Discovery of Novel Serum Biomarkers for Diagnosing and Staging Alzheimer's Disease

Dipti Jigar Shah

Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Chemistry Commons

BYU ScholarsArchive Citation
Shah, Dipti Jigar, "Discovery of Novel Serum Biomarkers for Diagnosing and Staging Alzheimer’s Disease" (2014). All Theses and Dissertations. 5268.
https://scholarsarchive.byu.edu/etd/5268

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Discovery of Novel Serum Biomarkers for Diagnosing
and Staging Alzheimer’s Disease

Dipti Jigar Shah

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Steven W. Graves, Chair
Matthew R. Linford
David V. Dearden
Adam T. Woolley
Milton L. Lee

Department of Chemistry and Biochemistry
Brigham Young University
June 2014

Copyright © 2014 Dipti Jigar Shah
All Rights Reserved
ABSTRACT

Discovery of Novel Serum Biomarkers for Diagnosing and Staging Alzheimer’s Disease

Dipti Jigar Shah
Department of Chemistry and Biochemistry, BYU
Doctor of Philosophy

Alzheimer’s disease (AD) is an untreatable neurologic disease affecting more than 5 million Americans, most over 60 years of age. Protein plaques and neurofibrillary tangles typify AD brain pathology and are thought to cause the progressive dementia and brain shrinkage observed in AD. Currently there are no methods to diagnose the disease at a time before damage becomes irreversible.

Biochemical tests for AD using cerebrospinal fluid analysis or neuroimaging are not yet sufficiently sensitive and specific, and they are invasive. This points to a need for a more easily applied and more sensitive diagnostic test. Although the gross anatomical changes are localized to the brain, AD is likely to involve changes throughout the body. As a result of this, changes in the abundance of certain biomolecules present in the circulation system are likely to occur. Consequently, a serum proteomics approach able to measure such changes, when applied to AD, would likely find quantitative changes in relevant molecules that can help diagnose the disease correctly, ideally early in the disease process.

The goal of this work was to discover and validate novel diagnostic serum biomarkers for AD. For biomarker discovery and validation, we used a novel serum proteomics approach involving reversed phase capillary-liquid chromatography-electrospray ionization-quadrupole-time of flight mass spectrometry. Our samples were protein depleted, which helped us survey low molecular weight species in the serum without ion suppression from larger proteins like albumin. We were able to observe more than 8000 molecular species in a single run. The overall project was comprised of four studies: (i) discovery of novel potential serum AD markers, (ii) blinded validation of diagnostically promising biomarkers found in the initial study, with their further chemical identification, (iii) exploring gender-based serum AD biomarkers, and (iv) discovery of biomarkers that distinguish early versus moderate stage AD.

In the first study, the approach found 38 significant (p < 0.05) biomarkers and 21 near significant (p = 0.05 to 0.099) biomarkers. On using the forward selection approach, we built multi-marker panels with specificities and sensitivities higher than 80%.

The second study reports on a blinded validation study that was performed on a new set of serum samples. We focused on the 13 most promising AD biomarkers found as part of the
initial study. We successfully validated 4 of these biomarkers that showed highly significant statistical p-values. As part of this study, research was conducted to identify these 4 biomarkers, which was accomplished using tandem mass spectrometry with fragmentation experiments.

The third study used data from the initial study but looked at gender specific biomarkers. We found 31 significant and near significant serum AD biomarkers for women, 16 for men, and 25 that were gender independent. Multi-marker panels of AD biomarkers for women or men had sensitivities of >60% and specificities >85%.

In the fourth study, cases with moderate AD were compared to cases with very mild or mild AD to find novel biomarkers that could be used for staging. We found 44 significant and near significant biomarkers that were quantitatively different between mild and severe AD.

In conclusion, we were successful in accomplishing the goal of this work of finding, validating and identifying novel serum biomarkers that diagnose AD.

Keywords: Alzheimer’s disease, low-molecular weight serum proteome, lipid biomarkers, gender-based biomarkers, diagnosis, stage comparisons, validation, biomarker identification.
ACKNOWLEDGEMENTS

First and foremost, I wish to thank my mentor and advisor, Dr. Steven Graves. He has allowed me to work on this interesting and exciting biomarker discovery-based research on Alzheimer’s disease. His constant support, vision and ideas have helped shape this project. I also want to express my deep gratitude to our collaborator, Dr. John Kauwe, for his invaluable contribution from the inception to the successful completion of this project. I offer special thanks to Dr. W. Evan Johnson for carrying out complex statistical analyses on my data. I would also like to thank Dr. Craig Thulin for his significant teachings and insights into the field of proteomics, and for giving his valuable time to us every week. Bruce Jackson, the manager of the mass spectrometry center, has been key in helping us with the mass spectrometer, especially when the primary instrument broke down, and assisting whenever we had questions and doubts about its operation. I am very thankful to him for all that he has done and taught us.

I am also grateful to my seniors, Dr. Karen Merrell and Dr. Tanielle Alvarez, for teaching me many things during the course of my Ph.D. work, and also for their direct contribution to some aspects of my work. I want to acknowledge Swati Anand’s contribution to my project towards the end, when I needed help the most. I have had several undergraduate students work with me. I thank them for their tremendous help in analyzing my samples and making this lab really fun to work in. These include Frederick Rohlfing, Jesse Cobell, Trent Tipton, Taylor Turnbull, Jordan Salmon, Ryan Egbert, Sydney Young, Jeff Olson, Matthew Ellsworth, Casey Stinnett, Spencer Ruben, Tyler Mower, Hunter Schone, Hunter Wright, Jackson King, Caitlin Nichols and Tasha McGhie. My coworkers Dr. Moana Hopoate, Dr. Jie Ma, Komal Kedia and Ying Ding will always be part of my fun-filled and long-lasting memories I built in the lab.
I also want to give credit to the amazing professors in this department as well, whose classes gave me meaningful direction and insight into my research. These include Dr. Milton Lee who taught me separations, Dr. Daniel Austin and Dr. David Dearden who instructed me on the subject of mass spectrometry, Dr. John Prince who made me look at proteomics in a very different way in his class on computational proteomics and Dr. Matt Linford for his lessons in surface chemistry. I deeply appreciate all my committee members Dr. Graves, Dr. Lee, Dr. Woolley, Dr. Linford and Dr. Dearden for their support, encouragement, critical reviews and guidance that helped me blossom so well in my research. My special thanks goes to my teacher and mentor Prashant Kurle, who taught me years ago, but whose teachings continue to touch and inspire me. He brought out the true seeker in me.

In my family, I would like to thank my mother and father for their dreams of seeing me very well educated and my brother for being my strongest supporter. Their belief in me inspired me to work harder everytime. My husband has been a very key figure in this achievement. His love and motivation helped me finish this project with great success. Without my in-law’s strong encouragement, this would not have been possible. I have been blessed with some very terrific friends who have stood by me and always felt proud of my achievements. Last, but not least, my grandparent’s and God’s blessings have propelled me forward in the midst of hardships. I express my heartfelt gratitude to my entire family.
This dissertation is dedicated to my parents Daksha Shah and Rajnikant Shah, for giving me the best foundation. Their unconditional love and blessings put me on the path to where I am today.
# TABLE OF CONTENTS

