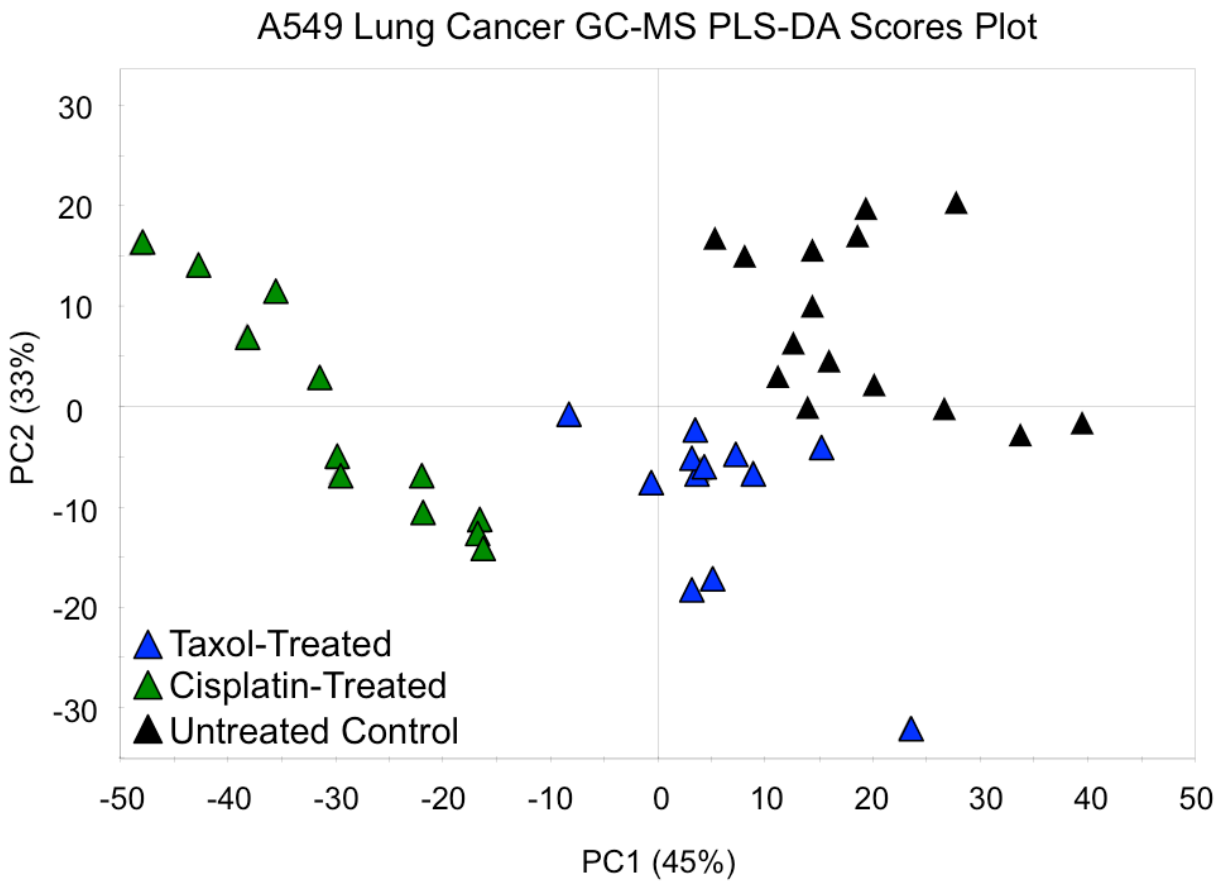

Figure 4.4. Scores plot resulting from PLS-DA showing the relationships between A549 lung cancer samples treated with the high doses of taxol and cisplatin and untreated controls (all time points included) analyzed using GC-MS. Samples with similar metabolic profiles group together and away from dissimilar samples.

The PLS-DA scores plot for the MCF7 breast cancer samples treated with the high dose of taxol or cisplatin as well as the untreated control samples collected at various time points up to seven days post-dose and analyzed using GC-MS is shown in Figure 4.5. Principal component 1 (*x*-axis) accounted for 32% of the variation in this data set, and PC2 (*y*-axis) explained 14% of the data set variation, meaning that 46% of the variance in this sample set was visualized in one plot, which was less than the corresponding plot from the A549 lung cancer sample set (Figure 4.4). The cisplatin-treated, taxol-treated, and untreated control sample groups also do not cluster as clearly in the MCF7 results compared to the A549 results. The taxol-treated and untreated samples substantially overlap, and an untreated sampled is also grouped with the cisplatin-treated samples. This indicates that more principal components are necessary to differentiate the treatment groups. Drug treatment-dependent changes were still observed in the GC-MS data collected from MCF7 breast cancer cells which are more complex (fewer metabolites correlated for combination into principal components) compared to the A549 lung cancer cell results. Examining the metabolites responsible for sample clustering in the scores plot may explain why some differentially-treated samples had similar metabolite profiles, and would point to biochemical pathways perturbed by taxol- and cisplatin-treatment in MCF7 breast cancer cells.

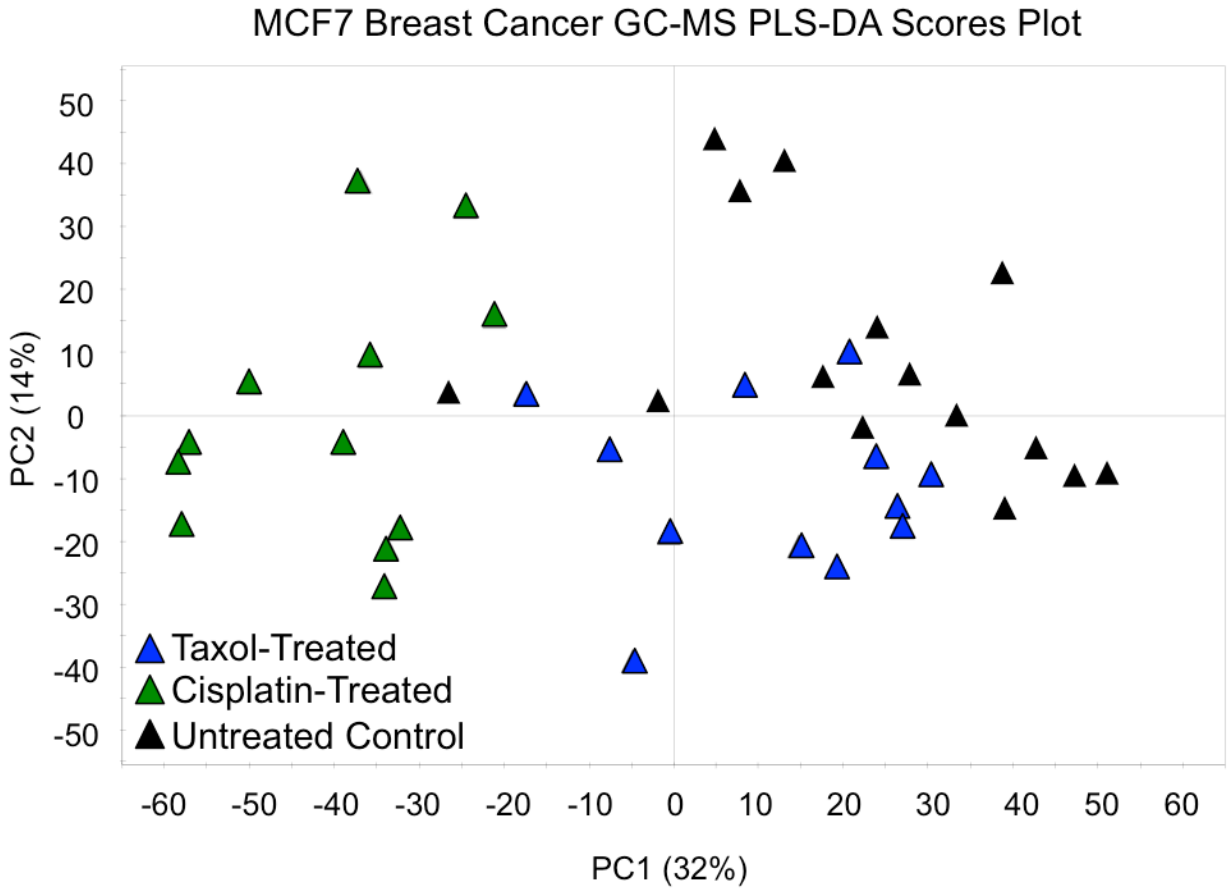

Figure 4.5. Scores plot resulting from PLS-DA showing the relationships between MCF7 breast cancer samples treated with the high doses of taxol and cisplatin and untreated controls (all time points included) analyzed using GC-MS. Samples with similar metabolic profiles group together and away from dissimilar samples.

The PLS-DA scores plot calculated using HPLC-MS data from the same A549 lung cancer cell samples analyzed using GC-MS data in Figure 4.4 is shown in Figure 4.6. Principal component 1 (*x*-axis) accounted for 66% of the variation in the data set, and PC2 (*y*-axis) explained 16% of the data set variation, meaning that 82% of the variance in this sample set is visualized in one plot. Similar to the PLS-DA results using the GC-MS data, the cisplatin-treated, taxol-treated, and untreated control sample groups all clustered together and away from differently treated samples. This indicated substantial treatment-dependent metabolite profile

changes, which were not as dependent on the time post-dose that the sample was collected. Differential metabolite profiles were detected in A549 lung cancer cells using both GC-MS and HPLC-MS analyses, and identification of the metabolites with changing levels after treatment with taxol or cisplatin will point to biochemical pathways involved in the cytotoxic response.

