2017

Proteomic and Metabolomic Analyses of Biological Systems with Liquid Chromatography Tandem Mass Spectrometry and Ion Mobility Mass Spectrometry

Megan M. Maurer

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/6179

This Dissertation is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Graduate Theses, Dissertations, and Problem Reports by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
Proteomic and Metabolomic Analyses of Biological Systems with Liquid Chromatography Tandem Mass Spectrometry and Ion Mobility Mass Spectrometry

Megan M. Maurer

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Analytical Chemistry

Stephen J. Valentine, Ph.D., Chair
Suzanne Bell, Ph.D.
Glen P. Jackson, Ph.D.
Carsten Milsmann, Ph.D.
Timothy R. Nurkiewicz, Ph.D.

Department of Chemistry

Morgantown, West Virginia
2017

Keywords: Proteomics, Metabolomics, Ion Mobility, Hydrogen/Deuterium Exchange
Copyright 2017 Megan M. Maurer
Abstract

“Proteomic and Metabolomic Analyses of Biological Systems with Liquid Chromatography Tandem Mass Spectrometry and Ion Mobility Mass Spectrometry”

by Megan M. Maurer

Proteomic and metabolomic analyses provide information about altered metabolic processes in plants and animals. This information can be used to assess the impact of toxicant exposure or diseases on biological systems. In this dissertation, development of an ion mobility spectrometry - mass spectrometry (IMS-MS) strategy has been evaluated in addition to utilizing traditional liquid chromatography tandem mass spectrometry (LC-MS/MS) techniques for metabolite and protein analysis. Traditional LC-MS/MS techniques have been applied to the study of: 1) the molecular mechanisms responsible for altered microvasculature function as a result of pulmonary TiO$_2$ exposure in rats; and 2) the effectiveness of intracerebroventricular (ICV) injection of streptozotocin (STZ) to mimic early metabolic changes found in tauopathies in mice exhibiting P301L or human wild-type tau.

Inhalation exposure to TiO$_2$ nanoparticles (NPs) has been shown to produce a pulmonary inflammatory response, to induce oxidative and nitrosative biomarkers in the systemic microcirculation, and to influence downstream microvascular dysfunction. However, the molecular mechanisms relating pulmonary inflammation with the systemic microvascular dysfunction in addition to differences in responses of vascular tissues have yet to be fully elucidated. The first study examined plasma from rats exposed to TiO$_2$ NPs. A total of 58 proteins were identified by at least 2 unique peptides and found in at least 3 samples, and 23 metabolites were identified through ChemSpider matches and filtered for endogenous compounds. The compounds were then analyzed by principal component analysis (PCA) and 29 proteins and 18 metabolites were found to contribute most to the separation in PC1. These proteins and metabolites were then input into Ingenuity Pathway Analysis (IPA). IPA revealed 13 canonical pathways as being significant (p ≤ 0.05) based on the input proteins, but none were found to be significantly up or down regulated (z ≥ |2|) based on fold differences of the input proteins.

The second study examining plasma, aorta, and small resistance vasculature tissue from rats exposed to TiO$_2$ NPs was performed in order to gain further insight into the pathway activation mechanisms as well as to determine if responses differ based on tissue structural and functional differences. Congruent with the previous study, acute phase response signaling, LXR/RXR activation, and FXR/RXR activation emerge as being significant pathways (p ≤ 0.05) in the aorta and plasma; however, none were found to be significantly up or down regulated (z ≥ |2|). ILK signaling, D-myo-inositol (1,3,4)-trisphosphate biosynthesis, and 1D-myo-inositol hexakisphosphate biosynthesis II (mammalian) pathways are observed to be significant (p ≤ 0.05) in the vasculature, but none were found to be significantly up or down regulated (z ≥ |2|).

P301L and human wild-type (WT) tau mice were administered an intracerebroventricular (ICV) injection of control vehicle buffer (Veh) or streptozotocin (STZ) to mimic early metabolic changes found in tauopathies, including Alzheimer’s disease and frontotemporal dementia. Brain hemispheres were analyzed from 6 sample
cohorts: Control (CT)-Veh, CT-STZ, P301L-Veh, P301L-STZ, WT-Veh, and WT-STZ. Bottom-up proteomic analyses were utilized to identify differentially abundant proteins within the brain proteome and the biopathways altered as a result of ICV-STZ treatment. An ANOVA was used to determine the top 50 significant proteins among the 6 sample cohorts. Biopathway analysis of the top 50 proteins revealed 49 biological pathways as being significant (p ≤ 0.05), but none were significantly up or down-regulated (z ≥ |2|) among the cohorts. 14-3-3 Mediated Signaling was found to be the most significant pathway among the cohorts. Protein Kinase A Signaling pathway was also found to be significant and had an associated z-score, although it was not found to be significantly up or down-regulated in any of the comparisons. Proteome and pathway changes were observed as a result of ICV-STZ administration; however, none were found to be significantly up or down-regulated.

LC-MS/MS is a widely used method for metabolite identification. However, there are challenges with compound identification due to matrix effects, difficulty in separating isomers and isobars, and long analysis times. IMS has proven to be useful in separating isomer and isobar ions according to differences in mobilities through an inert buffer gas. IMS coupled with MS also offers advantages in speed requiring a fraction of the time used in condensed-phase separations. Despite such advantages, IMS-MS suffers from decreased peak capacity relative to LC-MS/MS. The addition of gas-phase hydrogen/deuterium exchange (HDX) to IMS-MS could allow for further differentiation among compounds as there can be differences in the incorporation of deuterium. Activating the ion to induce hydrogen/deuterium (HD) scrambling in the drift tube followed by HDX has demonstrated deuterium uptake differences among different species. In proof-of-principle experiments, these techniques are here shown to be highly reproducible (drift time CVs <2.0% and isotopic pattern RMSDs of <1%) while demonstrating an overall increase in peak capacity. Furthermore, this strategy is faster than LC-MS/MS offering the potential for complete complex mixture analysis in as little as 20 seconds.
Dedication

For my grandfather, John Maurer, Ph.D. Chemistry. It will be an honor to be called Dr. Maurer. I hope I am able to live up to the legacy of the original.

“…life must be prepared for and attacked with precision; that losing is only part of your preparation for the next victory; and that when you cross the finish line, you shouldn't have to look back, because you know as the gun sounds, you did your best.”
– Jim Jones, Founder Troopers Drum and Bugle Corps
Acknowledgments

This work could have not been completed without the support of many people. First and foremost, my family has always supported and encouraged me, even when I have done crazy things like backpack the Maryland section of the Appalachian Trail, by myself. My other family at the Troopers Drum and Bugle Corps has always been supportive. I grew up in this organization and I would not be who I am today without them. Honor, Loyalty, Dedication. My friends near and far who supported and encouraged me. Thanks for helping me stay positive.

My advisor, Steve Valentine, always gave me honest feedback and allowed me to work independently as much as possible. It has given me the confidence to move forward in my career as an independent researcher. My current and former lab mates, Jim Arndt, Greg Donohoe, Mahdier Khakinejad, Samaneh Ghassabi Kondalaji, Hossein Maleki, Kushani Attanayake, and Sandra Majuta, shared laughter, tears, triumphs, and sorrows with me. It's such an honor to have you as my academic family. Suzanne Bell, Glen Jackson, and Carsten Milsmann, served on my committee and provided appreciated input. Tim Nurkiewicz provided samples for the TiO2 exposed rat projects, served on my committee, and provided me with invaluable advice in applying for grants and finishing this dissertation.

Miranda Reed provided mouse brain tissue samples and was extremely patient while we completed the work. Justin Legleiter, fellow California wine snob and graduate studies chair, supported and encouraged while I was applying for external fellowships. The BioNano Research Facility (BNRF) housed the Orbitrap I used for my LC-MS/MS work. Without the BNRF management (Huiyuan Li, Greg Donohoe, and Qi Zeng), much
of this work would have not been possible. Trina Wafle, director of the Shared Research Facilities, gave me the opportunity to work as a technician in the BNRF, experience that will be very useful in my new role at Arizona State University.

The C. Eugene Bennett Department of Chemistry provided financial support in the form of teaching assistantships and research awards. I was also partially supported an NSF grant (CHE-1553021) as a research assistant. Becky Secrist and Brenda Prentiss do so much work behind the scenes to keep the chemistry department going and really do not get thanked enough. Thank you, thank you, thank you! Finally, I want to thank Kym Scott and WVU Community Chorus for the beautiful music making and 3 hours a week that were not dedicated to science. Music brings together people from all walks of life and I will truly miss our comradery.
# Table of Contents

Abstract ii  
Dedication iv  
Acknowledgements v  
Table of Contents vii  
List of Tables ix  
List of Figures x  
List of Abbreviations xi  
List of Files xiv

Chapter 1 – Introduction  
1.1 Introduction 1  
1.2 References 9

Chapter 2 - Comparative Plasma Proteomic and Metabolomic Studies of Pulmonary TiO₂ Nanoparticle Exposure in Rats using Liquid Chromatography Tandem Mass Spectrometry 14  
2.1 Introduction 15  
2.2 Materials and Methods 18  
2.2.1 Samples 18  
2.2.2 Proteomics extractions and analysis procedures 20  
2.2.3 Metabolomics extractions and analysis procedures 22  
2.2.4 Informatics 22  
2.3 Results and Discussion 28  
2.3.1 Identified proteins and peptides 28  
2.3.2 Protein comparison 31  
2.3.3 Metabolite comparison 34  
2.3.4 Biopathway analysis 36  
2.4 Conclusion 45  
2.5 References 46

Chapter 3 - Proteomic and Metabolomic Comparative Analyses of Plasma and Vasculature Tissue from TiO₂ Nanoparticle Exposed Rats 51  
3.1 Introduction 52  
3.2 Materials and Methods 54  
3.2.1 Samples 54  
3.2.2 Proteomic and metabolomic extraction and analysis 56  
3.2.3 Informatics 60
<table>
<thead>
<tr>
<th>Chapter 4 - Proteomic and Metabolomic Analysis of Brain Tissue Following Intracerebroventricular Administration of Streptozocin in Mouse Models of Tauopathies by LC-MS/MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
</tr>
<tr>
<td>4.2 Methods</td>
</tr>
<tr>
<td>4.2.1 Samples</td>
</tr>
<tr>
<td>4.2.2 Proteomics extractions and analysis procedures</td>
</tr>
<tr>
<td>4.2.3 Metabolomics extractions and analysis procedures</td>
</tr>
<tr>
<td>4.2.4 Informatics</td>
</tr>
<tr>
<td>4.3 Results</td>
</tr>
<tr>
<td>4.4 Discussion</td>
</tr>
<tr>
<td>4.5 Conclusion</td>
</tr>
<tr>
<td>4.6 References</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5 – LC-MS/MS metabolomic analyses shortcomings</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 LC-MS/MS metabolomic analyses technique and hardware shortcomings</td>
</tr>
<tr>
<td>5.2 LC-MS/MS metabolomic analyses software shortcomings</td>
</tr>
<tr>
<td>5.3 Improvements on the horizon</td>
</tr>
<tr>
<td>5.4 References</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6 – Ion Mobility, Hydrogen/Deuterium Exchange, and Isotope Scrambling: Tools for Metabolite Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
</tr>
<tr>
<td>6.2 Experimental</td>
</tr>
<tr>
<td>6.3 Results and Discussion</td>
</tr>
<tr>
<td>6.4 Conclusions</td>
</tr>
<tr>
<td>6.5 References</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7 – Future Directions</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Proteomic Analyses</td>
</tr>
<tr>
<td>7.1.1 TiO$_2$ NP exposed rats</td>
</tr>
<tr>
<td>7.1.2 ICV-STZ treated mice</td>
</tr>
<tr>
<td>7.1.3 Multidimensional protein identification technology</td>
</tr>
<tr>
<td>7.1.4 Post translational modifications</td>
</tr>
<tr>
<td>7.1.5 Isotopically labeled protein quantitation</td>
</tr>
<tr>
<td>7.2 Metabolomic Analyses</td>
</tr>
<tr>
<td>7.3 IMS-MS Instrumentation</td>
</tr>
<tr>
<td>7.4 IMS-MS Automation and Software</td>
</tr>
<tr>
<td>7.5 References</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1 - Protein inputs for Ingenuity Pathway Analysis 26
Table 2.2 - Metabolite inputs for Ingenuity Pathway Analysis 27
Table 2.3 - Top statistically significant results from Ingenuity Pathway Analysis 41
Table 3.1 - Top proteins and metabolites by tissue 63
Table 3.2 - Top protein locations and functions by tissue 66
Table 4.1 - Significant canonical pathways 104
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Interplay of genes, proteins, and metabolites</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Protein location and function</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>PCA scatter plot for proteins</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>PCA scatter plot for metabolites</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Network from Ingenuity Pathway Analysis</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Ingenuity Pathway Analysis significant canonical pathways</td>
<td>39</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Venn diagram for proteins by tissue type</td>
<td>68</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Venn diagram for metabolites by tissue type</td>
<td>70</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>IPA significant canonical pathways by tissue type</td>
<td>72</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Similarity in ILK Signaling and LXR/RXR Activation pathways</td>
<td>77</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Edwards-Venn diagram of proteins from each cohort</td>
<td>97</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Top 20 protein locations and functions</td>
<td>99</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Dendrogram and heatmap of the top 50 proteins as determined by ANOVA</td>
<td>100</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Dendrogram and heatmap of the top 25 metabolites as determined by ANOVA</td>
<td>102</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Cloud plot of organic metabolites from tau mice</td>
<td>123</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>– Model peptides with different deuterium uptake patterns</td>
<td>141</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>– IMS-MS profile of BSA digest trials</td>
<td>145</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>– Isotopic distribution of [M+2H]$^2^+$ bradykinin ions</td>
<td>146</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>– Model peptides with different HD scrambling patterns</td>
<td>149</td>
</tr>
<tr>
<td>Figure 6.5</td>
<td>– Ion fragmentation spectra of [M+3H]$^3^+$ KKDDDDDIKKII peptide after deuterium uptake and HD scrambling</td>
<td>151</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>– Structure of the ICAT tag</td>
<td>169</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>– Structure of the iTRAQ tag</td>
<td>171</td>
</tr>
</tbody>
</table>
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetonitrile</td>
<td>ACN</td>
</tr>
<tr>
<td>Alzheimer's Disease</td>
<td>AD</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Automatic gain control</td>
<td>ACG</td>
</tr>
<tr>
<td>BioNano Research Facility</td>
<td>BNRF</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>BSA</td>
</tr>
<tr>
<td>Bronchoalveolar lavage fluid</td>
<td>BALF</td>
</tr>
<tr>
<td>Capillary electrophoresis</td>
<td>CE</td>
</tr>
<tr>
<td>Collision induced dissociation</td>
<td>CID</td>
</tr>
<tr>
<td>Control</td>
<td>CT</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>CRP</td>
</tr>
<tr>
<td>Cyclic inosine monophosphate</td>
<td>cyclic IMP</td>
</tr>
<tr>
<td>Dalton</td>
<td>Da</td>
</tr>
<tr>
<td>Difference gel electrophoresis</td>
<td>DIGE</td>
</tr>
<tr>
<td>Dithiothreitol</td>
<td>DTT</td>
</tr>
<tr>
<td>Drift tube ion mobility spectroscopy</td>
<td>DT-IMS</td>
</tr>
<tr>
<td>Electrical low pressure impactor</td>
<td>ELPI</td>
</tr>
<tr>
<td>Electron capture dissociation</td>
<td>ECD</td>
</tr>
<tr>
<td>Electron transfer dissociation</td>
<td>ETD</td>
</tr>
<tr>
<td>Electron volts</td>
<td>eV</td>
</tr>
<tr>
<td>Endothelin 1</td>
<td>ET-1</td>
</tr>
<tr>
<td>Engineered nanomaterial</td>
<td>ENM</td>
</tr>
<tr>
<td>Exponentially Modified Protein Abundance Index</td>
<td>emPAI</td>
</tr>
<tr>
<td>False discovery rate</td>
<td>FDR</td>
</tr>
<tr>
<td>Farnesoid X receptor/retinoid x receptor</td>
<td>FXR/RXR</td>
</tr>
<tr>
<td>Fibronectin</td>
<td>FN1</td>
</tr>
<tr>
<td>Fourier transform mass spectrometry</td>
<td>FTMS</td>
</tr>
<tr>
<td>Frontotemporal dementia</td>
<td>FTD</td>
</tr>
<tr>
<td>Gas chromatography - mass spectrometry</td>
<td>GC-MS</td>
</tr>
<tr>
<td>High collision induced dissociation</td>
<td>HCD</td>
</tr>
<tr>
<td>High density lipoprotein</td>
<td>HDL</td>
</tr>
<tr>
<td>Human Metabolome Database</td>
<td>HMDB</td>
</tr>
<tr>
<td>Hydrogen/deuterium exchange</td>
<td>HDX</td>
</tr>
<tr>
<td>Hydrophilic interaction chromatography</td>
<td>HILIC</td>
</tr>
<tr>
<td>Ingenuity Pathway Analysis</td>
<td>IPA</td>
</tr>
<tr>
<td>Interleuken 1</td>
<td>IL-1</td>
</tr>
<tr>
<td>Interleuken 6</td>
<td>IL-6</td>
</tr>
<tr>
<td>Intracerebroventricular</td>
<td>ICV</td>
</tr>
<tr>
<td>Iodoacetamide</td>
<td>IAM</td>
</tr>
</tbody>
</table>
Ion mobility spectrometry - mass spectrometry: IMS-MS
Isobaric tag for relative and absolute quantitation: iTRAQ
Isotope coded affinity tags: ICAT
Kyoto Encyclopedia of Genes and Genomes: KEGG
Liquid chromatography -tandem mass spectrometry: LC-MS/MS
Liver x receptor/retinoid X receptor: LXR/RXR
Loadings matrix: LM
Low density lipoprotein: LDL
Mass to charge ratio: m/z
MassBank of North America: MoNA
Matrix assisted laser desorption/ionization in-source decay: MALDI ISD
Methanol: MeOH
Mitochondrial ribonucleic acid: mRNA
Myeloperoxidase: MPO
Nanoparticle: NP
National Institutes of Standards and Technology: NIST
Nitric oxide: NO
Normalized collision energy: NCE
Nuclear magnetic resonance: NMR
Phosphate-buffered saline: PBS
Polymophonuclear leukocytes: PMN
Post translational modification: PTM
Potassium ethylenediaminetetraacetic acid: K2 EDTA
Principal component analysis: PCA
Principle component 1: PC1
PubChem chemical identifier: PubChem CID
Quadrupole - time of flight: Q-ToF
Reactive oxygen species: ROS
Root mean square deviation: RMSD
Scanning mobility particle sizer: SMPS
Stable-isotope labelling by amino acids in cell culture: SILAC
Streptozotocin: STZ
Strong anion exchange: SAX
Strong cation exchange: SCX
Tandem mass spectrometry: MS2 or MS/MS
Traveling wave ion mobility spectrometry: TWIMS
Trifluoroacetic acid: TFA
Tumor necrosis factor: TNF
Two dimensional: 2D
Ultra-high pressure liquid chromatography: UHPLC
Ultraviolet: UV
<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Veh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild Type</td>
<td>WT</td>
</tr>
</tbody>
</table>
List of Files

Chapter 2 Metabolite data supplemental
Chapter 2 Protein data supplemental
Chapter 2 Cytoscape network
Chapter 3 Aorta metabolite supplemental
Chapter 3 Aorta protein supplemental
Chapter 3 Plasma metabolite supplemental
Chapter 3 Plasma protein supplemental
Chapter 3 Vasculature metabolite supplemental
Chapter 3 Vasculature protein supplemental
Chapter 4 Metabolite data supplemental
Chapter 4 Protein data supplemental
Chapter 4 Cytoscape network
Chapter 1

Introduction
1.1 Introduction

Life has evolved complex biological processes, organizations, and functions based on a hierarchy of genes, proteins, and metabolites. Genes contain the long term “data” storage of the system; however, genes can be activated, inactivated, or even changed due to environmental response or random mutations [1]. Proteins, encoded by genes, can be thought of as the short term “data” storage systems within the body. Proteins respond to available resources and stimuli to drive metabolic processes within the body. This includes acting as enzymes for catabolic and anabolic processes of metabolites and activating and deactivating genes to induce or inhibit transcription of messenger RNA (mRNA). Metabolites not only act as building blocks for energy consumption and energy storage, they also serve as messengers within the body to promote activation and deactivation of proteins [1]. The interplay of this hierarchal system helps an organism adapt to different stressors including disease and toxicants as demonstrated in Figure 1.1. The result exposure to stressors can be measured as an overall physiological outcome or the changes in organism on a molecular level [1].

By linking the physiological outcomes to operative underlying molecular mechanisms, researchers can gain a better understanding of biological processes which can lead to improved treatments. Furthermore, the information can assist policy makers in making recommendations or enacting legislation to improve public health. There is a myriad of analytical instrumentation available to study organisms on a molecular level. Proteins and metabolites have been studied using instrumentation such as nuclear magnetic resonance (NMR) spectroscopy [2–5], capillary electrophoresis (CE) [6–9], liquid chromatography (LC) [10–13], CE with mass spectrometry (CE-MS) [14–
**Figure 1.1** – Stress, toxicants, or diseases can affect the genes, proteins, and metabolites. Genes encode proteins, but proteins can activate or inactivate genes. Proteins acting as enzymes direct catabolic or anabolic processes of metabolites, but metabolites can serve as secondary messengers for proteins. The result of a stressor could lead to pathological differences or other biological outcomes.
Gas chromatography (GC) has been used alone [22–25] and in combination with MS [26–29] for metabolite analysis; however, it is not well suited for protein or peptide analysis.

Each technique has its own advantages and disadvantages for comparative ‘omics analyses. NMR is useful for determining structures of compounds and is inherently quantitative; however, it lacks sensitivity and therefore is not appropriate for determination of low abundance species (as reviewed in reference [30]). CE offers many variations of the technique (e.g., zone electrophoresis, isoelectric focusing, gel electrophoresis, and micellar electrokinetic chromatography), requires low volumes of solvents, and it has also been coupled to mass spectrometers (as reviewed in reference [31]). However, Joule heating and diffusion can decrease the resolution of CE [31]. GC greatly improves signal-to-noise ratios, works well for volatile compounds, but derivatization is often required for sample analysis and analysis of proteins or peptides is not routine (as reviewed in reference [30]). LC also offers variations of the technique [normal phase, reverse phase, hydrophobic interaction chromatography (HILIC), and ion pairing], and recent reduction of column particle size has reduced the amount of solvents required and analysis times (as reviewed in reference [30]). Additionally, LC works well for many different compound classes [30]. Unfortunately, LC-MS techniques suffer from matrix effects due to mobile phases or sample solvents and have difficulty separating isobars and isomers (as reviewed in reference [19]). Furthermore, analysis times require several minutes to hours for completion and there is run to run variation in compound retention time [19].

For this work LC-MS/MS was selected as the primary means of protein and
metabolite analysis due to its versatility with regard to the types compounds that can be analyzed, and the availability of a LC system coupled to a high resolution mass spectrometer (Q Exactive, Thermo Fisher, San Jose, CA). Bottom up proteomic and metabolomics analyses by LC-MS/MS were used to determine what occurs at the molecular level when rats are exposed to TiO$_2$ nanoparticles (Chapters 2 and 3) and when tau mice are intracerebroventricularly (ICV) administered streptozotocin (STZ, Chapter 4). Proteomic and metabolomic mass informatics software were utilized to identify proteins and metabolites in the MS/MS spectra. Statistical analysis revealed the proteins and metabolites that were significantly different between exposed and control groups (TiO$_2$ rats) or cohorts (tau mice). Significant proteins and metabolites were then input into biopathway analysis software in order to elucidate the metabolic pathways that are affected as a result of the exposure or treatment.

Protein bioinformatic software suites automate the analysis of MS and MS/MS raw data by using various algorithms. For example, the Molecular Weight Search or the MOWSE algorithm (later known as Mascot) [32] and the Sequest algorithm [33] were developed in the early 1990’s as a means to identify proteins by using mass spectral matching to the theoretical spectra. The Mascot approach uses a scoring process coupled to a proprietary statistical analysis to provide a confidence limit for each spectral match [32,34]. Sequest utilizes a cross correlation algorithm to deduce the quality of one match [33,34]. These identification techniques became the basis of bottom up proteomics, identification of proteins by mass spectral analysis of the proteolytically cleaved proteins [35]. The advantage of bottom up proteomics is the ability to identify many different proteins within a mixture and to achieve relative quantitation of the proteins [35]. A
number of protein analysis software programs in use today employ one or both of the above algorithms for protein identification. Protein analysis in this work was performed with Proteome Discoverer (v. 1.4.0.288; Thermo Fisher, San Jose, CA) using the Sequest algorithm, although the Mascot algorithm is also available for analysis in this program.

Metabolite bioinformatics software programs typically employ proprietary algorithms for spectral identification based on exact mass matching or by fragment matching [36]. Sieve (Thermo Fisher, San Jose, CA) identifies compounds by comparing experimental MS exact mass information to exact masses within compound libraries. Other programs such as Mass Frontier (Thermo Fisher, San Jose, CA), Compound Discoverer (Thermo Fisher, San Jose, CA), XCMS Online (Scripps Research Institute, La Jolla, CA), and Elements for Metabolomics (Proteome Software, Portland, OR) rely on MS/MS data in addition to MS exact mass matches for compound identification. Sieve, Mass Frontier, Compound Discoverer, XCMS Online, and Elements for Metabolomics were utilized in this work for metabolite identification in Chapters 2, 3, and 4; however, there was low confidence in the results. An analysis of these programs and others is provided in Chapter 5.

Although LC-MS/MS is widely used for proteomic and metabolomic analyses, LC requires minutes to hours for separation of proteins and metabolites and separation of isomers and isobars by LC are challenging [19]. On the other hand, ion mobility spectrometry (IMS) instrumentation is a gas-phase separation strategy that occurs on the millisecond timescale and can be useful for separating isobars, isomers, and conformers (as reviewed by reference [37]). Separation is achieved based on differences in ions’ collisional cross sections (CCS). Overall, ions that are more compact and more highly
charged will traverse the mobility region faster than those ions that are more elongated and have lower charge states [37,38]. These advantages of ion mobility are attractive for the migration of bottom up proteomic and metabolomic analyses to an IMS-MS platform.