TITLE PAGE ................................................................. i
ABSTRACT ................................................................................ ii
ACKNOWLEDGEMENTS ........................................................... iv
TABLE OF CONTENTS ............................................................. vii
LIST OF TABLES ................................................................. xiii
LIST OF FIGURES ................................................................. xiv
LIST OF ABBREVIATIONS ........................................................ xvii
CHAPTER 1 - INTRODUCTION .................................................. 1
  1.1 Alzheimer’s Disease (AD) .................................................... 1
    1.1.1 History .............................................................................. 1
    1.1.2 Pathology and Research Progress ................................. 2
    1.1.3 Prevalence and Impact ................................................... 2
  1.2 AD Biomarkers ................................................................. 5
    1.2.1 Inflammatory Biomarkers ................................................. 6
    1.2.2 Genetic Biomarkers ........................................................ 7
    1.2.3 Neuroimaging Biomarkers .............................................. 7
    1.2.4 Neurophysiological Biomarkers ...................................... 8
    1.2.5 Tissue Biomarkers .......................................................... 9
    1.2.6 CSF Biomarkers ............................................................. 9
    1.2.7 Conclusions ...................................................................... 10
    1.2.8 Why Blood Biomarkers .................................................. 11
    1.2.9 Current Blood Biomarkers .............................................. 12
1.3 Proteins and Lipids as Biomarkers ................................................................. 12
  1.3.1 Proteomics ................................................................................................ 12
    1.3.1.1 Studying Low Molecular Weight Serum Proteins Using Approaches That Involve High
            Abundant Protein Depletion ................................................................................. 16
    1.3.1.2 Rationale for Acetonitrile Precipitation ....................................................... 18
    1.3.1.3 Uniqueness of This Approach ........................................................................ 19
  1.3.2 Lipidomics ................................................................................................. 21
    1.3.2.1 Glycerophospholipids .................................................................................. 21
  1.4 Mass Spectrometry Based Biomarker Discovery ......................................... 26
    1.4.1 General Workflow .......................................................................................... 26
    1.4.2 My Approach ................................................................................................ 28
    1.4.3 Electrospray Ionization-Quadrupole-Time-Of-Flight Mass Spectrometer: Most Frequently
            Used Instrument for Biomarker Discovery ........................................................ 28
      1.4.3.1 Electrospray Ionization ............................................................................... 32
      1.4.3.2 Quadrupole as Mass Filter ......................................................................... 35
      1.4.3.3 Time-of-Flight Mass Spectrometer .............................................................. 37
  1.5 Layout of the Dissertation ............................................................................. 41

CHAPTER 2 – INITIAL STUDY OF NOVEL SERUM BIOMARKERS FOR ALZHEIMER’S
DISEASE ........................................................................................................... 42
  2.1 Abstract ........................................................................................................ 42
  2.2 Introduction .................................................................................................. 43
  2.3 Materials and Methods ................................................................................. 47
    2.3.1 Sample Collection and Study Population ....................................................... 47
    2.3.2 Ethics Statement .......................................................................................... 48
    2.3.3 Sample Processing ....................................................................................... 48
2.3.4 Analysis of Sera on a Capillary-LC-ESI-QTOF Mass Spectrometry System ........................................ 49
2.3.5 Time Normalization of the MS Data .................................................................................................. 50
2.3.6 Selecting Candidate Biomarkers ..................................................................................................... 51
2.3.7 Normalization of the Candidate Biomarkers ..................................................................................... 51
2.3.8 Biostatistical Analysis ....................................................................................................................... 52
2.4 Results .................................................................................................................................................. 53
2.4.1 Evaluating Potential Batch Effects .................................................................................................. 53
2.4.2 Candidate Serum AD Diagnostic Biomarkers .................................................................................... 53
2.4.3 Assessment of Diagnostic Similarities between Biomarkers ............................................................. 57
2.4.4 Evaluation of the Biomarkers in Multi-marker Panels ....................................................................... 57
2.4.5 Cluster Analysis ................................................................................................................................. 58
2.5 Discussion ................................................................................................................................................. 63

CHAPTER 3 - VALIDATION AND CHARACTERIZATION OF ALZHEIMER’S DISEASE

BIOMARKERS ................................................................................................................................................. 66
3.1 Abstract .................................................................................................................................................. 66
3.2 Introduction ............................................................................................................................................ 67
3.3 Materials and Methods .......................................................................................................................... 68
3.3.1 Procurement and Storage of Samples ............................................................................................... 68
3.3.2 Validation of Biomarkers .................................................................................................................. 68
3.3.3 Fragmentation and Chemical Identification Studies ........................................................................... 70
3.3.3.1 Peptide Fragmentation Experiments Employing Applied Biosystem’s QSTAR Pulsar I Quadrupole Orthogonal Time-of-Flight Mass Spectrometer .............................................................................. 70
3.3.3.2 Peptide Sequencing via the Mascot Search Engine ..................................................................... 71
3.3.3.3 De Novo Sequencing of Peptides .................................................................................................. 72
3.3.3.4 Lipid Identifications Employing Applied Biosystem’s QSTAR Pulsar I Quadrupole
5.1 Abstract .............................................................................................................................................. 115

5.2 Introduction ....................................................................................................................................... 116

5.2.1 Onset of Alzheimer’s Disease .................................................................................................. 116

5.2.2 Review of Current Biomarkers for Early Diagnosis ............................................................... 116

5.2.2.1 CSF biomarkers.................................................................................................................... 116

5.2.2.2 Plasma biomarkers ............................................................................................................... 118

5.2.2.3 Platelet biomarkers ............................................................................................................ 119

5.2.2.4 Structural biomarkers ........................................................................................................ 120

5.2.2.5 Alzheimer Associated protein (ALZAS) .......................................................................... 121

5.2.3 Alzheimer’s Disease is Multifactorial ...................................................................................... 122

5.2.4 Unknown Mechanisms Involved.............................................................................................. 124

5.3 Materials and Methods ...................................................................................................................... 125

5.4 Results ............................................................................................................................................... 126

5.4.1 Discovery of Candidate Serum Biomarkers for Early Stage AD ............................................. 126

5.4.2 Assessment of Diagnostic Similarities between Biomarkers ................................................... 127

5.4.3 Evaluation of the Biomarkers in Multi-marker Panels ............................................................ 127

5.4.4. Cluster Analysis ...................................................................................................................... 130

5.5 Discussion .......................................................................................................................................... 130

CHAPTER 6 - CONCLUDING REMARKS ............................................................................. 136

6.1 Summary ............................................................................................................................................ 136

6.2 Limitations ......................................................................................................................................... 137

6.3 Future Research ................................................................................................................................. 141

REFERENCES ........................................................................................................................... 144

CHAPTER 7 - APPENDIX ........................................................................................................ 165

7.1 Quadrupole Mass Spectrometer .................................................................................................... 165
7.1.1 Equations of Motion

7.1.2 Stability Diagram

7.2 Time-of-Flight Mass Spectrometer
LIST OF TABLES

Table 1.1. Summary of the history of research into Alzheimer’s disease and related disorders. 3

Table 1.2. Serum biomarkers for AD. 13

Table 2.1. Biomarkers considered for biostatistical analysis with significant (< 0.05) or near significant (< 0.1) p values. 54

Table 2.2. Set of 13 biomarkers that appeared in 10% or more of all the multi-marker models. 59

Table 3.1. Student’s t-test results showing p values obtained for the unnormalized validation data. Different stages are compared and p values for them are listed in the table as well. Trends indicating whether values were higher in cases or controls were also studied. Observation was made if they followed the same or opposite trend. Finally, in the last column, indication is given as to whether or not the biomarker passed the validation study. 80

Table 3.2. Table of calculated and observed b and y ions, confirming the sequence L/IAENR for the biomarker peptide 602.3 (charge state 1). 83

Table 4.1. Biomarkers that appeared in 10% of all male specific multi-marker models. 101

Table 4.2. Biomarkers that appeared in 10% of all female specific multi-marker models. 105

Table 5.1. List of 20 significant (p value < 0.05) and 24 near-significant (p value between 0.05 and 0.1) biomarkers when raw and normalized data were considered. 128