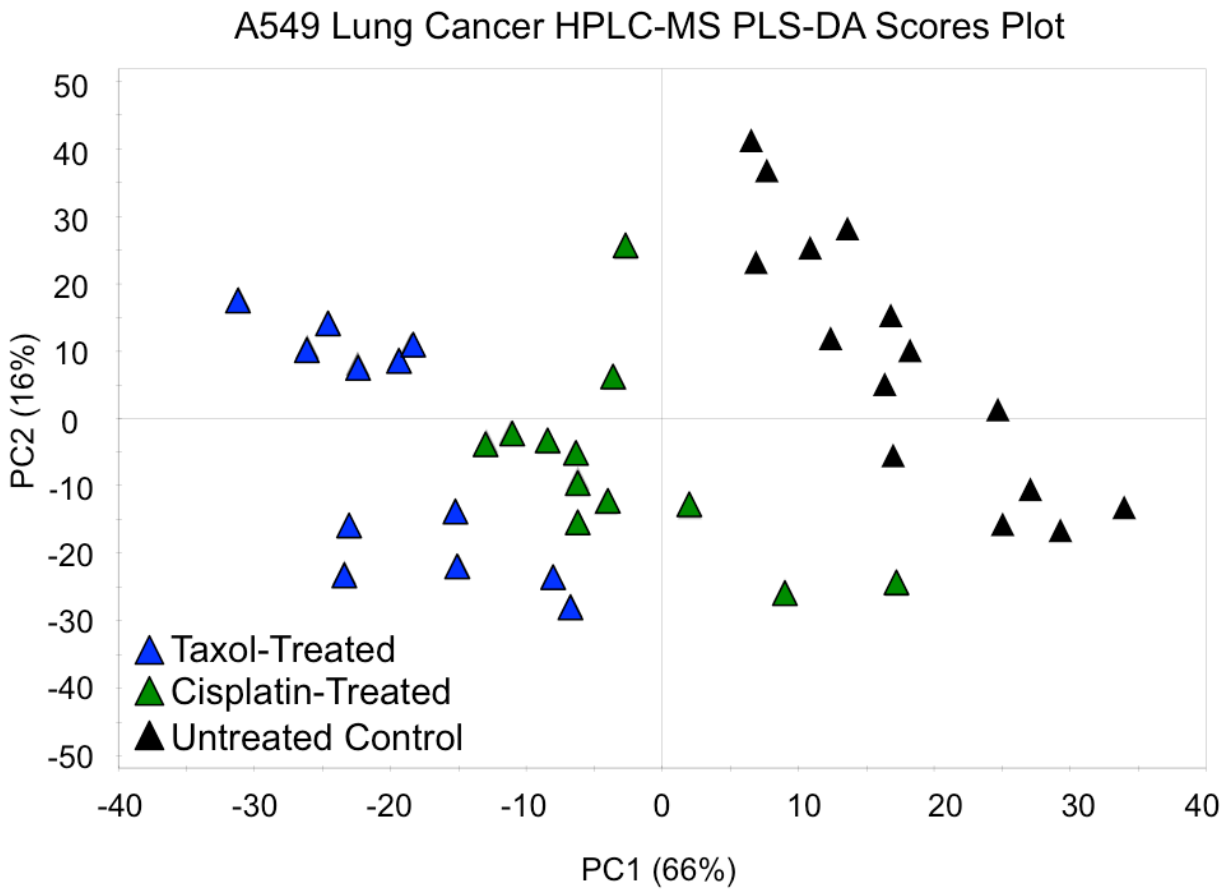

Figure 4.6. Scores plot resulting from PLS-DA showing the relationships between A549 lung cancer samples treated with the high doses of taxol and cisplatin and untreated controls (all time points included) analyzed using HPLC-MS. Samples with similar metabolic profiles group together and away from dissimilar samples.

The PLS-DA scores plot from HPLC-MS analyses of the same MCF7 breast cancer samples that were presented in Figure 4.5 after GC-MS analysis is shown in Figure 4.7. Principal

component 1 (x-axis) accounted for 40% of the variation in this data set, and PC2 (y-axis) explained 32% of the data set variation, meaning that 72% of the variance in this sample set was visualized in one plot. Similar to the PLS-DA scores plot from GC-MS analysis, there was some overlapping between the differentially treated sample groups. The untreated samples clustered together and overlapped with a couple of the cisplatin-treated samples, as observed in Figure 4.5 as well. The taxol-treated and cisplatin-treated samples also overlap substantially, more so in the HPLC-MS results compared to the GC-MS results. The MCF7 breast cancer cells may still have treatment-dependent metabolite profile changes, however, more focused multivariate statistical comparisons (*e.g.* between taxol-treated and cisplatin-treated samples only) are necessary to identify significantly changing metabolite levels in these samples.

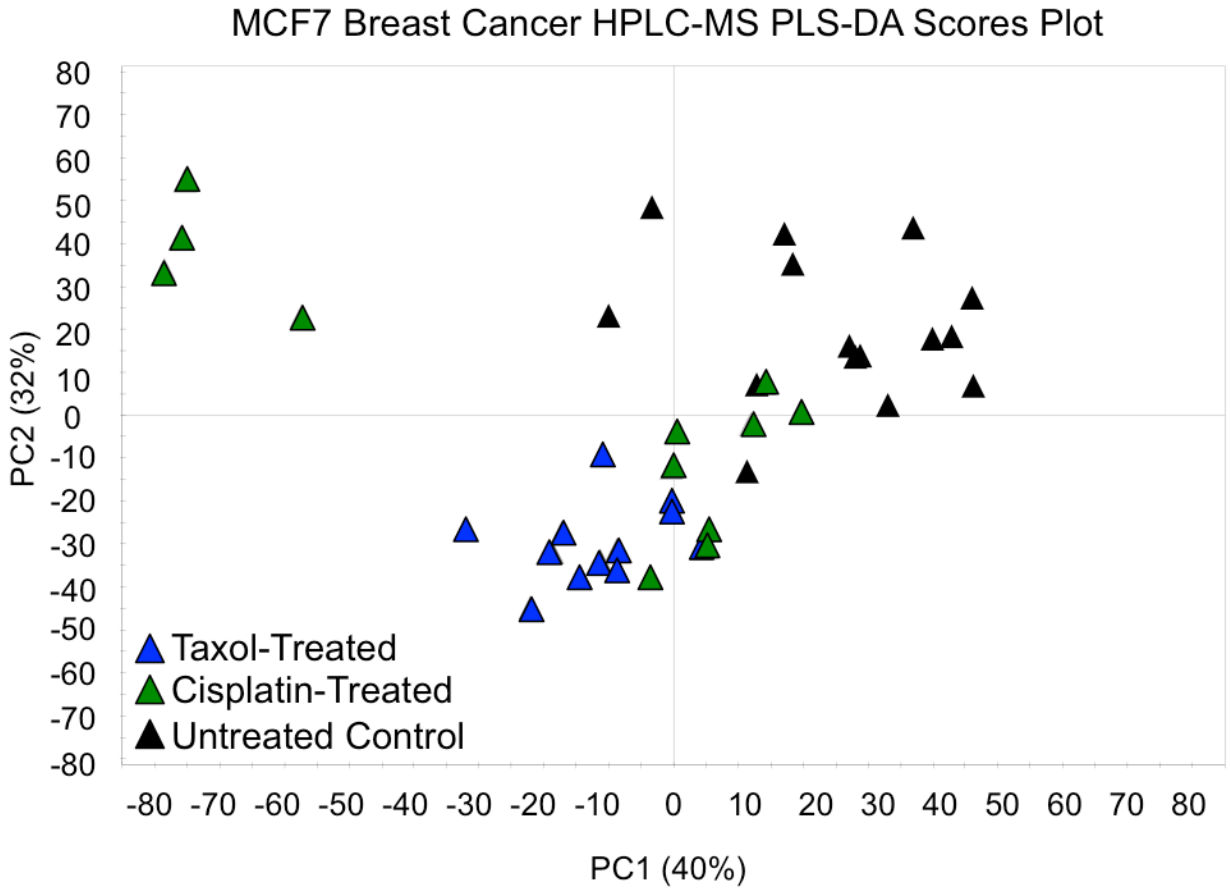

Figure 4.7. Scores plot resulting from PLS-DA showing the relationships between MCF7 breast cancer samples treated with the high doses of taxol and cisplatin and untreated controls (all time points included) analyzed using HPLC-MS. Samples with similar metabolic profiles group together and away from dissimilar samples.

In summary, the PLS-DA scores plot results of metabolite profiles measured using both GC-MS and HPLC-MS showed sample clusters that were treatment-dependent in A549 lung cancer cells. Identifying metabolites contributing to sample positions may indicate the biochemical pathways involved in cytotoxic response of lung cancer cells to taxol- and cisplatin-treatment. The PLS-DA results for metabolomic analyses of MCF7 breast cancer cells treated with taxol or cisplatin and untreated controls were not as clearly grouped as the A549 results. More specific multivariate statistical data analyses, such as orthogonal partial least squares

discriminant analyses (OPLS-DA) which remove all sources of variance in the data set that are not correlated to the Y variable, may reveal treatment-distinguishing metabolites in the MCF7 samples. In addition, comparing the taxol-treated and cisplatin-treated samples only (excluding the untreated control samples from the mathematical models) at each time point separately may uncover more metabolites that distinguish the two treatments, rather than distinguishing treated and untreated cell samples. Finally, supervised multivariate statistical models need to be validated to ensure that they are describing significant metabolite level changes and not spurious results in the underlying analytical variance in the data set. Therefore, the A549 lung cancer and MCF7 breast cancer GC-MS and HPLC-MS data sets were statistically analyzed further using OPLS-DA in order to isolate metabolite levels that differentiate taxol- and cisplatin-treatment.

4.4.2 Orthogonal Partial Least Squares-Discriminant Analyses

The orthogonal partial least squares discriminant analysis (OPLS-DA) method is another type of supervised multivariate pattern recognition model. The X and Y matrices are the same as in PLS-DA models, and the goal is still to find the metabolites (RT-*m/z* pairs; X variables) whose abundances contribute most to discriminating samples based on class assignment (input Y variable). The difference is that in OPLS-DA the systematic variation in the X matrix that is not correlated (orthogonal) to the Y matrix is removed from the analysis. This reduces the complexity of the model while preserving the interpretation of the variables (metabolites) correlated with sample class. As all sample variation in the data set is removed except for metabolite peak area changes correlated with class assignment, only one principal component

is calculated. This makes scores plots, like those discussed above resulting from PLS-DA, less useful because the samples are perfectly grouped according to class. The loadings plot displaying the metabolite signals correlated with sample class (*i.e.* taxol or cisplatin treatment) is also different in OPLS-DA, and an example is presented in Figure 4.8. The principal component is plotted on the *x*-axis, where metabolite position is primarily dictated by abundance, as more abundant compounds have more influence on the mathematical model. The degree of correlation of each variable with the class assignment, called p(corr), is the *y*-axis. The OPLS-DA shown in the example used a discriminant Y variable where samples treated with taxol were assigned a value of 1 and samples treated with cisplatin were assigned a value of 0. Therefore, metabolites (variables) whose abundances were significantly increased in taxol-treated samples compared to cisplatin-treated samples were linearly correlated with the Y variable and therefore displayed positive values on the p(corr) *y*-axis. Conversely, metabolites whose abundances were significantly increased in cisplatin-treated samples relative to taxol-treated samples were inversely correlated with the Y variable and had negative positions on the p(corr) *y*-axis. Note that metabolite abundance does not play a role in the evaluation of correlation, only the significance of the difference in metabolite levels between the two sample classes, which is a strength of the OPLS-DA approach. The resulting OPLS-DA loadings plot is called an “S” plot due to the spatial distribution of the metabolites in the plot. Metabolites highly linearly or inversely correlated with sample class are positioned at the upper and lower limits of the *y*-axis and are chiefly responsible for discriminating sample class.