Gas-phase hydrogen/deuterium exchange (HDX) is an additional technique that can be coupled with IMS-MS, and has been shown to be useful to study small molecules [39-41] and to study protein structure and different conformations [42-46]. Gas-phase HDX occurs through the introduction of a deuterium containing gas into the drift tube. Ions react with the gas and the hydrogens on the protein or analyte of interest exchange with the deuterium via a relay mechanism [47]. Hydrogens that are considered labile (charged sites and heteroatom sites) will exchange. Those at charge sites will exchange first, followed by other side chain heteroatoms, and then finally those on the amides of the backbone [48]. The amount of deuterium incorporation and the resulting mass spectral uptake pattern are dependent on the overall rate of exchange [49], which for peptides and proteins depends on the primary amino acid sequence [50]. Hydrogen/deuterium (HD) scrambling occurs when added energy redistributes the incorporated deuteriums to sites less accessible to HDX and hydrogens to sites more accessible to HDX. The hydrogens relocated to more accessible sites can undergo HDX resulting in further deuterium uptake. Like HDX, HD scrambling is highly dependent on the ion characteristics [50-55], which may provide an additional parameter to identify metabolites and proteins. In this work proof-of-principle analyses of the utilization IMS-MS, HDX, and HD scrambling with HDX for rapid compound identification in the absence of LC separation are presented in Chapter 6.

This work also provides future directions for the projects described in Chapters 2,
3, 4, and 6. Future work includes multidimensional separation techniques, analysis of post translational modifications (PTMs), isotopic labeling, and the development of new instrumentation. Multidimensional separation techniques would be of interested because of their ability to resolve proteins and metabolites that cannot be separated with the reversed phased separation alone [56]. Analyzing for protein PTMs can offer additional information about protein activation, inactivation, or other functions [57]. For example, phosphorylation PTMs activate or inactivate proteins and ubiquitin PTMs can mark proteins for degradation [57]. Isotopic labeling techniques afford a multiplex advantage which reduces the error associated with run to run variability in ionization efficiency and instrument performance and allows for a more accurate relative quantitation of proteins [58]. Finally, new IMS instrumentation in development has the potential for great impact on decreasing analysis times for complex mixture analysis following the strategy outlined in Chapter 6.
1.2 References


Chapter 2

Comparative plasma proteomic and metabolomic studies of pulmonary TiO$_2$ nanoparticle exposure in rats using liquid chromatography tandem mass spectrometry

The work in the chapter was published in part in the Journal of Proteomics.


2.1 Introduction

Over the last several years, the production of engineered nanomaterials (ENM) has increased dramatically. Here, nanoparticles (NP) are defined as species exhibiting a single dimension that is less than 100 nm in size. Such species find usage in a wide variety of products ranging from electronic and cosmetic to medical and pharmaceutical [1]. As an indication of their increasing usage, the number of consumer products sold in the United States containing ENM as well as the work force involved in the production of such products are expected to increase by more than an order of magnitude from 2008 to 2020 [2]. It has recently been estimated that to support this industry ~50,000 tons of TiO$_2$ nanoparticles alone are manufactured per year [3,4]. With this acceleration in NP production comes a greater need to investigate the impact of increased exposure on human health.

At the turn of the 21$^{st}$ century, investigative studies began associating air pollution particulate matter exposure to cardiovascular events [5–9]. Since that time, significant resources [10,11] have been dedicated to studies aimed at determining the effects of particulate matter exposure on the cardiovascular system [12–14]. Concerned with the limited applicability of macrovascular studies, a number of recent investigations have focused on the effects of particle exposure on microvasculature function [15–25]. With regard to NP exposure, such efforts have led to the development of three primary hypotheses describing processes by which exposure affects microvascular function [26]. These include systemic inflammation and changes in autonomic influence. A third process involves direct tissue-NP contact resulting from particle translocation from the lungs to the systemic microvasculature [26].
These mechanistic hypotheses generate questions regarding the roles of various molecular interactions in altering normal microvascular function. A number of experiments aimed at determining molecules associated with macrovascular dysfunction indicate that particle exposure has been shown to elicit a systemic inflammatory response. For example, circulating levels of tumor necrosis factor (TNF alpha), interleukin 1 (IL-1 beta), interleukin 6 (IL-6), endothelin 1 (ET-1), and c-reactive protein (CRP) are observed to be elevated for human subjects after exposure to particles [27–29]. Analyses of bronchoalveolar lavage fluid (BALF) from animal studies have demonstrated TiO₂ NP exposure increases inflammatory cells, such as neutrophils and lymphocytes; moreover, the observed increase of inflammatory cells was found to be dependent on days after exposure [30–33], dose [32–34], and method of introduction (whole body inhalation vs. intratracheal instillation) [33]. Gene expression analyses of lung tissue have also shown immune and inflammatory responses as a result of TiO₂ particle exposure [34–36]. Furthermore, markers associated with oxidative and nitrosative stresses in the microcirculation increase [20–22,25].

From such studies a better understanding that particle exposure with microvascular dysfunction is emerging. Briefly, local inflammation in the lungs (resident macrophages and epithelial cells) due to interaction with NPs results in recruitment and activation of polymorphonuclear leukocytes (PMNs) [26]. Additionally, the local response results in increased levels of cytokines and chemokines in the blood. The circulating PMNs and cytokines result in the release of additional molecules associated with inflammation, some of which may enhance leukocyte adhesion in the microvasculature. Attachment of leukocytes to the vascular endothelium is believed to
result in myeloperoxidase (MPO) deposition. Production of reactive oxygen species (ROS) increases local oxidative stress [25]. This ultimately decreases bioavailable nitric oxide (NO). Additionally, by-products of MPO-associated reactions result in decreased endothelial NO production. The marked decrease in local NO levels results in impaired arteriolar reactivity and further venular leukocyte adhesion.

Despite the success in determining relevant molecular responses (i.e., those potentially affecting the microvasculature) resulting from particle exposure, many factors remain to be elucidated. Currently the connection between pulmonary inflammation and the downstream microvascular effects is not well understood [26]. For example, it is unknown if secondary inflammatory markers released by activated PMNs (occurring at the lungs) affect leukocyte adherence and reactive stress within the microvasculature. Additionally, the role of secondary inflammatory markers in endothelial activation is not well known. Finally, the source of reactive stresses and their relationship to inflammatory markers and vascular activation parameters is unclear.

The current study employs comparative proteomic and metabolomic analyses to begin addressing the potential connections between pulmonary inflammation and microvascular dysfunction from TiO$_2$ NP exposure by examining protein dysregulation. Proteomic and metabolomic studies were conducted using liquid chromatography with tandem mass spectrometry (LC-MS/MS) under positive ion detection. Comparisons of assigned protein and metabolite abundances between plasma samples obtained from rats exposed to TiO$_2$ NP's and a control cohort reveal a number of species that may be used to distinguish sample types. These molecules have been used in biopathway analyses to generate a molecular regulation network. The comparative proteomics, as well as the
biopathway analyses, are discussed in light of recent findings associating NP exposure and microvascular dysfunction.

2.2 Materials and Methods

2.2.1 Samples

Experimental animals. Male Sprague Dawley rats (175 - 250 g) were purchased from Hilltop Laboratories (Scottdale, PA), and housed in laminar flow cages under controlled temperature and humidity conditions and a 12 hr light/12 hr dark cycle at the West Virginia University Health Sciences Center vivarium. Food and water were provided ad libitum. All animals were acclimated for 72-hours before nanomaterial inhalation exposure. To ensure that all methods were performed humanely and with regard to alleviation of suffering, all procedures were approved by the Institutional Animal Care and Use Committee of the West Virginia University.

ENM. Nano-titanium dioxide (TiO$_2$) powder was obtained from Evonik (Aeroxide TiO$_2$, Parsippany, NJ). This ENM is a mixture composed of anatase (80%) and rutile (20%) TiO$_2$, with a primary particle size of 21 nm, and a surface area of 48.08 m$^2$/g [24,37,38]. The nano-TiO$_2$ was prepared for aerosolization by drying, sieving, and storing the powder [19,24].

Inhalation Exposure. The Nanoparticle Aerosol Generator and exposure system used for the current experiments have previously reported and described (U.S. Patent #8,881,997) [19,24,39,40]. The system was developed specifically for rodent, whole-body nanoparticle inhalation exposures. Briefly, the apparatus was developed with a vibrating fluidized bed, a Venturi vacuum pump, cyclone separator, impactor and mixing
device, an animal housing chamber, and real-time monitoring devices. Aerosols were generated by allowing a high velocity air stream to pass through the vibrating fluidized bed and into the Venturi vacuum pump; drawing air and the TiO$_2$ nanoparticles as it passes. Aerosols enter the cyclone separator, which is gated to removed agglomerates >400 nm at an input flow rate of 60 L/min of clean dry air before entering the exposure chamber.

Size distribution, mean aerodynamic diameter, and relative mass concentration of the aerosols were monitored in real time (Electrical Low Pressure Impactor (ELPI), Dekati, Tempere, Finland). The particle size distribution was also measured in real-time with a Scanning Mobility Particle Sizer device (SMPS; TSI Inc., St. Paul, MN). The real-time size distribution and mass concentration of the aerosols were monitored between 7-1000 nm with these devices. The count median aerodynamic diameter of the aerosols was ~160 nm. Once the steady-state aerosol concentration was achieved, exposure duration was adjusted to achieve a calculated deposition of 30 µg/animal. Animals were exposed for 4-6 hours per day for at a target mass concentration of 6.0 mg/m$^3$ or to filtered air (0 mg/m$^3$, control).

**Blood collection and plasma isolation.** 24 hours after ENM inhalation exposure, rats were anesthetized by an intraperitoneal injection of thiobutabarbitral sodium (100 mg/kg) and placed on a heating pad to maintain a rectal temperature of 37 ºC. The trachea was intubated to ensure a patent airway, and the right carotid artery was cannulated to measure arterial blood pressure and sample whole blood. Whole blood was collected via the carotid artery cannula from live animals directly into Vacutainers containing K$_2$ EDTA (BD Biosciences, San Jose, CA). Blood samples were immediately centrifuged for
10 minutes at 2500 g. The plasma layer (supernatant) was then transferred into 1.5 mL microcentrifuge tubes. These tubes were snap frozen in liquid nitrogen before being placed at -80 ºC for long term storage.

2.2.2 Proteomics extraction and analysis procedures

**Immunodepletion.** A Seppro IgY14 immunodepletion kit was utilized (Sigma-Aldrich, St. Louis, MO) for immunodepletion of the 14 most abundant proteins in plasma. Prior to depletion, 20 µL of plasma was diluted with 500 µL of 1× Dilution buffer, added to a Corning Costar SpinX 0.45 µm column (Sigma-Aldrich, St. Louis, MO), and centrifuged at 10000 g for 1 minute. 500 µL of the diluted plasma was immunodepleted using Seppro IgY14 spin columns according to the manufacturer’s protocol. Immunodepleted samples were frozen at -80 ºC overnight and then dried down in a vacuum concentrator at 35 ºC for 8 hours. Samples were reconstituted in 1 mL of 200 mM ammonium bicarbonate with 6 M urea in water for further sample preparation.

**Bradford Analysis.** Total protein concentration of the immunodepleted samples was determined using a Bradford Assay. Standards of bovine serum albumin (0, 0.25, 0.5, 0.75, 1.0, 1.5, and 2.0 µg/µL) were prepared and all samples and standards were analyzed in duplicate. Protein concentrations were determined at 595 nm on an Epoch Biotek UV-vis spectrophotometer (BioTek, U.S., Winooski, VT). The average sample concentration was calculated and the amount of sample needed to react with 2 µL of 0.01 M dithiothreitol (DTT) at a concentration of 1:40 (protein:DTT) was determined.

**Protein reduction, alkylation, and digestion.** Plasma protein extractions were prepared as outlined by Valentine, *et al.* with modifications [41]. The calculated aliquot
of sample was diluted to a volume of 1 mL with 200 mM ammonium bicarbonate with 6 M urea. 2 µL 0.01M of DTT was added and the samples were incubated at 37 °C for 2 hours. Samples were placed on ice and 4 µL of 0.01 M iodoacetamide (IAM) was added. Samples were incubated at 4 °C for 2 hours in the dark. 2 µL of 0.01 M cysteine was added and samples were incubated at room temp for 30 minutes. 1 mL of 25 µg/mL trypsin was added and samples were incubated at 37 °C for 18 hours. After trypsin digestion, samples were stored at -80 °C overnight and then dried in a vacuum concentrator for 8 hours at 35 °C. Samples were reconstituted in 150 µL sample buffer (95/5 H2O/ACN + 0.5% TFA) and desalted using Pierce C18 Spin columns and manufacturer’s protocol. After desalting, samples were dried at 35 °C and reconstituted in 50/50 H2O/ACN with 0.1% formic acid for LC-MS/MS analysis.

**Proteomic analysis by LC-MS/MS.** A Thermo Scientific Q Exactive mass spectrometer with an Easy-nLC 1000 nano-LC was utilized (Thermo Fisher, San Jose, CA). 18 µL of sample was injected onto a Thermo Scientific EASY-spray PepMap C18, 3 µm, 75 µm x 150 mm column (Thermo Fisher, San Jose, CA) with the column temperature set to 35 °C. Mobile phases were A – H2O with 0.1% formic acid, and B – ACN with 0.1% formic acid. At a flow rate of 20 µL/min, the sample gradient was as follows: 40-minute linear gradient from 2% to 60% B; 20-minute linear gradient to 90% B and hold for 10 minutes; 5-minute linear gradient back to initial conditions and hold for 5 minutes. MS settings were: positive mode; m/z range 150-2000; MS resolution 70,000 and automatic gain control (AGC) 1E6; data dependent MS/MS top 5 with S/N 5, resolution of 140,000, AGC 1E5, normalized collision energy (NCE) 30, and isolation window 4.0 m/z; spray voltage 1.80 kV; and S lens 60 V. Samples were analyzed in a
random order.

2.2.3 Metabolomics extraction and analysis procedures

Sample preparation. A modified Bligh-Dyer extraction was used [42]. 750 µL 1:2 hexane:MeOH was added to 200 µL of plasma to precipitate proteins. An additional 250 µL of hexane was added along with 250 µL of water and mixed well. Samples were centrifuged at 10000 rpm for 5 minutes at 4°C. After centrifugation a layer of precipitated proteins was observed between the layers and at the bottom of the tube. The organic and aqueous fractions were carefully collected to avoid any collection of precipitant. Fractions were dried down with a vacuum concentrator and reconstituted with 50/50 H₂O/ACN with 0.1% formic acid and then filtered through a nylon 0.2 µm filter prior to LC-MS/MS analysis.

Metabolite analysis by LC-MS/MS. A Thermo Scientific Q Exactive mass spectrometer with an Accela UHPLC was used (Thermo Fisher, San Jose, CA). 18 µL of sample was injected onto a Phenomenex Synergi 4µ Fusion RP–80 2.0 x 100 mm column (Phenomenex, Torrance, CA). Mobile phases were A – H₂O with 0.1% formic acid, and B – ACN with 0.1% formic acid. Using a flow rate of 200 µL/min, the sample gradient was as follows: hold 2% B for 2 minutes; 60-minute linear gradient to 90% B and hold for 15 minutes; 5-minute linear gradient back to initial conditions and hold for 10 minutes. MS settings: positive mode; scan type, MS/MS top 10 with S/N 5; resolution of 140,000; spray voltage 1.80 kV; capillary temperature 275°C, and S lens 60.
**Proteomics.** Raw data files were analyzed with Proteome Discoverer (v. 1.4.0.288; Thermo Fisher, San Jose, CA) using the Sequest algorithm to search against the rat.fasta database. Precursors were selected in a retention time window from 0 to 60 minutes with a mass range from 150 to 5000 Da. Charge states examined were +1 to +4. The mass analyzer was set to FTMS, activation type to HCD, ionization source to nanospray, and polarity to positive mode. The Sequest algorithm searched for peptides ranging in length from 6 to 144 residues, allowing for 2 missed cleavages. A single dynamic modification of carbamidomethyl was selected and the maximum modifications per peptide was set at 3. Percolator confidence intervals were set at 0.05 and the decoy database search was set at 0.1 for relaxed target and 0.05 for strict target. All other default settings were used. Protein identifications are noted by their Uniprot Accession identifier.

Protein comparisons based on the total number of high-confidence peptide ion assignments were carried out using the Access Database software (Microsoft Office 2007, Microsoft Corporation, Redmond, WA). Additional abundance comparisons were performed by first determining an abundance score based on the exponentially modified protein abundance index (emPAI) technique described previously [43,44]. The emPAI score (S) is computed as [44]

\[
S = 10^{N_{\text{obs}}/N_{\text{exp}}} - 1,
\]

where \(N_{\text{obs}}\) is the total number of peptide ion assignments that are observed for a specified protein and \(N_{\text{exp}}\) is the expected number of digest peptides from the protein sequence. For these comparisons, \(N_{\text{exp}}\) is determined using the protein sequence length and the frequency of occurrence of Lysine (~7.2%) and Arginine (~4.2%) amino acid residues.
The number of expected tryptic digest peptides for a given protein was estimated to be the product of the sequence length and the sum of occurrence frequencies (expressed as a decimal) of the two amino acid residues.

Differences in protein abundance were estimated based on comparisons of the individual emPAI scores. The top 91 proteins in the highest abundance (at least 3 sample hits) across all samples were chosen to supply the y-values (protein emPAI scores) for principal component analysis (PCA) using the JMP Statistical software package (v. 11.2.0, SAS Institute Inc., Cary, NC) [46]. The PCA report includes a loadings matrix value (LM), which demonstrates the degree of effect that a y-variable has on the principle component. LMs range from 0 to 1, where values closer to 1 signify a greater influence on the separation associated with a component. The LMs were then used to determine proteins exhibiting the greatest degree of variation between the control and exposed samples. The proteins that contributed the most (|LM| > 0.5) to the separation in principle component 1 (PC1) were input into QIAGEN’s Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) to determine molecular pathways influenced by pulmonary nanoparticle exposure.

Metabolomics. Raw data files were analyzed with Sieve (v. 2.1.377; Thermo Fisher, San Jose, CA). Data files were aligned at a minimum intensity of 1000 and a maximum retention time shift of 0.2 minutes. The data were framed by group (control and exposed) with 5000 maximum frames and a retention time width of 2.5 minutes. m/z values analyzed ranged from 75 to 1000 Da over retention times from 0 to 60 minutes. Results were identified using the ChemSpider KEGG database with an m/z tolerance of 10. Final Sieve results were screened to include only endogenous compounds as
pharmaceuticals and phytochemicals identifications are inclined to be false positives. Results from all samples were compiled and normalized for total metabolite content in the sample. PCA of the normalized metabolite data was performed using JMP (v. 11.2.0, SAS Institute Inc., Cary, NC) [46]. Similar to the proteomics data, metabolites that contributed the most to the separation in PC1 (absolute value > 0.5) were input into IPA (Qiagen, Hilden, Germany).

Biopathway Analysis. Biological pathways associated with the differentially expressed proteins and metabolites were generated using the Ingenuity Pathway Analysis. The biological pathways were obtained from a molecular input list comprised of those proteins and metabolites exhibiting the greatest ability to distinguish the two sample types based on PCA. For each protein and metabolite, the fold difference between the average normalized emPAI score (protein) or peak intensity (metabolite) of the exposed samples to the average normalized emPAI score (protein) or peak intensity (metabolite) of the control samples was also input into IPA in order to determine up or down regulation of pathways (see Table 2.1 for proteins and Table 2.2 for metabolites). To generate the biological pathways, the core analysis function in the Ingenuity software was performed on the input molecules. The Ingenuity Knowledge Base was used as the reference set for analysis considering direct and indirect relationships. Molecules and relationships were restricted to rat or undefined species. The number of molecules in networks was restricted to 70 with 10 maximum networks. The top 3 canonical pathways that were found to be most significant based on a calculated p score have been selected for discussion (see section 2.3.4 Biopathway analysis).

It should be noted that there are limitations associated with using biopathway
### Table 2.1 – Protein inputs for Ingenuity Pathway Analysis

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Symbol</th>
<th>Fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>P55159</td>
<td>Serum paraoxonase/arylesterase</td>
<td>PON1</td>
<td>*</td>
</tr>
<tr>
<td>P31211</td>
<td>Corticosteroid-binding globulin</td>
<td>SERPINA6</td>
<td>*</td>
</tr>
<tr>
<td>P04937-4</td>
<td>Isoform 4 of Fibronectin</td>
<td>FN1</td>
<td>-14.28</td>
</tr>
<tr>
<td>G3V8D4</td>
<td>Apolipoprotein C-II</td>
<td>APOC2</td>
<td>-6.63</td>
</tr>
<tr>
<td>P08934-2</td>
<td>Isoform LMW of Kiniogen-1</td>
<td>KNG1</td>
<td>*</td>
</tr>
<tr>
<td>P09006</td>
<td>Serine protease inhibitor A3N</td>
<td>SERPINA3N</td>
<td>-2.52</td>
</tr>
<tr>
<td>P14480</td>
<td>Fibrogen beta chain</td>
<td>FGB</td>
<td>-1.54</td>
</tr>
<tr>
<td>P20059</td>
<td>Hemopexin</td>
<td>HPX</td>
<td>1.47</td>
</tr>
<tr>
<td>P14046</td>
<td>Alpha-1-inhibitor 3</td>
<td>A1</td>
<td>3</td>
</tr>
<tr>
<td>P12346</td>
<td>Serotransferrin</td>
<td>TF</td>
<td>1.38</td>
</tr>
<tr>
<td>E9PSV0</td>
<td>Protein C4-2</td>
<td>C4A/C4B</td>
<td>-4.62</td>
</tr>
<tr>
<td>F1LR92</td>
<td>Serine protease inhibitor A3M</td>
<td>SERPINA3M</td>
<td>-3.95</td>
</tr>
<tr>
<td>D3ZME7</td>
<td>HscB mitochondrial iron-sulfur cluster co-chaperone</td>
<td>HSCB</td>
<td>**</td>
</tr>
<tr>
<td>P01048</td>
<td>T-kininogen 1</td>
<td>MAP1</td>
<td>-2.17</td>
</tr>
<tr>
<td>D4A7D9</td>
<td>Anionic trypsin-2</td>
<td>PRSS2</td>
<td>*</td>
</tr>
<tr>
<td>P04639</td>
<td>Apolipoprotein A-1</td>
<td>APOA1</td>
<td>1.56</td>
</tr>
<tr>
<td>Q7TQ11</td>
<td>Vitronectin</td>
<td>VTN</td>
<td>-2.29</td>
</tr>
<tr>
<td>D4A6E3</td>
<td>Murinoglobulin-1</td>
<td>MUG1</td>
<td>5.01</td>
</tr>
<tr>
<td>Q5XIP6</td>
<td>Flap endonuclease 1</td>
<td>FEN1</td>
<td>**</td>
</tr>
<tr>
<td>G3V9J1</td>
<td>Protein LOC297568</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>D3ZFC6</td>
<td>Inter-alpha-trypsin inhibitor heavy chain family, member 4</td>
<td>ITIH4</td>
<td>-2.69</td>
</tr>
<tr>
<td>B2RYM3</td>
<td>Inter-alpha trypsin inhibitor, heavy chain 1</td>
<td>ITIH1</td>
<td>-1.73</td>
</tr>
<tr>
<td>P08932</td>
<td>T-kininogen 2</td>
<td>KNG1/KNG1</td>
<td>1</td>
</tr>
<tr>
<td>F1LK5</td>
<td>Protein Paqr9</td>
<td>PAQR9</td>
<td>*</td>
</tr>
<tr>
<td>P02767</td>
<td>Transthyretin</td>
<td>TTR</td>
<td>1.58</td>
</tr>
<tr>
<td>D3ZKV1</td>
<td>Monocarboxylate transporter 2</td>
<td>Slc16a7</td>
<td>*</td>
</tr>
<tr>
<td>P02770</td>
<td>Serum albumin</td>
<td>ALB</td>
<td>1.43</td>
</tr>
<tr>
<td>Q68FT8</td>
<td>Protein Serpinf2</td>
<td>SERPINF2</td>
<td>1.62</td>
</tr>
<tr>
<td>Q68FP1-2</td>
<td>Isoform 2 of Gelsolin</td>
<td>GSN</td>
<td>-3.10</td>
</tr>
</tbody>
</table>

*indicates the protein was not detected in the exposed samples. **indicates the protein was not detected in the control samples.

Table 2.1 – List of proteins used as inputs for IPA as determined from PCA analysis (|LM| > 0.5 in PC1). Where a – Uniprot protein Accession number, b – Uniprot protein name/description, c – Uniprot protein symbol, and d – the ratio of the average emPAI score for the exposed samples to the average emPAI score for control samples. * indicates the protein was not detected in the exposed samples. ** indicates the protein was not detected in the control samples.
<table>
<thead>
<tr>
<th>CID(^a)</th>
<th>Molecule(^b)</th>
<th>Fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5280903</td>
<td>Sphingosine 1-phosphate</td>
<td>1.47</td>
</tr>
<tr>
<td>193429</td>
<td>Alpha-phocaecholic acid</td>
<td>-2.16</td>
</tr>
<tr>
<td>19069</td>
<td>Cyclic IMP</td>
<td>-2.36</td>
</tr>
<tr>
<td>13711</td>
<td>2'-Deoxycytidine</td>
<td>1.67</td>
</tr>
<tr>
<td>997</td>
<td>Phenylpyruvic acid</td>
<td>1.66</td>
</tr>
<tr>
<td>5281128</td>
<td>8-Hydroxylinoleic acid</td>
<td>-2.55</td>
</tr>
<tr>
<td>445995</td>
<td>Geranyl diphosphate</td>
<td>-1.85</td>
</tr>
<tr>
<td>6057</td>
<td>L-Tyrosine</td>
<td>1.64</td>
</tr>
<tr>
<td>6140</td>
<td>Docosahexaenoate</td>
<td>1.69</td>
</tr>
<tr>
<td>159623</td>
<td>3-Keto-lithocholic acid</td>
<td>1.72</td>
</tr>
<tr>
<td>1175</td>
<td>Uric Acid</td>
<td>1.72</td>
</tr>
<tr>
<td>6443013</td>
<td>(13S)-Hydroxyoctadecadienoic acid</td>
<td>-2.11</td>
</tr>
<tr>
<td>444899</td>
<td>Arachidonic acid</td>
<td>1.07</td>
</tr>
<tr>
<td>5753</td>
<td>Corticosterone</td>
<td>1.49</td>
</tr>
<tr>
<td>10917</td>
<td>(-)-L-Carnitine</td>
<td>-1.13</td>
</tr>
<tr>
<td>6140</td>
<td>L-Phenylalanine</td>
<td>1.10</td>
</tr>
<tr>
<td>213144</td>
<td>Butyrylcarnitine</td>
<td>-1.12</td>
</tr>
<tr>
<td>5312830</td>
<td>(9S)-Hydroxyoctadecadienoic acid</td>
<td>-1.10</td>
</tr>
</tbody>
</table>

**Table 2.2 – Metabolite inputs for Ingenuity Pathway Analysis**

<table>
<thead>
<tr>
<th>CID(^a)</th>
<th>Molecule(^b)</th>
<th>Fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5280903</td>
<td>Sphingosine 1-phosphate</td>
<td>1.47</td>
</tr>
<tr>
<td>193429</td>
<td>Alpha-phocaecholic acid</td>
<td>-2.16</td>
</tr>
<tr>
<td>19069</td>
<td>Cyclic IMP</td>
<td>-2.36</td>
</tr>
<tr>
<td>13711</td>
<td>2'-Deoxycytidine</td>
<td>1.67</td>
</tr>
<tr>
<td>997</td>
<td>Phenylpyruvic acid</td>
<td>1.66</td>
</tr>
<tr>
<td>5281128</td>
<td>8-Hydroxylinoleic acid</td>
<td>-2.55</td>
</tr>
<tr>
<td>445995</td>
<td>Geranyl diphosphate</td>
<td>-1.85</td>
</tr>
<tr>
<td>6057</td>
<td>L-Tyrosine</td>
<td>1.64</td>
</tr>
<tr>
<td>6140</td>
<td>Docosahexaenoate</td>
<td>1.69</td>
</tr>
<tr>
<td>159623</td>
<td>3-Keto-lithocholic acid</td>
<td>1.72</td>
</tr>
<tr>
<td>1175</td>
<td>Uric Acid</td>
<td>1.72</td>
</tr>
<tr>
<td>6443013</td>
<td>(13S)-Hydroxyoctadecadienoic acid</td>
<td>-2.11</td>
</tr>
<tr>
<td>444899</td>
<td>Arachidonic acid</td>
<td>1.07</td>
</tr>
<tr>
<td>5753</td>
<td>Corticosterone</td>
<td>1.49</td>
</tr>
<tr>
<td>10917</td>
<td>(-)-L-Carnitine</td>
<td>-1.13</td>
</tr>
<tr>
<td>6140</td>
<td>L-Phenylalanine</td>
<td>1.10</td>
</tr>
<tr>
<td>213144</td>
<td>Butyrylcarnitine</td>
<td>-1.12</td>
</tr>
<tr>
<td>5312830</td>
<td>(9S)-Hydroxyoctadecadienoic acid</td>
<td>-1.10</td>
</tr>
</tbody>
</table>

**Table 2.2 – List of metabolites used as inputs for IPA as determined from PCA analysis (|LM| > 0.5 in PC1). Where \(a\) – PubChem CID number, \(b\) – PubChem molecule name, and \(c\) – the fold difference of the average peak intensity for the exposed samples to the average peak intensity for control samples.
analysis software to determine relevant biological pathways based on differentially expressed genes, and protein and small molecule abundance differences [47]. For example, in using pathway topology software such as IPA, biopaths are dependent on the type of cell and condition or disease state analyzed. The results are a snapshot of the biological system at a given time point and are not able to fully characterize the dynamic states of the system. Furthermore, the inter-dependence of pathways may not be realized due to weak links between pathways [47]. Despite the limitations, the information gained from biopathway analysis is useful as a starting point in understanding cause and effect relationships resulting from biological stressors.