Table 5.2. Set of 13 biomarkers that appeared in 9% or more of all the multi-marker models. 131
LIST OF FIGURES

Figure 1.1. Comparison between (1 μg) non-protein depleted serum samples and acetonitrile (ACN) treated protein depleted samples 20

Figure 1.2. Structures of the different classes of Glycerophospholipids 22

Figure 1.3. Structure of 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine, commonly known as PC 36:4 24

Figure 1.4. Schematic representation of the fluid mosaic model of cell membranes. Reprinted from (Carter 2011, Singer and Nicolson 1972) with permission 25

Figure 1.5. Workflow for biomarker investigation in plasma/serum samples using global mass spectrometry 27

Figure 1.6. Quadrupole time-of-flight mass spectrometer 30

Figure 1.7. Electrospray ionization setup 34

Figure 1.8. Operation of a quadrupole in the X-Z and Y-Z plane 36

Figure 1.9. Quadrupole as a band pass filter. Reprinted from (Ginsbach and Dunnivant 2008) with permission 38

Figure 1.10. TOFMS with two electrodes for space focusing 40

Figure 2.1. Multimarker model (ROC curve) for biomarkers 531.3, 1568.2, 804.6, 602.3 and 708.3 60

Figure 2.2. Multimarker model (ROC curve) for biomarkers 1618.2, 804.6, 531.3, 602.3 and 708.3 61

Figure 2.3. Biomarker cluster analysis of 13 AD biomarkers subset 62

Figure 3.1 MS/MS spectrum revealing fragment ions of molecule m/z 602.31. 84
Figure 3.2 MS/MS spectra revealing fragment ions of molecule m/z 804.55 (second spectrum zoomed in to show fragments 520.3342 and 542.3211).

Figure 3.3 MS/MS spectrum revealing fragment ions of molecule m/z 874.59.

Figure 3.4 MS/MS spectrum revealing fragment ions of molecule m/z 804.53.

Figure 3.5. Receiver operator characteristic curve with AUC 0.7865 for diagnosis of AD using all 4 validated biomarkers 602.3, 804.6, 874.6, 804.5. The curve represents biomarkers that looked different between non-demented controls (CDR = 0) vs. any stage AD (CDR = 0.5, 1.0, 2.0).

Figure 3.6. Receiver operator characteristic curve with AUC 0.8502 for early diagnosis of AD using biomarkers 602.3 and 804.6. The curve represents biomarkers that looked different between non-demented controls (CDR = 0) vs. mild stage AD (CDR = 0.5).

Figure 3.7. Receiver operator characteristic curve with AUC 0.8494 for early diagnosis of AD using all 4 biomarkers 602.3, 804.6, 874.6 and 804.5. The curve represents biomarkers that looked different between non-demented controls (CDR = 0) vs. mild stage AD (CDR = 0.5).

Figure 4.1. Multimarker model (ROC curve) for biomarkers 804.6 and 1618.2 for identification of AD in men.

Figure 4.2. Multimarker model (ROC curve) for biomarkers 574.3, 1170.0, 810.6 and 804.6 for identification of AD in men.

Figure 4.3. Biomarker cluster analysis of 10 best male specific AD biomarkers.

Figure 4.4. Multimarker model (ROC curve) for biomarkers 675.6, 531.3, 513.3 and 515.3 for identification of AD in women.

Figure 4.5. Multimarker model (ROC curve) for biomarkers 531.3, 513.3, 1576.2 and 515.3 for identification of AD in women.
Figure 4.6. Multimarker model (ROC curve) for biomarkers 675.6, 1516.2, 921.4 and 790.6 for identification of AD in women.

Figure 4.7. Multimarker model (ROC curve) for biomarkers 1107.5, 1516.2, 921.4 and 790.6 for identification of AD in women.

Figure 4.8. Multimarker model (ROC curve) for biomarkers 790.6, 1516.2 and 921.4 for identification of AD in women.

Figure 4.9. Biomarker cluster analysis of 10 best female specific AD biomarkers.

Figure 5.1. Possible subgroups of Alzheimer's disease based on the risk factors.

Figure 5.2. Multimarker model (ROC curve) for biomarkers 552.3, 744.4 and 662.4.

Figure 5.3. Multimarker model (ROC curve) for biomarkers 808.8, 552.3, and 571.4.

Figure 5.4. Biomarker cluster analysis of 13 AD biomarkers subset.

Figure 7.1. Stability diagram showing stable solutions of Mathieu’s equation in the X direction.

Figure 7.2. Stability diagram showing stable solutions of Mathieu’s equation in the Y direction.

Figure 7.3. Overlaid stability diagram showing the stable regions at the intersection points.

Figure 7.4. Magnified image of the stability region.

Figure 7.5. Concept of time-of-flight mass spectrometer.
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-DE, 2-DGE</td>
<td>Two dimensional gel electrophoresis</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>ACT</td>
<td>A1-antichymotrypsin</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>ADRC</td>
<td>Alzheimer's disease research centre</td>
</tr>
<tr>
<td>AELO</td>
<td>AD cases with low Aβ&lt;sub&gt;1–42&lt;/sub&gt;, high incidence of APOE ε4, and late onset</td>
</tr>
<tr>
<td>ALZAS</td>
<td>Alzheimer's associated protein</td>
</tr>
<tr>
<td>ApoE4</td>
<td>Apolipoprotein E-e4 gene</td>
</tr>
<tr>
<td>APP</td>
<td>Amyloid precursor protein</td>
</tr>
<tr>
<td>APP&lt;sub&gt;SWE&lt;/sub&gt;</td>
<td>APP with swedish mutation</td>
</tr>
<tr>
<td>ASI</td>
<td>Aerospray ionization</td>
</tr>
<tr>
<td>ATEO</td>
<td>AD cases with low Ab&lt;sub&gt;1–42&lt;/sub&gt;, high tau, and early onset</td>
</tr>
<tr>
<td>ATURO</td>
<td>AD with low Ab&lt;sub&gt;1–42&lt;/sub&gt;, high tau, high ubiquitin, and recent onset</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under curve</td>
</tr>
<tr>
<td>Aβ</td>
<td>Amyloid beta</td>
</tr>
<tr>
<td>Aβ 42</td>
<td>42-amino acid amyloid beta peptide</td>
</tr>
<tr>
<td>BACE-1</td>
<td>Beta-secretase 1</td>
</tr>
<tr>
<td>BBB</td>
<td>Blood brain barrier</td>
</tr>
<tr>
<td>BCA</td>
<td>Bicinchoninic acid assay</td>
</tr>
<tr>
<td>CAA</td>
<td>Cerebral amyloid angiopathy</td>
</tr>
<tr>
<td>CDR</td>
<td>Clinical dementia rating</td>
</tr>
<tr>
<td>CJD</td>
<td>Creutzfeldt-Jakob disease</td>
</tr>
<tr>
<td>cLC</td>
<td>Capillary-liquid chromatography</td>
</tr>
<tr>
<td>CNR2</td>
<td>Cannabinoid receptor 2</td>
</tr>
<tr>
<td>CRP</td>
<td>C reactive protein</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>E</td>
<td>Electric field</td>
</tr>
<tr>
<td>EH</td>
<td>Electrohydrodynamic ionization</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme linked immunosorbent assay</td>
</tr>
<tr>
<td>ERP</td>
<td>Event related potential</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>ESI-TOF</td>
<td>Electrospray ionization time of flight</td>
</tr>
<tr>
<td>FA</td>
<td>Formic acid</td>
</tr>
<tr>
<td>FD</td>
<td>Field desorption</td>
</tr>
<tr>
<td>FIB</td>
<td>Fast atom bombardment</td>
</tr>
<tr>
<td>FTDP-17</td>
<td>Frontotemporal dementia with parkinsonism-17</td>
</tr>
<tr>
<td>GPLs</td>
<td>Glycerophospholipids</td>
</tr>
<tr>
<td>GSS</td>
<td>Gerstmann-Sträussler-Scheinker disease</td>
</tr>
</tbody>
</table>
CHAPTER 1 - INTRODUCTION

1.1 Alzheimer’s Disease (AD)

1.1.1 History

In 1906, German physician Dr. Alois Alzheimer came across a curious case of a 51-year-old woman, Auguste Deter, in the mental asylum of Frankfurt am Main. The woman was completely delirious and had alternating versions of reality. She showed jealousy towards her husband and thought that her life was in constant danger from someone unknown. She screamed for hours at end and would suffer from auditory hallucinations. Her memory was seriously impaired as well. She would see certain objects and forget them the next moment. Her reading and writing abilities had deteriorated to meaningless levels. She no longer remembered the use of certain objects and was baffled at her surroundings all the time (Stelzmann, Norman Schnitzlein et al. 1995).