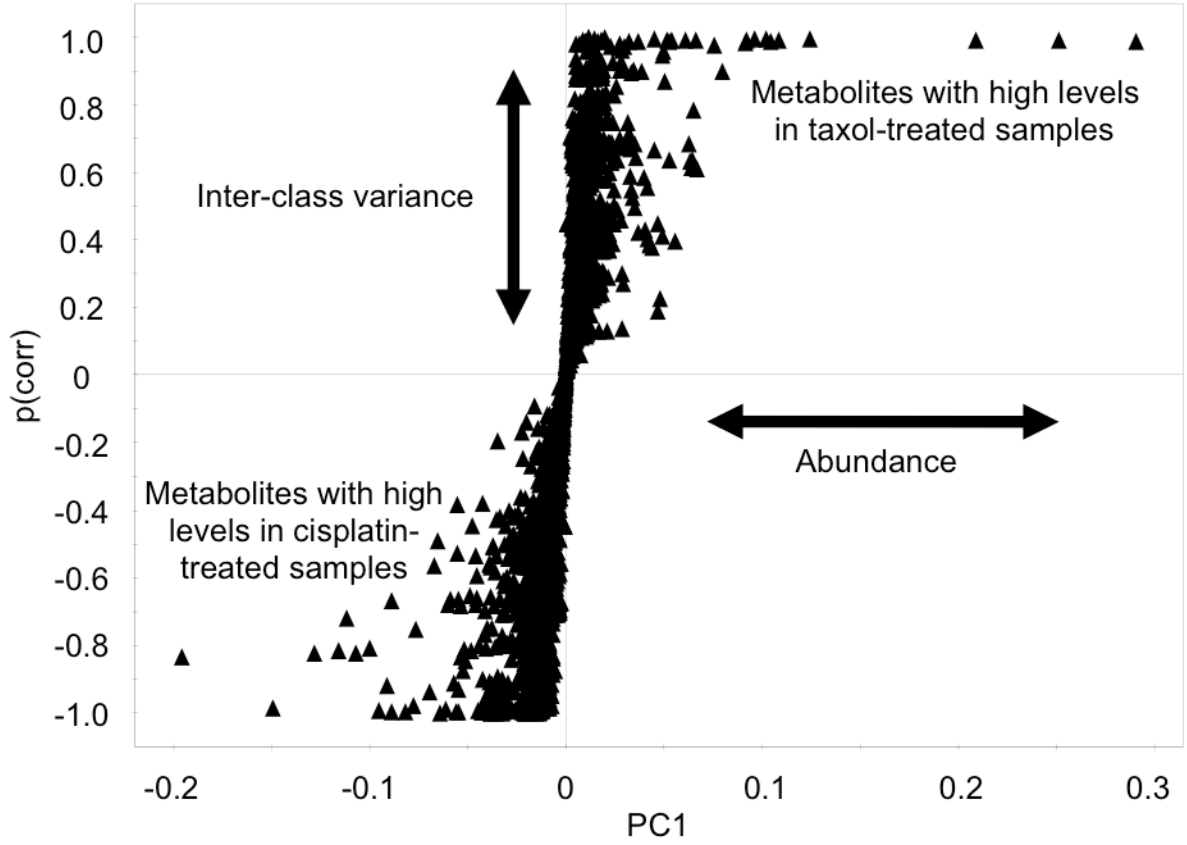

Figure 4.8. An example OPLS-DA loadings "S" plot showing metabolites detected in GC-MS spectra that differentiate A549 lung cancer cells seven days post-dosing with the high doses of taxol and cisplatin.

An OPLS-DA model was calculated to compare taxol- and cisplatin-treatment at each time point post-dose individually for each cancer cell type (A549 lung cancer and MCF7 breast cancer) from both GC-MS and HPLC-MS data sets. The OPLS-DA models isolated the peaks most responsible for differences between treatment effects for identification and association with biochemical responses to anticancer drug treatment. However, as each OPLS-DA model contained only six samples (*e.g.* A549 cells treated with the high dose of taxol or cisplatin in biological triplicate, collected two days post-dose, analyzed using GC-MS), model validation was essential to ensure the models were describing metabolite abundance variations and not analytical variance in the data set.

4.4.3 OPLS-DA Model Validation

While OPLS-DA is a powerful chemometric method for highlighting the metabolites correlated with anticancer drug treatment, the models must be validated to ensure that the correlations are not due to random analytical variance. Supervised statistical models for data sets containing more variables (measured metabolite peak areas) than samples are susceptible to overfitting, which occurs when statistical models describe the random noise in the data instead of the interesting sample relationships superimposed on the noise, which is undesirable. The OPLS-DA models comparing the metabolite profiles (measured using either GC-MS or HPLC-MS) of A549 lung cancer cells or MCF7 breast cancer cells treated with taxol or cisplatin were validated using the leave-one-out cross-validation method. The leave-one-out cross-validation method leaves one sample out of the population as the testing sample (*e.g.* one of the replicate A549 samples treated with cisplatin on day 2 post-dose is left out of the analysis) and uses the rest of the samples as the training set (*e.g.* the other two replicate A549 samples treated with cisplatin on day 2 post-dose as well as all three replicate A549 samples treated with taxol on day 2 post-dose) to fit an OPLS-DA model using a Y variable indicating sample class (*e.g.* treated with taxol or cisplatin). The calculated OPLS-DA model is then used to assign the testing sample to one of the sample classes. The model validation result is a root mean squared error of prediction (RMSEP), which is the likelihood that the testing sample will be mis-classified. The OPLS-DA model validation process is repeated using each sample in the population as the testing sample once, and the RMSEP values for each validation model are averaged to yield the mean RMSEP value for the OPLS-DA model. A value of 0.5 for the mean RMSEP indicates a 50% chance of sample mis-classification using the model being validated,

which is equivalent to random chance. Lower RMSEP values signify more robust mathematical models with higher predictive ability. Tables reporting the leave-one-out cross-validation results for the OPLS-DA models of A549 lung cancer and MCF7 breast cancer cells treated with the high dose of either taxol or cisplatin at various time points post-dose are displayed below. Table 4.1 lists the mean RMSEP values for OPLS-DA models generated for GC-MS metabolite profiles, and Table 4.2 shows the mean RMSEP results for models generated using the metabolite profiles measured using HPLC-MS.

Table 4.1. Model validation results for OPLS-DA of GC-MS data at each time point post-dose with taxol or cisplatin (RMSEP = root mean squared error of prediction).

A549		MCF7	
Time Post-Dose	Mean RMSEP	Time Post-Dose	Mean RMSEP
14h	0.277	1d	0.139
2d	0.409	3d	0.179
4d	0.261	5d	0.403
7d	0.255	7d	0.221

Table 4.2. Model validation results for OPLS-DA of HPLC-MS data at each time point post-dose with taxol or cisplatin (RMSEP = root mean squared error of prediction).

A549		MCF7	
Time Post-Dose	Average RMSEP	Time Post-Dose	Average RMSEP
14h	0.383	1d	0.439
2d	0.407	3d	0.455
4d	0.423	5d	0.385
7d	0.382	7d	0.407

The OPLS-DA model validation results listed in Table 4.1 indicated that metabolite profiles generated from GC-MS analyses of A549 lung cancer cells and MCF7 breast cancer cells treated with the high dose of either taxol or cisplatin and analyzed at various time points post-dose had varying levels of predictive ability. All mean RMSEP values were less than 0.5 (50%

chance an unknown sample will be mis-classified), indicating that all of the OPLS-DA models were more accurate than random chance for classifying an unknown sample correctly. The mean RMSEP values for A549 cells ranged from 0.255 (seven days post-dose) to 0.409 (two days post-dose). An examination of the raw GC-MS data for the A549 samples collected two days post-treatment with either taxol or cisplatin did not reveal a reason for the decreased predictive ability of the OPLS-DA model for that data subset. The same was true for the MCF7 cell samples collected five days post-treatment with the high dose of either taxol or cisplatin, which yielded a mean RMSEP value of 0.403, indicating a poor predictive ability for the OPLS-DA model generated from that data subset. The lowest mean RMSEP value (highest predictive ability) for an OPLS-DA model of GC-MS results comparing differentially treated MCF7 cells was 0.139, generated from samples collected one day post-dose. There was no meaningful relationship between the mean RMSEP value and the time point that the samples were collected. With the exceptions of the A549 cells sampled two days post-dose and the MCF7 cells collected five days-post dosing, all OPLS-DA models generated using GC-MS data displayed < 28% chance of mis-classifying an unknown sample. This means that identifying the metabolites contributing to the significant differences between taxol- and cisplatin-treated lung and breast cancer cells highlighted by the robust OPLS-DA models will be useful for elucidating the differential cytotoxic responses of cancer cells to anticancer drug treatment.

The mean RMSEP results from validating OPLS-DA models calculated using HPLC-MS metabolomics data for A549 and MCF7 cells treated with the high dose of either taxol or cisplatin in Table 4.2 are higher in comparison to the GC-MS results in Table 4.1. Mean RMSEP values ranged from 0.382 (seven days post-dose) to 0.423 (four days post-dose) in the A549

cells, and 0.385 (five days post-dose) to 0.455 (three days post-dose) in the MCF7 cells. All of the OPLS-DA models calculated using the HPLC-MS data had a > 30% chance of mis-classifying an unknown sample, indicating poor predictive ability using HPLC-MS data. Due to these OPLS-DA model validation results, the HPLC-MS results were excluded from further consideration as too few metabolites varied significantly between taxol- and cisplatin-treatment to draw conclusions about metabolic phenotype measured using HPLC-MS. The leave-one-out cross-validation results for metabolite profiles measured using GC-MS indicated that the OPLS-DA models were robust, so efforts were undertaken to identify the metabolites and the associated biochemical pathways disturbed by anticancer agents in lung and breast cancer cells.

4.5 Biochemical Pathways Perturbed by Chemotherapeutic Agents

Successful sample classifications visualized using multivariate statistical procedures are generally associated with a complex pattern of many metabolite signals, all of which contribute to the model in varying amounts. Nevertheless, some of the important metabolites can be tentatively identified. The 200 highest correlated metabolites (RT-*m/z* pairs) in both the positive and negative direction in the OPLS-DA of the GC-MS metabolite profiles were selected for attempted identification, regardless of abundance. The high abundance metabolites were more successfully annotated compared to the low abundance metabolites, as the low signal-to-noise (S/N) ratios made mass spectral comparisons with the database difficult. Spectra for a handful of the higher abundance metabolites did not match any reference spectra and remained unidentified as well, as is frequently the case for such metabolomic analyses. Table 4.3 lists the annotated metabolites determined by OPLS-DA to differentiate treatment types in A549 lung

cancer and MCF7 breast cancer cells. The levels of the metabolites that were annotated were examined over time and in all treatment types (untreated control, taxol-treated, and cisplatin-treated) for relationships. Ability to differentiate treated from control and the two treatment types in a manner that was consistent over time identified the most important metabolites with which to assess the biochemical pathways affected by anticancer drug treatment.

Table 4.3. Annotations for metabolites that differentiate treatments according to OPLS-DA and their retention times (RT) in the GC-MS spectra of A549 lung cancer and MCF7 breast cancer cells. The slight shift in retention times between the A549 and MCF7 results is due to column maintenance (*e.g.* clipping or replacement) between sample set analyses.