2.3 Results and Discussion

2.3.1 Identified proteins and peptides

All identified proteins and peptides were identified at the 95% confidence interval. The LC-MS/MS analysis employed here resulted in identification of 367 total proteins from all the samples; 260 proteins from control samples (n=7) and 207 proteins from exposed samples (n=7). For a full description of the protein complement, see the supplementary information “Chapter 2 protein data supplemental.” 8275 total peptides were found in all the samples with 1361 unique peptides identified; 1114 and 881 unique peptides were from control and exposed samples, respectively. Lengths of peptides observed ranged from 6 to 47 for both the control and exposed samples. Peptide charge states for the control samples have been identified as [M+4H]⁴⁺, [M+3H]³⁺, [M+2H]²⁺, and [M+H]⁺, and comprise ~7.9%, ~36.9%, ~44.9%, and ~10.2%, respectively, of the total peptide compliment. Similarly, the same peptide charge states for the exposed
samples were found to be \( \sim 7.6\% \), \( \sim 39.2\% \), \( \sim 47.7\% \), and \( \sim 5.4\% \), respectively of the total peptide compliment. Using the ratio of the total number of high confidence peptides to the total number of strict target decoy database peptide hits yielded an actual false discovery rate (FDR) of 3.8\% [48].

367 total proteins were identified from all the samples; 78 proteins were identified by at least 2 unique peptides. Control samples (n=7) had 260 total proteins and 72 proteins with at least 2 unique peptide hits, and exposed samples (n=7) had 207 total proteins and 61 proteins with at least 2 unique peptide hits. In order to determine the FDR of proteins, peptide hits were randomly removed equal to the number of false positives (3.8\% of 8275) [49]. The number of proteins identified by the number of remaining peptides was divided by the number of total proteins identified and subtracted from 100\%. This resulted in and FDR of 2.3\%. For a full description of the protein complement, see the supplementary information “Appended emPAI scores” and “emPAI scores from Access” tabs in “Chapter 2 protein data supplemental.”

The protein locations and functions were found using the QuickGO browser [50]. The 367 identified proteins represent 195 unique locations within the cell of 833 total locations identified. The top 19 protein locations totaling 500 hits and their location counts are shown in Figure 2.1A. 319 unique protein functions were identified from a total of 833 functions. The top 18 functions totaling 308 hits and their individual function counts are shown in Figure 2.1B. For a complete list of the locations and functions see supplemental material “Protein locations” and “Protein biofunctions” tabs in “Chapter 2 protein data supplemental.”

The dynamic range of proteins in rat plasma is similar to that of humans,
Figure 2.1 – A. The top 19 protein locations, totaling 500 hits, along with the counts for all identified proteins in control and exposed samples. B. The top 18 protein functions totaling 308 hits for all identified proteins in control and exposed samples.
approximately 9 orders of magnitude [51]. Immunodepletion columns are commonly used to remove a large portion of the most abundant proteins within plasma; for example, albumin and IgG are reported to comprise 80-85% of the total plasma protein. The immunodepletion process does not completely deplete the targeted compounds; in our analysis albumin was still found to comprise a significant fraction of the protein content. Even after immunodepletion the dynamic range of the serum proteins is still several orders of magnitude; however, the detection of mid and low abundant compounds is improved [51].

2.3.2 Protein comparison

emPAI scores were calculated (see equation 1 in Methods: Informatics above) for each identified protein in each sample. The resulting scores and associated proteins were input into a database (Microsoft Access) for query analysis. The compiled emPAI scores for each protein were normalized to the total emPAI value for each sample. Proteins that had 3 or more sample hits were selected to be used for inputs for PCA. Proteins with 2 or fewer samples hits were considered to be indeterminate because they were found in only less than 15% of the samples, whereas proteins with 3 or more sample hits are found in approximately 20% of the samples. See supplementary information for emPAI scores used for PCA inputs on the “Selection of proteins” tab in “Chapter 2 protein data supplemental.” As a result of this filtering, a total of 58 proteins were selected to be used as inputs for PCA (“JMP import” tab supplementary material) and IPA (“IPA input” tab supplementary material) in “Chapter 2 protein data supplemental.”

The PCA of proteins shows a separation of the control and exposed samples
primarily along principal component 2 (PC2, Figure 2.2). Additionally, a smaller deviation is observed along PC1 where the majority of the exposed and control samples are observed to exhibit positive and negative PC1 values, respectively. One exposed sample is shown to have clustered with the control samples; however, the sample was not excluded because the emPAI scores for 43 out of 58 proteins (74.1%) were within 1 standard deviation of the mean exposed emPAI score. Another exposed sample appears to be an outlier, but it was not excluded because the emPAI scores for 57 out of 58 proteins (98.3%) were within 1 standard deviation of the mean exposed emPAI score. 38 proteins were shown to contribute to the separation in PC1 and PC2 at LMV values > 0.8 and subsequently, these proteins were used as inputs for IPA. Table 2.1 shows the top 20 protein inputs based on the highest summed emPAI scores. Complete loadings matrix values of PC1 and PC2 and the calculated loadings matrix vector can be found in supplementary information on the “JMP loadings results” tab in “Chapter 2 protein data supplemental.”

A student’s t test was also performed on the emPAI scores for the 58 proteins (see supplementary information “JMP import” tab in “Chapter 2 protein data supplemental”). 7 proteins were found to be significantly different (p ≤ 0.05); 6 of these 7 proteins were found to be significant in the separation of the groups by PCA. The 7 significant proteins are: inter-alpha trypsin inhibitor, heavy chain 1 (accession B2RYM3), protein Itih4 (accession D3ZFC6), apolipoprotein A-I (accession P04639), alpha-2-macroglobulin (accession P06238), serine protease inhibitor A3N (accession P09006), serotransferrin (accession P12346), hemopexin (accession P20059). The 7 proteins and their fold differences were used separately as inputs for IPA (see supplementary information “IPA
Figure 2.2 – PCA scatter plot, generated in JMP, is derived from the emPAI scores of the 91 most-abundant proteins. Control samples are shown in red and exposed samples are shown in blue. Sample regions are shown as semi-transparent shading of the correlating color.
import significant proteins” tab in “Chapter 2 protein data supplemental”). The resulting biopathway analysis revealed the same top 3 canonical pathways as being significant (p ≤ 0.05) as with the 38 proteins from PCA. There were differences in the z scores compared to the IPA of 38 proteins, but none were considered to be significantly up or down regulated (see supplementary information “IPA output significant proteins” tab in “Chapter 2 protein data supplemental”). The IPA results from the 38 protein inputs represent a more in depth analysis and the remainder of this manuscript will focus on the results for the 38 proteins.

2.3.3 Metabolite comparison

524 unique mass peaks were found in the aqueous fraction. 124 aqueous metabolites were identified by ChemSpider KEGG search and are listed by their PubChem CID number on the “Aqueous Peak IDs” tab in “Chapter 2 metabolite data supplemental.” 299 unique mass peaks were found in the organic fraction. 70 organic metabolites were identified with a ChemSpider KEGG search and are listed on the “Organic Peak IDs” tab in “Chapter 2 metabolite data supplemental.” All identified metabolites were found in both the control (n=5) and exposed (n=6) sample groups. After screening identified metabolites for endogenous mammalian compounds and duplicate compounds, a total of 19 compounds from the aqueous fraction and 4 compounds from the organic fraction were identified. The metabolites and their intensities were used as inputs for PCA. 47% of the variation of the data points can be explained by PC1 (Figure 2.3). The 18 metabolites that contributed the most to the separation in PC1 at an absolute value of ≥ 0.5, and were used for IPA (Table 2.2). The top 6 metabolites (absolute value
Figure 2.3 – PCA scatter plot, generated in JMP, is derived from the normalized peaks areas of the identified endogenous metabolites. Control samples are shown in red and exposed samples are shown in blue. Sample regions are shown as semi-transparent shading of the correlating color.
≥ 0.83) were identified as: sphingosine 1-phosphate (CID 5280903); alpha-phocaecholic acid (CID 193429); cyclic IMP (CID 19069); 2'-deoxycytidine (CID 13711), phenylpyruvic acid (CID 997), and 8-hydroxylinoleic acid (CID 5281128). A complete list of the identified metabolites and their PCA loadings scores can be found in the supplementary information on the “PCA loadings matrix” tab in “Chapter 2 metabolite data supplemental.”

2.3.4 Biopathway analysis

In order to better understand the molecular differences between the two sample groups, a biopathway analysis (see Section 2.2.4, Methods: Informatics above) has been conducted using molecular inputs obtained from the proteomics and metabolomics data. Proteins exhibiting weighted (see section 2.3.2 Protein comparison above) |LM| values >0.5 from the PCA (29 proteins) were combined with the metabolites exhibiting |LM| PC1 values >0.5 from the PCA (18 molecules) and input into IPA to search for interactions with the Ingenuity Knowledge Base molecules. IPA created a network based on the input proteins; metabolites are not considered in the networks, but are used to determine canonical pathways. Input proteins that were found to interact with database proteins were deemed network eligible and a network was generated such that the interactions of the network eligible proteins were maximized. The network was given a score based on the probability of observing the network by chance. The network generated by IPA as determined by the 29 protein and 18 metabolite inputs was given a score of 39. The network is shown in Figure 2.4 and a list of the network molecules and their relationships can be found in supplemental material in “Chapter 2 Cytoscape
Figure 2.4 – Network 1 from IPA showing the inter-relationships between relevant proteins. Solid lines indicate direct relationships and dashed lines indicate indirect relationships. Loops represent feedback systems. The inter-relationships shown are known from the IPA database. The 18 input molecules are highlighted in green (up regulated) and red (down regulated). The network hubs are FN1, IL6, Insulin, and IL1.
network.”

17 input molecules were found in the network (Figure 2.4), comprising of approximately 1/4 of the molecules in the network. Of the 17 input molecules, 7 are up regulated and 10 are down regulated. The network was characterized by Cytoscape [52]. There are 72 nodes with no isolated nodes, 36 multi-edge node pairs, and 59 self-loops. The network diameter, or the largest distance between two nodes is 8 and the network radius, or the minimum non-zero eccentricities of the nodes is 1. The characteristic path length, or the average shortest path length is 2.994 and the number of nearest neighbors is 5.056. The number of connected components is 2, indicating strong connectivity; 27 of the 63 molecules in the network (~43%) have 4 or more connections. The network hubs are IL6, FN1, and Insulin. IPA functional analysis of this network found the top functions associated to be free radical scavenging, cell movement, and cardiovascular system development and function, organismal development. Canonical pathways were generated based on the 47 input proteins and metabolites. The most significant pathways, those with p values <0.05, are found in Figure 2.5. A z score for the pathway is determined from the predicted up or down regulation of a molecule. Those that are predicted to be up regulated are given a +1 (Nₚ) and those that are predicted to be down regulated are given a -1 (Nₜ). The Nₚ and Nₜ are summed to give N. The z score is then calculated by equation 2 [53]:

\[ z = \frac{x}{\sigma_x} = \frac{N}{\sqrt{N}} \]  

(2)

where a z score of ≥ |2| is considered significant. Biopathy analysis of the 47 proteins and metabolites indicated 1 canonical pathway (acute phase response signaling) to have a negative z score, or is down regulated, but not significantly.
Figure 2.5 – IPA canonical pathways generated from input molecules. The bars indicate the p-value of the pathway; blue bars indicate a negative z-score (down regulation of the pathway) and orange bars indicate a positive z-score (up regulation of the pathway). The orange boxes indicate the ratio of the number of input molecules to total number of molecules in the pathway.
The top 3 canonical pathways based on p value were found to be acute phase response signaling, liver x receptor/retinoid X receptor (LXR/RXR) activation, and farnesoid X receptor/retinoid x receptor (FXR/RXR) activation. Biopathway analysis implicated the top 3 biological functions based on the p value as synthesis of reactive oxygen species, activation of cells, and vasculogensis. Synthesis of reactive oxygen species and vasculogensis had negative z scores (down-regulated) and activation of cells had a positive z score (up-regulated). However, the associated z scores were not large enough to be considered significantly up or down regulated, nor were any other calculated z scores from the biological functions considered to be significantly up or down regulated. The top toxicological functions based on p value were proximal renal tubule toxicity, fibrosis of the heart, decreased synthesis of albumin, and fibrosis of the left ventricle. There are no associated z scores with the toxicological functions. The associated molecules, p-values, and z scores for each pathway and function mentioned above are listed in Table 2.3. See supplementary information for complete bioanalysis data in “Chapter 2 protein data supplemental.” We will focus the remainder of our discussion on the canonical pathways found to be most significant from our 47 input proteins and metabolites.

LXR/RXR activation plays a role in regulating lipid metabolism as well as the catabolism of cholesterol to bile acids. FXR/RXR activation regulates many metabolic processes including the regulation of bile, lipids, and glucose and is often complementary and reciprocal to LXR/RXR activation. It is well known that changes in lipid and glucose metabolism are associated with many diseases including type 2 diabetes, obesity, and cardiovascular disease [54]. The p values of these pathways suggest that these
### Table 2.3 – Top statistically significant results from Ingenuity Pathway Analysis

<table>
<thead>
<tr>
<th>Ingenuity Canonical Pathways</th>
<th>P value</th>
<th>Z score</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Phase Response Signaling</td>
<td>3.61E-19</td>
<td>-1.633</td>
<td>ITIH4, C4A/C4B, FN1, TF, FGB, APOA1, SERPINF2, SERPINA3, HPX, ALB, TTR</td>
</tr>
<tr>
<td>LXR/RXR Activation</td>
<td>1.41E-18</td>
<td>0.000</td>
<td>ITIH4, C4A/C4B, APOC2, TF, VTN, APOA1, SERPINF2, HPX, ALB, TTR</td>
</tr>
<tr>
<td>FXR/RXR Activation</td>
<td>2.69E-18</td>
<td>-</td>
<td>ITIH4, C4A/C4B, APOC2, TF, VTN, APOA1, SERPINF2, HPX, ALB, TTR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biofunction Categories</th>
<th>Diseases or Functions</th>
<th>P value</th>
<th>Z score</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Radical Scavenging</td>
<td>Synthesis of Reactive Oxygen Species</td>
<td>7.11E-09</td>
<td>-0.277</td>
<td>ALB, APOA1, APOC2, arachidonic acid, corticosterone, FN1, ITIH4, KNG1, SERPINA3, TF, uric acid</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction</td>
<td>Activation of Cells</td>
<td>4.31E-08</td>
<td>0.192</td>
<td>ALB, APOA1, arachidonic acid, C4A/C4B, FN1, KNG1, L-phenylalanine, L-tyrosine, SERPINA1, SERPINF2, TF, TTR, uric acid, VTN</td>
</tr>
<tr>
<td>Cardiovascular System Development and Function, Organismal Development</td>
<td>Vasculogenesis</td>
<td>3.61E-07</td>
<td>-0.239</td>
<td>APOA1, arachidonic acid, corticosterone, FN1, Kng1/Kng1I1, SERPINA3, Serpina3c/Serpina3m, TF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tox Function Categories</th>
<th>Diseases or Functions</th>
<th>P value</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renal Damage, Renal Tubule Injury</td>
<td>Proximal Tubular Toxicity</td>
<td>1.87E-06</td>
<td>ALB, FGB, FN1, HPX, KNG1</td>
</tr>
<tr>
<td>Cardiac Fibrosis</td>
<td>Fibrosis of Heart</td>
<td>3.00E-04</td>
<td>APOA1, corticosterone, FN1</td>
</tr>
<tr>
<td>Decreased Levels of Albumin</td>
<td>Decreased Synthesis of Albumin</td>
<td>9.98E-04</td>
<td>ALB</td>
</tr>
<tr>
<td>Cardiac Fibrosis</td>
<td>Fibrosis of Left Ventricle</td>
<td>9.98E-04</td>
<td>APOA1</td>
</tr>
</tbody>
</table>

Table 2.3 - Top statistically significant canonical pathways, biological functions, and toxicological functions as determined from the 29 proteins and 18 metabolites input into Ingenuity Pathway Analysis.
pathways are significant in relation to TiO$_2$ NP exposure; however, our data does not indicate whether these pathways are significantly up or down regulated.

Other researchers have studied the effect of NP exposure on lipid and cholesterol levels. Bourdon and coworkers exposed C57BL/6 mice to carbon NPs via intratracheal instillation and analyzed the high density lipoprotein (HDL), low density lipoprotein (LDL), and cholesterol levels at 1, 3, and 28 days post exposure [55]. They found a significant decrease in plasma HDL at 3 and 28 days post exposure, an increase in plasma LDL at 28 days post exposure, and a marginal increase in hepatic cholesterol at 28 days post exposure. Another research group investigated the effect of nickel hydroxide NP via whole body inhalation exposure for 5 hours per day, 5 days per week for a duration of 1 week or 5 months on apoprotein E-deficient mice [56]. In the examination of plasma LDL and cholesterol concentrations, there was found to be no significant difference between exposed and control samples despite the marked increase in atherosclerosis. Additionally, Hu and coworkers exposed apoprotein E-deficient mice to TiO$_2$ NPs at 3 different exposure levels via intratracheal instillation for 6 weeks [57]. The plasma HDL levels were decreased in the high exposure group compared to control and the plasma organic lipid ratios were increased in the high and medium exposure groups relative to the control. The medium and high exposure groups also showed a significant increase in the progression of atherosclerosis.

Although it cannot be discounted that the aforementioned studies have differences in the type of NP exposure and model organism used, these studies suggest that lipid and cholesterol metabolism are affected as a result of NP exposure. The biopathway analysis of our data indicated that the LXR/RXR activation and FXR/RXR activation pathways
are significantly related to the proteins found to contribute to exposed and control sample differences as elucidated by PCA. The results of the previous studies also suggest that significant differences in these pathways at later time points relative to exposure may also be observed. We plan to examine the metabolites, particularly lipids and cholesterol, and proteins at multiple time points post exposure in order to determine if there is a significant up or down regulation in the LXR/RXR activation and FXR/RXR activation pathways as a result of TiO$_2$ NP exposure.

LXR/RXR activation pathway also plays a role in the regulation of inflammation. Acute phase response signaling pathway is a rapid inflammatory response that can be initiated by tissue injury or trauma. It is interesting to note that there is a negative z score (down regulation) associated with acute phase response signaling, but it is not considered significant. Others have observed an up regulation of the acute phase response signaling pathway in response to TiO$_2$ NP exposure via gene expression analysis of mouse lung tissue [35,36]. Husian and coworkers examined differentially expressed genes from mouse lung tissue after intratracheal instillation of low, medium, and high concentrations of TiO$_2$ NPs at 1, 3, and 28 days post exposure. Biopathway analysis of the differentially expressed genes revealed a significant up regulation in the acute phase response signaling and LXR/RXR activation pathways at 1 day post exposure and at all doses [36]. In a separate paper, Saber and coworkers examined the BALF for inflammatory cells in these same mice. The increase in the inflammatory cells found in BALF correlated to an increase in the up regulation of acute phase response signaling and LXR/RXR activation pathways, except at the low dose concentration. The low dose exposure compared to the control did not show a significant difference in the inflammatory cell concentration in
BALF at any time point [32]; however, the genes associated with the inflammatory response were found to be perturbed indicating that inflammatory signaling has been initiated, but without a measurable change in the inflammatory cells [36]. Halappanavar and coworkers also noted an up regulation of the acute phase response pathway as a result of gene expression analysis of mouse lung tissue and an increase in inflammatory cells in BALF as a result of whole body inhalation of TiO$_2$ NPs [35]. While there are noted differences in these studies and ours, such as the species used, amounts of TiO$_2$ NPs exposure, and tissue examined, parallels can be drawn from these studies to our results. We hypothesize that while acute phase response signaling and LXR/RXR activation pathways are observed to be highly associated with the input proteins from plasma, a significant up regulation in these pathways should be observed in vasculature and lung tissues. Future comparative proteomic studies of plasma, vasculature, and lung tissue will be performed to test this hypothesis.

As with the previous studies noted time dependent changes in lipid and cholesterol metabolism [55–57], it would be also prudent to examine different time points post exposure to determine when, or if, a significant up regulation pathway(s) related to the inflammatory response are observed in plasma. Demonstrated by the work of Gustafsson and coworkers, the BALF inflammatory cell concentrations of TiO$_2$ exposed rats via intratracheal instillation were examined at 1, 2, 8, 16, 30, and 90 days post exposure [30]. The authors demonstrated an increase in the inflammatory cells peaking at approximately 16 days post exposure. This work suggests that although an analysis of plasma in the present study demonstrated a non-significant down regulation of the acute phase response at 24 hours post exposure, additional time points are needed in order to
observe changes in the regulation of the inflammatory response.

2.4 Conclusion

In order to determine the molecular pathways involved after pulmonary exposure to TiO$_2$ NPs, comparative proteomic and metabolomic analyses were performed. 29 proteins and 18 metabolites were identified as being able to distinguish between the two sample cohorts. Ingenuity Pathway Analysis of the 29 proteins and 18 metabolites indicated the significant pathways associated with these molecules to be acute phase response signaling, LXR/RXR activation, and FXR/RXR activation. The acute phase response signaling pathways was found to be down regulated, although not significantly. The top 3 significant biological functions were synthesis of reactive oxygen species, activation of cells, and vasculogenesis. Synthesis of reactive oxygen species and vasculogenesis biological functions were down regulated (negative z scores) and the activation of cells biological function was up regulated (positive z score), but none of these biological functions were significantly changed.

We have planned further studies to examine larger cohorts and perform metabolomic and proteomic analyses of plasma, vascular, and lung tissue. Furthermore, we will examine different amounts of NP exposure to examine the differences in molecular response and will examine responses at time points from hours to days post exposure to determine molecular response changes over time.
2.5 References


Chapter 3

Proteomic and Metabolomic Comparative Analyses of Plasma and Vascular Tissue from TiO$_2$ Nanoparticle Exposed Rats

This is the pre-peer reviewed in part version of the following article:


which will be published in final form at http://www.sciencedirect.com/science/journal/22129685.
3.1 Introduction

Since the turn of the 21st century, an increasing body of work has revealed that pulmonary exposure to particulate matter from ambient air pollution produces adverse cardiovascular effects [1–7]. Because engineered nanomaterial (ENM) aerosols share similar, if not identical particle size distributions with ultrafine particulate matter, an increasing concern exists to aerosolized EMN's or nanoparticles (NP) may produce similar outcomes. NPs were once limited to occupational exposures; however, exposures have become much more ubiquitous due to their use in sunscreens [8–10], cosmetics [11–13], paints [14–17], in vivo drug delivery systems [18–21], and environmental remediation solutions [22–24].

Several studies have examined the physiological translocation and distribution of the TiO$_2$ NPs after exposure. Rats were exposed to TiO$_2$ aerosols for 1 hour and then the distribution of particles within the lung tissues were examined at 1 hour and 24 hours post exposure [25]. The researchers found that approximately 24% of the TiO$_2$ NPs had crossed the epithelium barrier of the lungs with no significant difference between the time points. They further concluded that the NPs themselves could move between tissues freely [25]. An earlier study by Ferin, et al. examined the pulmonary retention of TiO$_2$ NPs and fine particles (FP) after single exposure to NPs or FPs or after repeated exposures to NPs or FPs for 12 weeks [26]. The authors concluded that the particle size, dose, and dose rate influenced the migration of particles to the interstitium with the NPs being more likely to migrate [26].

The physical presence of NPs activates clearance mechanisms within the body and has been shown to produce systemic inflammatory responses. TiO$_2$ NP exposure in
animals has previously been shown to increase neutrophils and lymphocytes in bronchoalveolar lavage fluid (BALF) [27–31]. This increase was found to be dependent on method of introduction (whole body inhalation vs. intratracheal instillation) [30], dose [29–31], and post-exposure duration [27–30]. Gene expression analyses of animal lung tissue have demonstrated TiO$_2$ particle exposure stimulates immune and inflammatory responses [31–33]. Further animal studies have also shown an increase in markers associated with oxidative and nitrosative stresses in the microcirculation [34–37].

It has also been demonstrated that different tissues will exhibit a difference in response to NPs at the same post-exposure time point. Halappanavar, et al. examined the gene expression changes in mouse liver and lung tissue following whole body inhalation exposure to TiO$_2$ particles [32]. The authors only found 11 genes in the liver tissue showing differential expression between controlled and exposed cohorts; furthermore, the fold change of these genes was found to be less than 1.6. Within the lung tissue 353 genes were found to be differentially expressed and of these, 53 demonstrated fold changes greater than 1.5 [32]. Another study demonstrated genes within the acute phase response signaling pathway had a more pronounced fold difference in the lungs compared to the liver at 1, 3, and 28 days after exposure to carbon black NPs [38]. Although these studies have demonstrated differential responses in tissues, the heterogeneous effects of NP exposure on the microcirculation within functionally different organs has yet to explored.