After her death, her postmortem showed a uniformly degenerated brain without macroscopic focal degradation. Depositions of a pathological metabolic substance in or on 67% of all neurons of the cortex were observed along with entangling and clustering of neurofibrils. These depositions led to small aggregates distributed all over the cortex with the highest concentration in the upper layers. Many glial cells showed fibrous outgrowth accompanied by a proliferation of vessels and a growth on the endothelia. All this led Dr. Alzheimer to believe that this was an unusual and uncommon illness, and he published his findings in 1907. Emil Kraepelin, Dr. Alois Alzheimer’s mentor in Munich, finally named this illness “Alzheimer’s disease” in 1910 (Stelzmann, Norman Schnitzlein et al. 1995, Small and Cappai 2006).
1.1.2 Pathology and Research Progress

The post-mortem findings of Dr. Alzheimer were the stepping-stone to highlight the pathological and neurological effects of AD in the patient’s brain. In an AD brain, plaques are found outside the brain cells while neurofibrillary tangles are seen inside the brain cells. Plaques interrupted synapses, breaking the signals between the brain cells. Tangles damage brain cells by obstructing the transport of food and energy around the brain cells. Thus, progressive memory loss and brain degeneration are hallmarks of AD. Despite these consistent anatomical changes, there are no approved ways to diagnose the disease in time for effective treatment and appropriate dietary supplementation.

More than a century later, vast advances have been made to the understanding of this dreaded disease. The pathological and neurological effects of AD have been well documented although methods are still being researched to provide an early diagnosis that may allow for treatments to be tested with the hope that they will be able to prevent or slow down the disease in the long term. The progress made in the research of Alzheimer’s is traced chronologically in Table 1.1.

1.1.3 Prevalence and Impact

Despite all of the research, AD is the most common form of dementia and affects 20-30 million people worldwide. In the US alone, one out of nine develops AD and over 5.2 million people in the US already suffer from AD as of 2013. The number is predicted to grow to 13.8 million by 2050. AD is the sixth leading cause of death in the country and AD fatalities have increased 68% between 2000 and 2010. In that same period, deaths attributed to heart disease, which is the number one cause of death in the country, have decreased 16% (German, Gurnani et al. 2007, Association 2013).
Table 1.1. Summary of the history of research into Alzheimer’s disease and related disorders.

<table>
<thead>
<tr>
<th>Year</th>
<th>Related Developments</th>
<th>Alzheimer's Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1902</td>
<td>Improved silver stains</td>
<td></td>
</tr>
<tr>
<td>1906</td>
<td>Alzheimer’s case history of Auguste Deter</td>
<td></td>
</tr>
<tr>
<td>1910</td>
<td>Alzheimer’s disease named</td>
<td></td>
</tr>
<tr>
<td>1922</td>
<td>Lewy body described</td>
<td></td>
</tr>
<tr>
<td>1932</td>
<td>First hereditary case described</td>
<td></td>
</tr>
<tr>
<td>1962</td>
<td>L-dopa therapy in Parkinson's disease (Mayeux, Honig et al.)</td>
<td></td>
</tr>
<tr>
<td>1964</td>
<td>Ultrastructure of plaque and tangle by electron microscopy</td>
<td></td>
</tr>
<tr>
<td>1968</td>
<td>Recognition of prevalence of disease in the elderly</td>
<td></td>
</tr>
<tr>
<td>1976</td>
<td>Cholinergic deficit described</td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>Huntington's genetic linkage</td>
<td>Sequence of Amyloid-β (Aβ) from AD amyloid angiopathy</td>
</tr>
<tr>
<td>1984</td>
<td>Sequence of Aβ from Down's syndrome from amyloid angiopathy</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>Cloning of the prion gene</td>
<td>Sequence of Aβ from plaques</td>
</tr>
<tr>
<td>1986</td>
<td>Tau as major component of tangles</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>Cloning of amyloid precursor protein (APP) and localization to chromosome 21</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>Mutations in prion gene in Creutzfeldt-Jakob Disease (CJD)/ Gerstmann-Sträussler-Scheinker disease (GSS)</td>
<td>APP mutations in AD; a descriptive system of cataloguing the neuropathology determined</td>
</tr>
<tr>
<td>1990</td>
<td>Mutations in APP cause hereditary cerebral hemorrhage with amyloidosis-Dutch type (HCHWA-D); prion mutations cause neurodegeneration in mice</td>
<td>APP mutations increase Aβ42</td>
</tr>
<tr>
<td>1993</td>
<td>Apolipoprotein E-ε4 gene (ApoE4) associated with AD; cholinergic therapy approved for AD</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>APP transgenic mice made with plaque pathology;</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Event</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>Presenilins cloned as loci for early onset AD</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>Synuclein mutations identified in PD; Synuclein identified as major component of Lewy bodies</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>Tau mutations identified in Frontotemporal dementia with parkinsonism-17 (FTDP-17)</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>Presenilins identified as γ-secretase</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>Beta-secretase cloned; Aβ immunization in mice reduces amyloid pathology</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Mice with tangles made using FTDP-17 mutations</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>Mice with plaques and tangles made</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Aβ vaccine trials halted because of side effects</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>Pittsburgh Compound B (PIB) positron electron tomography (PET) ligand shown to bind to fibrillar brain Aβ</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>Dubois’s research criteria for AD (biomarker-based criteria, independent of dementia) published.</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Dubois’s lexicon defining AD (preclinical AD, prodromal AD, AD with dementia) published; cerebrospinal fluid (CSF) assay for Aβ oligomers developed</td>
<td></td>
</tr>
</tbody>
</table>
Another notable aspect to the statistics on AD are the number of caregivers. In 2012, 15 million family members and voluntary caregivers gave 17.5 billion hours of their time to care for patients with AD and other mental illnesses, a contribution worth more than 216 billion dollars. Apart from this valuation, the total payments made to health care professionals, hospitals, medical treatments, long term care and hospice services for AD patients came out to be 203 billion dollars for 2013. All this doesn’t take into account the mental and physical stress the caregivers go through. For family caregivers, to see a loved one go through memory loss, functional decline and mental outbursts is emotionally devastating. It starts taking a toll on their health too, increasing their own risk of chronic diseases and stress related disorders (Association 2013).