Metabolite	Abbreviation	A549 RT (min)	MCF7 RT (min)
Lactate	Lac	4.32	4.44
Alanine	Ala	4.56	4.68
Hydroxylamine	HA	4.65	- *
Valine	Val	5.27	5.39
Urea	Urea	5.42	-
Benzoic Acid	BA	5.47	-
Glycerol	Gol	5.64	5.75
Phosphate	Pho	5.67	5.78
Proline	Pro	5.83	5.95
Glycine	Gly	5.89	6.01
Serine	Ser	6.23	6.35
Threonine	Thr	6.43	6.55
β-Alanine	β-Ala	6.71	-
Malate	Mal	7.14	-
Aspartate	Asp	7.38	7.50
Glutamate	Glu	7.43	7.56
Cysteine	Cys	7.66	-
Creatinine	Cr	-	7.82
Phenylalanine	Phe	-	8.31
Glycerol Phosphate	GP	9.16	9.27
Phosphorylethanolamine	PE	-	9.43
Ornithine	Orn	9.52	9.63
Citrate	Cit	-	9.65
Tyrosine	Tyr	10.34	10.45
Inositol	Ino	11.40	11.49

* "-" indicates not identified by OPLS-DA to differentiate treatments in that cell type

4.5.1 Global Metabolite Variations in MCF7 Breast Cancer Cells in Response to Taxol- and Cisplatin-Treatment

Example GC-MS total ion chromatograms from MCF7 breast cancer cells treated with the high dose of either taxol or cisplatin for seven days are shown in Figure 4.9 in comparison to control, untreated cells (truncated to focus on retention time range of interest). As it was challenging to interpret the raw data for significant changes over time with respect to treatment, results from multivariate statistical OPLS-DA were used as the basis for interpretation. Breast cancer drug-responsive metabolites tentatively identified from OPLS-DA results included lactate, alanine, valine, glycerol, phosphate, proline, glycine, serine, threonine, aspartate, glutamate, creatinine, phenylalanine, glycerol phosphate, phosphorylethanolamine, ornithine, citrate, tyrosine, inositol, and inositol phosphate. The levels of these highlighted metabolites were then examined further to determine the response of MCF7 breast cancer cells to taxol- and cisplatin-treatment.

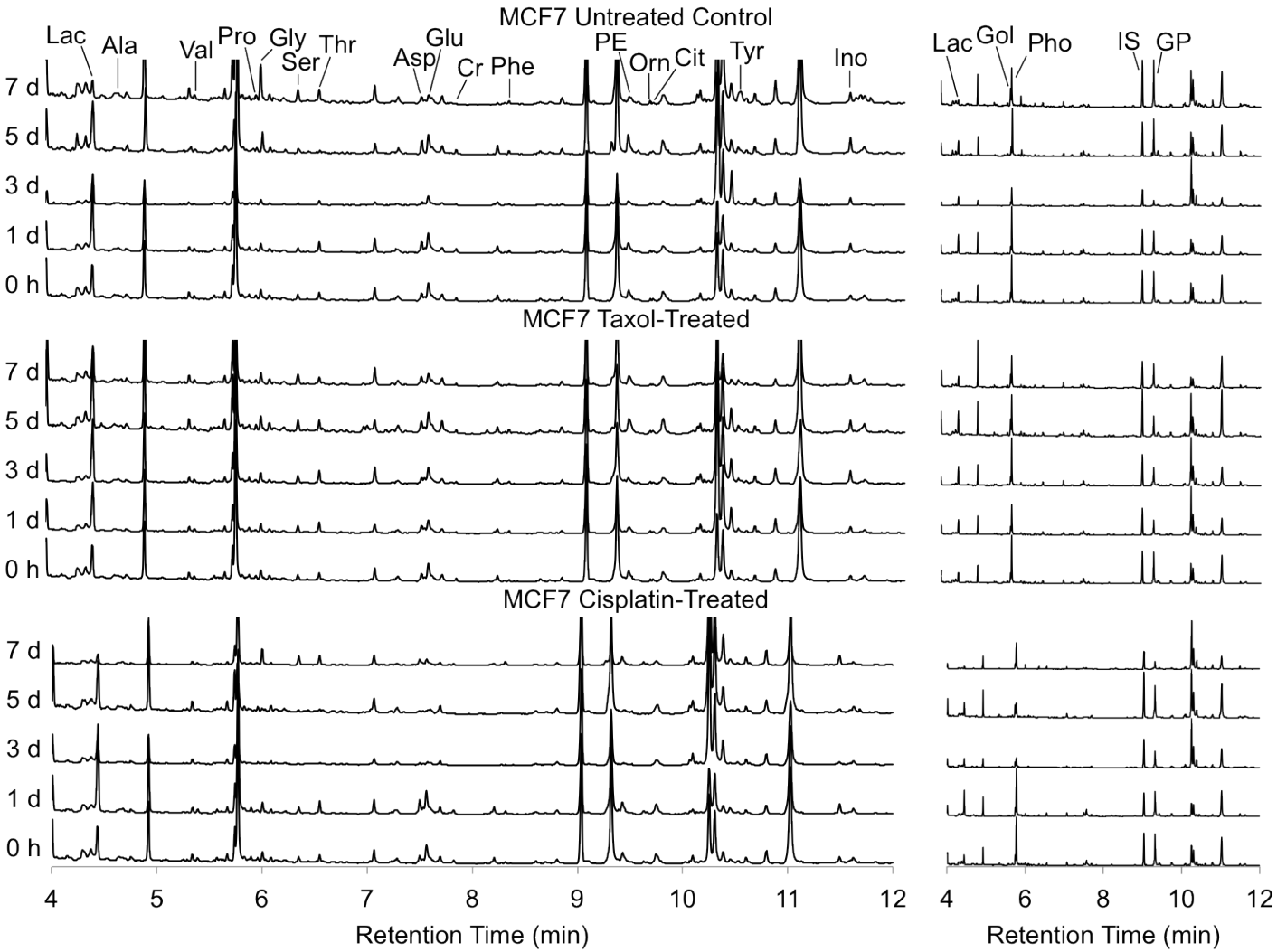

Figure 4.9. Example GC-MS total ion chromatograms of untreated control MCF7 cells as well as cells treated with taxol or cisplatin and analyzed over seven days (normalized and off-set for comparison). Chromatograms on the left have decreased scale to display lower abundance peaks, while the chromatograms on the right are shown at full scale for the higher abundance peaks. Peaks of some of the metabolites of interest are indicated (abbreviations listed in Table 4.3; IS indicates internal standard ribitol peak).

Most of the amino acids were present at relatively low levels, preventing significant differences among treatments over time due to low S/N ratio. This observation was expected as cancer cells characteristically have high translation activity to support replication, meaning that they use the amino acids as soon as they are generated and maintain low quantities of free amino acid pools (Yang *et al.* 2007). The same was true for the TCA cycle metabolites, with no statistically significant differences that could be attributed to such low compound levels. Previous work has shown that in many cases metabolite fluxes are high, resulting in unchanged or only slightly altered metabolite pool levels as they are rapidly consumed as soon as they are generated (Yang *et al.* 2007, Yang *et al.* 2008, Richardson *et al.* 2008). Fatty acid metabolism is know to be of interest in breast cancer due to changes glycerophospholipid metabolism (Yang *et al.* 2007, Richardson *et al.* 2008), however, some fatty acid contamination in the GC-MS results was observed in blank samples and prevented drawing conclusions about the fatty acids detected in this data set. Nevertheless, several of the metabolites of interest from the OPLS-DA did have significant abundance changes over time, especially when base peak or high-abundance fragment ions with high S/N ratios in the MS data were used for interpretation. These metabolites were used to draw conclusions about the response of MCF7 breast cancer cells to anticancer drug treatment.

Relative levels of creatinine (measured in enol form using characteristic fragment ion *m/z* 115.2 of the trimethylsilyl (TMS) derivative) in MCF7 breast cancer cells decreased after treatment with both taxol and cisplatin relative to untreated MCF7 cells in the GC-MS metabolite profiles. Malignant breast cells have higher creatine levels compared to normal breast tissue, and creatine and creatinine are directly metabolically linked in the creatine

pathway (KEGG database (http://www.genome.jp/kegg/), Bathen *et al.* 2007, Yang *et al.* 2007, Bayet-Robert *et al.* 2010b), with creatinine being a dehydrated version of creatine. Bayet-Robert and coworkers (2010a, 2010b) also showed that treating MCF7 cells with docetaxel, an analog of taxol, resulted in decreased levels of creatine compared to untreated cells. The creatinine results in this study are consistent with what has been observed in the literature for creatine and likely reflect general damage to the cells and/or repair mechanisms caused by anticancer drug treatment. This response of MCF7 breast cancer cells to anticancer drug treatment was not observed in A549 lung cancer cells, demonstrating the uniqueness of the cellular response of different cancers to therapeutic agents.

Breast cancer cells characteristically have active glycolysis (similar to many different cancers) and choline phospholipid metabolism relative to healthy breast tissue. Anticancer drug treatment may disrupt these high-functioning pathways, among others (Oakman *et al.* 2011). Previous studies of breast cancers have shown that alterations in glycerophospholipid metabolism, reflected in levels of phosphatidylcholines, phosphatidylethanolamines, and their derivatives, are some of the principal effects of the cancer phenotype and chemotherapeutic treatment (Ting *et al.* 1996, Bathen *et al.* 2007, Yang *et al.* 2007, Yang *et al.* 2008, Bayet-Robert *et al.* 2010b). Lipid components were also highlighted in the present study as responding to anticancer drug treatment. Relative levels of glycerol (measured in the form of characteristic fragment ion *m/z* 205.1 of the TMS derivative) in the MCF7 cells were significantly decreased after cisplatin-treatment relative to untreated control and taxol-treated cells (Figure 4.10). Free glycerol is a component in glycerolipid synthesis in cells and is also a possible (though rare) glycerophospholipid headgroup. The glycerol backbones of glycerophospholipids are derived

from a glycolysis intermediate in the form of glycerol phosphate instead of free glycerol. Also, previous reports of alterations in glycerophosphocholine and glycerophosphoethanolamine levels in breast cancer cells reflect catabolism of intact glycerophospholipids (*i.e.* lipase activity). Glycerol levels in this study indicate that glycerolipid metabolism is altered as a result of cisplatin treatment in MCF7 cells, possibly as a biosynthesis repair mechanism after cellular damage resulting from treatment. To the authors' knowledge this is the first time glycerolipid metabolism has been reported to be involved in chemotherapeutic treatment response in breast cancer cells, though it has been implicated in the evidence for glycerophospholipid metabolism alterations studied extensively in the literature.