The current study employs comprehensive comparative proteomics and metabolomics utilizing LC-MS/MS instrumentation in order to better understand the effect of NP exposure on the response of different levels of arteriolar vasculature.
Proteomic and metabolomics analyses were performed on plasma, aorta, and resistance vascular tissues obtained from rats exposed to TiO$_2$ NPs or filtered air (control). Comparisons of assigned protein and metabolite abundances between sample and tissue types reveal a number of species that may be useful in determining differences in tissue response. These molecules have been used in biopathway analyses to determine biopathways that have been altered as a result of NP exposure. The comparative proteomics and metabolomics, in addition to biopathway analyses, are discussed in light of recent findings associating NP exposure and microvascular dysfunction.

3.2 Materials and Methods

3.2.1 Samples

Experimental animals. Male Sprague Dawley rats (175 - 250 g) were purchased from Hilltop Laboratories (Scottsdale, PA). The animals were housed in laminar flow cages under controlled temperature and humidity conditions and a 12 hr light/12 hr dark cycle at the West Virginia University Health Sciences Center vivarium. Food and water were provided *ad libitum*. Animals were acclimated for 72-hours prior to nanomaterial inhalation exposure. All procedures were approved by the Institutional Animal Care and Use Committee of the West Virginia University (Animal Welfare Assurance Number A3597-01).

Nanoparticles. Nano-titanium dioxide (TiO$_2$) powder was purchased from Evonik (Aeroxide TiO$_2$, Parsippany, NJ). This mixture is composed of anatase (80%) and rutile (20%) TiO$_2$, with a primary particle size of 21 nm, and a surface area of 48.08 m$^2$/g [39–41]. The nano-TiO$_2$ was prepared for aerosolization by drying, sieving, and storing.
the powder [39,42].

**Inhalation Exposure.** The NP aerosol generator and exposure system utilized here has previously been described elsewhere (U.S. Patent #8,881,997) [39,42–44]. Briefly, the apparatus was developed with a vibrating fluidized bed, a Venturi vacuum pump, cyclone separator, impactor and mixing device, an animal housing chamber, and real-time monitoring devices. Aerosols were generated by allowing a high velocity air stream to pass through the vibrating fluidized bed and into the Venturi vacuum pump; drawing air and the TiO$_2$ nanoparticles as it passes. Aerosols enter the cyclone separator, which is gated to remove agglomerates >400 nm at an input flow rate of 60 L/min of clean dry air before entering the exposure chamber.

The mean aerodynamic diameter, size distribution, and relative mass concentration of the aerosols were monitored in real time (Electrical Low Pressure Impactor (ELPI), Dekati, Tempere, Finland). The particle size distribution was measured in real-time with a Scanning Mobility Particle Sizer device (SMPS; TSI Inc., St. Paul, MN). These devices allow for monitoring of the real-time size distribution and mass concentration between a range of 7 nm - 10 μm. The median aerodynamic diameter of the aerosols was approximately 160 nm. Once the steady-state aerosol concentration was achieved, the duration of the exposure was adjusted to achieve a calculated pulmonary deposition of 30 μg/animal. This resulted in exposure times of 4-6 hours per day for a target mass concentration of 6.0 mg/m$^3$ or filtered air (0 mg/m$^3$, control).

**Plasma and tissue isolation.** Rats were anesthetized 24 hours post inhalation exposure with an intraperitoneal injection of thiobutabarbitial sodium (100 mg/kg). A rectal temperature of 37 °C was maintained by placing the animal on a heating pad and
the trachea was intubated to ensure a clear airway. Whole blood was collected through a
cannula in the right carotid artery directly into Vacutainers containing potassium EDTA
(BD Biosciences, San Jose, CA). Blood samples were immediately centrifuged for 10
minutes at 2500 g and then the supernatant (plasma layer) was transferred into 1.5 mL
microcentrifuge tubes. Prior to vascular tissue collection, the hearts were removed from
the anesthetized rats, a quick and efficient means of euthanasia consistent with
recommendations of the Panel on Euthanasia of the American Veterinary Medical
Association. Tissue samples were then harvested and transferred into 1.5 mL
microcentrifuge tubes. Tubes for both plasma and vascular tissue were immediately flash
frozen in liquid nitrogen and then placed at -80 °C for long term storage.

3.2.2 Proteomic and metabolomic extraction and analysis

**Plasma immunodepletion.** A Seppro IgY14 immunodepletion kit (Sigma-
Aldrich, St. Louis, MO) was utilized to immunodeplete the 14 most abundant proteins in
plasma. 20 μL of plasma was first diluted with 500 μL of 1× Dilution buffer. The
diluted plasma was added to a Corning Costar Spin X 0.45 μm column (Sigma-Aldrich,
St. Louis, MO) and centrifuged at 10000 g for 1 minute. The filtrate was added to the
Seppro IgY14 spin columns and immunodepleted following the manufacturer’s protocol.
Immunodepleted samples were frozen overnight at -80 °C and then placed in a vacuum
concentrator at 35 °C for 8 hours. Dried immunodepleted samples were reconstituted in
1 mL of 200 mM ammonium bicarbonate with 6 M urea in water for protein digestion. A
total of 14 control samples and 15 exposed samples were used for plasma protein
analysis.
**Tissue preparation.** 100 µL of 1× PBS buffer was added to each frozen tissue sample and homogenized with a pestle. An additional 100 µL of 1× PBS buffer was used to rinse the pestle. Tissue samples were further processed by sonication on ice for 10 minutes. 400 µL of acetonitrile was then added to precipitate the proteins. Samples were vortexed well and then centrifuged at 1500 g and 4°C for 1 hour for protein and metabolite separation. Supernatant was transferred to another tube for metabolomic analyses (see below) and the precipitate was used for proteomic analysis. 1 mL of 200 mM ammonium bicarbonate with 6 M urea in water was added; samples were vortexed well and sonicated in an ice bath for 10 minutes. Samples were centrifuged at 1500 g at 4°C for 15 minutes to separate the remaining tissue from the supernatant containing proteins. A total of 13 control aorta samples, 11 exposed aorta samples, 15 control vasculature samples, and 15 exposed vasculature samples were prepared for protein analysis.

**Bradford Analysis.** The Bradford Assay was used to determine the total protein concentration from aliquots the immunodepleted plasma samples and the tissue extractions. All samples and BSA standards (0, 0.25, 0.5, 0.75, 1.0, 1.5, and 2.0 µg/µL) were analyzed in duplicate. Protein concentrations were determined at 595 nm on an Epoch Biotek UV-vis spectrophotometer (BioTek, U.S., Winooski, VT). The average sample concentration from the 2 replicates was calculated and the amount of sample extract required to react with 2 µL of 0.01 M DTT at a concentration of 1:40 (protein:DTT) was determined.

**Protein reduction, alkylation, and digestion.** Plasma and tissue protein reduction, alkylation, and digestion were performed as previously described [45].
Briefly, the calculated aliquot of sample was diluted to a volume of 1 mL with 200 mM ammonium bicarbonate with 6 M urea. 2 μL of 0.01M DTT was added to reduce the samples and incubated at 37 °C for 2 hours. The samples were placed on ice and 4 μL of 0.01 M iodoacetamide (IAM) was added for alkylation at 4 °C for 2 hours in the dark. 2 μL of 0.01 M cysteine was added to quench the alkylation reaction and samples remained at room temp for 30 minutes. 0.5 mL of 50 μg/mL trypsin was added to digest the proteins for 18 hours at 37 °C. Digested samples were removed from the incubator and stored at -80 °C overnight. Samples were then dried down for 8 hours at 35 °C using a vacuum concentrator. Sample residues were reconstituted with 150 μL of 5% in water (v/v) with 0.5% (v/v) TFA and desalted with Pierce C18 Spin columns according to the manufacturer’s protocol. Desalted samples were once again dried for 8 hours at 35 °C in a vacuum concentrator and reconstituted in 50% ACN in water (v/v) with 0.1% (v/v) formic acid for LC-MS/MS analysis.

**Metabolite fractionation.** 300 μL of hexane was added to the metabolite fraction (supernatant from protein precipitation step above). Samples were vortexed well and then centrifuged for 5 minutes at 1500 g. The top hexane layer was transferred to another tube for analysis of the organic constituents, and the bottom layer retained the aqueous constituents. The organic and aqueous fractions were dried down with a vacuum concentrator and reconstituted with 50% ACN in water with 15 mM ammonium acetate and then filtered through a nylon 0.2 μm filter prior to LC-MS/MS analysis.

**Peptide and metabolite analysis by LC-MS/MS.** A Thermo Scientific Accela UHPLC coupled to a Thermo Scientific Q Exactive Orbitrap mass spectrometer (Thermo Fisher, San Jose, CA) housed at the WVU BioNano Research Facility was used for LC-
MS/MS proteomic analyses. 18 µL of sample was injected onto a Phenomenex Synergi 4u Fusion C18, 4 µm, 2.0 mm x 100 mm column (Phenomenex, Torrance, CA) maintained at ambient temperature.

Proteomic gradient separations utilized water with 0.1% (v/v) formic acid as mobile phase A and acetonitrile with 0.1% (v/v) formic acid as mobile phase B. The gradient separation using a flow rate of 200 µL/min was as follows: initial 2% B; linear gradient to 60% B over 40 minutes; linear gradient to 90% B over 20 minutes; and hold for 10 minutes. MS settings were: positive mode; m/z range 150-2000; MS resolution 70,000 and automatic gain control (AGC) 1E6; data dependent MS/MS top 10 with signal to noise ratio (S/N) 5, resolution of 17,500 and AGC 1E5, normalized collision energy (NCE) 30, and isolation window 4.0 m/z; spray voltage 1.80 kV; capillary temperature 275°C; and S lens 60 V. Samples were analyzed in a random order.

Metabolite gradient separations used water with 15 mM ammonium acetate as mobile phase A and acetonitrile with 15 mM ammonium acetate as mobile phase B. The flow rates were 200 µL/min for organic and aqueous fraction analysis. For the organic fractions the sample gradient was as follows: initial 50% B, hold for 2 minutes, followed by a linear gradient over 25 minutes to 100% B, and hold for 2 minutes. For the aqueous fractions the sample gradient started at 2% B, was held for 2 minutes, followed by a linear gradient over 25 minutes to 50% B, and was held for 2 minutes. MS settings were the same for both organic and aqueous fractions: positive and negative mode; m/z range 60-750; MS resolution 70,000 and AGC 3E6; data dependent MS/MS top 15 with S/N 5, resolution of 17,500, AGC 1E5, NCE 30, and isolation window 4.0 m/z; spray voltage 3.80 kV; capillary temperature 320°C; and S lens 61 V. Samples were analyzed in a
random order within their respective fraction.

3.2.3 Informatics

Proteomics. Raw data files were analyzed with Proteome Discoverer (v. 1.4.0.288; Thermo Fisher, San Jose, CA) to search against the rat.fasta database. Precursor masses ranging from 150 to 5000 Da were selected from a time range of 0 to 62 minutes. Charge states examined were [M+H]$^+$ to [M+4H]$^{4+}$ using the FTMS mass analyzer, HCD activation, and nanospray ionization source. Peptide lengths ranging from 6 to 144 residues allowing for 2 missed trypsin cleavages were searched with the Sequest algorithm. A dynamic carbamidomethyl modification with 3 modifications maximum per peptide was used. Percolator confidence intervals were set at $p \leq 0.05$ for aorta and plasma samples and $p \leq 0.10$ for vasculature samples (see section 3.3 Results and Discussion for why this confidence level was used). Identified peptide spectra were searched against the decoy database containing reversed protein sequences. A $p \leq 0.10$ relaxed target and $p \leq 0.05$ strict target were used. All other default identification settings were used. Proteins here and in supplementary material are noted by their Uniprot Accession identifier.

Excel Spreadsheet software (Microsoft Office 2007, Microsoft Corporation, Redmond, WA) was utilized to compile proteins identified from Proteome Discoverer. Protein abundance scores were determined using the exponentially modified protein abundance index (emPAI) technique described previously [46]. The emPAI score ($S$) is computed as [46]:

$$S = 10^{\frac{N_{obs}}{N_{exp}}} - 1,$$  

(1)
where \( N_{\text{obs}} \) is the total number of observed peptide ion assignments for a specified protein and \( N_{\text{exp}} \) is the expected number of peptides from the protein sequence. For these comparisons, \( N_{\text{exp}} \) is estimated using the product of the protein sequence length and sum of the frequency of occurrence of trypsin cleavable amino acid residues, lysine (~7.2%) and arginine (~4.2%) [47]. This emPAI score represents relative protein abundances in statistical and biopathway analysis.

Proteins and their emPAI scores were compiled with Access Database software (Microsoft Office 2007, Microsoft Corporation, Redmond, WA). The compiled proteins were further narrowed by selecting for proteins identified by at least 2 unique peptides and then those found in at least 5 samples. A student’s t test was performed using Excel software (Microsoft Office 2007, Microsoft Corporation, Redmond, WA) to determine which proteins were significantly different (\( p \leq 0.1 \)) between control and exposed samples. Proteins that were found to be significantly different in relative abundance were used as inputs for Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity) [48]. A more relaxed p value was used for the t test in order to maximize the number of protein inputs into IPA.

**Metabolomics.** Raw data files were analyzed with Elements for Metabolomics (v. 1.2.1 Proteome Software, Portland, OR). The mixed search mode was utilized to scan \([M + H]^+\) [\(M - H\)], and \([M + NH_3]^+\) ions with \(m/z\) of 50 to 750 Da in a retention time window from 0 to 62 minutes. The parent ion tolerance was set at 20 ppm and the fragment ion tolerance was set at 0.5 ppm. Libraries searched were NIST, LipidBlast, and Human Metabolome Database. Putative assignments were made based on those metabolites with an \(MS^2\) score greater than 0.7. Identified metabolites were further
filtered by those that were found in at least 4 samples. A student’s t test was performed to determine which metabolites were significantly different ($p \leq 0.1$) between control and exposed samples. Finally, metabolites were filtered further to exclude phytochemicals, pharmaceuticals, and other non-endogenous compounds. The metabolites were combined with the proteins and input into IPA. As with the protein inputs, a more relaxed $p$ score was used in order to maximize the number of metabolite inputs for biopathway analysis.

**Biopathway Analysis.** Biological pathways associated with the proteins and metabolites exhibiting significant abundance differences were determined using IPA. The biological pathways for each sample type were acquired from 30 input proteins and 0 metabolites for aorta samples, 21 input proteins and 2 metabolites for plasma samples, and 23 input proteins and 0 metabolites for vasculature samples. The ratio of the average normalized emPAI score of the exposed samples to the average normalized emPAI score of the control samples was calculated to determine the protein fold difference for each sample type. The fold difference for the metabolites was found in the same manner but rather the averaged peak intensities were used. The fold differences for each protein and metabolite were used in pathways analysis in order to determine if any pathways have been up or down-regulated as a result of NP exposure. 10 proteins exhibiting the largest fold difference for each tissue type and the 2 plasma metabolites are listed in Table 3.1. For the full complement of data, see Chapter 3 protein and metabolite supplementary data for all tissues.

The core analysis function in the Ingenuity software was used to generate canonical pathways based on input proteins and metabolites. The Ingenuity Knowledge
# Table 3.1 – Top proteins and metabolites by tissue

<table>
<thead>
<tr>
<th>Aorta Accession</th>
<th>Description</th>
<th>Fold Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1M6T5</td>
<td>Uncharacterized protein</td>
<td>9.48</td>
</tr>
<tr>
<td>P50137</td>
<td>Transketolase</td>
<td>-5.11</td>
</tr>
<tr>
<td>P01026</td>
<td>Complement C3</td>
<td>4.86</td>
</tr>
<tr>
<td>D4AE7B3</td>
<td>Helicase senataxin</td>
<td>4.61</td>
</tr>
<tr>
<td>P63259</td>
<td>Actin</td>
<td>-4.24</td>
</tr>
<tr>
<td>Q01129</td>
<td>Decorin</td>
<td>-4.05</td>
</tr>
<tr>
<td>A0A0G2JUH8</td>
<td>Sorting nexin-13</td>
<td>-4.04</td>
</tr>
<tr>
<td>G3V7K3</td>
<td>Ceruloplasmin</td>
<td>4.00</td>
</tr>
<tr>
<td>B2RZC9</td>
<td>Kielin/chordin-like protein</td>
<td>3.70</td>
</tr>
<tr>
<td>P04636</td>
<td>Malate dehydrogenase</td>
<td>-3.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasma Accession/ HMDB ID</th>
<th>Description</th>
<th>Fold Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>D4ABJ9</td>
<td>Acyl-CoA desaturase 4</td>
<td>6.62</td>
</tr>
<tr>
<td>Q63472</td>
<td>Potassium voltage-gated channel subfamily H member 1</td>
<td>6.42</td>
</tr>
<tr>
<td>F1LS89</td>
<td>Ryanodine receptor 2</td>
<td>5.96</td>
</tr>
<tr>
<td>P04642</td>
<td>L-lactate dehydrogenase</td>
<td>-4.95</td>
</tr>
<tr>
<td>P14480</td>
<td>Fibrinogen beta chain</td>
<td>4.13</td>
</tr>
<tr>
<td>Q03626-2</td>
<td>Murinoglobin-1</td>
<td>3.77</td>
</tr>
<tr>
<td>Q5IOI1</td>
<td>Serine/threonine-protein kinase RIO2</td>
<td>3.69</td>
</tr>
<tr>
<td>P04276</td>
<td>Vitamin D-binding protein</td>
<td>3.56</td>
</tr>
<tr>
<td>D4AB56</td>
<td>Uncharacterized protein</td>
<td>3.17</td>
</tr>
<tr>
<td>F1LSD3</td>
<td>Integrin beta</td>
<td>-3.03</td>
</tr>
<tr>
<td>HMDB08097</td>
<td>PC(18:1(9Z)/14:0)</td>
<td>-2.91</td>
</tr>
<tr>
<td>HMDB01043</td>
<td>Arachidonic acid</td>
<td>2.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vasculature Accession</th>
<th>Description</th>
<th>Fold Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3ZEIH1</td>
<td>Protocadherin Fat 4</td>
<td>-4.26</td>
</tr>
<tr>
<td>A0A0G2JYL5</td>
<td>Axonemal dynein heavy chain 2</td>
<td>-2.22</td>
</tr>
<tr>
<td>D4A2D3</td>
<td>E3 ubiquitin-protein ligase MYCBP2</td>
<td>2.12</td>
</tr>
<tr>
<td>P42335</td>
<td>Inositol-trisphosphate 3-kinase B</td>
<td>-1.89</td>
</tr>
<tr>
<td>F1LRU2</td>
<td>Axonemal dynein heavy chain 17</td>
<td>-1.77</td>
</tr>
<tr>
<td>F1M789</td>
<td>Myosin-13</td>
<td>-1.72</td>
</tr>
<tr>
<td>F1M6V1</td>
<td>Heterochromatin protein 1-binding protein</td>
<td>1.57</td>
</tr>
<tr>
<td>Q64612</td>
<td>Receptor-type tyrosine-protein phosphatase V</td>
<td>-1.55</td>
</tr>
<tr>
<td>Q6ED65-3</td>
<td>Echinoderm microtubule-associated protein</td>
<td>1.46</td>
</tr>
<tr>
<td>G3V9Y1</td>
<td>Myosin, heavy polypeptide 1</td>
<td>-1.43</td>
</tr>
</tbody>
</table>

Table 3.1 - Top 10 proteins for each tissue type exhibiting the greatest fold difference. The 2 metabolites that were found to be significantly different in plasma and their fold difference are shown here.
Base with both direct and indirect relationships was used as the analysis reference set. Molecules and relationships were restricted to the rat species and the number of molecules in networks was restricted to 35 with 25 maximum networks. The top significant canonical pathways based on a calculated p score have been selected for discussion below.

3.3 Results and Discussion

**Proteins.** There were 51218 individual peptides identified in the aortic samples (23003 in the exposed samples and 28214 in the control samples). Aortic peptides ranged in length from 7 to 45 residues for the exposed samples and 7 to 34 residues for the control samples. The charge states of the aortic peptides were comprised of approximately <1% [M+H]$^{1+}$, 31% [M+2H]$^{2+}$, 64% [M+3H]$^{3+}$, and 5% [M+4H]$^{4+}$ ions for both exposed and control samples. There were 90824 individual peptides identified in the plasma samples (49674 in the exposed samples and 41149 in the control samples). Plasma peptides ranged in length from 6 to 47 residues for the exposed samples and 6 to 42 residues for the control samples. The charge states of the plasma peptides were comprised of approximately <1% [M+H]$^{1+}$, 80% [M+2H]$^{2+}$, 39% [M+3H]$^{3+}$, and 1% [M+4H]$^{4+}$ ions for exposed samples and <1% [M+H]$^{1+}$, 72% [M+2H]$^{2+}$, 28% [M+3H]$^{3+}$, and <1% [M+4H]$^{4+}$ ions for control samples. At the p ≤ 0.05 level, there were 450 individual peptides identified in the vasculature samples (272 in exposed samples and 178 in control samples). When the confidence level was decreased to the p ≤ 0.1 level, 187936 individual vasculature peptides were identified (92446 in the exposed and 95490 in the control samples). Due to this great increase in identification, analysis was
continued at the p ≤ 0.1 for the vasculature peptides and proteins. Vasculature peptides identified at the p ≤ 0.1 level ranged in length from 6 to 42 residues for the exposed samples and 6 to 43 residues for the control samples. The charge states of these vasculature peptides were comprised of approximately 75% [M+H]$^{+1}$, 18% [M+2H]$^{+2}$, 7% [M+3H]$^{+3}$, and >1% [M+4H]$^{+4}$, with similar distributions for the exposed and control samples.

A total of 3914 proteins were identified in aortic extracts (1650 in exposed and 2264 in control samples), corresponding to a total of 913 unique proteins (386 in exposed and 527 in exposed samples). A total of 4668 proteins were identified in the plasma extracts (2631 in exposed and 2037 in control samples), corresponding to a total of 1313 unique proteins (732 in exposed and 581 in exposed samples). At the p ≤ 0.05 level, a total of 271 proteins were identified in the vasculature extracts (136 in the exposed samples and 135 in the control samples). This corresponds to 107 unique total proteins (65 in the exposed samples and 57 in the control samples). When the confidence level was decreased to p ≤ 0.1, a total of 52538 proteins were identified in the vasculature extracts (25176 in the exposed and 27362 in the control), corresponding to 2058 unique proteins (922 in exposed and 1136 in control). It is unclear as to what caused the significantly lower number of proteins identified at the at the p ≤ 0.05 level. For the statistical and biopathway analysis, the complement found at the p ≤ 0.1 level was used.

The locations and functions of the identified proteins in all groups were found using the Quick GO database (European Bioinformatics Institute, Hinxton, Cambridge, UK) [49]. The top 10 locations and functions for each sample type can be found in Table 3.2. The full complement of location and function can be found in Chapter 3 protein
Table 3.2 – Top protein locations and functions by tissue

<table>
<thead>
<tr>
<th>Aorta Location</th>
<th>Count</th>
<th>Plasma Location</th>
<th>Count</th>
<th>Vasculature</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membrane</td>
<td>111</td>
<td>Nucleus</td>
<td>324</td>
<td>Nucleus</td>
<td>86</td>
</tr>
<tr>
<td>Cytoplasm</td>
<td>108</td>
<td>Membrane</td>
<td>289</td>
<td>Membrane</td>
<td>83</td>
</tr>
<tr>
<td>Nucleus</td>
<td>101</td>
<td>Cytoplasm</td>
<td>228</td>
<td>Cytoplasm</td>
<td>74</td>
</tr>
<tr>
<td>Integral component of</td>
<td></td>
<td>Integral component of</td>
<td>186</td>
<td>Integral component</td>
<td>57</td>
</tr>
<tr>
<td>membrane</td>
<td></td>
<td>membrane</td>
<td></td>
<td>membrane</td>
<td></td>
</tr>
<tr>
<td>Plasma membrane</td>
<td>55</td>
<td>Plasma membrane</td>
<td>157</td>
<td>Plasma membrane</td>
<td>39</td>
</tr>
<tr>
<td>Cellular component</td>
<td>50</td>
<td>Extracellular region</td>
<td>131</td>
<td>Extracellular</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exosome</td>
<td></td>
<td>exosome</td>
<td></td>
</tr>
<tr>
<td>Extracellular exosome</td>
<td>46</td>
<td>Extracellular space</td>
<td>74</td>
<td>Intracellular</td>
<td>23</td>
</tr>
<tr>
<td>Nucleoplasm</td>
<td>38</td>
<td>Extracellular exosome</td>
<td>67</td>
<td>Nucleolus</td>
<td>19</td>
</tr>
<tr>
<td>Extracellular region</td>
<td>34</td>
<td>Mitochondrion</td>
<td>64</td>
<td>Mitochondrion</td>
<td>17</td>
</tr>
<tr>
<td>Extracellular space</td>
<td>31</td>
<td>Cellular</td>
<td>64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aorta Function</th>
<th>Count</th>
<th>Plasma Function</th>
<th>Count</th>
<th>Vasculature</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal ion binding</td>
<td>81</td>
<td>ATP binding</td>
<td>112</td>
<td>Nucleotide binding</td>
<td>50</td>
</tr>
<tr>
<td>Molecular function</td>
<td>55</td>
<td>Metal ion binding</td>
<td>103</td>
<td>ATP binding</td>
<td>45</td>
</tr>
<tr>
<td>ATP binding</td>
<td>46</td>
<td>Protein binding</td>
<td>90</td>
<td>Metal ion binding</td>
<td>33</td>
</tr>
<tr>
<td>Nucleotide binding</td>
<td>40</td>
<td>DNA binding</td>
<td>84</td>
<td>Protein binding</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poly(A) RNA binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleic acid binding</td>
<td>33</td>
<td>Molecular function</td>
<td>75</td>
<td>Zinc ion binding</td>
<td>23</td>
</tr>
<tr>
<td>Zinc ion binding</td>
<td>33</td>
<td>Nucleotide binding</td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrolase activity</td>
<td>33</td>
<td>Nucleic acid binding</td>
<td>44</td>
<td>DNA binding</td>
<td>22</td>
</tr>
<tr>
<td>Protein binding</td>
<td>29</td>
<td>Hydrolase activity</td>
<td>42</td>
<td>Molecular function</td>
<td>21</td>
</tr>
<tr>
<td>DNA binding</td>
<td>25</td>
<td>Transferase activity</td>
<td>42</td>
<td>Transferase activity</td>
<td>19</td>
</tr>
<tr>
<td>Transferase activity</td>
<td>22</td>
<td>Zinc ion binding</td>
<td>40</td>
<td>Nucleic acid binding</td>
<td>16</td>
</tr>
</tbody>
</table>

*Table 3.2.* Top 10 locations and functions for proteins in each tissue type.
supplementary material for each tissue.