1.2 AD Biomarkers

Development of medical therapies for AD has been difficult because subjects with earlier stage disease cannot be radiologically or biochemically identified. Over the last decade, there has been substantial research done on identifying biomarkers that would aid in early diagnosis of AD. The research can broadly be categorized into the following fields: genomics, proteomics, structural and functional imaging and neurophysiology (Cedazo-Minguez and Winblad 2010)

The defining characteristics of an ideal biomarker is as follows:

1. A biomarker must dependably differentiate pathological cases from normal subjects at an early stage of the disease.

2. The patient should incur the least (possible) discomfort, risk and attendant expenses so that more samples can be collected to allow serial research studies to observe various stages of the disease as it progresses.
3. The biomarker must undergo uniform validation in various populations of AD patients.

4. The biomarker should be able to forecast the emergence of psychological symptoms and should be validated in potential clinical trials (preferably with follow-up to postmortem to allow concrete diagnosis).

The current status of all biomarkers for AD to date is summarized in the following sections.

1.2.1 Inflammatory Biomarkers

AD brains exhibit evidence of inflammatory processes. A serine protease inhibitor, α1-antichymotrypsin (ACT), that is found associated with Aβ in senile/neuritic plaques, is a thoroughly researched inflammatory marker. Preliminary studies of ACT have yielded variable results from unchanged ACT to increased ACT in AD serum or CSF. Further studies that involved a larger sample population concluded an overall increase in the plasma and CSF ACT levels of patients with AD, and an inverse correlation between the levels and cognitive capabilities. The general consensus was that increased levels of ACT signified increased risk of dementia, vascular dementia (Fjell, Walhovd et al. 2010) and AD but was not able to differentiate between the different types of dementia (Nielsen, Minthon et al. 2007).

The findings from cytokine studies for AD patients have been inconclusive as well. Independent studies have found mixed results for CSF and plasma/serum interleukin-6 (IL-6) levels in AD patients and also the use of their levels to differentiate between VD and AD. These results have been attributed to the basic challenges faced in the evaluation of cytokines in AD because cytokine concentration changes significantly over time and can vary based on the individual’s genetic background, multiple inflammatory processes in the system, anti-inflammatory drugs use and environmental factors (Craig-Schapiro, Fagan et al. 2009).
1.2.2 Genetic Biomarkers

An alternative approach to identifying biomarkers in mental disease is the use of transcriptional profiling of genes in patient samples. This approach can potentially predict changes at a much earlier stage than changes observed for protein alterations. Also, the uniformity of whole blood RNA allows transcriptome studies to be done easily compared to both proteome and metabolic studies. Initial studies of AD patients compared with controls found 20 genes out of 3200 that showed a different expression. The expression patterns of 33 genes which included APP, insulin degrading enzyme, histone cluster 1, H3e (HIST1H3E) and cannabinoid receptor 2 (CNR2) were further studied over a period of one year in correlation to AD. The results demonstrated a negative correlation of HIST1H3E and a slight positive correlation of CNR2 to the Mini-mental State Examination (MMSE) score. These results hold out promise of gene expression profiling as a way to find biomarkers for early AD (Cedazo-Minguez and Winblad 2010).

1.2.3 Neuroimaging Biomarkers

Functional and molecular neuroimaging have shed light into brain structure and physiology. This has permitted investigation of specific proteins and aggregates. This is hard to achieve even via brain autopsy. For instance, the distinctive type of degeneration in cortical and hippocampal areas in cases of severe AD are difficult to quantify even during an autopsy. Nevertheless, the biggest drawback is that brain atrophy cannot be measured until the latter stages of the disease when it appears. Additionally, neurodegeneration and cognitive decline are also found in other brain diseases and tissue volume changes in the brain are not indicative of the pathology of the disease. In the end, all the information on brain degeneration and volume changes in late stage
AD cannot be used to any real advantage if it does not make early stage, ante-mortem analysis possible (Perrin, Fagan et al. 2009).

There have been efforts to make neuroimaging more sensitive and more specific. Among the five dyes: [18F]FDDN, 18F-BAY94-9172, 11C-SB-13, 11C-BF-227 and 11C-PIB used to date for imaging Aβ plaques, only [18F]FDDNP may be retained by neurofibrillary tangles (Small and Cappai 2006). As yet, no compound that selectively binds to aggregates of tau has been found (Perrin, Fagan et al. 2009). When microglia go through changes with activation, the expression of benzodiazepine receptors increases. Thus, it can be utilized as a target for radiological compounds like [11C]DAA1106 (Zhang, Maeda et al. 2004), [11C]vinpocetine (Vas, Shchukin et al. 2008) and [11C](R)-PK11195. Only the last of these has been reported in studies of Alzheimer's disease in humans, in conjunction with PIB (Perrin, Fagan et al. 2009). Their value then is largely unknown.

1.2.4 Neurophysiological Biomarkers

During the transition between healthy aging and AD, various quantitative electroencephalography (qEEG) components, as well as event related potential (ERP) activity, are altered. For example, anomalies in P600 and N400, two ERP components, link the conversion from memory loss related MCI to AD. A vagus nerve stimulation (VNS) technique was recently established to attempt early diagnosis of dementia. The consideration of these markers is important because of their simplicity and cost-effectiveness in identifying early AD patients that would allow timely measures to be taken (Cedazo-Minguez and Winblad 2010). However, at this time these tests are not applied generally.
1.2.5 Tissue Biomarkers

Amyloid-β is a peptide fragment of amyloid precursor protein produced by sequential activities of the β-secretase and γ-secretase enzymes. It shows up in different lengths with amino acid residues ranging from 38 to 43. The peptides most frequently involved in the formation of plaques associated with AD are Aβ40 and Aβ42. In addition, other proteins are associated with the neurofibrillary tangles that are also seen in AD. The two most common are tau and phosphorylated forms of tau, and these have emerged as chief diagnostic and prognostic CSF biomarkers. The amyloid cascade hypothesis of AD considers Aβ42 as the central peptide leading to Aβ aggregation (Hardy and Selkoe 2002, Perrin, Fagan et al. 2009).

1.2.6 CSF Biomarkers

Because the cerebrospinal fluid communicates with aspects of the brain, the same proteins as described in the preceding section have been investigated as biomarkers for AD. Some of the well-documented changes in these amyloid peptides in patients with diagnosed AD include decreased CSF levels of the 42-amino acid beta amyloid peptide (Aβ 42). This is thought to indicate a change from a soluble form in the CSF to its accumulation and aggregation in plaques in the central nervous system. Initial studies of APP levels were too varied to be considered useful as a potential biomarker for AD. Also, elevated CSF levels of microtubule-linked protein tau are considered a generic indicator of brain damage, but the hyperphosphorylation of tau that seems to be observed mostly in the brains of AD patients makes increased CSF levels of phosphorylated tau (p-tau) a more promising and specific indicator for AD. These markers are used commercially to aid in AD diagnosis. Reduced levels of CSF Aβ 42, increased total tau (t-tau) and tau phosphorylated at residue 181 (p-tau181) have the potential to be an early diagnostic
indicators of AD in persons with mild cognitive impairment (Okonkwo, Mielke et al. 2011). These findings also suggest that AD pathological degradation starts much earlier than dementia. It’s still being determined as to when these CSF alterations actually take place during the early phases of AD before the symptoms set in. The challenges faced are in the procurement of samples from normal, middle-aged people who will eventually develop AD in their older age (Ringman, Coppola et al. 2012).

1.2.7 Conclusions

The only conclusive diagnosis of AD is post-mortem microscopic examination of the brain. At present, AD diagnosis includes detailed assessment of the patient’s personal and family medical histories and administering neuropsychological tests to evaluate cognitive function, as well as precluding other sources of dementia such as vitamin and mineral deficiencies, low thyroid function, infections, cancer and depression and excluding possibilities of other brain-degenerative dementias like frontotemporal, Lewy-body and CJD.