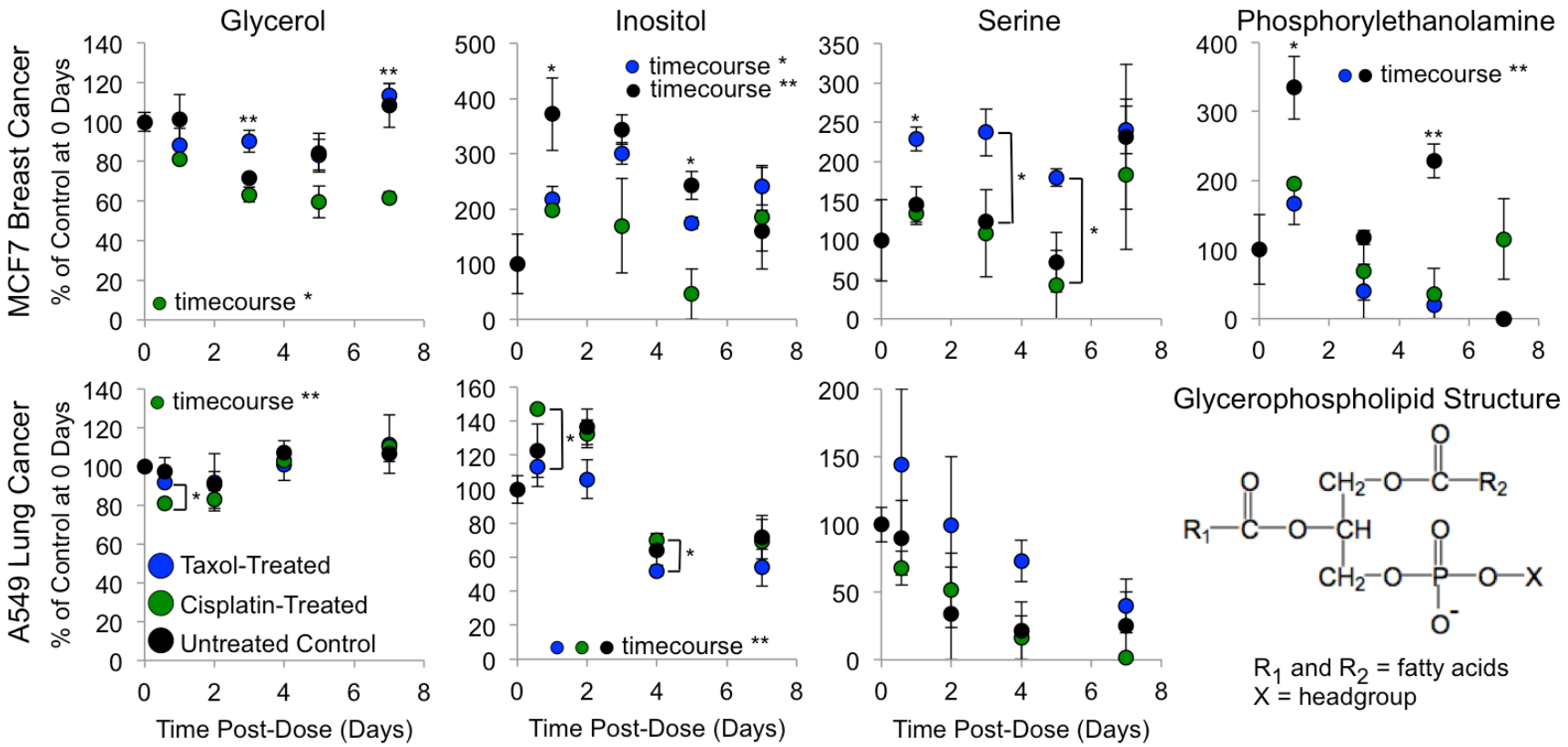

Figure 4.10. Glycerol, inositol, serine, and phosphorylethanolamine relative levels over time in taxol- and cisplatin-treated and control A549 lung cancer cells and MCF7 breast cancer cells detected in GC-MS metabolite profiles. Glycerol relative levels are decreased after cisplatin-treatment relative to taxol-treatment and control over time in MCF7 cells, but only at the 14 h time point in A549 cells. Inositol and phosphorylethanolamine levels decreased post-treatment with both taxol and cisplatin relative to control in MCF7 cells over time, which was not observed in A549 cells (phosphorylethanolamine was not highlighted in OPLS-DA of A549 cells). Serine relative levels are significantly increased in taxol-treated cells relative to cisplatin-treated and control MCF7 cells over time, while no significantly different levels were observed in A549 cells (n = 3, error bars are standard error; one-way ANOVA and Student's t-test results: * indicates $p < 0.05$, ** indicates $p < 0.01$). Post-treatment differences in glycerol, inositol, serine, and phosphorylethanolamine relative levels suggest that glycerolipid and glycerophospholipid metabolism is affected by anticancer drug treatment in a cell type-dependent manner.

Changes in glycerophospholipid metabolism as part of the cisplatin and taxol mechanisms of action in MCF7 breast cancer cells were also observed in the GC-MS data. Glycerophospholipid biosynthesis requires five components: glycerol phosphate (derived from a glycolytic intermediate), acetyl-CoA, two fatty acids (acyl chains), and a headgroup (*e.g.* choline, ethanolamine, serine, inositol, or glycerol). Inositol (measured in the form of characteristic fragment ion *m/z* 305.2 of the TMS derivative), serine (measured in the form of characteristic fragment ion *m/z* 204.1 of the TMS derivative), and phosphorylethanolamine (measured in the form of characteristic fragment ion *m/z* 299.1 of the TMS derivative) were all highlighted as significantly changing in MCF7 cells post-treatment (Figure 4.10). Phosphorylethanolamine, which is the ethanolamine headgroup with a phosphate attached, showed significantly decreased relative levels after both cisplatin- and taxol-treatment relative to control. This finding is consistent with increased production of glycerophospholipids in the breast cancer cells as a repair mechanism after treatment with anticancer agents. Inositol and serine are both involved in several intracellular metabolic pathways, including those of glycerophospholipid synthesis, making their levels in Figure 4.10 more challenging to interpret. Inositol relative levels significantly decreased in both taxol- and cisplatin-treated cells relative to controls, and levels were lower in cisplatin-treated compared to taxol-treated cells. Previous researchers have observed that inositol levels are increased in breast cancer cells relative to healthy breast tissue (Bathen *et al.* 2007, Yang *et al.* 2007), consistent with inositol playing an important role in breast cancer metabolism that needs to be better understood. Serine relative levels are significantly increased in taxol-treated cells relative to cisplatin-treated and control, which is different from the responses of the other two headgroups of interest. As serine is a non-

essential amino acid that is synthesized by cells for several functions, including conversion into cysteine that goes into glutathione, a flux-based analysis using isotopic tracers may be necessary to determine which metabolic pathways are most relevant to the anticancer drug response. In summary, the levels of free inositol, serine, and phosphorylethanolamine pools in the MCF7 cells are suggestive of sensitivity of glycerophospholipid metabolism to anticancer drug treatment, and the results are distinct from those seen in A549 cells.

The metabolomics results for MCF7 breast cancer cells demonstrate anticancer drug treatment effects on metabolite pools that suggest effects on both glycerolipid and glycerophospholipid metabolism. These pathways are essential to the survival and replication of cancer cells, and are perhaps even more important in breast cancer due to the high fat content of breast tissue. The temporal dependence of metabolite levels suggests that the treated MCF7 cells are repairing themselves and/or displaying adaptive mechanisms as a result of anticancer drug treatment, as increases in biosynthesis are implied. It is worth noting that most of the treatment-differentiating metabolites highlighted in MCF7 cells are specialized, *i.e.* not directly in the main biochemical pathways of carbon metabolism and building block synthesis, but are related. Similarly, complex lipids were not detectable using GC-MS, but effects on cellular lipid biosynthesis can still be inferred through this global metabolomics approach. These results highlight the advantages of global, rather than targeted, metabolomics in leading to novel anticancer drug target and mechanism of action information that may have gone overlooked if only the effects to central carbon metabolism were examined.

4.5.2 Global Metabolite Variations in A549 Lung Cancer Cells in Response to Taxol- and Cisplatin-Treatment

Example GC-MS chromatograms from A549 lung cancer cells treated with either taxol or cisplatin for seven days as well as control, untreated cells are shown in Figure 4.11 (truncated to focus on RT range of interest). Similar to the MCF7 breast cancer cells, results from OPLS-DA were used to determine substantial changes over time due to the complexity of the raw data. Lung cancer drug-responsive metabolites identified from OPLS-DA results included lactate, alanine, hydroxylamine, valine, urea, benzoic acid, glycerol, phosphate, proline, glycine, serine, threonine, β-alanine, malate, aspartate, glutamate, cysteine, serotonin, glycerol phosphate, ornithine, tyrosine, inositol, and inositol phosphate. These findings were consistent with results reported by Fan *et al.* (2009) where most of these metabolites were found to be present in all lung tissue, so modulation due to cancer and anticancer drug treatment is expected. The relative levels of these metabolites of interest were further examined to determine the response of A549 cells to treatment with taxol and cisplatin.