The Venn diagram in Figure 3.1 depicts the relationship of proteins within each sample type from a total of 3001 identified proteins; 707 from aorta, 1045 from plasma, and 1588 from vasculature. The Venn Diagram Generator from the Bioinformatics and Evolutionary Genomics group at Ghent University (http://bioinformatics.psb.ugent.be/webtools/Venn/) [50] was utilized to generate the diagram. A list of common and unique proteins for each group and comparison can be found in Chapter 3 protein supplementary material for each tissue. There were 13 proteins that were found in all 3 sample types; IPA analysis of these proteins revealed the significant canonical pathways (p < 0.05) as “sucrose degradation V (mammalian)” and “agrin interactions at neuromuscular junction.” There were 130 proteins found to be shared between plasma and vasculature sample types. IPA analysis of these proteins revealed “cell cycle control of chromosomal replication” and “cell cycle: G2/M DNA damage checkpoint regulation” were found to be the top significant canonical pathways. There were 72 proteins found to be similar between aorta and vasculature samples. Among the top significant pathways were found to be “adipogenesis pathway” and “tRNA charging.” Finally, there were 111 proteins found in both plasma and aorta samples. The top 3 canonical pathways as found by IPA analysis of these proteins were “FXR/RXR activation,” “LXR/RXR activation,” and “acute phase response signaling.”

Metabolites. A total of 2242 metabolites were identified in the aqueous fraction across all aorta samples, 531 of these metabolites (294 in exposed and 237 in control) had MS² scores. The aorta organic fraction compliment contained a similar number of identified metabolites, 2297, with 541 metabolites (295 exposed and 246 in control)
Figure 3.1 - A Venn diagram demonstrating the overlap of all identified proteins between sample types. 13 proteins were found to be common among all sample types. There were 130 proteins in common in plasma and vasculature sample types. 72 proteins were found to overlap between aorta and vasculature samples. There were 111 proteins in common between aorta and plasma samples. Ingenuity Pathway Analysis was used to determine the significant canonical pathways from these overlapping proteins.
having a MS² score. The combined plasma samples contained a total 1965 metabolites in the aqueous fraction with 470 metabolites (272 in exposed and 199 in control) with a MS² score. The plasma organic fraction contained 6634 total identified metabolites, and 825 metabolites (431 in exposed and 394 in control) with a MS² score. The combined small vasculature samples had 2147 metabolites identified in the aqueous fraction with 517 metabolites (272 from exposed and 245 from control) with MS² scores. Finally, the vasculature organic fraction compliment contained 2030 identified metabolites, with 480 metabolites (236 from exposed and 244 from control) given MS² scores.

The Venn diagram in Figure 3.2 shows the relationship of the identified metabolites in aorta, vasculature, and plasma. There were a total of 116 combined unique aqueous and organic metabolites identified in aorta tissue, 180 combined unique aqueous and organic metabolites identified in plasma, and 105 combined unique aqueous and organic metabolites identified in vasculature. 99 metabolites were common to all three tissue types, 10 were common between aorta and plasma, but none were common between aorta and vasculature or plasma and vasculature. 6 unique metabolites were found in the vasculature, 7 unique metabolites were found in the aorta, and there were 71 unique metabolites in the plasma.

Of the identified metabolites in all tissue types, disregarding those that did not have MS² scores and were identified as phytochemicals, pharmaceuticals, or other non-endogenous compounds, only 2 were found to be significantly different between control and exposed samples. The low number of significantly different identified metabolites could be due to the metabolite identification software. Using metabolite identification software to identify metabolites in LC-MS/MS analyses is complicated due to matrix
Figure 3.2 - A Venn diagram demonstrating the overlap of all identified aqueous and organic metabolites between sample types. 99 metabolites were found to be common among all sample types. There were 10 metabolites in common in plasma and aorta sample types, but there was no overlap between aorta and vasculature or plasma and vasculature samples.
effects from LC mobile phases or sample solution, differences in fragmentation energies between the library and experimental conditions, and isobaric and isomeric interferences [51,52]. These interferences and the drawbacks to metabolite identification software are discussed in greater detail in Chapter 5.

**Biopathway results.** Proteins and metabolites that were found to be significantly different between exposed and control samples by a student’s t test for each tissue type were input into IPA for biopathway analysis. IPA revealed similar results for the aorta and plasma samples; however, different, yet related results for vasculature. Aorta and plasma samples both have “acute phase response signaling,” “LXR/RXR activation,” and “FXR/RXR activation” as being in the top 5 significant pathways. These pathways and their associated $-\log(p \text{ value})$ can be found in Figure 3.3. The full list of significant pathways and the associated p values can be found in Chapter 3 protein supplementary material for aorta and plasma. These 3 pathways were found to be the top 3 pathways in previous analysis of plasma [45]. “Acute phase response signaling” is a rapid inflammatory response to injury or trauma [53]. “LXR/RXR activation” helps to regulate inflammation in addition to regulating lipid metabolism and cholesterol catabolism [54]. “FXR/RXR activation” is complementary to “LXR/RXR activation” and works to regulate bile, lipids, and glucose levels [55]. Dysregulations of the “acute phase response signaling” and “LXR/RXR activation” pathways are expected as many researchers have noted increased inflammatory markers as a result of NP exposure [32,33,56–59]. However, perhaps less expected are observed changes in pathways associated with lipid and cholesterol metabolism.

Recent studies have demonstrated that nanoparticle exposure in rats significantly
Figure 3.3 - Significant canonical pathways as determined by IPA and the associated $-\log(p\ value)$. Acute phase response signaling, LXR/RXR activation, and FXR/RXR activation were found to be significant for aorta and plasma. ILK signaling, D-myo-inositol (1,3,4)-trisphosphate biosynthesis, and 1D-myo-inositol hexakisphosphate biosynthesis II (mammalian) were found to be significant for vasculature tissue. The p values for the ILK signaling pathway in aorta and plasma were found not to be significant.
alters lipid levels [60,61]. In a 2015 study by Sulaiman, et al., rats orally exposed to gold-silver nanoparticles were shown to have significant changes in serum lipid levels relative to the control (p < 0.05); however, these changes did not correlate with dosage [60]. For example, rats exposed to 10 mg/kg gold-silver nanoparticles had increased levels of total cholesterol, triglycerides, glycerol, and LDL-C relative to the control group. Conversely, rats exposed to 50 and 100 mg/kg of gold-silver nanoparticles had decreased levels of total cholesterol, triglycerides, and glycerol relative to the control group, but no significant difference in LDL-C levels. Furthermore, there was no significant difference found between control and 10 mg/kg gold-silver nanoparticle exposed rats in HDL-C levels, but the 50 mg/kg gold-silver nanoparticle exposed rats were significantly higher than the control cohort and the 100 mg/kg gold-silver nanoparticle exposed rats were significantly lower than the control [60].

Recently, Adeyemi, et al., exposed rats (oral) to platinum nanoparticles and similarly there were significant alterations (p < 0.05) in serum lipid levels that did not correlate with dose [61]. In this study, all exposure levels had increased levels of total cholesterol relative to control. However, the 10 mg/kg and 50 mg/kg platinum nanoparticle exposed rats exhibited significantly increased levels of triglycerides and glycerol and the 100 mg/kg exposed rats exhibited significantly decreased levels of triglycerides and glycerol. Furthermore, there was no significant difference found in HDL-C levels for 10 mg/kg exposed rats, but there was significant increase in HDL-C levels for 50 mg/kg and 100 mg/kg exposed rats. For LDL-C levels, there was no difference for the 100 mg/kg exposed rats, but there was a significant increase in the 10 mg/kg and 50 mg/kg exposed rats [61].
These two recent studies further support the idea that NP exposure can alter the “LXR/RXR activation” and “FXR/RXR activation” pathways, but these alterations appear to be dose and nanoparticle type dependent [60,61]. As these pathways are related to liver function (cholesterol and bile regulation), it is unclear if the nanoparticles themselves are effecting liver function or if it is a cascading biological effect. This present study and a previous study [45] did not indicate the “LXR/RXR activation” and “FXR/RXR activation” pathways as being significantly up or down regulated. Because at 24 hours post exposure significant alterations in these pathways are not detected, it may be hypothesized that examining time points prior to or after 24 hours would present significant alterations.

Biopathway analysis for the vascular tissue samples revealed the top 5 pathways as “ILK signaling,” “methionine degradation of homocysteine,” “D-my-o-inositol (1,3,4)-trisphosphate biosynthesis,” “cysteine biosynthesis III,” and “1D-my-o-inositol hexakisphosphate biosynthesis II (mammalian).” This work will focus on the “ILK signaling” pathway and the two inositol pathways. These three pathways and their associated \(-\log(p\text{ value})\) are shown in Figure 3.3. The full list of significant pathways and the associated p values can be found in the “Chapter 3 Vasculature protein supplementary material.” ILK signaling has been shown to regulate the transcription of cyclo-oxygenase II (COX-2), an enzyme that when upregulated promotes production of prostaglandins leading to inflammation [62]. Previous histological analyses has indicated that prostaglandins contribute to the dysfunction and vasodilatory response to acetylcholine in coronary arterioles [34]. The inositol biosynthesis pathways help to regulate calcium stores and channels within the cells which are important as secondary messengers for
signal transduction [63,64]. Other researchers have also noted changes in calcium signaling due to TiO\textsubscript{2} NP exposure [33,65].

In work performed by Sun, et al. an intratracheal instillation of TiO\textsubscript{2} NPs for 90 consecutive days at different doses examined genes and proteins in lung tissue related to inflammation [59]. The researchers found a dose dependent response for many inflammatory markers including COX-2, nuclear factor-κB (NF-κB), and TNF-α, all of which are in the “ILK signaling” pathway. Both the genes and proteins for COX-2, NF-κB, and TNF-α were found to be significantly increased in relation to increased dose of TiO\textsubscript{2} NPs. Additionally, the researchers noted TiO\textsubscript{2} NP exposure increased the production of reactive oxygen species and lipid peroxidation resulting in activation of inflammatory pathways [59]. Although the TiO\textsubscript{2} NP exposure times and total amounts differ drastically between the aforementioned study and the present study, TiO\textsubscript{2} NPs appear to elicit similar responses for acute and chronic exposures. Such findings suggest that future studies on the long term effects of acute and chronic exposures would be prudent.

Work performed by Husain, et al. examined the effect of low, medium, and high doses of TiO\textsubscript{2} NPs via intratracheal instillation at 1, 3, and 28 days post exposure on the gene expression and subsequent biopathway analysis of lung tissue [33]. At 3 days after exposure, the “ILK signaling” and “calcium signaling” pathways were found to be significant in all exposures. Many of the down-regulated genes identified in the day 3 analysis are known to regulate calcium homeostasis and muscle contraction resulting in changes in muscle contraction and ion concentration within the lungs. However, there was little to no inflammation noted in the lung tissue at low and medium doses. The
authors concluded that inflammation may not be the underlying cause of changes in ion signaling, but rather the result of particle retention in the lungs [33].

Although the significant pathways found for vasculature in this study differ from those found in plasma and aorta, there are similarities in the functions of these pathways. “ILK signaling” could be considered a pre-inflammatory response, or a pathway that might be activated prior to the acute phase response signaling pathway. In “ILK signaling” COX-2 transcription is increased to promote prostaglandin production, which in turn stimulates inflammation. Prostaglandin synthesis can also be controlled through the “LXR/RXR activation” pathway regulation of NF-κB mediated transcription of COX-2. Figure 3.4 demonstrates the similarities in the signaling cascades within these 2 pathways. The “D-myo-inositol (1,3,4)-trisphosphate biosynthesis” and “1D-myo-inositol hexakisphosphate biosynthesis II (mammalian)” pathways regulate calcium stores required for certain signaling cascades to occur. For example, the PI3K signaling cascade within the “acute phase response” pathway is dependent on calcium signaling for proper function. The differences observed in the significant pathways from the vasculature compared to the significant pathways from plasma and aorta may be a result of time-dependent tissue response due to the proximity of the tissue to the initial point of exposure. Future work will examine vascular tissue more distal from the lungs and heart in addition to the plasma, aorta, and vascular tissue close to the heart to determine if there are other early response pathways involved.

3.4 Conclusion

There were observed differences in the significant pathways as determined by
Figure 3.4 - The signaling cascade within the ILK Signaling pathway which leads to activation of NF-kB to promote transcript on COX-2 is highlighted in blue. The signaling cascade with LXR/RXR activation which leads to activation of NF-kB is highlighted in red. Direct interactions between proteins are depicted by solid lines and indirect interactions are depicted by dashed lines. COX-2 promotes the production of prostaglandins which result in inflammation. ECM – extracellular matrix proteins; ILK – integrin linked protein kinase; AKT – protein kinase B; LPB - lipopolysaccharide binding protein; LPS – lipopolysaccharide; LY96 – lymphocyte antigen 96 protein; CD14 - cluster of differentiation 14 protein; TLR4 – toll-like receptor 4 protein; NF-κB – nuclear factor-κB; and COX-2 – cyclooxygenase-2.
biopathway analysis of extracted proteins and metabolites from aorta, vasculature, and plasma from rats exposed to TiO$_2$ NPs. In aorta and plasma, the “acute phase response signaling,” “LXR/RXR activation,” and “FXR/RXR activation” pathways were found to be significant and are consistent with a previous analysis of plasma from TiO$_2$ exposed rats [45]. These pathways are related to inflammation and changes in lipid and cholesterol metabolism. However, analysis of the resistance vasculature revealed “ILK signaling,” “D-myoinositol (1,3,4)-trisphosphate biosynthesis,” and “1D-myoinositol hexakisphosphate biosynthesis II (mammalian)” as being significant. These pathways are associated with changes in inflammation through production of prostaglandins and changes in calcium signaling and storage within the cells. Despite these outward differences between the significant pathways observed in the vasculature compared to the plasma and aorta, the observed pathways have similar functions. The differences may result from temporal differences in response due to the proximity of the tissue to the initial site of injury.

The differences in the significant pathways in the tissue types suggest the type of response elicited results from not only the tissue distance to the initial site of NP introduction, but also the difference in the structure and function of the tissue. This response is expected to change over time and by tracking these changes throughout different sections of the vascular tree, it may be able to discern the systemic effect of NP exposure. Future studies will examine the effect over time and the responses of will analyze structurally and functionally different sections of the vascular tree and additional organs such as the liver, kidney, heart, brain, and nervous system.
3.5 References


Chapter 4

Proteomic and Metabolomic Analysis of Brain Tissue Following Intracerebroventricular Administration of Streptozotocin in Mouse Models of Tauopathies by LC-MS/MS

4.1 Introduction

Tauopathies are a group of neurodegenerative disorders associated with pathological misfolding and aggregation of the tau proteins [1] and include diseases such as Alzheimer’s (AD) and frontotemporal dementia (FTD). Although there are reported mutations of tau that lead to neurodegenerative diseases, such as FTD [2] and Pick’s disease [3], AD has not been reported to contain mutated forms of tau. Mouse models used to study AD often rely on mutated forms of tau predominate in other tauopathies. For example, the rTg(TauP301L)4510 mouse expresses a proline-to-leucine mutation at the 301\textsuperscript{st} amino acid residue (P301L) of tau typically found in FTD. In order to create a more AD-like model, increased hippocampal and neocortical expression of P301L tau is driven by a forebrain-specific calcium- and calmodulin-dependent protein kinase-II promoter system [4]. These P301L mutated mice present with tau hyperphosphorylation, neurofibrillary tangles, neuron loss, and memory impairment [4], all of which match tau-related changes observed in AD. However, this model is not phenotypically accurate as tau mutations are not observed in AD. Even among individuals with FTD, only approximately 10\% of patients have a familial form of the disease associated with a single gene mutation, such as the P301L mutation. Most cases of FTD are considered sporadic, or not inherited. Thus, while transgenic models of familial AD and FTD are useful, understanding how they differ from sporadic forms of these diseases is critical.

Recent studies have demonstrated early stages of AD and FTD are preceded or accompanied by deficits in glucose utilization and energy metabolism [5–8]. Intracerebroventricular (ICV) infusions of streptozotocin (STZ) have been shown to produce insulin-resistance and decreased glucose metabolism in the brain, as well as a
number of other molecular and pathological changes that mimic AD and FTD, including tau hyperphosphorylation [9]. However, the effects of ICV-STZ have only been examined in rodents that have the rodent form, not the human form, of tau. Here, ICV-STZ administration has been explored as a means to initiate early metabolic changes observed in the brain of mice expressing the human P301L tau mutation and the wild-type human tau in order to create a more phenotypically relevant mouse model of AD and FTD. This work is interested in the changes within the brain proteome as a whole rather than changes observed in tau protein as a result of the ICV-STZ administration.

Comparative bottom-up proteomic analyses of brain tissue were employed in order to determine if ICV-STZ treatment in wild-type tau mice could be used an alternative tauopathy model to one which relies on tau genetic mutations (P301L mice). Brain hemispheres, minus the hippocampus, were analyzed to determine the proteins and biopathways affected as a result of the genetic background and/or treatment. Proteomic analyses were performed on high resolution LC-MS/MS instrumentation. Biopathway analysis was performed using a pairwise approach to compare cohorts. Comparisons between cohorts revealed the Protein Kinase A signaling pathway as being down-regulated, although not significantly, in 4 of the 6 comparisons. Furthermore, the 14-3-3 Mediated Signaling pathway was the most significant pathway based on the input proteins; however, it was not found to be up or down-regulated in any comparison. There is evidence to support that ICV-STZ treatment has an effect on protein abundances and signaling cascades within human tau mouse models.

4.2 Methods
4.2.1 Samples

**Subjects and Treatments.** Mouse tissues samples were generated through a collaborative effort with the Reed lab in conjunction with other work. The generation of rTg(TauP301L)4510 (hereafter called P301L) and rTg(TauWT)21221 (hereafter called WT) mice has been described elsewhere [10]. The P301L, WT, and genetic control mice used in this study had unrestricted access to food and water and were group housed in a temperature- and humidity-controlled colony room with a 12-hour light/dark cycle. To suppress human tau expression during development, dams were placed on doxycycline water (40 ppm) 2 weeks prior to breeding, and experimental offspring continued doxycycline until they were approximately 2.5 months of age. On the day of ICV-STZ surgery, doxycycline water was replaced with regular water and transcription of human tau was allowed to occur for approximately 4 months [11]. All experimental procedures were approved by the West Virginia University Animal Care and Use Committee (Animal Welfare Assurance Number A3597-01).

**Surgery and ICV-STZ Administration.** On the same day mice were removed from doxycycline, mice received either a single, bilateral infusion of ICV-STZ (3 mg/kg; 2.5 µl/side, hereafter called STZ) dissolved in 0.05 M citrate buffer vehicle (pH 4.5) or vehicle (hereafter called Veh) into the lateral ventricles (AP: -0.5 mm, ML: +/- 1.0 mm, DV: -2.5 mm). Microinfusion (GenieTouch Syringe Pump, Kent Scientific, Connecticut, USA) of STZ or Veh was performed with a 28-gauge injector attached to a 25 µl Hamilton syringe at a steady rate of 1 µl/min for 2.50 min per side with an additional 5 min for diffusion.

**Dissection.** At 6 months of age, after 4 months of human tau expression, mice
were euthanized with carbon dioxide. The brain hemispheres were separated, the hippocampus removed for other analyses, and the remaining hemispheres were immediately frozen on dry ice, and stored in the freezer at -80 °C. Due to problems with sample preparation and storage, we were only able to secure limited numbers of samples for analysis: CT-STZ (n=4), CT-Veh (n=2), P301L-STZ (n=5), P301L-Veh (n=3); WT-STZ (n=5) and WT-Veh (n=6). CT-Veh samples were not used for statistical analysis due to the small sample size.

4.2.2 Proteomics extraction and analysis procedures

Sample extraction. Brain tissue samples were manually homogenized with 100 µL of 1× PBS buffer and a pestle. The pestle was rinsed with an additional 100 µL of 1× PBS buffer and the samples were sonicated on ice for 10 minutes. Proteins were precipitated with 400 µL of acetonitrile followed by vortexing and centrifugation for 1 hour at 1500 g and 4°C. Supernatant was used for metabolomic analysis described in section 4.2.3, and the solids were used for proteomic analysis. 1 mL of 6 M urea with 200 mM ammonium bicarbonate in water was added to the solids. Samples were vortexed and sonicated on ice for 10 minutes. Samples were centrifuged for 15 minutes at 1500 g at 4°C. The Bradford assay was performed on an aliquot of the supernatant to determine total protein content of the sample.

Bradford Analysis. 5 µL of sample and standards of bovine serum albumin (BSA; 0, 0.25, 0.5, 0.75, 1.0, 1.5, and 2.0 µg/µL) were added to a 96 well plate in duplicate. 245 µL of Bradford reagent was added to each well and allowed to react at room temperature for 5 minutes. Protein concentrations were determined at 595 nm on an
Epoch Biotek UV-vis spectrophotometer (BioTek, U.S. Winooski, VT). The average sample concentration was calculated and the amount of sample extract required to react with 2 µL of 0.01 M dithiothreitol (DTT) at a ratio of 1:40 (protein:DTT) was determined.

**Protein reduction, alkylation, and digestion.** Protein extractions were prepared as previously described [12]. Briefly, the sample aliquot was diluted to 1 mL with 200 mM ammonium bicarbonate with 6 M urea. Disulfide bonds were reduced by adding of 2 µL 0.01M of DTT and incubating at 37°C for 2 hours. Samples were removed from the incubator and immediately placed on ice. Samples were alkylated by adding 4 µL of 0.01 M iodoacetamide (IAM) and incubating at 4°C for 2 hours in the dark. Excess IAM was quenched with 2 µL of 0.01 M cysteine and incubated at room temp for 30 minutes. Proteins were digested with 0.5 mL of 0.05 mg/mL trypsin and incubated at 37°C for 18 hours. Following trypsin digestion, the samples were stored at -80°C overnight. Samples were then dried for 8 hours at 35°C in a vacuum concentrator. Samples were reconstituted in 150 µL sample buffer [5% (v/v) ACN with 0.5% (v/v) TFA in water] and desalted with Pierce C18 Spin columns following the manufacturer’s protocol. After desalting, samples were dried at 35°C and reconstituted in 50% (v/v) ACN with 0.1% (v/v) formic acid in water for LC-MS/MS analysis.

**Proteomic analysis by LC-MS/MS.** A Thermo Scientific Accela UHPLC coupled to a Thermo Scientific Q Exactive Orbitrap mass spectrometer (Thermo Fisher, San Jose, CA) housed at the WVU BioNano Research Facility was used for LC-MS/MS proteomic and metabolomic analyses. 18 µL of sample digest was injected onto a Phenomenex Synergi 4u Fusion C18, 4 µm, 2.0 mm x 100 mm column (Phenomenex,
Torrance, CA) at ambient temperature. UHPLC separations utilized water with 0.1% (v/v) formic acid as mobile phase A and acetonitrile with 0.1% (v/v) formic acid as mobile phase B. The gradient at a flow rate of 200 μL/min was: initial 2% B; linear gradient to 60% B over 40 minutes; linear gradient to 90% B over 20 minutes; hold for 10 minutes. Column was re-equilibrated by returning the gradient 2% B over 5 minutes and holding for 10 minutes. MS settings were: m/z range 150-2000; MS resolution 70,000 and automatic gain control (AGC) 1E6; and positive mode. The MS/MS settings were: data dependent MS/MS of the top 10 peaks with S/N 5, resolution of 17,500 and AGC 1E5, normalized collision energy (NCE) 30, and isolation window 4.0 m/z. The spray voltage was set to 1.80 kV, capillary temperature set to 275°C and the S lens set to 60 V. Samples were analyzed in a random order.

4.2.3 Metabolomics extraction and analysis procedures

Metabolite fractionation. 300 μL of hexane was added to the supernatant from section 4.2.2. Samples were vortexed well and centrifuged at 1500 g for 5 minutes. The hexane layer was transferred to another tube (organic fraction). The organic and aqueous fractions were dried down with a vacuum concentrator and reconstituted with 50% ACN in water with 15 mM NH₄OH. Extracts were filtered through a nylon 0.2 μm filter prior to LC-MS/MS analysis.

Metabolite analysis by LC-MS/MS. 18 μL of sample was injected onto a Phenomenex Synergi 4u Fusion C18, 4 μm, 2.0 mm x 100 mm column (Phenomenex, Torrance, CA) with the column at ambient temperature and analyzed by a Thermo Q Exactive Orbitrap. Metabolite gradient separations used water with 15 mM NH₄OH as
mobile phase A and acetonitrile with 15 mM NH$_4$OH as mobile phase B. The flow rates were 200 μL/min for organic and aqueous fraction analysis. For the organic fractions the sample gradient was as follows: 50% B hold for 2 minutes followed by a linear gradient over 25 minutes to 100% B and hold for 2 minutes. For the aqueous fractions the sample gradient started at 2% B and was held for 2 minutes. This was followed by a linear gradient over 25 minutes to 50% B and held for 2 minutes. MS settings were the same for both organic and aqueous fractions: negative mode; m/z range 60-750; MS resolution 70,000 and AGC 3E6; data dependent MS/MS top 15 with S/N 5, resolution of 17,500, AGC 1E5, NCE 30, and isolation window 4.0 m/z; spray voltage 3.80 kV; capillary temperature 320°C; and S lens 61 V. Samples were analyzed in a random order within their respective fraction.

4.2.4 Informatics

Proteomics. Proteome Discoverer (v. 1.4.0.288; Thermo Fisher, San Jose, CA) was utilized to analyze raw proteomics data. Proteome Discoverer searched for peptides in a retention time window ranging from 0 to 62 minutes with masses ranging from 150 to 5000 Da and charge states from [M+H]+ to [M+4H]4+. The mass analyzer settings were FTMS with an electrospray ionization source, higher-energy collisional dissociation (HCD) ion activation, in positive mode. Identified peptides were searched against the mouse.fasta database. The Sequest algorithm searched for matching peptides with a minimum of 6 and a maximum of 144 residues, allowing for 2 missed cleavages. A carbamidomethyl fixed modification (from IAM capping) was selected for a maximum of 3 modifications per peptide. The Percolator algorithm was used to identify incorrect
spectral identifications using a decoy database and confidence intervals at p ≤ 0.05. Percolator attempts to match the experimentally identified peptide sequence against a reversed protein database to determine the false discovery rate (FDR) of the peptides. The decoy database search significance was set to p ≤ 0.1 for a relaxed target and p ≤ 0.05 for a strict target. All other default settings were used.

Approximate protein abundance scores were calculated using the exponentially modified protein abundance index (emPAI) [13]. The emPAI score (S) is determined by [13]:

\[ S = 10^{\frac{N_{\text{obs}}}{N_{\text{exp}}}} - 1, \]

where \( N_{\text{obs}} \) is the total ion assignments for an observed peptide of a specified protein and \( N_{\text{exp}} \) is the expected number of peptides produced from a protein sequence. \( N_{\text{exp}} \) is determined by multiplying total number of residues in a protein sequence by the approximate abundance of lysine (~7.2%) and arginine (~4.2%) [14]. This emPAI score is used here to approximate protein abundances for statistical and biopathway analysis.