Brain imaging and CSF can assist in discriminating later stage and potentially earlier stage AD from other dementias and neurodegenerative disorders. This conclusion is based on the evidence that the AD brain has damaged and/or affected learning and memory areas, shows atypical deposits of Aβ plaques, accumulation of apoE and neurofibrillary tangles (NFTs) of protein tau and α-synuclein, and is characterized by diminished glucose metabolism. CSF testing shows decreased concentrations of Aβ peptides and increased concentration of tau protein in cases of AD, although this is often in later stage AD (Mucke 2009). While NFTs and Aβ plaques typify the disease, they do not entirely characterize AD. Neuronal, axonal and synaptic impairment/loss together with inflammation may be even earlier events in the disease process,
some believing they start decades before the first symptoms of Alzheimer’s surface (Braak and Braak 1997, Perrin, Fagan et al. 2009, Price, McKeel Jr et al. 2009). NFTs and Aβ plaques occur only about 10-15 years before the neurons of the hippocampus and medial temporal lobe are wiped out, leading to the onset of the clinically significant cognitive decline (Petersen, Doody et al. 2001, Perrin, Fagan et al. 2009). Most of the tests, including imaging and CSF testing, used to diagnose AD biomarkers, cannot accurately predict disease before substantial damage is present. In a disease like AD, this provides little value.

1.2.8 Why Blood Biomarkers

Over the last 10 years, increasing focus has been placed on blood-based biomarkers. The CSF biomarkers and neuroimaging techniques are sensitive and accurate. Aβ-42, total tau and hyperphosphorylated tau proteins in cerebrospinal fluid are specific AD biomarkers with moderate sensitivity that can differentiate people with MCI who progressed to AD and who did not. However, there are limitations to clinical implementations of these techniques. The lumbar puncture technique for CSF sample collection is considered invasive and has become unpopular in several countries. Also, collecting samples at different stages of the disease is often not possible, which impacts monitoring (Henriksen, O’Bryant et al. 2014).

Hence, blood biomarkers would have a distinct advantage compared with CSF biomarkers because of the relative ease of collecting samples. Also, these samples can be regularly obtained at a hospital or the residence of the patient to enable economically viable checkups and stage-by-stage monitoring of the disease (Lista, Faltraco et al. 2013). Because of this, early diagnosis of AD may become possible which will benefit efforts to find better treatment and improve the lives of the patients. This is especially critical since AD is becoming more and more globally
1.2.9 Current Blood Biomarkers

Biomarkers in plasma for AD have been researched for the last decade but have not yielded consistent results because of the limitations imposed by the blood brain barrier (BBB). Proteins and metabolites from the brain often have limited and variable access to the general circulation, and even if they cross the BBB, they become significantly diluted when they enter the plasma. It is hard to detect biomarkers closely linked to brain pathology, including small metabolites, lipophilic molecules and even molecules with specific transporters. These biomarkers appear to differ from person to person. A conscious effort has been made to focus most of the research on measuring the disease-specific molecules, which in AD’s case is beta-amyloid (Aβ). Total plasma Aβ or Aβ-42 levels are found to be elevated in inherited AD, but inconsistency in the results allowed only poor diagnosis of AD. Another reason for the inconclusive evidence is that there are other sources of Aβ such as platelets which make Aβ in plasma a non-tissue specific marker. Studies have also shown that simultaneous use of medications, including insulin, also affects the concentration of Aβ in plasma (Cedazo-Minguez and Winblad 2010). Many research studies have been done to find proteomic-based biomarkers. Table 1.2 lists the biomarkers found to date.

1.3 Proteins and Lipids as Biomarkers

1.3.1 Proteomics

The word proteome was first used to classify a set of proteins in a genome in a particular
Table 1.2. Serum biomarkers for AD (*Serum proteomics studies, **Only when AD only and AD with cerebral amyloid angiopathy (CAA) are combined, NEV = No estimate of variance). (Source: Zabel, Schrag et al. 2012).

<table>
<thead>
<tr>
<th>Identified protein</th>
<th>Protein function</th>
<th>Publications</th>
<th>Reported change in AD (initial finding)</th>
<th>Current result (LC-MS/MS follow up results)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alpha2-macroglobulin</strong></td>
<td>Protease inhibitor</td>
<td>(Giometto, Argentiero et al. 1988)</td>
<td>No change</td>
<td>Increased</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Morini et al. 1995)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Parnetti et al. 1995)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Hye, Lynham et al. 2006)</td>
<td>↑ p = 0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Hye, Lynham et al. 2006)</td>
<td>↑ p = 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td><strong>Alpha1-antichymotrypsin</strong></td>
<td>Protease inhibitor</td>
<td>(Matsubara, Amari et al. 1988)</td>
<td>↑ p &lt; 0.001</td>
<td>Trend toward increase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Matsubara, Hirai et al. 1990)</td>
<td>↑ p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Brugge, Katzman et al. 1992)</td>
<td>↑ p &lt; 0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Hinds, Kukull et al. 1994)</td>
<td>↑ p &lt; 0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Morini et al. 1995)</td>
<td>↑ p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Parnetti et al. 1995)</td>
<td>↑ p &lt; 0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Lieberman, Schleissner et al. 1995)</td>
<td>↑ p &lt; 0.00001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Altstiel, Lawlor et al. 1995)</td>
<td>↑ p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Pedrini et al. 2000)</td>
<td>↑ p &lt; 0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Pedrini et al. 2000)</td>
<td>↑ p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(McIlroy, Vahidassr et al. 2000)</td>
<td>↑ p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(DeKosky, Ikonomovic et al. 2003)</td>
<td>↑ p = 0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Nielsen, Minthon et al. 2007)</td>
<td>↑ p &lt; 0.0.05</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Function</td>
<td>Reference</td>
<td>Change</td>
<td>p-Value</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>Alpha1-antitrypsin</td>
<td>Protease inhibitor</td>
<td>(Porcellini, Davis et al. 2008)</td>
<td>↑ p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Furby, Leys et al. 1991)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Pirttila, Mehta et al. 1994)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zhang, Barker et al. 2004)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Giometto, Argentiero et al. 1988)</td>
<td>↑ p &lt; 0.001</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Matsubara, Amari et al. 1988)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Morini et al. 1995)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Parnetti et al. 1995)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Nielsen, Minthon et al. 2007)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Liao, Yu et al. 2007)</td>
<td>↑ p = 0.0003</td>
<td></td>
</tr>
<tr>
<td>Complement C3</td>
<td>Inflammation</td>
<td>(Giometto, Argentiero et al. 1988)</td>
<td>↑ p &lt; 0.005</td>
<td>Increased</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td>p = 0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Complement C4</td>
<td>Inflammation</td>
<td>(Giometto, Argentiero et al. 1988)</td>
<td>↑ p &lt; 0.005</td>
<td>Increased</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td>p = 0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Hye, Lynham et al. 2006)</td>
<td>↓ p = 0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>C1 inhibitor</td>
<td>Complement regulation</td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td>Trend toward increase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>↑ p = 0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>↑ p = 0.029</td>
<td></td>
</tr>
<tr>
<td>Complement factor H</td>
<td>Complement regulation</td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Hye, Lynham et al. 2006)</td>
<td>↑ p = 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Serum amyloid P</td>
<td>Immune regulation</td>
<td>(Hye, Lynham et al. 2006)</td>
<td>↑ p = 3E-04</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Verwey, Schuitemaker et al. 2008)</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>CD5 antigen</td>
<td>Immune regulation</td>
<td>(Hye, Lynham et al. 2006)</td>
<td>↓ p = 0.029</td>
<td>No change</td>
</tr>
</tbody>
</table>