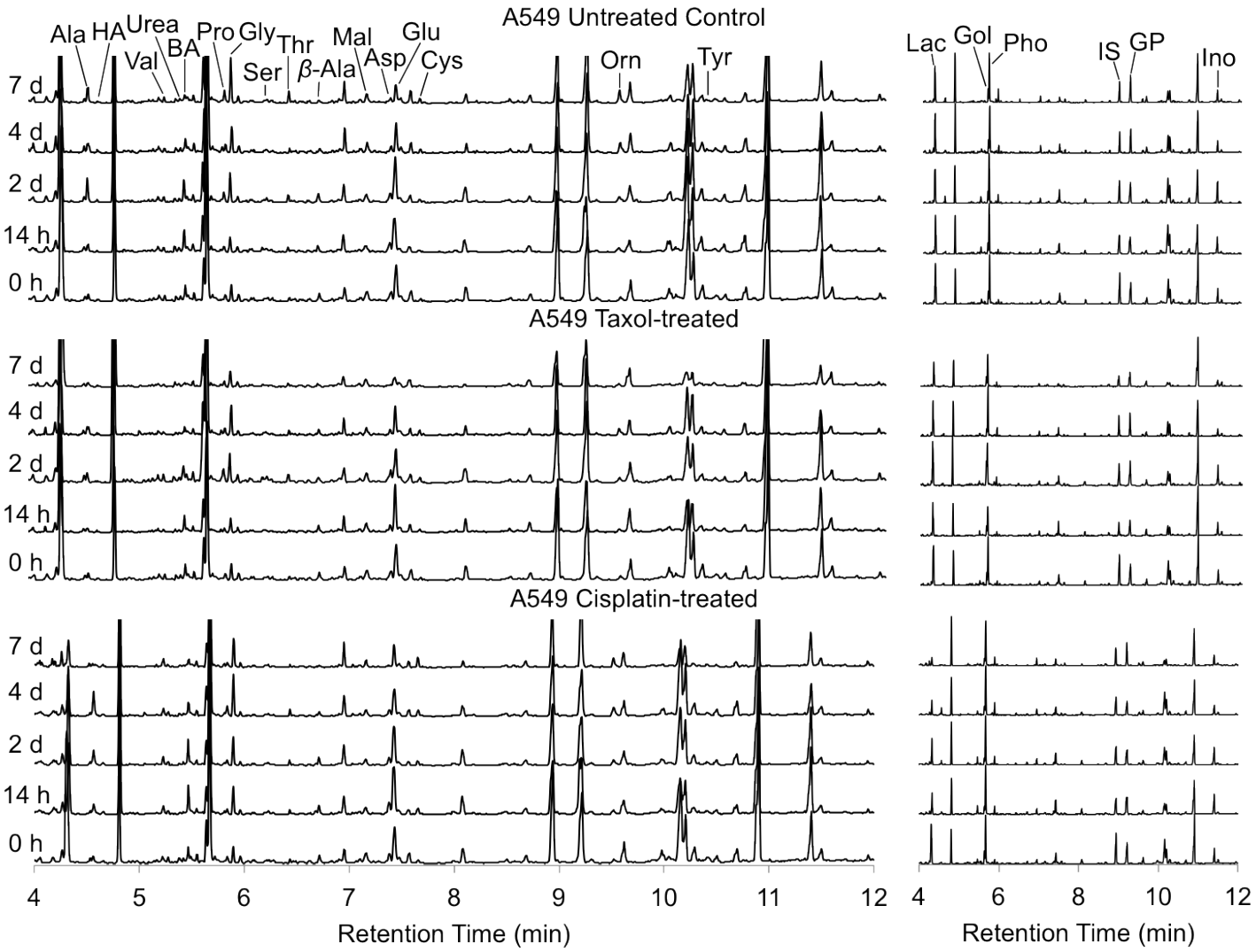

Figure 4.11. Example GC-MS total ion chromatograms of control A549 lung cancer cells as well as cells treated with taxol or cisplatin and analyzed over seven days (normalized and off-set for comparison). Chromatograms on the left have decreased scale to display lower abundance peaks, while the chromatograms on the right are shown at full scale for the higher abundance peaks. Peaks of some of the metabolites of interest are indicated (abbreviations listed in Table 4.3; IS indicates internal standard ribitol peak).

Similar to the results from the MCF7 breast cancer cells, most amino acids and TCA cycle metabolites were present at relatively low levels due to high metabolic fluxes, preventing significant differences among treatments over time in the A549 cells. However, several of the metabolites of interest from the OPLS-DA did exhibit significant differences between taxol- and cisplatin-treatment, and these metabolites were used to draw conclusions about the response of A549 cells to anticancer drug treatment and comparisons to the observed responses in MCF7 cells, discussed below.

Relative levels of the non-essential amino acids cysteine, glutamate, and glycine all displayed significant changes (ANOVA $p < 0.01$) over the time course of the study in both the taxol- and cisplatin-treated A549 cells (Figure 4.12). Relative cysteine levels (measured in the form of characteristic fragment ion *m/z* 220.1 of the TMS derivative) were significantly increased (ANOVA $p < 0.01$) after cisplatin-treatment relative to taxol-treated and control on days 4 and 7 post-dose (~1000% and ~1600% of control at day 0, respectively). Glutamate (characteristic fragment ion *m/z* 156.1 of the TMS derivative) relative levels were significantly increased (ANOVA $p < 0.05$) to ~130% of control in cisplatin-treated A549 cells relative to the other two treatment groups at 4 days post-dose. Glycine (characteristic fragment ion *m/z* 174.1 of the TMS derivative) relative levels were significantly increased after cisplatin-treatment to ~200% of control 14 hours after treatment (ANOVA $p < 0.05$), ~360% of control 4 days after treatment ((ANOVA $p < 0.01$), and ~340% of control 7 days after treatment (*t*-test $p < 0.01$). Cysteine is a reactive oxygen species (ROS) scavenger that helps control the redox state of the cell, both on its own and as a component of glutathione (GSH). Glutathione (γ-Glu-Cys-Gly) is a tripeptide, synthesized from glutamate, cysteine, and glycine in the cellular cytoplasm, which

plays a key role in maintaining the cellular redox potential (Griffith and Meister 1985). It is notable that all three precursors of GSH increased upon cisplatin treatment. Glutathione levels have been shown to be characteristically higher in cancer cells relative to normal cells. Bayet-Robert and coworkers (2010b) observed higher GSH levels in MCF7 breast cancer, PC3 prostate cancer, and 143B bone cancer cells compared to normal human fibroblasts. High GSH levels are necessary to counteract the production of ROS due to the partial inhibition of mitochondrial oxidative phosphorylation, which is also a characteristic trait of cancer cells (Wallace 2005). While the chemistry and molecular biology of ROS and cellular redox regulation has been the subject of extensive study, understanding of the biochemical disorders induced by oxidative stress is far from complete (Bayet-Robert *et al.* 2010c). It has also been well established that upregulation of GSH in cancer cells contributes to cisplatin resistance in that cisplatin is inactivated by covalent GSH adduct formation, and subsequently removed from the cell (Balendiran *et al.* 2004). Odenheimer and Wolf (1982) determined that cisplatin forms adducts with both cysteine and GSH *in vivo* over several days post-dose. Hosking and coworkers (1990) discovered that high levels of GSH positively correlate with cellular resistance to cytotoxic effects of radiation and chemotherapy, including cisplatin, in many human cancers. Meijer and coworkers (1990) demonstrated that continuous exposure to cisplatin leads to an increase in GSH content in previously cisplatin-sensitive lung cancer cells, and this mechanism has been implicated in tumor resistance to cisplatin. The GSH-cisplatin adduct retains electrophilic characteristics that confer reactivity toward protein nucleophiles, and this mechanism has been proposed to activate ROS generation, creating the need for higher levels of ROS scavengers including cysteine and GSH, to maintain normal physiological functions (Olas and Wachowicz

1996). In summary, these results suggest that cysteine and GSH are both functioning as ROS scavengers and forming cisplatin adducts in A549 cells in response to cisplatin treatment. In comparison, the MCF7 redox state is not perturbed as much in response to cisplatin treatment.

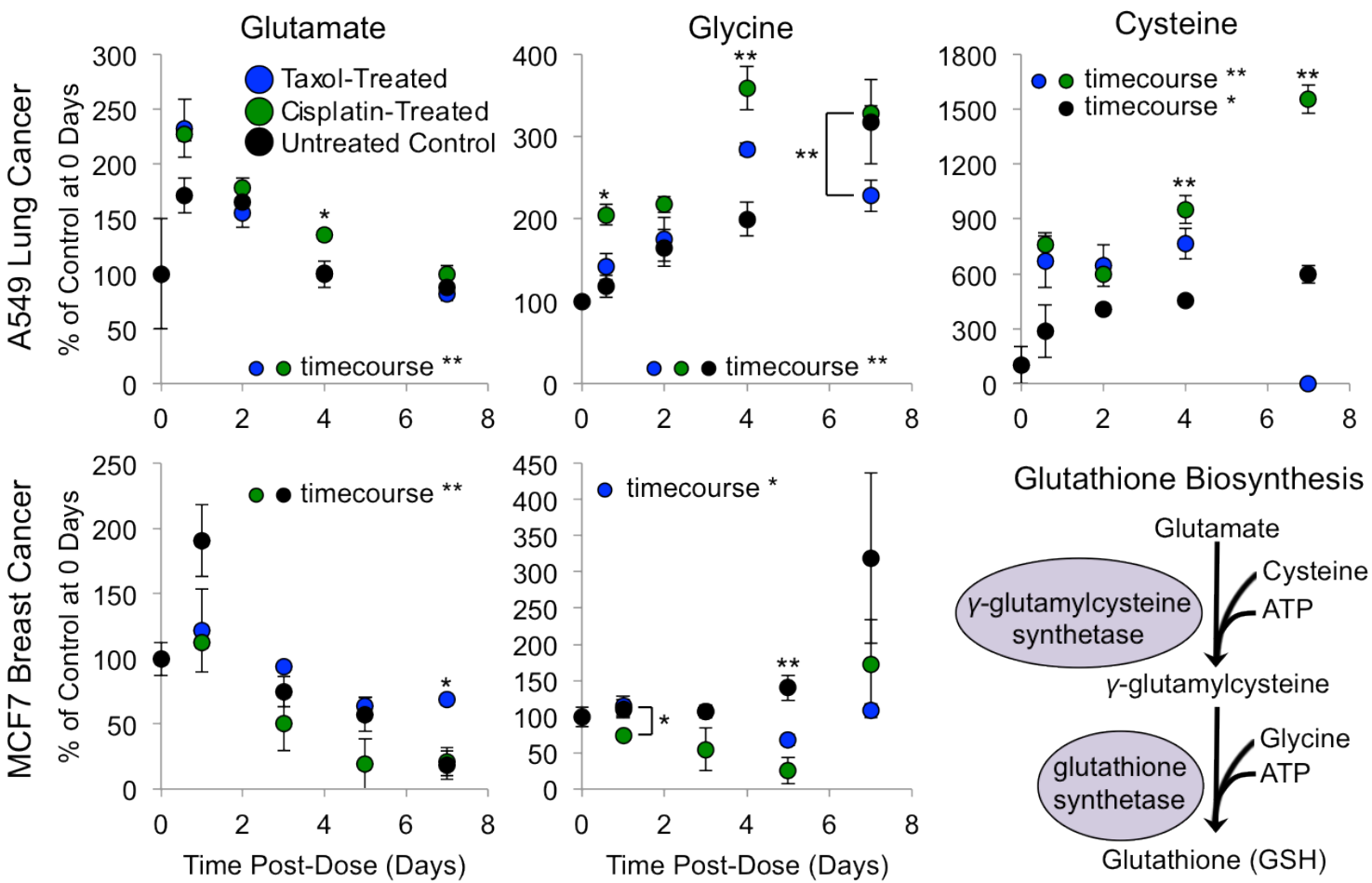

Figure 4.12. Glutamate, glycine, and cysteine relative levels over time in taxol- and cisplatin-treated and control A549 lung cancer cells and MCF7 breast cancer cells. Levels of all three metabolites are increased after cisplatin-treatment relative to taxol-treatment and control over time in A549 cells. The opposite is true in MCF7 cells, as glutamate and glycine levels are decreased after cisplatin-treatment compared to taxol-treated and control samples (cysteine was not highlighted in OPLS-DA of MCF7 cells) (n = 3, error bars are standard error; Student's t-test and one-way ANOVA results: * indicates $p < 0.05$, ** indicates $p < 0.01$). Post-treatment differences in glutamate, glycine, and cysteine levels in both A549 and MCF7 cells suggest that glutathione metabolism is affected by anticancer drug treatment in a cell type-dependent manner. ATP = adenosine triphosphate

Relative levels of ornithine, a metabolite in the amino acid class and a central part of the urea cycle, increased significantly ($p < 0.01$) over time in taxol-treated, cisplatin-treated, and untreated control A549 lung cancer cells (Figure 4.13). Ornithine (measured in the form of characteristic fragment ion *m/z* 142.1 of the TMS derivative) relative levels increased over time in the untreated A549 cells up to ~700% of control by day seven. In contrast, cisplatin-treated cells displayed relative increases in ornithine levels initially, up to ~450% of control through day four, but showed a smaller increase up to only approximately ~500% of control at day seven. The taxol-treated cells showed the lowest increases in ornithine over time compared to control, with the maximum increase at ~350% on day four and staying the same on day seven. Results from ANOVA indicated the three treatment groups varied significantly at two, four, and seven days post-treatment (days two and four, $p < 0.05$; day seven, $p < 0.01$). In comparison, ornithine was also highlighted by OPLS-DA in MCF7 cells, however, no statistically significant differences over time or significant differences in levels in the three different treatment groups were observed. Another urea cycle metabolite, urea (measured using the TMS derivative M-15 ion *m/z* 189.1), was highlighted as treatment distinguishing in A549 cells. While differences in urea levels were not statistically significant due to its low abundance and large variance of urea peak area in the GC-MS results (Figure 4.13), relative urea levels were higher in A549 cells treated with taxol compared to cisplatin-treated and control on days two through seven. These ornithine and urea results point to urea cycle disruption upon anticancer drug treatment in A549 cells. Post-urea cycle ornithine is converted into polyamines through the enzyme ornithine decarboxylase (ODC), which is known to be up-regulated in cancer due to the use of polyamines in cell cycle growth and repair (Gerner *et al.* 2004). Bayet-Robert and coworkers

(2010b) observed increased levels of polyamines in several types of cancer cells relative to normal human fibroblasts. After treatment with docetaxel (analog to taxol), the cancer cells displayed a significantly decreased level of polyamines, suggesting their use to repair the cells after injury by anticancer drug treatment. Fan and coworkers (2005) also observed evidence of urea cycle and polyamine metabolism perturbations in A549 cells treated with the anticancer drug selenite.