Proteins identified by at least 2 unique peptides were compiled with Access Database software (Microsoft Office 2007, Microsoft Corporation, Redmond, WA). Proteins that were found in at least half of the samples were selected for statistical analysis. One-way ANOVA and hierarchical clustering statistical analysis of all cohorts, except for the CT-Veh cohort (see section 4.2.1), were performed using MetaboAnalyst (v. 3.0, McGill University, www.metaboanalyst.ca) [15] by using the selected proteins and their emPAI scores as inputs. The one-way ANOVA first identified the individual proteins that were significant between the groups. Then, the top 50 proteins identified by the one-way ANOVA were ranked by their calculated f and p values, indicating the
degree of significance of the overall comparison, and then used to create the hierarchical cluster. The cluster was generated with Euclidean distance to measure the similarity parameter and Ward clustering to minimize the sum of the squares between two clusters. The top 50 proteins were additionally used as inputs for biopathway analysis. All proteins used here are identified by their Uniprot Accession number.

**Metabolomics.** Raw data files were analyzed with XCMS Online (Scripps Research Institute, xcmsonline.scripps.edu) using a custom parameter based on the parameters in the UPLC/Orbitrap analysis, but only changing the polarity to negative [16]. Putative assignments were made based on score and approximate compound polarity. A standard mixture of selected polar analytes was analyzed by LC-MS/MS in order to confirm the identity of selected compounds (data not shown). Elution order of organic compounds was based on work by Kroumova [17] and Peters [18]. Peak areas were normalized with a normalization constant based on the protein concentrations as determined by the Bradford analysis (Section 4.2.2). Normalized metabolite data was then analyzed by MetaboAnalyst using hierarchical clustering. MetaboAnalyst did not perform data imputation, filtering, or normalization prior to clustering. The heatmap was generated with Euclidean distance and Ward clustering. The top 25 metabolites identified by ANOVA from the cluster analysis were used as inputs for IPA analysis.

**Biopathway Analysis.** Molecular regulation networks were generated using the Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, www.ingenuity.com) [19]. Data was compared in a pairwise fashion; P301L STZ vs. P301L Veh; WT STZ vs WT Veh; P301L STZ vs. CT STZ; WT STZ vs. CT STZ; P301L STZ vs WT STZ; and P301L Veh vs. WT Veh. The regulation network was obtained from a molecular input list comprised
of the top 50 proteins and top 25 metabolites identified by one-way ANOVA. The core analysis function was used on the input molecules to generate the canonical pathways and regulation networks. Direct and indirect relationships were considered to generate networks comprising of 35 molecules including endogenous chemicals. All data sources and levels of confidence were considered in mouse brain tissues with all mutations. The most statistically significant canonical pathway and one pathway with altered regulation have been selected for discussion.

4.3 Results

Tau mice protein results. All identified proteins and peptides were analyzed at the 95% confidence interval. The LC-MS/MS analysis employed here resulted in the identification of 29218 total peptides. 3952 peptides (1125 unique) were identified in the CT-STZ group, 2151 peptides (835 unique) in the CT-Veh group, 5765 peptides (1279 unique) in the P301L-STZ group, 4624 peptides (1241 unique) in the P301L-Veh group, 6559 peptides (1329 unique) in the WT-STZ group, and 6167 peptides (1401 unique) in the WT-Veh group. The length of the peptides ranged from 6 to 42 (CT-STZ), 6 to 43 (CT-Veh, WT-STZ), 6 to 44 (P301L-STZ), and 6 to 47 (P301L-Veh, WT-Veh). Peptide charge states for the total peptide complement for all samples has been identified as [M+4H]^{4+}, [M+3H]^{3+}, [M+2H]^{2+}, and [M+H]^{+}, and comprise ~5.3%, ~31.1%, ~60.6%, and ~3.0%, respectively, with similar distributions for each sample cohort. The ratio of the total number of high confidence peptides to the total number of strict target decoy database peptide hits yielded an actual false discovery rate (FDR) of 2.5% [20].

18485 total proteins were identified in all the samples which can be sorted into
5258 unique proteins for all samples. 2353 of these were identified by at least 2 unique peptides in all samples. CT-STZ (n=4) samples had a total of 773 proteins and 346 proteins (160 unique) were identified by at least 2 unique peptides. CT-Veh (n=2) samples had a total of 379 proteins identified and 167 proteins (118 unique) identified by at least 2 unique peptides. P301L-STZ (n=5) samples had a total of 1087 proteins identified and 462 proteins (168 unique) were identified by at least 2 unique peptides. P301L-Veh (n=3) samples had a total of 780 proteins identified at 360 proteins (176 unique) were identified by at least 2 unique peptides. WT-STZ (n=5) samples had 1115 proteins identified and 518 proteins (182 unique) were identified with at least 2 unique peptide hits. WT-Veh (n=6) had 1124 proteins identified and 500 proteins (205 unique) were identified with at least 2 unique peptide hits. In order to determine the FDR of proteins, peptide hits were randomly removed equal to the number of false positives (2.5% of 29218) [21]. The number of proteins identified by the number of remaining peptides was divided by the number of total proteins identified and subtracted from 100%. This resulted in an FDR of 4.2%.

A symmetric Edwards-Venn diagram was generated using the InteractiVenn web based tool [22]. Figure 4.1 shows the resulting diagram demonstrating the overlap of proteins in each group. There were 95 proteins that were found in every group. CT-STZ samples contained 11 unique proteins, no unique proteins in the CT-Veh cohort, 8 unique proteins in the P301L-STZ samples, 11 unique proteins in P301L-Veh samples, 16 unique proteins in the WT-STZ samples, and 19 unique proteins in the WT-Veh samples. “Chapter 4 protein supplemental” on the “Venn Diagram” tab contains lists of the proteins that were within each group from the Edwards-Venn diagram generated with the
Figure 4.1 - The symmetric Edwards-Venn diagram (generated from InteractiVenn, H. Heberle, et al. 2015.) depicting the overlap of proteins between each cohort and proteins unique to each cohort. The horizontal orange rectangle represents the CT-STZ cohort which has 11 unique proteins. The vertical red rectangle represents the CT-Veh cohort and contains no unique proteins. The grey circle represents the P301L-STZ cohort and contains 8 unique proteins. The green double lobed shape represents the P301L-Veh cohort which has 11 unique proteins. The purple 4 lobed shape represents the WT-STZ group and contains 16 unique proteins. The light blue 8 lobed shape represents the WT-Veh group with 19 unique proteins. There are 95 proteins that are contained in all 6 groups.
Venn Diagram Generator from the Bioinformatics and Evolutionary Genomics group at Ghent University (http://bioinformatics.psb.ugent.be/webtools/Venn/) [23]. The protein locations and functions were found using the QuickGO browser [24]. The 1164 uniquely identified proteins across all samples represent 452 distinctive locations within the cell of 3491 total locations identified. The top 20 protein locations totaling 1905 hits and their location counts are shown in Figure 4.2a. 779 distinct protein functions were identified from a total of 3220 functions. The top 20 functions totaling 1203 hits and their individual function counts are shown in Figure 4.2b. For a complete list of the locations and functions see the “Chapter 4 protein supplemental” data “Protein locations” and “Protein functions” tabs.

Proteins identified by at least 2 unique peptides and found in at least half the samples were used as inputs for hierarchical clustering analysis using MetaboAnalyst [15]. A one-way ANOVA was used to calculate F and p values for all the input proteins except from those in the CT-Veh cohort due to the small sample size. Proteins were sorted based on their descending f values (increasing p values), and the top 50 were selected as inputs for IPA. The biopathway analysis was performed in a group pairwise fashion, as outlined above (see section 4.2.4). Figure 4.3 shows the dendrogram of the top 50 proteins identified from the one-way ANOVA. The color represents the relative emPAI scores for the proteins in each sample.

**Tau mice metabolomics results.** In the aqueous metabolite fraction, a total of 19178 MS/MS spectral features were identified, corresponding to 2670 possible identifications across all retention times. In the organic fraction, a total of 16403 MS/MS spectral features were identified, corresponding to 1656 possible identifications across all
Figure 4.2 – A. The top 20 protein locations as determined from the QuickGO database. Cytoplasm, membrane, and extracellular exosome were the 3 most likely locations of the identified proteins. B. The top 20 protein functions as determined from the QuickGO database. Protein binding, nucleotide binding, and metal ion binding were the 3 most likely functions of the identified proteins.
Figure 4.3 - Dendrogram and heatmap generated from the hierarchical analysis function in MetaboAnalyst. The top 50 significant proteins as determined by ANOVA are shown here. The dendrogram shows 6 main branches of the cluster delineated by heavy black lines and branch colors.
retention times. Endogenous metabolites were only considered for potential peak matches. Putative assignments were further refined by XCMS score and the polarity of the compound. The organic compounds were additionally identified based on elution order as determined by Kroumova [17] and Peters [18]. 52 aqueous compounds and 18 organic compounds were identified in this manner. 70 total metabolites were used for hierarchical clustering inputs in MetaboAnalyst software and the top 25 metabolites were identified by ANOVA. The cluster of metabolites can be found in Figure 4.4.

**Biopathway results.** IPA results showed 14-3-3 mediated signaling, phagosome maturation, cell cycle G2/M DNA damage checkpoint regulation, and remodeling of epithelial adherens junctions, and glycolysis I as being the top 5 significant pathways (p ≤ 0.05) in all comparisons, except glycolysis I was not in the top 5 in the WT STZ vs CT STZ comparison. There were no z-scores associated with these pathways, and thus it is unknown if these pathways are up or down-regulated. IPA also determined the protein kinase A signaling pathway to be significant (p ≤ 0.05) as well as calculated a z-score for this pathway in all 6 comparisons. A z-score of -0.447 was calculated for both P301L STZ vs. P301L Veh and WT STZ vs. WT Veh comparisons indicating that in the respective STZ induced cohort the protein kinase A signaling pathway was down-regulated compared to the respective Veh cohort. However, this z-score does not indicate a significant down regulation (z is significant if z ≥ |2|). Z-scores of -1.000 were calculated for the P301L STZ vs. CT STZ and WT STZ vs. CT STZ comparisons, also indicating a non-significant down regulation of the protein kinase A pathway in the diseased group compared to the control. P301L Veh vs. WT Veh and P301L STZ vs. WT STZ both had positive z-scores (0.447 and 1.342 respectively) indicating the
Figure 4.4 - Dendrogram and heatmap generated from the hierarchical analysis function in MetaboAnalyst. The top 25 significant metabolites as determined by ANOVA are shown here. The dendrogram shows 2 main branches of the cluster. The upper branch is comprised of non-polar molecules and the bottom branch contains mostly polar molecules.
pathway is up-regulated, but not significantly, in the P301L cohorts compared to the WT cohorts. Table 4.1 lists the canonical pathways, the associated p value for each comparison, and the input molecules relevant to the pathway.

A combined regulation network was created from the top network in all comparisons (Figure 4.5) with the following related diseases and functions: cell death and survival; neuronal disease; cell to cell signaling and interaction; and nervous system development and function. 25 proteins were input into this network; 9 proteins were increased on average in abundance in the comparisons and 16 proteins were decreased on average in abundance in the comparisons. The combined network was analyzed using Cytoscape [25]. There are 67 nodes, no isolated nodes, 42 multi-edge nodes, and 24 self-loops. The network diameter, or the largest distance between two nodes is 10 and the network radius, or the minimum non-zero eccentricities of the nodes is 1. The characteristic path length, or the average shortest path length is 3.238 and the number of nearest neighbors is 7.373. The number of connected components is 1, indicating strong connectivity; 32 of the 67 molecules in the network (47.8%) have 4 or more connections. The network hubs are tumor protein p53 (TP53) and huntingtin protein (HTT) with 38 and 25 connections, respectively. HTT as a network hub in this analysis is not a surprise as Huntingtin’s disease also involves tau dysregulation. The Cytoscape network can be found in the “Chapter 4 Cytoscape network” supplementary material.

4.4 Discussion

Metabolite Dendrogram. Figure 4.4 shows the dendrogram of the metabolites containing 2 main branches. The upper branch consists of non-polar compounds and the
Table 4.1 Significant canonical pathways

<table>
<thead>
<tr>
<th>Pathway</th>
<th>P301L STZ vs P301L Veh</th>
<th>WT STZ vs WT Veh</th>
<th>P301L STZ vs CT STZ</th>
<th>WT STZ vs CT STZ</th>
<th>P301L Veh vs WT Veh</th>
<th>P301L STZ vs WT STZ</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-3-3-Mediated Signaling</td>
<td>1.66E-07</td>
<td>1.66E-07</td>
<td>5.62E-08</td>
<td>5.62E-08</td>
<td>1.66E-07</td>
<td>1.66E-07</td>
<td>TUBB4A,YWHAG,TUBB3,TUBA4A,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>YWHAZ,YWHAB,GFAP</td>
</tr>
<tr>
<td>Phagosome Maturation</td>
<td>1.23E-05</td>
<td>1.23E-05</td>
<td>5.89E-06</td>
<td>5.89E-06</td>
<td>1.23E-05</td>
<td>1.23E-05</td>
<td>TUBB4A,PRDX6,TUBB3,TUBA4A,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATP6V1E1</td>
</tr>
<tr>
<td>Cell Cycle: G2/M DNA Damage Checkpoint</td>
<td>2.63E-04</td>
<td>2.63E-04</td>
<td>1.70E-04</td>
<td>1.70E-04</td>
<td>2.63E-04</td>
<td>2.63E-04</td>
<td>YWHAZ,YWHAB,YWHAB</td>
</tr>
<tr>
<td>Regulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remodeling of Epithelial</td>
<td>7.76E-04</td>
<td>7.76E-04</td>
<td>5.01E-04</td>
<td>5.01E-04</td>
<td>7.76E-04</td>
<td>7.76E-04</td>
<td>TUBB4A,TUBB3,TUBA4A</td>
</tr>
<tr>
<td>Adherens Junctions</td>
<td>1.29E-03</td>
<td>1.29E-03</td>
<td>4.68E-02</td>
<td>4.68E-02</td>
<td>1.29E-03</td>
<td>1.29E-03</td>
<td>GAPDH,GPI</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>4.27E-03</td>
<td>4.27E-03</td>
<td>1.38E-02</td>
<td>1.38E-02</td>
<td>4.27E-03</td>
<td>4.27E-03</td>
<td>YWHAZ,GNB1,PPP3CA,YWHAZ,YWHAB</td>
</tr>
<tr>
<td>Protein Kinase A Signaling - p value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Kinase A Signaling - Z score</td>
<td>-0.447</td>
<td>-0.447</td>
<td>-1.000</td>
<td>-1.000</td>
<td>0.447</td>
<td>1.342</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1 – Above, the top 5 significant pathways as determined by Ingenuity Pathway Analysis, the associated p value for each comparison, and the input molecules relevant to the pathway. Below, the p values and z scores for the only pathway to be identified as being up or down-regulated, although not significantly, as determined by Ingenuity Pathway Analysis.
Figure 4.5 – Ingenuity Pathway Analysis combined network from the individual comparison networks pertaining to the diseases and functions: cell death and survival; neuronal disease; cell to cell signaling and interaction; and nervous system development and function. Solid lines indicate direct relationships and dashed lines indicate indirect relationships. Loops represent feedback systems. The inter-relationships shown are known from the IPA database. The 25 input molecules are highlighted in green (increased in abundance) and red (decreased in abundance). The network hubs are tumor protein p53 (TP53) and huntingtin protein (HTT).
lower branch consists of polar compounds. MetaboAnalyst software is not able to distinguish classes of compounds and creates the dendrograms only based on input intensities. It was serendipitous that the polar and non-polar metabolites grouped together. The non-polar compounds tended to be higher in intensity in the CT-STZ cohort, and the polar compounds tended to be higher in intensity in the P301L-STZ cohort.

**Significant Metabolites.** ANOVA analysis revealed 4 compounds as being significantly different across the groups (p < 0.05, see Chapter 4 Metabolite supplemental); myristic acid, pyroglutamic acid, isobutyric acid, and creatinine. Myristic acid plays a role in cell signaling and membrane integrity. Pyroglutamic acid is found in large amounts in the brain and is thought to be involved with a storage mechanism for glutamic acid [26], but pyroglutamic acid also has negative implications in neurodegenerative diseases such as Alzheimer’s [27]. Isobutyric acid is a short branched chain fatty acid and is more often found in the gastrointestinal tract as a break down product of gut metabolism. Creatinine is the break down product of creatine phosphate from the muscle and shuttled to the blood to be metabolized in the kidneys for elimination. Isobutyric acid and creatinine appear to be suspect identifications; however, there was no alternate compound for creatinine identified and the alternate compound to isobutyric acid was given as butyric acid. Because this work examines brain tissue, creatinine may not be expected to be present in these samples as well as isobutyric acid. Herein lies a major problem with untargeted metabolite identification. Biologically relevant compounds have been identified, but generally are not associated with the tissue type examined in this study. The actual identities of these compounds may be the
tentative assignments here, or not at all.

Although the metabolites were input into IPA with the identified proteins, few pathways contained the metabolites as molecules of importance. Furthermore, in the top 5 identified pathways and the 1 pathway with a z-score no input metabolites were identified as being relevant to the pathway.

**Protein Dendrogram.** The dendrogram in Figure 4.3 has 6 regions related by similar relative protein intensity within various sample groups. The proteins within each region were analyzed with IPA in order to determine the most likely pathway (as determined by p value) that relates these proteins together. It should be noted that the pathway, although related to the overall changes observed in the organism, is only the most significant pathway for the small subset of proteins and not for the 50 input proteins.

The top group (blue brackets) in the dendrogram consisting of 5 proteins showing similar relative protein intensity in the P301L-STZ and P301L-Veh cohorts. The most significant pathway for these proteins is the Glutamate Receptor Signaling pathway (p = 0.02). The second group (yellow-green brackets) comprised of 13 proteins is defined by those proteins that are relatively more intense in the P301L-Veh and WT-STZ cohorts. The nNOS (neuronal nitric oxide synthase) Signaling in Neurons pathway (p = 0.0007) was found to be the most significant in this group. The third group (green brackets) contains 5 proteins that have higher relative intensities in CT-STZ and WT-Veh cohorts. These proteins showed the Glutathione Redox Reactions I pathway (p = 0.007) as being the most significant. For the fourth group (brown brackets) which consists of 12 proteins that are relatively more intense in CT-STZ and WT-STZ cohorts, the Glutamine
Biosynthesis I pathway (p = 0.001) was determined to be the most significant. The fifth group (orange brackets) contains 5 proteins that are relatively more intense in the WT-STZ cohort. The most important pathway way is found to be Glutathione Redox Reactions I pathway (p = 0.005). Finally, the sixth group (purple brackets) contains 10 proteins that are typically are more abundant in all cohorts. Hypusine Biosynthesis pathway (p = 0.0006) was found to be the most significant pathway for this set of proteins.

**Significant proteins.** Previous studies have demonstrated the expression of the vesicular glutamate transporter 1 (vGLUT1, Accession Q3TXX4) is increased and the expression of glutamate transporter 1 (GLT-1, Accession P43006) is decreased in P301L mice relative to WT and control mice [10]. This present analysis did not identify vGLUT-1 or GLT-1 proteins; however, the amino acid transporter, glial high affinity glutamate transporter (SLC1A2, Accession A2APL7), was found mostly in the P301L samples and only in one WT-Veh sample. SLC1A2 was determined to be in the top 50 proteins, but it not found to be significantly different in the one-way ANOVA (p ≥ 0.05). Like the vGLUT and GLT-1 proteins, the SLC1A2 functions to transport glutamate released from the cells across the plasma membranes and clear glutamate from the extracellular space at the synapses. Low glutamate concentrations in the extracellular fluid must be maintained, particularly in the synapse, in order to facilitate synaptic transmission; moreover, high concentrations of extracellular glutamate is neurotoxic [28].

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Accession P16858) was found to be in the top 50 proteins in the one-way ANOVA, but GAPDH was not found to be significantly different between any group (p≤ 0.05). GAPDH has a well-defined role
in the glycolysis pathway; however, many recent studies have shown other functions including regulation of gene expression, acting as a phosphotransferase, and interacting with small molecules such as p53 and glutathione (as reviewed by reference [29]). More relevant to this current body of work, GAPDH has also been shown to interact with proteins associated with neurodegenerative diseases including Huntington's and Alzheimer's (as reviewed by references [29,30]).

Another significant protein found by one-way ANOVA is glial fibrillary acidic protein (GFAP, Accession P03995-2, p= 0.008). The Fisher’s LSD revealed that the GFAP protein in the CT-STZ cohort was significantly different from P301L-STZ, P301L-Veh, and WT-Veh cohorts, but not significantly different from WT-STZ cohort. GFAP in the WT-STZ cohort was significantly different from the P301L-Veh and WT-Veh groups, but not significantly different from the P301L-STZ cohort. GFAP has been shown to anchor the glutamate/aspartate transporter (GLAST) and excitatory amino acid transporter 1 (EAAC1) in the membrane [31]. GFAP was once thought to provide support for astrocyte cells; however, an emerging body of work suggests GFAP plays a role in regulation the blood brain barrier and signal transduction pathways (as reviewed by reference [31]). Increased GFAP expression is known to occur with brain damage and neuronal degeneration [31]. Dabir et al. demonstrated a regiospecific expression of GFAP in P301L and WT mice with the protein being more highly expressed in the spinal cord, followed by the brainstem, and the lowest expression in the cortex [32]. Furthermore, the expression of GFAP was found to increase in both the P301L and WT mice compared to the controls. This is in contrast to the significant decreases in expression of GLT-1 and GLAST proteins in the spinal cord and brainstem [32]. GLAST
and EAAC1 proteins were not identified in this present study, likely due to the data dependent mass spectral analysis.

**Significant pathways.** The Protein Kinase A Signaling Pathway utilizes neurotransmitters and hormones to activate processes within the body related to growth, memory, and metabolism [19]. Within this pathway protein kinase A (PKA) phosphorylates glycogen synthase kinase-3 (GSK3) in order to inactivate GSK3 and prevent phosphorylation of the tau protein and consequent neurodegeneration [19]. In this present study, there was an observed down-regulation of this pathway in the P301L STZ and WT STZ cohorts compared to CT STZ cohorts, as well as in the P301L STZ and WT STZ cohorts with respect to the Veh control cohorts. The Protein Kinase A Signaling pathway was found to be up-regulated, but not significantly, as a result of the STZ treatment in the P301L and WT cohorts. By inference it appears that the Protein Kinase A Signaling Pathway is most down-regulated in the WT cohort, but also still down-regulated in the P301L cohort compared to the controls.

14-3-3 Mediated Signaling was found to be the most significant pathway in all comparisons. This pathway that utilizes 14-3-3 adapter and scaffold proteins to bind to specific phosphoserine and phosphothreonine motifs in order to inhibit or activate proteins, provide structural support, or to translocate other proteins [19]. Relevant to this study, Li and coworkers examined the binding affinity of 14-3-3γ isomer (Accession A8IP69) to GFAP as they are known to play protective roles in neurodegenerative diseases [33]. GFAP was shown to associate with 14-3-3γ under phosphorylated and hyperphosphorylated conditions, but no association was found under dephosphorylated conditions [33]. Since 14-3-3γ was found in all of the samples and GFAP was found in
70 percent of the samples, it is not unexpected that this was the most significant pathway reported by biopathway analysis.

### 4.5 Conclusions

Comprehensive proteomics and metabolomic analyses were performed on genetic control, P301L, and WT mice brains treated with ICV-STZ or a vehicle buffer control in order to determine how the brain proteome and biological pathways are altered as a result of the genetic background and STZ treatment. A one-way ANOVA of control, P301L, and WT mice brains treated with ICV-STZ or a vehicle buffer control was used to select the top 50 proteins and 25 metabolites by calculated F and p values. IPA analysis of these proteins and metabolites revealed 49 of significant pathways across all comparisons (p ≤ 0.05) and 1 of these pathways as being up or down-regulated (Protein Kinase A signaling). There was an observed a down-regulation of this pathway, although not significant, in 4 of the 6 comparisons. The Protein Kinase A signaling pathway regulates various functions within the body including those related to memory as known inhibition of this pathway contributes to cognitive decline. The 14-3-3 Mediated Signaling pathway was found to be the most significant pathway based on the input proteins and metabolites; however, it was not found to be up or down-regulated. This pathway has been found to employ adapter and scaffold proteins in order to inhibit or activate proteins, provide structural support, or to translocate other proteins. It also has been found to be key in GFAP regulation.

There is evidence to support that ICV-STZ has an effect on protein and metabolite levels and signaling cascades related to AD and FTD within the brains of P301L and WT
tau mouse models used. In order to fully understand the effect of ICV-STZ treatment on the human tau mouse models, additional studies are required to assess the effect of tau pathology progression and cognitive function.
4.6 References


Chapter 5

LC-MS/MS metabolomic analyses shortcomings
5.1 – LC-MS/MS metabolomic analyses technique and hardware shortcomings

Metabolite compound identification can be performed with a variety of platforms including: gas chromatography (GC); capillary electrophoresis (CE); nuclear magnetic resonance (NMR); and liquid chromatography (LC). Each technique has its advantages and disadvantages. GC works well for volatile compounds, but often compounds must be derivatized prior to sample analysis to improve volatility or to add a functional group so a compound may be detected with a particular detector [1]. GC - mass spectrometry (MS) was once thought to be the gold standard for metabolite identification; however, it has recently come into question for thermal degradation of compounds, leading to misidentification of compounds [2].

CE provides the advantage of increased peak capacity and sensitivity over LC; unfortunately, the low flow rates used in CE are challenging to couple to MS [3]. In more recent years, improved interfaces have emerged to couple CE to mass spectrometers such as the porous tip sheathless interface, the flow-through microvial interface, the electro-osmotic flow driven sheath-liquid interface, and the porous emitter sheathless interface [4]. As a result, CE has become a more popular technique for metabolite identification particularly in the clinical metabolomics realm [4–8].

NMR spectroscopy has the advantage of allowing researchers to minimally process samples prior to analysis [9]. NMR is also highly reproducible, non-destructive, quantitative, is useful for untargeted analysis [9]. Recent studies have utilized NMR successfully to study metabolites in urine [10–14], serum [11,15–18], cancerous cells [19–22], and cerebrospinal fluid [23–26]. However, the major drawback to using NMR for metabolite analysis is that it is not as sensitive as GC-MS or LC-MS and many low
abundance species are not identified [27].

Metabolite compound identification with LC separation is not without its complications either. Identification in complex mixtures often relies on retention time matching, UV or fluorescence profile matching, and/or mass fragment matching to known standards [28]. Matrix effects can often complicate compound identification, for example, by changing the retention time or forming adducts with substances within the matrix [28]. Untargeted metabolomic analyses are challenging due to not knowing what metabolites may be of interest in a sample [29,30]. In this regard, using a systems biology approach and examining the genes and proteins in addition to the metabolites, researchers can anticipate what metabolites may be of interest.