14
<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>Reference</th>
<th>Change</th>
<th>p-value</th>
<th>Change</th>
<th>p-value</th>
<th>Change</th>
<th>p-value</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceruloplasmin</td>
<td>Copper chaperone</td>
<td>(Giometto, Argentiero et al. 1988)</td>
<td>↑ p &lt; 0.001</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td>p = 0.717</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Licastro, Morini et al. 1995)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Kessler, Pajonk et al. 2006)</td>
<td>↓ p = 0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Squitti, Quattrochi et al. 2006)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Squitti, Ventriglia et al. 2007)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transthyretin</td>
<td>Thyroxin/retinol transport</td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td>Trend toward increase</td>
<td>p = 0.182</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clusterin</td>
<td>Lipid transport</td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>↑ p = 0.019</td>
<td>No change</td>
<td>p = 0.243</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Liao, Yu et al. 2007)</td>
<td>↓ p = 0.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↓ p = 0.0061</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein A-1</td>
<td>Lipid transport</td>
<td>(Zhang, Barker et al. 2004)</td>
<td>No change</td>
<td>No change</td>
<td>p = 0.243</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Liu, Hu et al. 2006)</td>
<td>↑ p &lt; .0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein B100</td>
<td>Lipid transport</td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td>No change</td>
<td>p = 0.616</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein E</td>
<td>Lipid transport</td>
<td>(Scacchi, Gambina et al. 1999)</td>
<td>No change</td>
<td>No change</td>
<td>p = 0.545</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zhang, Barker et al. 2004)</td>
<td>↑ NEV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gelsolin</td>
<td>Actin-binding protein</td>
<td>(Cutler, Akuffo et al. 2008)</td>
<td>No change</td>
<td>No change</td>
<td>p = 0.656</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin D-binding protein</td>
<td>Vitamin D metabolite transport</td>
<td>(Liao, Yu et al. 2007)</td>
<td>↑ p = 0.0015</td>
<td>No change</td>
<td>p = 0.346</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter-alpha-trypsin inhibitor</td>
<td>Acute Phase Protein</td>
<td>(Matsubara, Amari et al. 1988)</td>
<td>No change</td>
<td>Increased*</td>
<td>p = 0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Liao, Yu et al. 2007)</td>
<td>↑ p = 0.0016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑ p = 0.0045</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑ p = 0.0030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
time and space. The composition of these proteins varies from cell to cell and tissue to tissue and is dependent on various cellular conditions such as pH. The primary structure is of course the specific amino acid sequence encoded by a specific gene, although on occasion there can be splice variants of a single gene and there are frequently somewhat variable numbers and types of post-translational modifications (PTMs). These factors account for the very large number of unique proteins observed compared to the number of their equivalent genes. These reasons mark both the potential and challenge in the consideration of proteins to be biomarkers for the diagnosis and staging of neurodegenerative diseases (Lista, Faltraco et al. 2013).

This comparatively novel approach to finding biomarkers has diverged from the more traditional approach of investigating only those molecules believed to play a role in the pathology of AD. Current proteomic approaches for biomarker discovery use an unbiased, global profiling of proteins expressed in bodily fluids to discover new biomarkers.

1.3.1.1 Studying Low Molecular Weight Serum Proteins Using Approaches That Involve High Abundant Protein Depletion

Protein depletion strategies reduce the dynamic range and complexity of serum proteins to enable detection of low abundance molecules. After precipitating high abundance serum proteins, success has been reported in the observation of approximately 10,000 (representing a possible majority of the blood proteome) different lower molecular weight, low abundance peptide disease markers, which are highly indicative of an underlying cell or tissue state (Adkins, Varnum et al. 2002, Tirumalai, Chan et al. 2003, Ray, Reddy et al. 2011). These may often be generated from proteolytic cleavage of larger molecules in blood (Davis, Auger et al. 2007).

Albumin, the most abundant large protein in blood routinely acts as a carrier protein to
transport physiologically crucial biomolecules like hormones, cytokines and lipoproteins (Tirumalai, Chan et al. 2003, Veenstra, Conrads et al. 2005). Hence, many of these low abundant, low molecular weight proteins in serum form biomarker-carrier complexes by non-covalent linkage with large proteins (Tirumalai, Chan et al. 2003, Chertov, Biragyn et al. 2004, Araujo, Petricoin et al. 2008). This process facilitates clearing of unbound low abundance smaller biomarkers in the bloodstream through the kidneys while larger carrier proteins and bound smaller peptides are generally not filtered or are much more gradually eliminated from the body (Araujo, Petricoin et al. 2008).

To avoid ion suppression by larger proteins in the mass spectrometer system, prior depletion of these large proteins is necessary. This helps study the low abundance and low mass molecular species in serum (Ray, Reddy et al. 2011). Now, since several of the small and low abundance proteins/peptides of interest may be in a complex with large carrier proteins, specific methodologies must be undertaken to curtail biomarker loss during protein depletion experiments.

Formerly, a variety of protein depletion methods have been used. Some of these include: 1) employing antibody-based affinity columns to selectively remove the most abundant serum proteins including albumin, haptoglobulins, transferrins, α-1-antitrypsin, α-1-acid glycoprotein, hemopexin, and α-2-macroglobulin (Tirumalai, Chan et al. 2003, Chertov, Biragyn et al. 2004, Veenstra, Conrads et al. 2005); and 2) use of solid phase extraction involving dye-ligand binding, antibody-based, metal chelating, and ion-exchange techniques (Ray, Reddy et al. 2011). Notwithstanding the commercially available protein depletion apparatuses, the important point is that an ideal depletion approach should eliminate highly abundant proteins as well as lead to the dissociation of low abundance, smaller peptides from carrier proteins (Tirumalai, Chan et al.
Centrifugal ultrafiltration has been commonly used in the area of proteomics in the past. It did not necessarily involve use of specific solvent conditions to unlink large protein-biomarker complexes. Besides, use of molecular weight cutoff filters restricts analysis to the low molecular weight proteome, lower than a specific molecular weight cut-off (Tirumalai, Chan et al. 2003). Hence, the risk remains of losing some important molecules that are larger than the cut-off filter and pore size (Merrell 2009, Alvarez 2013).

1.3.1.2 Rationale for Acetonitrile Precipitation

A traditional protein precipitation method uses an organic solvent, e.g., acetonitrile. It has been reported to effectively precipitate large molecular weight, highly abundant proteins such as serum albumin (Chertov, Biragyn et al. 2004). Research has shown that acetonitrile also disrupts the bond between small molecules and large proteins, thus allowing the peptides and small proteins to be visible during mass spectrometric analysis; it is greatly superior to other approaches (Merrell, Southwick et al. 2004, Alvarez 2013).

In previous studies done in my lab, experiments were conducted to study the effect of acetonitrile precipitation on serum proteins. Comparisons were made between untreated cLC-MS chromatograms and those treated with acetonitrile. The results demonstrated that the post-treatment total-ion chromatogram (TIC) had increased numbers of molecular species compared to pre-treatment ones. Panels A and B in Figure 1.1 show the difference clearly. Also the mass spectrum of the untreated serum protein (Panel C) shows an encompassing band that has peaks between 1300-1400 that signifies higher molecular weight species such as albumin or related compounds (60-70 kDa), thereby obscuring many molecular species. A larger number of species
can also be seen in regions of lower m/z (Panel D) that further confirms that organic solvent precipitation of serum proteins releases smaller, protein-bound species. Although the higher molecular weight species in the serum are lost by the acetonitrile treatment, Panel E shows that highly abundant albumin can also create the same effect, i.e., mask the larger proteins. Also, as seen in Panel C and D, acetonitrile treatment results in a larger number of smaller proteins of low molecular weight that are more observable. Also, considering the fact that the majority of growth, regulatory, stimulatory and pathology-responsive factors are small proteins, the probability of finding valid disease biomarkers increases after these treatments that reveal the smaller molecular species. In summary, acetonitrile pretreatment allows a wider spectrum of proteins and peptides in serum to be studied that in turn can increase the likelihood of discovering diagnostic patterns and informative biomarkers.