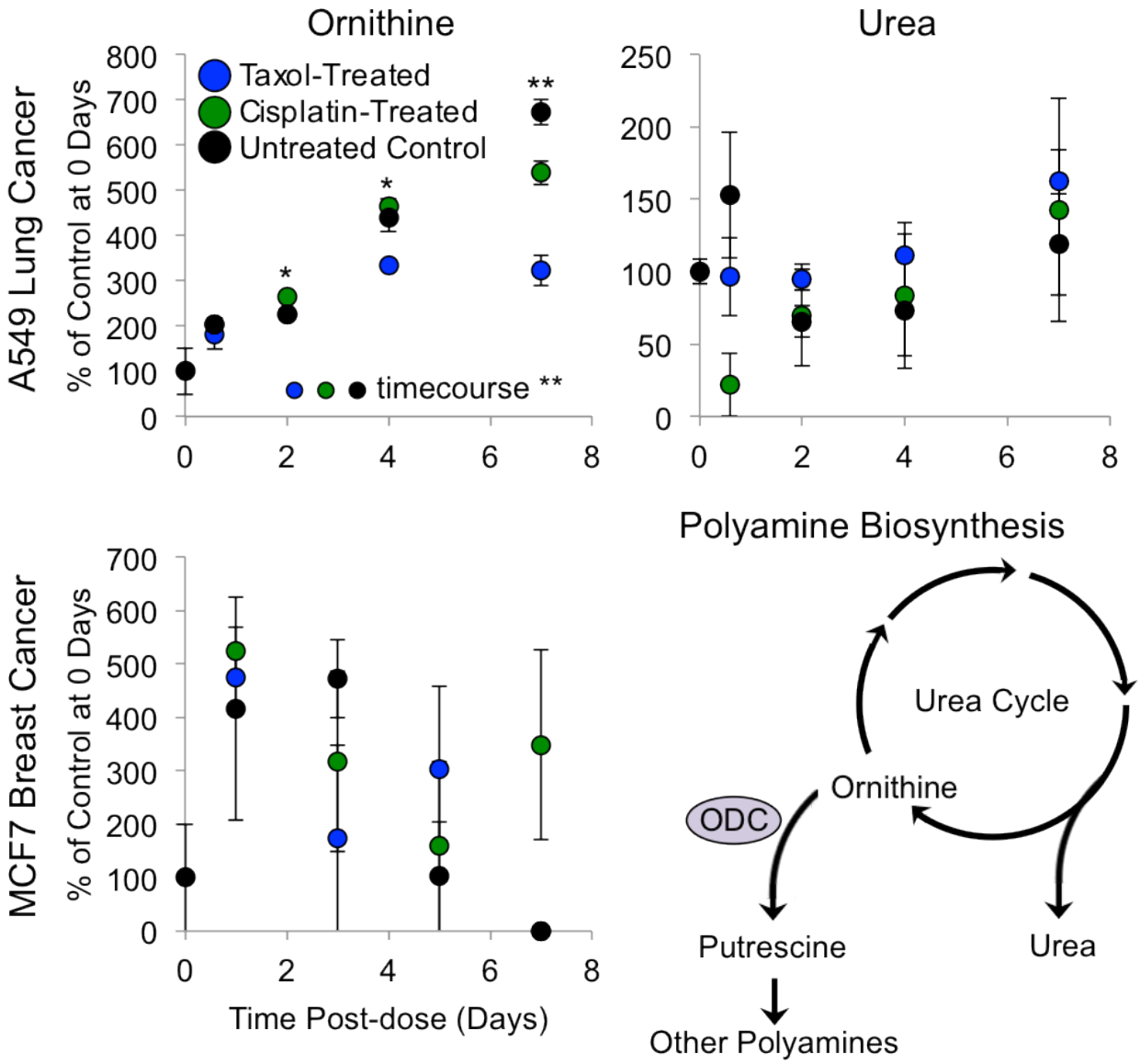

Figure 4.13. Ornithine and urea relative levels determined using GC-MS at various post-dose times in taxol- and cisplatin-treated and control A549 lung cancer cells and MCF7 breast cancer cells. Ornithine levels in taxol- and cisplatin-treated cells decrease compared to control over time in A549 cells, while no significant differences are observed over time in MCF7 cells. Urea levels in treated A549 cells increased relative to control over time, though not significantly (n = 3, error bars are standard error; one-way ANOVA results: * indicates $p < 0.05$, ** indicates $p < 0.01$). The ornithine and urea levels in A549 cells suggest that fluxes through the urea cycle and polyamine metabolism shown are affected by anticancer drug treatment. ODC = ornithine decarboxylase

From the results shown in Figure 4.13, relative levels of urea increased after taxol-treatment, consistent with increased activity of the urea cycle. Conversely, ornithine, the end-

product of the urea cycle, displayed decreased relative levels after treatment with both taxol and, to a lesser extent, cisplatin. This finding is consistent with ornithine being consumed downstream of the urea cycle, as is the case for conversion of ornithine to polyamines. There are at least two observations in the literature consistent with the depletion of ornithine to make higher levels of polyamines in cancer cells after anticancer drug treatment. The first is that the cancer cells need polyamines to help repair the damage caused by anticancer drug treatment. Russell and coworkers (1968) determined that polyamines have important functions in rapidly growing tissue, including tumors, and Luk and coworkers (1980) showed an increase in ODC activity during the recovery of intestinal mucosa after chemotherapy injury, which took longer if ODC was inhibited. Another explanation is that polyamines inhibit the apoptotic process in cancer cells and help them survive anticancer drug treatment. Hsu and coworkers (2008) found that overexpression of ODC caused leukemia cells to become resistant to taxol- and cisplatin-induced apoptosis. Polyamines were not identified in the GC-MS data, perhaps due to high metabolic fluxes or metabolite export to the media that resulted in low abundances of free metabolite pools for detection. In summary, alterations of metabolites like ornithine and urea suggest the urea cycle and associated polyamine metabolism were disrupted in A549 lung cancer cells as a result of anticancer drug treatment, with the effect more pronounced after taxol-treatment compared to treatment with cisplatin. These effects were not observed in the MCF7 cells exposed to the same treatments. Metabolite profiling has identified the urea cycle and polyamine metabolism as being affected by taxol and cisplatin treatment, which points us toward potential target mechanisms for anticancer drug cytotoxicity and cellular responses that may enhance cancer cell survival. As part of urea cycle takes place in the mitochondria, and

mitochondria are the key to apoptosis and cell death, it is a promising target for anticancer therapies or a novel way to monitor new anticancer drug efficacy.

4.6 Conclusions and Future Directions

The aim of this work was to use MS-based metabolomics to uncover biochemical disorders induced by taxol and cisplatin in A549 lung cancer and MCF7 breast cancer cells to help identify metabolic targets and mechanisms of cytotoxicity. The results revealed differential metabolic alterations in response to anticancer drug treatment in the different cancer cell types that may be assigned to molecular targets. These metabolic endpoints may serve the design of novel anticancer drugs with the same targets but improved efficacy and selectivity.

In this study, lipid metabolism was perturbed in MCF7 breast cancer cells by treatment with both taxol and cisplatin. Energy metabolism was affected in A549 lung cancer cells in response to anticancer drug treatment. Taxol-treatment led to increased urea cycle activity, while cisplatin attacked nucleophilic molecules such as cysteine and GSH, necessitating increased production to maintain cellular redox balance. This work demonstrates a new way to elucidate the mode of cytotoxicity of potential cancer therapeutic agents to screen drug candidates for further development. Metabolomics techniques revealed metabolic networks that are altered by drug treatment, which lead the way for novel hypotheses regarding the biochemical mechanisms of action of taxol and cisplatin as well as promising new anticancer drug targets for breast and lung cancer cells. The utility of high-throughput metabolomics for screening cancer therapeutics for mode of action and activity information in order to select the most promising lead compounds for further development was demonstrated as well.

The systematic design of experiments (DOE) used for this work was a full factorial experimental design that varied multiple experimental factors simultaneously in order to assess interactions between the varied factors as well as individual factor effects. Three factors were studied at various levels for effects on metabolic activity: cancer cell type (breast or lung), anticancer drug treatment (taxol or cisplatin), and time post-dose (14 hours, 2 days, 4 days, and 7 days for A549 lung cancer cells; 1 day, 3 days, 5 days, and 7 days for MCF7 breast cancer cells). The full factorial DOE permitted more efficient sample preparation and analysis compared to the approach of varying one factor at a time (OFAT), as it required fewer experimental conditions to be examined for metabolic effects. Examining the interactions between the varied factors also answers the question: Which factors are the dominating ones for affecting metabolic response? As the experimental results (GC-MS and HPLC-MS data) were multivariate, projection analyses (PLS-DA and OPLS-DA) were required to interpret the results of the study by visually grouping the samples into classes (Trygg *et al.* 2006). The most dominant factor for distinguishing sample results was cell type, followed by drug treatment, and the time post-dose had the smallest influence on sample differentiation. As lung tissue and breast tissue have very different physiological functions, the inherent metabolism differences were reflected in the measured metabolomics data and overwhelmed the other changing sample factors (data not shown). Drug treatment also resulted in significant metabolic response differences, which was the aim of this study. Treatment-dependent changes in cellular biochemistry detected using a metabolomic approach enabled the assessment of *in vitro* toxicity of chemotherapeutic agents in cancer cells.

REFERENCES

4.7 References

American Cancer Society (2012). Cancer facts & figures 2012. American Cancer Society, Atlanta, GA. Available from: http://www.acs.org.