Library matches for ion fragments from LC-MS/MS analyses are difficult because ion fragmentation energies differ from instrument to instrument due to dependence on instrument geometry [31–33]. The resulting ion fragmentation pattern of the same compound at the same normalized collision energy will often differ from a quadrupole to an ion trap instrument, for example. Additionally, matrix effects from the sample, extraction medium, or mobile phases and even the type of collision gas used can influence how a metabolite will fragment after LC separation [31–33]. There are databases that contain fragmentation patterns of various metabolites at different collision activation voltages, in different collision gasses, and on different instruments [34]. Unfortunately, these databases currently contain a small amount of compounds and rely on users to supplement the database with their own compounds. Further complicating matters is that ion fragmentation patterns and the intensities of the fragment peaks may not be reproducible from one matrix to another [35].
LC separations also have difficulty in separating isomeric and isobaric interferences [29,30]. Multistage tandem mass spectrometry (MS^n) is helpful for discerning isomeric and isobaric species, but is not amenable for all compounds [34]. Ion mobility spectrometry (IMS) can be useful in separating isomeric and isobaric species [36–41]. Ion mobility is a gas-phase separation technique which allows ions to separate by their collisional cross section (CCS) and charge state. Like LC-MS/MS libraries, IMS libraries are in their infancy and CCS values contained within are dependent on the type of buffer gas used among other factors [42–44]. Furthermore, inconsistencies are noted with calculated CCS depending on the type of mobility separation employed. For example, drift tube (DTIMS) CCS values may differ from those calculated with traveling wave (TWIMS) devices [45]. TWIMS devices must be calibrated and if the calibrant used has substantially different properties than the analyte of interest, discrepancies may arise in the calculated CCS compared to the actual CCS [45]. A new IMS-MS approach is introduced in this work employing mobility separations to improve peptide and small molecule identification (Chapter 6).

5.2 – LC-MS/MS metabolomic analyses software shortcomings

Multiple software programs were utilized in our metabolomics work, with the goal that subsequent versions would be able to more confidently identify the metabolites within these samples. One of the first software programs used for this work was Sieve (v. 2.2.58 SP2, Thermo Fisher, San Jose, CA), followed by Mass Frontier (v. 7.0.5.9 SR3, Thermo Fisher, San Jose, CA), and Compound Discoverer (v. 1.0.0.692, Thermo Fisher, San Jose, CA). Also used in this work was the free software XCMS Online (v. 3.05
Scripps Research Institute, La Jolla, CA), and a trial version of Elements for Metabolomics (v. 1.2.1 Proteome Software, Portland, OR).

Sieve software has the advantage of being able to search the ChemSpider library and the KEGG pathways. Raw data files from Thermo instruments can be directly uploaded into Sieve for analysis and batch processed. Sieve also allows the user to perform spectral alignments and statistical analysis including PCA. One drawback is that compounds are matched based on exact mass matching rather than fragment ion matches. Users must pour through data and decide if the compound assignment is relevant to the analysis or not. For example, for the metabolomics analysis presented in Chapter 2 (TiO$_2$ rat plasma), the data from Sieve had to be sequentially examined in order to remove any compounds identified as pharmaceuticals or phytochemicals because of their propensity to be false positive assignments. After evaluating the data for false positives and irrelevant compounds, approximate polarities and elution orders from literature sources were utilized to further refine the results. This proved to be a long and arduous task.

Mass Frontier software has the capability to identify compounds by MS/MS fragmentation. Currently, the library associated with Mass Frontier contains over 2000 compounds with both positive and negative ion mode libraries. Mass Frontier also has the capability of allowing one to search identified compounds by class, spectral tree similarities, molecular formula, $m/z$, and retention time. Despite the advantages of this software over Sieve, it is only useful for analysis of pure compounds [46].

Compound Discoverer 1.0 has a similar layout to Proteome Discoverer where users can create custom workflows to meet specific data analysis needs. Users can directly import raw Thermo data files into the program for analysis. The BioNano
Research Facility (BNRF) at West Virginia University purchased this as an upgrade for Mass Frontier with the promise to fill the gap for untargeted analysis. Unfortunately, the database from Mass Frontier did not transfer over to Compound Discoverer 1.0 and instead requires the user to import or create user database. Databases from NIST and the Human Metabolome database were downloaded, but despite being the correct file type (.xml), the data within the files was not formatted correctly for integration into Compound Discoverer 1.0. Compound Discoverer 2.0, in addition to allowing users to create databases, has the added capability of searching mzCloud, KEGG, or Chemspider databases. This program can evaluate isotopic distributions and high mass resolution of compounds in order to determine elemental composition of compounds. Users can also perform descriptive statistics, PCA, and differential analysis on the results. The BNRF is currently evaluating an upgrade to Compound Discoverer 2.0.

XCMS Online is free online metabolite identification software and was used to identify metabolites found in the tau mice (Chapter 4). However, prior to sample analysis, data must be converted from the raw data format to mzML format, which is performed by a separate, but free program called MSConvert. Compounds are identified from the METLIN database, also maintained by the Scripps Institute, which has fragmentation spectra generated at 10, 20, and 40 eV. Although it is useful to have spectra generated at multiple fragmentation energies, the spectra may not always be comparable across platforms (see section 5.1). Compounds identified by XCMS Online are given a score based on how well the experimental result matches the library result. Reports are generated for both MS spectral matches and for MS² spectral matches. The output also includes PCA plots of the compounds for the samples. A cloud plot is also
provided with the output, such as that shown in Figure 5.1, which visualizes the chromatographic location of the ions, their masses, and their intensities.

Elements for Metabolomics is relatively new software, and one of the more attractive features of this program is that it allows for direct import of data files without file conversion from various MS vendors (e.g. Thermo and Waters). Features within the chromatograms are identified and matched against NIST, HMDB, and user-created libraries. Identified compounds are scored based on how well the experimental MS and MS/MS data matches the library data. One advantage is that there are fewer identified compounds that are unlikely matches. However, users are still required to examine the data to ensure that compound matches are coherent. For example, pharmaceutical compounds should not be included in the results if they were unlikely to be present in the model organism. This software program was used to identify metabolites in the plasma, vasculature, and aorta from TiO$_2$ exposed rats (Chapter 3).

Despite these various software programs, the confidence in metabolite identifications was still not sufficient to publish the metabolite data with the proteomics data (Chapters 2, 3, and 4). Furthermore, when the metabolite data was included with the proteins and input into Ingenuity Pathway Analysis (IPA), the resulting pathways only differed slightly in the p value compared to results using the proteins alone. That is, the input metabolites improved IPA results, but only marginally. Therefore, only the proteins were used as inputs for published results rather than the combination of proteins and metabolites.

5.3 Improvements on the horizon
Figure 5.1- Cloud plot of organic metabolites from tau mice. Dots show at what retention time and m/z ions were identified. The size of the dot indicates the intensity of the peak with larger sizes indicating higher intensities.
There has been a push within the metabolomics community to create better and more freely available databases. One such project is MassBank of North America (MoNA) created by the Fiehn laboratory at the University of California at Davis [47]. Users are able to upload their own characterized compounds into the database and can rate the quality and accuracy of the compound identification. The database contains experimental and in-silico spectra. Data are curated with metadata describing the origin of the data, which instrumentation was used to generate spectra, the ionization mode utilized, the compound exact mass, the precursor mass used for fragmentation, the type of precursor ion (e.g. \([\text{M+H}]^{+}\) ), and the level of fragmentation shown (e.g. MS\(^2\) or MS\(^3\) ). Metadata also includes an ion peak table containing the fragments and their intensities. Currently, the database contains over 205,000 spectra, separated into different groups according to the type of spectra (e.g. high resolution or low resolution), phytochemicals, human metabolites, and fatty acids. Although this database would appear to be very promising to analyze the data described in Chapters 2, 3 and 3, there was no tool available on the MoNA website to batch identify unknown compounds [47]. The database was also downloaded, but could not be incorporated into Compound Discoverer 1.0. Although the downloaded database was the correct type of file (.xml), the data was not formatted correctly within the file to be read by Compound Discoverer.

mzCloud is another freely accessible database for metabolite identification [48]. Like MoNA, mzCloud contains many spectra of compounds at different collision energies and MS\(^n\) levels. Data is curated using information associated with the compound name and synonyms, structure, annotated peaks, adduct information, and even the laboratory of origin. The mzCloud database is able to be searched either as a
standalone web-based database or it can be searched within Compound Discoverer 2.0 [48].

With improvements in software packages and databases, users will be able to more easily identify unknown and untargeted metabolites within their datasets. Projects like the MONA database encourage collaboration and free sharing among scientists in the pursuit of advancing science without cost prohibition. However, improvements must be made for the ease of incorporating these databases into existing software programs.

Identifying compounds by precursor ion fragmentation alone is not sufficient for high-confidence assignments. Additional parameters, such as retention time matching, are required for compound identification. One parameter that is gaining traction for compound identification is ion mobility in addition to fragmentation data [49]. As mentioned in section 5.1, IMS is able to separate isomers and isobars [36–41] for which co-elution is a hindrance to compound identification by LC-MS alone. IMS CCS compound libraries have already been established and are expanding [42–44]. Furthermore, additional techniques can be coupled with IMS-MS to garner even more information about an ion. For example, gas-phase hydrogen/deuterium exchange (HDX) coupled with IMS-MS has demonstrated different rates of exchange of labile hydrogens on ions of different compounds [50–55]. The extent of hydrogen/deuterium (HD) scrambling through ion activation has also been shown to differ according to the amino acid sequence of a peptide ion [56–59] and the type of charge carrier [56,60,61]. Because the amount of deuterium uptake and scrambling depends on the analyte characteristics, these are ideal analytical parameters to utilize for further characterization of compounds. Chapter 6 presents a proof-of-principle analysis in which HDX, HD scrambling with
HDX, and ion fragmentation are considered as compound identification parameters.
5.4 References


Chapter 6

Ion Mobility, Hydrogen/Deuterium Exchange, and Isotope Scrambling: Tools for Metabolite Identification


6.1 Introduction

Liquid chromatography tandem mass spectrometry (LC-MS/MS) is a technique often used for untargeted metabolite analysis [1,2]. It allows for concurrent separations of many different classes of compounds whereas techniques, such as GC-MS, are better suited for volatile compounds and often rely on a derivatization step prior to analysis [1]. LC-MS/MS is also more sensitive compared to other approaches such as NMR which may not be able to detect low abundance species in complex mixtures [1,2]. However, one of the challenges of LC-MS/MS is its ability to distinguish between isobars and isomers that are not well separated by LC [2,3]. Ion fragmentation is one technique used to identify differences in isomeric and isobaric species [4], but in the case of co-eluting compounds, it may be difficult to discern fragments from one species from the other. Additionally, fragmentation of isomeric species often produces ions of the same mass requiring multistage tandem mass spectrometry (MS^n) for identification [5–7]. Another approach gaining in usage for ion identification is the determination of molecular formula from accurate mass matches [8–11]. That said, accurate mass matching often requires separations steps to remove interfering isomeric and isobaric species [11,12]. It is also noted that this approach requires costly, high-end mass spectrometers.

Ion mobility spectrometry (IMS) has proven to be useful in separating isomeric and isobaric species [13–18]. IMS is a gas-phase separation strategy which is achieved through differences in ions’ collisional cross sections (CCS) and charge. Overall, ions that are more compact and more highly charged traverse the mobility region more rapidly than ions that are more elongated and have lower charge states [19,20]. Mobility separation of compounds occurs on the millisecond time scale, whereas LC separations
occur on the minutes to hours timescale. Thus, it is possible to separate ions in complex mixtures on much shorter timescales. That said, because of the correlations between ion CCS and mass, the combined IMS-MS peak capacity can be less than that of LC-MS [21].

Gas phase hydrogen/deuterium exchange (HDX) can be coupled to IMS to further differentiate ions based on deuterium incorporation (uptake). Gas-phase HDX is often used as a tool to examine protein ion structure and conformational changes [22–29], but it has also been used to study small molecules [30–34]. The exchange of hydrogens for deuterium atoms with D₂O reagent gas at labile sites (charge sites and other heteroatoms) occurs via a relay mechanism [35]. It has previously been shown that the overall rate of deuterium uptake varies in proteins, peptides, and small molecules [35–40], and in the case of peptides, the uptake is dependent upon the amino acid sequence [41]. The resulting mass spectra show differences in the isotopic distribution that could be advantageous to use in pattern recognition for small molecule identification. Such uptake patterns are dependent on the partial pressure of D₂O gas and the overall HDX reaction time [40]. Because at set partial pressures of D₂O and electric fields ions experience the same number of collisions with the reagent gas from sample to sample, the uptake patterns of individual molecules are expected to be reproducible. They therefore provide a means of compound identification.

Hydrogen/deuterium (HD) scrambling, or rearrangement of labile hydrogen and deuterium atoms occurs with the addition of energy which statistically reapportions the incorporated deuteriums. This can be troublesome when examining protein structures [41–43]. The phenomenon is described by the “mobile” proton theory in which
collisional activation results in significant proton mobilization [44,45]. This process can be minimized by other fragmentation techniques such as electron capture dissociation (ECD) [46,47], electron transfer dissociation (ETD) [48–50], and matrix assisted laser desorption/ionization in-source decay (MALDI ISD) [51,52]. Several studies have determined that HD scrambling on peptide ions should be dependent upon the amino acid sequence [41,44,53,54] and the type of charge carrier (H⁺ compared to Na⁺) [41,55,56] and is independent of the type of mass spectrometer used (Q-TOF vs. ion trap) [41]. Because HD scrambling has been shown to be dependent on the analyte itself, this characteristic may be exploited as an additional parameter to further differentiate metabolites without the use of LC separation. In the study described here, HD scrambling is used to repopulate some exchange sites with hydrogens which can then undergo gas-phase HDX resulting in an overall increase in the amount of incorporated deuteriums.

The reproducibility of IMS, HDX, and HD scrambling with HDX measurements is particularly important when using such parameters for compound identification. High-reproducibility provides the opportunity to “dial in” these parameters across many samples using select measurement conditions. For set separation and activation electric fields in the drift tube, the overall reproducibility of these measurements requires that the partial pressures of the buffer gases remain relatively constant within an experimental run and from run to run.

Herein, a strategy that utilizes IMS, HDX, and HD scrambling with HDX is proposed in order to rapidly distinguish compounds in a complex mixture. The approach is evaluated for its ability to provide unique parameters for identification as well as for its
overall reproducibility. The former evaluation addresses issues with regard to limited peak capacity while the latter indicates feasibility with regard to automated pattern matching. Although the demonstration reported here has been performed using relatively rudimentary control of reagent gas partial pressure and model systems, the ion distinguishing and reproducibility characteristics of the approach suggest the possibility of incorporation into routine ‘omics workflows in the future.

6.2 Experimental

Materials. LC-MS grade water, acetonitrile, formic acid, and acetic acid were purchased from Sigma-Aldrich (St. Louis, MO). D$_2$O (99.9% D) was purchased from Sigma-Aldrich (St. Louis, MO). Ammonium bicarbonate (≥ 99.5%), urea (ACS reagent grade), bovine serum albumin (BSA, 98.0%), dithiothreitol (DTT, ≥ 99.5%), iodoacetamide (IAM, ≥ 99%), cysteine (≥ 98.5%), trypsin (from bovine pancreas, TPCK treated) were also purchased from Sigma-Aldrich (St. Louis, MO). Bradykinin (≥95.0%), substance P (≥95.0%), and synthetic peptide KKDDDDDIKKII (KKD peptide, 94.1%) were all purchased from GenScript (Piscataway, NJ). These materials were used without subsequent purification.

Sample Preparation for reproducibility studies. BSA was solubilized in 200 mM ammonium bicarbonate with 6 M urea in water. BSA proteins were reduced with DTT, alkylated with IAM, and quenched with cysteine. The proteins were then digested with trypsin and desalted with Peirce C18 spin columns (Thermo Fisher, San Jose, CA). This sample preparation has been described in detail previously [57]. Digested samples were dried, reconstituted in 1:1 water: acetonitrile with 0.1% (v/v) formic acid, and then...
pooled for sample analysis. Pooled samples were loaded into a gas tight syringe, the syringe was placed into a syringe pump and sample was infused at 300 nL/min through the electrospray needle.

A mixture of model peptides containing 45 µg/mL bradykinin, 45 µg/mL substance P, and 450 µg/mL KKD peptide was generated in 1:1 water: acetonitrile with 0.1% (v/v) acetic acid. Samples were loaded into a gas tight syringe for direct infusion into the ion source.

**Data collection and analysis.** Data were collected on a home-built, dual-gating IMS device coupled to a Thermo LTQ Velos mass spectrometer (Thermo Fisher, San Jose, CA) which has been described in detail elsewhere [58]. The front gate was open for a duration of 150 µs and the back gate was opened at a delay with respect to the first gate for 200 µs. Ions were allowed to traverse the drift tube and then were collected in the ion trap for 100 ms prior to mass analysis for each delay time setting. Data were collected for 30 seconds at each drift time resulting in an averaged mass spectrum.

For sequential and random drift time data collection, a total of 3 sample sets were collected. For sequential drift time data collection, the delay times sequentially stepped through (5.0 ms, 5.2 ms, 5.4 ms, etc.) the drift time range. For random drift time data collection, the delay times were randomized (7.8 ms, 6.2 ms, 10.4 ms, etc.). The order of the type of acquisition was randomly assigned. The run order was: random, sequential, sequential, random, sequential, random. The helium buffer gas pressure within the drift tube was maintained at ~2.70 torr for reproducibility experiments. This pressure was maintained in the drift tube for the duration of the data acquisition for the 6 trails.

For the HDX and scrambling experiments, the partial pressures of He and D₂O
were maintained at ~2.51 torr and ~0.02 torr, respectively. The reservoir containing the D$_2$O was maintained at a temperature of 35°C and the tubing from the reservoir to the drift tube was heated to prevent D$_2$O condensation which could lead to partial pressure fluctuations. The HDX reproducibility sample sets using BSA digests were obtained by randomly stepping through the delay times with 3 trials for each sample type. HDX reproducibility studies using isotopic distributions after deuterium uptake were performed on the model peptide mix without drift selection. HDX scrambling with HDX reproducibility studies also were performed on the model peptide mixture without drift time selection. A total of 6 replicates were performed for each HDX and HD scrambling with HDX study using this sample. Between each trial, the drift tube voltage was removed and the drift tube gases were evacuated. Care was taken to return the gas partial pressures to the same approximate values by using the same leak valve settings (Granville Phillips, MKS Instruments, Andover, MA).

Data from each IMS-MS dataset were compiled into a 3-column file (drift time, m/z, and intensity). Peak intensity was normalized and then converted to log$_2$ scale. Two-dimensional (2D) contour plots of m/z verses drift time of the normalized, scaled data were constructed using DPlot (v. 2.3.5.3, HydeSoft Computing, LLC; Vicksburg, MS, USA). An in house peak picking algorithm was used to select the most intense features by examining the change in intensity in separate drift time and m/z regions (~1 ms by ~4 m/z units). The peaks with normalized intensities greater than 1 were extracted and then compared across the data collections. Data features corresponding to BSA digest peptide ions were compared across replicate datasets to determine the reproducibility in drift time. Coefficients of variation (CVs) were calculated for multiple
BSA digest peptide ions.

Isotopic envelope reproducibly from a drift time distribution with HDX of BSA digest peptides in a drift distribution with HDX was performed by a using an algorithm to extract drift times and m/z around each peptide of interest. The dimensions of the extraction box varied based on conformers proximity, the total number of isotopes, and the charge state of the peptide ion. A drift time range of not more than 1 ms wide was used. Ion intensities within the region were summed for each m/z bin (0.08333 units wide) in the extracted region and then normalized by the total intensity of the region. The root mean square deviation (RMSD) for each experimental run was compared to the group mean.

Isotopic distribution reproducibility for HDX and HD scrambling with HDX measurements for the model peptide mixture (bradykinin, substance P, and KKD peptide) was examined without drift selection. HD scrambling was induced by collisionally activating ions in the last portion of the drift tube (~0.9 m from the ion source). Continued HDX was then allowed to proceed in the remaining portion of the drift tube (~0.1 m). Six replicates of both the HDX and HD scrambling with HDX analyses were collected. Between each trial, the mass spectrometer was set to stand by, the applied voltage was removed, and the buffer gases were evacuated from the drift tube. Subsequently the gases were reintroduced manually by dialing the leak valves to their previous setting values. Thus, each replicate is essentially a new, "from scratch" data collection. Isotopic distribution reproducibility was determined as described above for the BSA digest. A separate demonstration examined the fragment ions produced by peptide ions undergoing HD scrambling with HDX; the fragment ions were produced by
collision-induced dissociation (CID) in the linear ion trap mass spectrometer. Isotopic distribution comparisons were performed (RMSD values) for some of the resulting fragment ions.

6.3 Results and Discussion

Deuterium uptake for distinguishing molecules. As mentioned above, the ability to utilize a distinguishing measurement relies on its probing of a separate physicochemical property of the ion. With regard to gas-phase HDX using D₂O reagent gas, the property is labile site accessibility which has been interpreted as accessibility to charge site and the ion’s surface (collisions with D₂O) \([35,40,50,59–61]\). With different arrangements of charge sites and heteroatoms, a large number of polar compounds should exhibit different HDX characteristics. To demonstrate the nature of gas-phase HDX as a compound-specific measurement, a model peptide mixture is here used to provide examples of different HDX behavior. Notably, the use of HDX to increase the overall peak capacity of the measurement in this manner is similar to seminal studies that employed parallel dissociation \([62,63]\) and drift shift reagents \([64–66]\) to disrupt the correlation between IMS and MS measurements.

Unique deuterium uptake patterns of the different peptide ions were observed for the model peptide mixture (Figure 6.1). For example, for the \([\text{M}+2\text{H}]^{2+}\) ions of substance P, little to no deuterium (<1) uptake is observed as shown in Figure 6.1A. In contrast, for \([\text{M}+2\text{H}]^{2+}\) bradykinin ions from the same experiment, Figure 6.1B shows that, on average, the \([\text{M}+2\text{H}]^{2+}\) bradykinin ions incorporate ~3 deuteriums. For the \([\text{M}+3\text{H}]^{3+}\) KKD peptide ions, the average deuterium uptake is ~12 deuteriums (Figure 6.1C).
Figure 6.1 – Model peptides showing different deuterium uptake patterns with D$_2$O added to the drift tube. A) [M+2H]$^{2+}$ ions of substance P. B) [M+2H]$^{2+}$ ions of bradykinin. C) [M+3H]$^{3+}$ ions of the synthetic peptide KKDDDDDDIIKIK. Buffer gas composition is labeled for conditions in which He (blue) and He + D$_2$O (orange) is used.
For these experiments, the partial pressure of D$_2$O is sufficient to exchange the rapidly exchanging sites (readily accessible) [36,67–70]. To examine the origin of the differences in deuterium incorporation (Figure 6.1), it is instructive to consider the numbers of exchangeable hydrogens on these peptide ions. There are 5 residues on substance P (RPKPQQFFGLM) containing side-chain exchange sites namely: 1 arginine; 1 lysine; 2 glutamines; and, 1 methionine. Therefore, including these sites and those at the backbone amide, amino- and carboxy-termini, and the two extra protons, the total number of exchangeable hydrogens is 24 for [M+2H]$^{2+}$ substance P ions. In comparison, there are 19 and 31 exchangeable hydrogens for the [M+2H]$^{2+}$ bradykinin and [M+3H]$^{3+}$ KKD peptide ions, respectively. In comparing the two doubly-charged ions, the percent deuterium incorporation for substance P and bradykinin is 4.2% and 15.8%, respectively. This translates into nearly a four-fold efficiency increase for the bradykinin ions. Therefore, although the ions have the same number of charges and nearly the same number of residues and exchangeable hydrogens, they exhibit significantly different deuterium incorporation capabilities. Presumably this is a reflection of differences in exchange site accessibility to charge sites and surface collisions (i.e., the ion structures) [71]. Overall, the triply-charged KKD peptide ions exhibit the greatest efficiency (38.7%) in gas-phase HDX (nearly double that of the bradykinin ion).

**Reproducibility of drift time in HDX experiments.** In order to take advantage of spectral matching for peptide ion assignment, it is requisite that drift time distributions resulting from experiments using the buffer gas mixture are highly reproducible. Indeed, such spectral reproducibility is a strength of metabolite assignment using GC-MS data where the fragmentation spectra from electron ionization (EI) are extremely consistent.
Admittedly, the instrument used in these experiments (described in Donohoe, et al.[58]) uses a longer gate time at the back of the drift tube compared to the front (200 µs vs. 150 µs). This is required because of the relatively lengthy second gate (~3 mm). That is, a longer time is required to allow the ions to fully traverse the distance of the second gate and any extra region caused by impinging fields during the off-transmit setting. Thus, in order to determine if gating differences effect the measurement reproducibility, data were collected by sequentially (5.0 ms, 5.2 ms, 5.4 ms, etc.) or randomly (7.8 ms, 6.2 ms, 10.4 ms, etc.) stepping through the drift times using a buffer gas of He only. Three replicates each of the sequential and randomized drift time acquisition were performed in a random order for a tryptic digest of BSA. The average CV for peptide ion peaks is determined to be 1.7 ± 0.5% (data not shown). This analysis serves as the benchmark against which the drift time reproducibility of peptide ions in the buffer gas mixture (He + D₂O) is compared.

There is a notable shift in the drift times for features that appears to slightly increase with time and is fairly consistent across the whole drift time distribution. This shift correlates with a slight increase in drift pressure over time which likely results from the relatively simplistic gas introduction system (see section 5.2 above). Although this can be corrected when calculating a collision cross section, it could present a challenge in rapid drift time database comparisons. That said, the reproducibility is similar to that reported in other IMS instruments [72–74]. It is noted that the high reproducibility observed for the random selection of delay times demonstrates analytical utility in that selection times may be used for targeted analyses.

Variability in buffer gas pressure is a greater concern with the buffer gas mixture
considering that only a small percentage of the composition is D$_2$O. Therefore, a small change in partial pressure could result in large changes in drift times. This could be a significant source of error especially for a system that utilizes an evaporative introduction system employing a leak valve such as is used here. The purpose of the following experiments was to determine how much the drift time reproducibility deviates from the benchmark values upon using the buffer gas mixture. Figure 6.2 provides an indication of the run-to-run variability in drift time for BSA digest peptide ions that undergo HDX in the drift tube. Dataset features for [M+2H]$^{2+}$ ions of the peptides RHPEYAVSVLLR and LGEYGFQNALIVRYTR show some difference in drift time (Figures 6.2A and 6.2B). However, the CV of these peaks across the 3 replicates is 1.2% and 1.3%, respectively (Figure 6.2C). Separate dataset features examined in the drift time distributions of the BSA sample were found to have average CVs of 1.4 ± 0.1% (Figure 6.2C). The reproducibility is not significantly different (CV$_{He} = 1.7$ ± 0.5%, CV$_{He + D2O} = 1.4$ ± 0.1%, p=0.68, α=0.05) when separately dialing in the buffer gas mixture components with the leak valves. This is still within the acceptable reproducibility range of IMS experiments.

**Reproducibility of deuterium uptake in HDX experiments.** As with the drift times, comparisons of HDX reactivity reproducibility can be achieved with spectral comparisons. Here, due to the presence of multiple isotopologue ions, a RMSD comparison has been employed. To best illustrate the reproducibility, results for [M+2H]$^{2+}$ bradykinin ions from the peptide mixture are presented in Figure 6.3. In these experiments, data were collected and the MS instrument was set to standby, the drift tube voltage was removed, and the buffer and reagent gases were evacuated between each run.
Figure 6.2 – IMS-MS profile of two BSA digest trials (A) and (B) demonstrating the run to run reproducibility of compounds in a complex mixture after addition of D₂O to the drift tube. Panel C shows the CV for select peptide ions. m/z and average drift time values are provided for the peptide ions. The charge state of each compound is also listed. CV values for Features 9 and 10 in (A) and (B) are shown in (C). Other features tentatively identified as 1- VTK; 2- FPK; 3- LGEGFQNALIVRYTR (3+); 4- YLYEIARR; 5- ECCDKPLEK; 6- DTHKSEIAHR; 7- LFTFADICTLPDEK; 8- SLGKVTRCCTKPESE; 9- RHPEYAVSVLLR; and 10- LGEGFQNIALVRYTR (2+).
Figure 6.3 — Isotopic distribution of \([M+2H]^2+\) bradykinin ions from 6 replicates of deuterium uptake. The RMSDs of the trials were less than 1%. See text for details.
by turning down (and then off) the leak valves. This was done in order to determine if high-reproducibility could be achieved from a “new” start of the instrument. Figure 6.3 shows the isotopic distribution of the \([M+2H]^{2+}\) bradykinin ions across the 6 trials. The RMSDs of the trials were less than 0.2%. The RMSDs of the other peptides within the model peptide mix were also determined for the \([M+2H]^{2+}\) KKD peptide and substance P ions; \([M+3H]^{3+}\) bradykinin, KKD peptide, and substance P peptide ions; and the \([M+4H]^{4+}\) KKD peptide ions. The RMSDs were all found to be less than 1% (data not shown).

A question arises as to whether or not comparisons of such isotopic distributions could be of utility for complex mixture analysis. Here, it is noted that the highly-reproducible drift time separation (Figure 6.2) is useful. That is, in mobility separation experiments, it is possible to select for dataset features and obtain the high reproducibility of the isotopic distributions demonstrated for the bradykinin ions. For example, mobility selection of the features associated with \([M+2H]^{2+}\) ions of the peptides RHPEYAVSVLLR and LGEYGFQNALIVRYTR (Figures 6.2A and 6.2B) have resulted in isotopic distribution RMSDs that are less than 1% (data not shown) in the triplicate analyses described above. Therefore, the mobility and reactivity information together can be used for compound matching provided that the isotopic distributions of other species in complex mixtures do not shift and interfere with that of the desired analyte.

**Isotopic distribution reproducibility after HDX and HD scrambling with HDX.** One issue in determining ion identities in the manner proposed here is whether or not there is a sufficient number of identifiers to obtain high-confidence assignments. Ideally increased parameterization (separate measurements) would be required to address
mixtures of greater complexity. A question arises as to whether or not additional measurements can be made with the existing instrumentation that would provide unique ion characterization. One approach could involve the use of HD scrambling followed by HDX. It is instructive to consider that the “mobile proton” model for peptide ion dissociation [44,45,75,76] accurately describes HD scrambling [41,77]. Collisional activation causes intramolecular proton transfer events where a severing of a chemical bond can occur at the site of proton transfer. Because deuterons can be mobilized as well as protons, HD scrambling can follow collisional activation. HD scrambling offers an opportunity in that hydrogens can be scrambled to accessible sites and deuteriums can be scrambled to sites less accessible by HDX. Therefore, more extensive deuterium uptake can occur as the newly populated sites with hydrogen undergo subsequent HDX.

To test the reproducibility of a process in which ions are subjected to HDX and then HD scrambling with subsequent HDX, experiments were performed for the peptide mixture sample as outlined in the Experimental section. As with the other studies, the drift voltage and activation voltage were removed between runs and the drift gases were evacuated and then reintroduced. Figure 6.4 shows the effect of HD scrambling with further HDX on the same peptide ions from the model peptide mixture shown in Figure 6.1. Figures 6.4A and 6.4B demonstrate that even with ion activation there is little to no further deuterium incorporation for the [M+2H]^{2+} substance P and bradykinin ions. Figure 6.4C shows that [M+3H]^{3+} KKD peptide ions incorporate significantly more (~3 to 4) deuteriums after scrambling. The isotopic distribution RMSDs for the 6 HD scrambling trials of the peptide ions shown in Figure 6.4 were less than 1% also demonstrating suitable reproducibility for routine use in ion identification analyses.
Figure 6.4 — Model peptides showing different deuterium uptake patterns with HDX. A) [M+2H]^{2+} ions of substance P. B) [M+2H]^{2+} ions of bradykinin. C) [M+3H]^{3+} ions of the synthetic peptide KKDDDDDIKIIK. The buffer gas composition is labeled for conditions in which He + D_{2}O (orange) and He + D_{2}O + activation is used.
Reproducibility in experiments using HDX and HD scrambling with HDX followed by CID. As a final ion descriptor, it is useful to consider ion fragmentation in conjunction with the HDX and HD scrambling with HDX processes. Such an approach derives conceptually from parallel dissociation methods performed using ion mobility devices [62,63]. Here, although CID results in HD scrambling, the fragmentation process would provide reproducible ion fragmentation spectra as well as isotopic distributions for fragment ions. Furthermore, given that HDX provides a sufficient shift in \( m/z \) value, ion fragmentation could allow for further disambiguation of sample peaks within the same drift time window.

The model peptide mixture was subjected to HDX combined with HD scrambling with HDX. Then a peak at a drift time of 6.7 ms and having \( m/z \) 526 corresponding to extensively deuterated \([M+3H]^{3+}\) KKD peptide ions was selected (mobility and \( m/z \)) and the precursor ions were subjected to CID in the linear ion trap at a normalized collision energy of 40. Three replicates were collected and, as before, between each replicate the drift and activation voltages were set to zero and the buffer gases evacuated from the drift tube. Figure 6.5A presents the MS/MS spectra for the \([M+3H]^{3+}\) KKD peptide ions. Notably the ion fragmentation spectra are very similar as may be expected from the highly reproducible process of CID [78]. For example, the relative intensities of the identified y-ion series are very similar across the fragmentation spectra. Figure 6.5B shows an expanded region of the fragmentation spectra encompassing the isotopic distribution of the \( y_{12}^{2+} \) fragment ion. The isotopic distribution was found to be highly reproducible (RMSDs < 1%) for the \( y_{12}^{2+} \) fragment ion (Figure 6.5B). Other identified fragment ions exhibited a similar RMSD (data values not shown).
Figure 6.5 – A) Fragmentation spectra for [M+3H]^{3+} KK peptide ions after HDX and collisional activation. The ion was drift selected at 6.7 ms and fragmented at 40 NCE. Several fragment ions have been labeled. B) Isotopic distribution reproducibility of the y_{12}^{2+} fragment ion.
Although the experiments described above provide proof-of-principle demonstrations of combined IMS-HDX-MS/MS approaches for obtaining information that could serve as unique identifiers for ions, the discussion has thus far not addressed the timescale of these measurements. Admittedly, the combination of IMS with the scanning MS approach results in a relatively long timescale measurement (e.g., MS are collected at separate delay time settings). The IMS-MS data reported here required timescales of tens of minutes for this instrumentation geometry. That said, all of the measurements shown here can be performed on an instrument that combines IMS with time-of-flight (TOF) MS where the MS measurement is nested in the IMS experiment [79]. Additionally, parallel CID can be conducted for such an instrument at no cost to experimental run time [62,63]. Thus it would be possible to make these measurements on the seconds timescale. Consider experiments in which drift times and \( m/z \) values could be recorded in 5 seconds for which He is employed as a drift buffer gas. Then these values could be recorded for 5-second experiments in which a He and D\(_2\)O buffer gas mixture is used. Next the \( m/z \) shift could be determined in a 5-second HD scrambling with HDX experiment. Another 5 second experiment could be used to perform parallel CID. Overall, 20 seconds of instrumentation time could produce these six pieces of ion identification information. This is a significant time savings compared with condensed-phase separations.

**6.4 Conclusions**

The experiments described here provide proof-of-principle examples of how IMS combined with gas-phase HDX and MS analysis techniques can be used to provide
information-rich datasets. Such datasets could have a number of unique ion descriptors to be used in identification efforts for compounds in mixtures such as those encountered in ‘omics analyses. The studies show that sufficient reproducibility can be obtained for these descriptors such that ion identification could be achieved through comparisons with values recorded for molecular standards. Although such an approach would require a Herculean effort to populate such databases, the rapidity of the measurements would allow for significant improvements in sample throughput which is a significant challenge for biomarker discovery efforts. In addition to the database challenge, scoring schemes would need to be devised for providing an assessment of confidence in ion matches. Although important, such strategies could benefit from algorithms developed previously. For example, scoring isotopic distribution matches could be achieved using an approach that is similar to the cross correlation approach employed by the SEQUEST software suite used in proteomics experiments [80]. Finally, although the peptide systems utilized here provided a simple means for proof-of-principle studies, the approach should be suitable for ions of other electrosprayed compounds. This would include negatively-charged species because the relay mechanism for exchange with D₂O [81,82] as well as intramolecular proton transfer [83] should be operative in these ions. The approach may even be extended to include other descriptors such as exchange of aliphatic hydrogens at elevated drift tube temperatures [84]. In the near future, studies will be presented that demonstrate the approach for other metabolomics compounds.
6.5 References


Chapter 7

Future Directions
7.1 Proteomic Analyses

7.1.1 TiO$_2$ NP exposed rats. The affected biopathways identified in Chapters 2 and 3 begin to shed light on the connection of systemic inflammation to the observed changes in the arterial vasodilation response. Observed differences in the significant pathways of vascular tissue compared to plasma and aorta tissue suggest the tissues can respond dissimilarly to the TiO$_2$ exposure. However, it is unclear if this is due to spatial differences from the location of NP introduction, to temporal differences in the tissue response, or to functionality differences of the tissues, all of which occur by different mechanisms of toxicity.

Future proteomic studies to examine spatial differences should include additional resistance vascular tissue further downstream from that examined in Chapter 3 in order to determine if the response of the vascular tree is similar throughout or if there are changes as a result to the proximity to the particle introduction site. Rats should also be examined at multiple time points after exposure to examine temporal deviations in response. Various studies have examined changes in protein concentrations or gene expression as a result of TiO$_2$ NP exposure at various time points. For example, studies have been conducted at 1, 2, 8, 16, 30, and 90 days post exposure [1]; at 1 and 70 days post exposure [2]; at 5 days post exposure [3]; and at 1, 3, and 28 days post exposure [4]. Examining variations in the protein levels over time would also allow for information to be gathered on what metabolic pathways are being altered over time and relate this information to the physiological response. This could provide information as to whether vasodilation response returns to normal, what metabolic processes are involved, and when this occurs post-exposure.
7.1.2 ICV-STZ treated mice. Although protein concentration changes were observed as a result of the ICV-STZ treatments, none were found to be significant. In the work presented here, mice were given a single dose of STZ at a concentration of 3 mg/kg. Cognitive analysis was performed at 3.5 months post treatment and the animals scarified at 4 months post treatment [5]. Other studies used the same dose but examined rats for cognitive changes at 14 days post ICV-STZ treatment [6] and at 30 days post ICV-STZ treatment [7]. A different study examining cognitive changes in rats also utilized 3 mg/kg concentrations of STZ, but the STZ was administered twice, 48 hours apart [8]. Yet other studies have administered large doses of STZ to rats, 55 mg/kg [9] and 150 mg/kg [10], to determine the effect on tau phosphorylation, as higher degrees of tau phosphorylation are more strongly associated with cognitive decline [11]. Future studies should examine if different doses, a singular dosage, or multiple dosages would induce the early metabolic changes observed in AD. Furthermore, examining the animals at different times post ICV-STZ treatment might demonstrate marked metabolic changes.

The bottom-up proteomic analyses performed on the TiO$_2$ exposed rats and the ICV-STZ treated tau mice can be supplemented by additional proteomic techniques. For example, multidimensional protein identification technology (MudPIT) utilizes strong cation exchange (SCX) coupled with reversed phased (RP) chromatography separations to enhance peptide identification [12,13]. Examination of protein post translational modifications (PTMs) can provide valuable clues as to the protein activation state or the progression of a disease state [14]. Furthermore, isotopic labeling techniques allow for multiplexing of samples reducing the overall instrument analysis time and decreasing the run to run variability [15].
7.1.3 Multidimensional protein identification technology. In 1999 the Yates group described direct analysis of large protein complexes (DALPC), where two types of chromatography, strong cation exchange (SCX) and reverse phase (RP), are used in tandem to separate peptides based on different chemical properties followed by mass spectral identification [12]. By using 2 orthogonal types of chromatography, the researchers were able to increase peptide separation aiding in the identification of more peptides [12]. Similar to gel-based techniques, the SCX column separates peptides based on the isoelectric point of the peptide determined by the amino acid sequence. Typically, buffer solutions of increasing strength are applied to the SCX column and the salt ions displace the bound peptides. Peptides with low isoelectric points will elute first and peptides with high isoelectric points will elute later. The peptides are then subject to RP separation eluting the peptides based on decreasing polarity. After elution from the RP column, peptides are analyzed by mass spectrometry [12]. In the DALPC iteration, fractions were collected and then loaded onto the RP column [12]; however, this was later refined to elute peptides directly from the SCX column onto the RP column, a technique known as MudPIT (Multidimensional protein identification technology) [13]. These techniques were applied to analyze the *Saccharomyces cerevisiae* ribosome [12] and yeast proteome [13] and were able identify novel proteins and low abundance proteins within the samples [12,13].

The MudPIT technique applied to the projects in this work would be of particular interest to identify low abundance proteins such as kinases and transcription factors. Additionally, since there is a large dynamic range in the concentration of plasma proteins even after immunodepletion [16], 2D orthogonal separations would be useful in
discerning plasma peptides not identified in 1D RP separations.

7.1.4 Post translational modifications. PTMs can affect the structure and function of proteins [14]. In addition to bottom-up proteomics methods, PTMs analysis would provide greater insight into understanding metabolic processes within a biological system. For example, phosphorylation is a common modification to activate or inactivate proteins by a kinase or a phosphorylase adding or removing, respectively, a phosphate group from a serine, tyrosine, or threonine residue. The relative ratio of a non-phosphorylated protein to a phosphorylated protein can provide clues as to whether a pathway is up or down-regulated [14]. Analysis of phosphorylated proteins and peptides requires an enrichment step prior to analysis by mass spectrometry due to their low abundance [17]. There are many different enrichment options available including fractionation techniques, such as hydrophilic interaction chromatography (HILIC), strong cation exchange (SCX), or strong anion exchange (SAX); or enrichment techniques based on compound affinity, antibodies, or covalent modifications (as reviewed in reference [17]).

Another PTM of interest is methionine oxidation, a modification which has been shown to scavenge reactive oxygen species (ROS), to mark proteins for degradation, and to take part in cell signaling [18]. Furthermore, methionine oxidation has been implicated in a number of diseases including emphysema, cancer, and Alzheimer’s (as reviewed in references [19,20]). However, methionine residues are readily oxidized when exposed to air and unless samples are collected and prepared under an inert atmosphere, results can be misinterpreted (as reviewed in reference [21]). As with phosphorylated proteins and peptides, there are several options available to researchers
for methionine oxidation enrichment including the use of antibodies, diagonal chromatography, or simply looking for 16 Da shifts in mass spectral data [21].

Deamidation is another PTM that occurs on asparagine and glutamine residue through removal of the side chain amine, converting these residues into aspartic acid and glutamic acid, respectively. These changes can result in altered protein functionality and have been implicated in human aging, Alzheimer’s disease, and cataract formation (as reviewed by [22]). However, the rates of deamidation are dependent on the amino acid residue sequence, the tertiary structure of the protein, and the environment that the protein experiences [22]. One of the consequences of deamidation is the formation of the isomers iso-aspartate and gamma-glutamate. Top down proteomic approaches, where intact proteins are ionized and then fragmented, and bottom-up approaches, like those used in this work, have been successfully used to characterize the residue-specific locations of deamidation by searching for iso-aspartate and gamma-glutamate residues [22].

7.1.5 Isotopically labeled protein quantitation. Relative protein quantitation is a fast and simple method of analyzing relative changes between control and exposed groups. However, changes in mass spectral ionization efficiencies from sample to sample due to matrix effects or electrospray conditions add uncertainty to the results. One method of combating this is through the use of isotopic labeling. Gel based labeling methods and isotopic labeling prior to mass spectral analysis are methods commonly used for relative quantitation.

Difference gel electrophoresis (DIGE) is a technique that employs different dyes that fluoresce at different wavelengths [23,24]. Dyes label 3 different protein sample
sets, typically, a control sample, a treated sample, and an internal standard [24]. However, a third sample or a mixture of the two samples may be substituted for the internal standard [23]. The 3 labeled samples are combined and analyzed by 2D gel electrophoresis. Proteins are first separated by their isoelectric point and then separated in the orthogonal direction by molecular weight [24]. The same protein from each sample will migrate to approximately the same area on the gel, with slight differences due to the different fluorescent tags. The gel is then analyzed 3 different times at 3 different wavelengths to measure the protein concentrations based on the fluorescence intensity. The advantage of this technique is that multiplexing samples removes the run to run variation in running and analyzing gels compared to single sample gels [24]. Furthermore, 2D gel analysis has demonstrated the ability to separate over 10,000 protein spots [25]. Although this multiplexing technique allows for simultaneous analysis of different samples, multiple gels are still needed for statistical power [24]. Additionally, 2D gels methods are labor intensive and excision of the protein spots is required for protein identification by mass spectrometry [23].

Isotope coded affinity tags (ICAT) and isobaric tags for relative and absolute quantitation (iTRAQ) serve as means to isotopically label extracted proteins from organisms prior to mass spectral analysis. The isotopic label allows for determination of the relative amount of a protein in different samples when using a multiplex approach. ICAT is a technique that covalently binds the isotopic tag to the sulfhydryl group on the cysteine residues [26]. The tag consists of 3 parts: the thiol specific reactive group attaches the tag to cysteine residues, the linker region which makes up the middle part of the tag and contains hydrogens along a saturated carbon chain for the “light” label and
deuteriums for the “heavy” label, and the biotin affinity tag for peptide enrichment by affinity columns prior to analysis (see Figure 7.1). The “light” and “heavy” samples are combined, digested, and then analyzed with a bottom-up proteomics approach. The monoisotopic mass of the labeled peptide will be shifted from the monoisotopic mass of the unlabeled peptide [26]. The result is the ability to perform relative quantitation of proteins in two different samples at the same time. The comparison of the abundance of the “heavy” and “light” labeled peptides occurs in the MS analysis and not the MS/MS analysis [26].

One caveat to using this tag in LC-MS/MS experiments is the deuterium effect, that is, changes in the elution time of compounds are observed due to the incorporation of deuteriums [27,28]. Deuterium atoms will interact more with the C18 stationary phase than hydrogens. In the case where the deuteriums are on polar functional groups, there is less interaction with the stationary phase, and little to no change in elution time is observed. However, when the deuterium atoms are located on non-polar groups, such as in these ICAT tags, there is the possibility for interaction with the stationary phase resulting in longer elution times compared to the hydrogen containing tag [27,28]. The same research group that developed the ICAT tags modified them later to replace the deuterium labels with $^{13}$C in the linker region to eliminate any observed deuterium effect [29]. Therefore, this technique can be employed without consideration of the deuterium effect.

iTRAQ was first developed as a means to expand isotopic labeling for quantitation purposes from a binary system to a quaternary system [30]. Like ICAT, iTRAQ contains 3 regions: a reporter group, which contains $^{13}$C and/or $^{15}$N isotopes
Figure 7.1 – The structure of the ICAT tag, which contains a biotin, a linker region, and the thiol specific reactive group. The thiol specific reactive group attaches the tag to the sulfhydryl group of cysteine residues. The X's in the linker region denote the locations of hydrogen for the “light” label or deuterium for the “heavy” label.
detectable upon MS/MS fragmentation; a mass balance group, which balances the mass of the reporter group so that the combined masses of the two regions are equal; and an amine reactive group, which attaches the tag to the peptide (Figure 7.2). Protein samples are digested and then labeled with one of 4 different iTRAQ tags on the N-terminal ends and the side chains of lysine residues. The samples are then combined for MS/MS analysis. Since the overall mass of each of the 4 labels is the same, the co-eluting labeled peptides will have the same mass in MS analysis. However, when the peptide is fragmented, the iTRAQ tag is cleaved resulting in a neutral loss of the balance group and charge retention on the reporter group. The reporter group will then be detected at 114, 115, 116, and 117 m/z and the peak areas of the reporter group correspond to the amount of peptide in the associated sample [30]. After its initial development as a 4-plex system, iTRAQ is also available as an 8-plex system [31].

One advantage to using isotopic labeling techniques is the increase in sample throughput due to a reduction the overall instrument usage time compared to analyzing samples individually. Another advantage is the improved statistical power as the run to run sample error is reduced. However, the advantages of multiplexing are lost if there is not a way to compare one multiplex to another. This is accomplished by duplicating samples between multiplexes or using a standard protein or peptide in each multiplex. Furthermore, multiplexing kits cost approximately $65 per sample and can be cost prohibitive for some laboratories.

### 7.2 Metabolomic Analyses

Although untargeted metabolomics by LC-MS/MS is touted as a means to
Figure 7.2 – The iTRAQ tag containing a reporter group, which incorporates $^{13}$C and/or $^{15}$N to give different masses, a mass balance group, which balances the mass of the reporter group, and the peptide reactive group to attach the tag to the peptide.
examine a wide-swath of compound classes, it is hindered by data analysis. There are some promising software products on the horizon, but their efficacy has yet to be determined. For a review of metabolite identification software used in this work and upcoming software, refer back to Chapter 5. Until LC-MS/MS metabolite identification software has reached the ease and reliability of current protein identification software, researchers can use proteomic results to guide metabolomic studies. By understanding affected protein pathways researchers can target metabolites acting as signaling molecules or secondary messengers within pathways of interest. In cases where there is limited sample available, a similar extraction scheme as used in Chapters 2 through 4 would prove useful in extraction of many different classes of metabolites. Unfortunately, even with targeted analysis, standards are required in order to determine the ionization efficiency of the compound, if it will co-elute with other compounds, and the fragmentation pattern of the compound on a particular instrument with the desired settings.

7.3 IMS-MS Instrumentation

The IMS strategy described in Chapter 6 captures ion mobility, hydrogen/deuterium exchange (HDX), scrambling, and fragmentation parameters of different compounds in order to identify compounds within a complex mixture without the use of chromatographic separations. Although the home-built instrument used in Chapter 6 (and described in detail in reference [32]) demonstrated acceptable analytical reproducibility of all parameters, the linear ion trap mass spectrometer was not the ideal type of mass spectrometer for this type of analysis. Approximately 30 minutes was
required to collect the data across the entire drift time distribution. Moreover, the ion trap only allowed for a 100 Da sized window of fragmentation, which would greatly increase the analysis time in order to collect all the fragmentation data for the entire drift distribution.

New IMS-MS instrumentation is being developed in the Valentine laboratory with a similar drift tube design, but replacing the trapping mass spectrometer with a dispersive time-of-flight (TOF) mass spectrometer. A dispersive mass spectrometer would allow for detection of all ions, not just the sequential collection of ions collected at different m/z values. Additionally, a TOFMS would be able to incorporate parallel fragmentation [33,34] and fragment all ions at once, unlike the 100 Da sized window in the linear ion trap. By using a dispersive mass spectrometer, the total analysis time for a full drift time scan is estimated to decrease from 30 minutes to 5 seconds, and thus, greatly increase sample throughput.

7.4 IMS-MS Automation and Software

The new strategy described in Chapter 6 paired with the new instrumentation will collect ion mobility, HDX, hydrogen/deuterium (HD) scrambling with HDX, and fragmentation data on small molecules and peptides. With such a large amount of data being collected and the current algorithms in use, it would be very tedious to compile and analyze the data generated. Additional algorithms and software need to be developed to help sift through all of the data. Furthermore, data from model proteins, peptides, and compounds will be used to create searchable databases of known mobilities, uptake patterns, and scrambling patterns. Similar to the development of RP retention time
prediction algorithms [35–37], algorithms for HDX and HD scrambling could be developed to predict how compounds in the same class, or peptides based on their amino acid sequence would react under various conditions. Furthermore, scoring algorithms could be developed to compare experimental data to theoretical or library data for compound matching. Existing software that may be of use is The Deuterator [38] and HD Desktop [39], which determine the percent incorporation of deuterium [38,39] and the later provides visualization and statistical analysis of data [39].

It is also important to examine how different post translational modifications (PTM) would affect deuterium uptake and scrambling patterns. For example, a PTM such as an amidation would not only add an additional charge site, but also create an additional location for deuterium exchange to occur. Phosphorylation of a serine, threonine, or tyrosine would also add additional labile hydrogens. Acetylation, on the other hand, has previously been shown to decrease the gas-phase basicity of arginine [40].

Once the new instrument is brought online and automated data analysis methods developed, it will be important to validate the instrument. This will be achieved by comparing collected ion mobility, HDX, and scrambling data from the new instrument to literature values [41–45]. Additionally, it would be worthwhile to compare the performance of the new instrument and strategy to currently used LC-MS/MS methods. Performance of the two methods will be compared in terms of the total analysis time, the number of proteins, peptides, and metabolites identified, and the accuracy of the identifications.
7.5 References


[28] A. Valleix, S. Carrat, C. Caussignac, E. Léonce, A. Tchapla, Secondary isotope effects in liquid chromatography behaviour of 2H and 3H labelled solutes and


Proteomic and Metabolomic Analyses of Biological Systems with Liquid Chromatography Tandem Mass Spectrometry and Ion Mobility Mass Spectrometry

Megan M. Maurer

Dissertation submitted to the Eberly College of Arts and Sciences at West Virginia University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry

C. Eugene Bennett Department of Chemistry

APPROVAL OF THE EXAMINING COMMITTEE

Stephen Valentine, Ph.D., Chair

Suzanne Bell, Ph.D.

Glen P. Jackson, Ph.D.

Capsten Milsmann, Ph.D.

Date: 02-20-17

Timothy R. Nurkiewicz, Ph.D.