Asiago, V. M, Alvarado, L. Z., Shanaiah, N., Gowda, G. A. N., Owusu-Sarfo, K., Ballas, R. A., Raftery, D. (2010). Early detection of recurrent breast cancer using metabolite profiling. Cancer Res. 70:8309-8318.

Balendiran, G. K., Dabur, R., Fraser, D. (2004). The role of glutathione in cancer. Cell Biochem. Funct. 22:343-352.

Bathen, T. F., Jensen, L. R., Sitter, B., Fjosne, H. E., Halgunset, J., Axelson, D. E., Gribbestad, I. S., Lundgren, S. (2007). MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. Breast Cancer Res. Treat. 104:181-189.

Bayet-Robert, M., Morvan, D., Chollet, P., Barthomeuf, C. (2010a). Pharmacometabolomics of docetaxel-treated human MCF7 breast cancer cells provides evidence of varying cellular responses at high and low doses. Breast Cancer Res. Treat. 120:613-626.

Bayet-Robert, M., Loiseau, D., Rio, P., Demidem, A., Barthomeuf, C., Stepien, G., Morvan, D. (2010b). Quantitative two-dimensional HRMAS ^{1}H-NMR spectroscopy-based metabolite profiling of human cancer cell lines and response to chemotherapy. Magnet. Reson. Med. 63:1172-1183.

Bayet-Robert, M., Lim, S., Barthomeuf, C., Morvan, D. (2010c). Biochemical disorders induced by cytotoxic marine natural products in breast cancer cells as revealed by proton NMR spectroscopy-based metabolomics. Biochem. Pharmacol. 80:1170-1179.

Becker, W. M., Kleinsmith, L. J., Hardin, J. (2006). *The World of the Cell* (6th Ed.). Pearson Education, Inc., San Francisco, CA.

Beloueche-Babari, M., Jackson, L. E., Al-Saffar, N. M. S., Workman, P., Leach, M. O., Ronen, S. M. (2005). Magnetic resonance spectroscopy monitoring of mitogen-activated protein kinase signaling inhibition. Cancer Res. 65:3356-3363.

Boccard, J., Grata, E., Thiocone, A., Gauvrit, J.-Y., Lanteri, P., Carrupt, P.-A., Wolfender, J.-L., Rudaz, S. (2007). Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding. Chemometr. Intell. Lab. 86:189-197.

Chung, Y.-L., Griffiths J. R. (2008). Using metabolomics to monitor anticancer drugs. Ernst Schering Found. Symp. Proc. 4:55-78.

Claudino, W. M., Quattrone, A., Biganzoli, L., Pestrin, M., Bertini, I., Di Leo, A. (2007). Metabolomics: available results, current research projects in breast cancer, and future applications. J. Clin. Oncol. 25:2840-2846.

DeBerardinis, R. J. (2008). Is cancer a disease of abnormal cellular metabolism? New angles on an old idea. Genet. Med. 10:767-777.

Di Leo, A., Claudino, W., Colangiuli, D., Bessi, S., Pestrin, M., Biganzoli, L. (2007). New strategies to identify molecular markers predicting chemotherapy activity and toxicity in breast cancer. Ann. Oncol. 18:xii8-xii14.

Fan, T. W. M., Lane, A. N., Higashi, R. M. (2004). The promise of metabolomics in cancer molecular therapeutics. Curr. Opin. Mol. Ther. 6:584-592.

Fan, T. W. M., Bandura, L. L., Higashi, R. M., Lane, A. N. (2005). Metabolomics-edited transcriptomics analysis of Se anticancer action in human lung cancer cells. Metabolomics 1:325-339.

Fan, T. W. M., Lane, A. N., Higashi, R. M., Farag, M. A., Gao, H., Bousamra, M., Miller, D. M. (2009). Altered regulation of metabolic pathways in human lung cancer discerned by ^{13}C stable isotope-resolved metabolomics (SIRM). Mol. Cancer 8:41-59.

Frickenschmidt, A., Frohlich, H., Bullinger, D., Zell, A., Laufer, S., Gleiter, C. H., Liebich, H., Kammerer, B. (2008). Metabonomics in cancer diagnosis: mass spectrometry-based profiling of urinary nucleosides from breast cancer patients. Biomarkers 13:435-449.

Gao, J., Zhao, H., Hylands, P. J., Corcoran, O. (2010). Secondary metabolite mapping identifies *Scutellaria* inhibitors of human lung cancer cells. J. Pharmaceut. Biomed. 53:723-728.

Gerner, E. W., Meyskens, F. L. (2004). Polyamines and cancer: old molecules, new understanding. Nat. Rev. Cancer 4:781-792.

Griffin, J. L., Shockcor, J. P. (2004). Metabolic profiles of cancer cells. Nat. Rev. Cancer 4:551-561.

Griffith, O. W., Meister, A. (1985). Origin and turnover of mitochondrial glutathione. P. Natl. Acad. Sci. USA 82:4668-4672.

Gu, L., Jones, A. D., Last, R. L. (2010). Broad connections in the Arabidopsis seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant. Plant J. 61:579-590.

Hosking, L. K., Whelan, R. D. H., Shellard, S. A., Bedford, P., Hill, B. T. (1990). An evaluation of the role of glutathione and its associated enzymes in the expression of differential sensitivities to antitumor agents shown by a range of human tumour cell lines. Biochem. Pharmacol. 40:1833-1842.

Hsu, P., Hung, H., Liao, Y., Liu, C., Tsay, G. J., Liu, G. (2008). Ornithine decarboxylase attenuates leukemic chemotherapy drugs-induced cell apoptosis and arrest in human promyelocytic HL-60 cells. Leukemia Res. 32:1530-1540.

Jordan, K. W., Adkins, C. B., Su, L., Halpern, E. F., Mark, E. J., Christiani, D. C., Cheng, L. L. (2010). Comparison of squamous cell carcinoma and adenocarcinoma of the lung by metabolomic analysis of tissue-serum pairs. Lung Cancer 68:44-50.

Kim, H. K., Wilson, E. G., Choi, Y. H., Verpoorte, R. (2010). Metabolomics: a tool for anticancer lead-finding from natural products. Planta Med. 76:1094-1102.

Lindon, J. C., Keun, H. C., Ebbels, T. M. D., Pearce, J. M. T., Holmes, E., Nicholson, J. K. (2005). The consortium for metabonomic toxicology (COMET): aims, activities, and achievements. Pharmacogenomics 6:691-699.

Luk, G. D., Marton, L. J., Baylin, S. B. (1980). Ornithine decarboxylase is important in intestinal mucosal maturation and recovery from injury in rats. Science 210:195-198.

Meijer, C., Mulder, N. H., Hospers, G. A. P., Uges, D. R. A., de Vries, E. G. E. (1990). The role of glutathione in resistance to cisplatin in human small cell lung cancer cell line. Brit. J. Cancer 62:72-77.

Merz, A. L., Serkova, N. J. (2009). Use of nuclear magnetic resonance-based metabolomics in detecting drug resistance in cancer. Biomarkers Med. 3:289-306.

Miller, J. N., Miller, J. C. (2005). *Statistics and Chemometrics for Analytical Chemistry* (5th Ed.). Pearson Education Limited, Harlow, UK, pp. 107-108, 110-111, 142, 167-169, 215-219.

Morvan, D., Demidem, A. (2007). Metabolomics by proton nuclear magnetic resonance spectroscopy of the response to chloroethylnitrosourea reveals drug efficacy and tumor adaptive metabolic pathways. Cancer Res. 67:2150-2159.

Nicholson, J. K. (2006) Global systems biology, personalized medicine, and molecular epidemiology. Mol. Syst. Biol. 2:52.

Oakman, C., Tenori, L., Biganzoli, L., Santarpia, L., Cappadona, S., Luchinat, C., Di Leo, A. (2011). Uncovering the metabolomic fingerprint of breast cancer. Int. J. Biochem. Cell B. 43:1010-1020.

Odenheimer, B., Wolf, W. (1982). Reactions of cisplatin with sulfur-containing amino acids and peptides I. Cysteine and glutathione. Inorg. Chim. A-Bioinor. 66:L41-L43.

Olas, B., Wachowicz, B. (1996). Cisplatin-induced changes in biological activity of blood platelets: thiol-related mechanisms. Anti-Cancer Drug. 7:476-482.

Rabik, C. A., Dolan, M. E. (2007). Molecular mechanisms of resistance and toxicity associated with platinating agents. Cancer Treat. Rev. 33:9-23.

Richardson, A. D., Yang, C., Osterman, A., Smith, J. W. (2008). Central carbon metabolism in the progression of mammary carcinoma. Breast Cancer Res. Treat. 110:297-307.

Rosenberg, B., Van Camp, L., Krigas, T. (1965). Inhibition of cell division in *Escherichia coli* by electrolysis products from a platinum electrode. Nature 205:698-699.

Rosenberg, B., Van Camp, L., Trosko, J. E., Mansour, V. H. (1969). Platinum compounds: a new class of potent antitumor agents. Nature 222:385-386.

Russell, D., Snyder, S. H. (1968). Amine synthesis in rapidly growing tissues: ornithine decarboxylase activity in regenerating rat liver, chick embryo, and various tumors. P. Natl. Acad. Sci. USA 60:1420-1427.

Taylor, S. L., Ganti, S., Bukanov, N. O., Chapman, A., Fiehn, O., Osier, M., Kim, K., Weiss, R. H. (2010). A metabolomics approach using juvenile cystic mice to identify urinary biomarkers and altered pathways in polycystic kidney disease. Am. J. Physiol. Renal Physiol. 298:F909-F922.

Ting, Y. L., Sherr, D., Degani, H. (1996). Variations in energy and phospholipid metabolism in normal and cancer human mammary epithelial cells. Anticancer Res. 16:1381-1388.

Trygg, J., Gullberg, J., Johansson, A. I., Jonsson, P., Moritz, T. (2006). Chemometrics in Metabolomics – An Introduction, in *Biotechnology in Agriculture and Forestry, Vol. 57: Plant Metabolomics*, K. Saito, R. A. Dixon, L. Willmitzer, eds, Springer, New York, NY, pp. 117-128.

Trygg, J., Holmes, E., Lundstedt, T. (2007). Chemometrics in metabonomics. J. Proteome Res. 6:469-479.

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., McPhail, A. T. (1971). Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. J. Am. Chem. Soc. 93:2325-2327.

Warburg, O., Wind, F., Negelein, E. (1927). The metabolism of tumors in the body. J. Gen. Physiol. 8:519-530.

Warburg, O. (1956). On the origin of cancer cells. Science 123:309-314.

Yang, C., Richardson, A. D., Smith, J. W., Osterman, A. (2007). Comparative metabolomics of breast cancer. Pac. Symp. Biocomput. 12:181-192.

Yang, C., Richardson, A. D., Osterman, A., Smith, J. W. (2008). Profiling of central metabolism in human cancer cells by two-dimensional NMR, GC-MS analysis, and isotopomer modeling. Metabolomics 4:13-29.