1.3.1.3 Uniqueness of This Approach

There was something regarding this method that I had not initially realized. In addition to the low molecular weight protein/peptide molecules, it also provides access to low molecular weight lipid molecules. These were mainly focused in the latter half of the useful chromatographic region. Many of these were glycerophospholipids and their oxidation products, because most other lipid classes do not easily become ionized without adding adducts to the processed specimens. These have proven to be reproducible and very important biomarkers during the course of my study. Since the approach involves lipids and specifically glycerophospholipids as well, the next section will focus on them and their usefulness as biomarkers.
Figure 1.1. Comparison between (1 μg) non-protein depleted serum samples and acetonitrile (ACN) treated protein depleted samples. A: total ion chromatograms (TIC) for serum proteins without protein depletion, B: TIC for acetonitrile precipitated sample, C: averaged mass spectra for the cLC elution taken from 24 to 25 min of a serum specimen prior to ACN precipitation, D: same as C but collected post ACN precipitation, and E: overlay of the mass spectra for species present between 25 and 26 min of elution for the two samples. Before ACN precipitation is shown in blue, after ACN precipitation is shown in red. Reprinted from (Merrell 2009) with permission. ©Karen Merrell, 2009.
1.3.2 Lipidomics

Lipidomics involves study of the molecular species of lipids and their distinguishing characteristics (Carrasco-Pancorbo, Navas-Iglesias et al. 2009). This field has attracted growing interest in the past few years. The interest has been fueled by the increased knowledge of how lipids interact with the body’s physiology and pathology, and progress made in the analytical instrumentation area that has enabled fast and extensive analysis.

1.3.2.1 Glycerophospholipids

The plasma membrane of every living cell is composed of phospholipids [also called glycerophospholipids (GPLs)], carbohydrates, proteins and cholesterol. The main element of the membrane, however, is GPLs. They are amphipathic in nature and have non-polar acyl chains attached to their polar head groups, which characterize the lipid subgroup or class. The polar head group has a phosphate group esterified to the sn-3 position of a glycerol backbone (Schiller, Süß et al. 2004). Phosphatidylcholine, phosphatidylethanolamines, phosphatidylserines, phosphatidylinositols and phosphatidylglycerols are the major head group classes (Wolf and Quinn 2008). Long-chain fatty acids, 14-24 carbons in length, acylate or alkylate the glycerol backbone at the sn-1 and sn-2 positions. Esterification of saturated or monounsaturated fatty acids is found in the sn-1 position of the glycerol backbone, and of polyunsaturated fatty acids (PUFAs) in the sn-2 position (Hermansson, Hokynar et al. 2011). The structural variety caused by the varying fatty acids and head groups leads to hundreds of lipid molecules (Figure 1.2).

Fahy et al. published a complete classification system for lipids, thereby establishing a general nomenclature (Fahy, Subramaniam et al. 2005). The classification of glycerophospholipids is done based on their polar head group and location of fatty acid residues.
A comprehensive classification system for lipids was published in 2005 by Fahy et al. to establish a universal nomenclature for lipids[142]. From this glycerophospholipids are recognised by their polar head group and type and location of their fatty acid residues. For instance, a choline containing head group will be termed a glycerophosphocholine, or simply phosphatidylcholine, if sn-1 and sn-2 contained palmitic acid (hexadecanoic acid) and arachidonic acid (5Z,8Z,11Z,14Z-eicosatetraenoic acid) respectively, would have a systemic name of 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine, commonly called PC(16:0/20:4(5Z,8Z,11Z,14Z)), which can then be further abbreviated to PC 36:4. The structure of this lipid is presented in figure 1.6, GPLs should always be drawn with the head group on the right[142].

Figure 1.2. Structures of the different classes of Glycerophospholipids. Reprinted from (Carter 2011) with permission. ©Claire Louise Carter, 2011.
Therefore, a choline at the head group will be termed a glycerophosphocholine. It would be termed a phosphatidylcholine, and if sn-1 and sn-2 contained palmitic acid (hexadecanoic acid) and arachidonic acid (5Z,8Z,11Z,14Z-eicosatrienoic acid), respectively; this lipid would have a systemic name of 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine, called PC(16:0/20:4(5Z,8Z,11Z,14Z)), which can then be further abbreviated to PC 36:4 (Figure 1.3).

GPLs are plentiful in cellular systems. Singer and Nicholson published the fluid mosaic model for cell membranes in 1972, where the membrane is a sandwich of phospholipids (Singer and Nicolson 1972). GLPs were initially assumed to be membrane barriers and sources of energy, but are now understood to interact with multiple physiological processes like cell signaling (Hermansson, Hokynar et al. 2011). GPLs hydrolyze the poly-unsaturated fatty acids (PUFAs), which are further metabolized into various biologically functional molecules. Figure 1.4 shown above displays the fluid mosaic model of the cell membranes as well as a summary of lipid involvement in physiological processes. Lipid prevalence has been found to be changed in very many pathological conditions like diabetes, Alzheimer’s, cancer, heart diseases, etc. (Kawata, Chitrannukroh et al. 1987, Williams, Coleman et al. 1991, Vered, Eugenio et al. 1993, Montine, Neely et al. 2002, Fathi, Isbel et al. 2004, Cunnane, Plourde et al. 2009, Florent-Béchard, Desbène et al. 2009, Gorjão, Azevedo-Martins et al. 2009, Huwiler and Pfeilschifter 2009, Beel, Sakakura et al. 2010, Han 2010, Reed 2011). Studies need to be done to understand the consistency of these changes as potential biomarkers for diseases (Carter 2011).
Figure 1.3. Structure of 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine, commonly known as PC 36:4. Reprinted from (Carter 2011) with permission. © Claire Louise Carter, 2011.
1.4 Mass Spectrometry Based Biomarker Discovery

1.4.1 General Workflow

Enormous advances in mass spectrometry (MS) techniques have made proteomics a potent tool to discover biomarkers. The initial steps in the process have generally included collecting samples and separating proteins by means of two dimensional gel electrophoresis (2-DE), liquid chromatography (LC) or protein-chip arrays. The next steps are to analyze the samples using MS or tandem MS and then making database searches to identity the protein (Craig-Schapiro, Fagan et al. 2009). This is summarized in Figure 1.5.

Figure 1.5 outlines the workflow for biomarker investigation in plasma/serum samples using global mass spectrometry. The first step involves having appropriate, carefully collected samples from well-documented patients and controls. These samples are then processed to remove proteins that are highly abundant or the proteins are separated into convenient size ranges using gel chromatography (Rami, Gomez-Anson et al. 2007), LC or high performance liquid chromatography (HPLC). The second step then is to identify proteins that are quantitatively different between the samples of the patients and controls using one or more mass spectrometry platforms or databases. Once a panel of potential biomarkers has been identified, the third step is to validate them, to define the ones that can be used for clinical diagnosis. Independent analytical approaches are preferable for validation studies such as antibody-based techniques like enzyme-linked immunosorbent assay (ELISA) or western blotting (Rappeport, Parkman et al. 1980) that are well known for their high sensitivity (Lista, Faltraco et al. 2013).
Figure 1.5. Workflow for biomarker investigation in plasma/serum samples using global mass spectrometry. Reprinted from (Lista, Faltraco et al. 2013) with permission. ©Elsevier, 2013.
1.4.2 My Approach

Specifically related to research on AD, diagnostic biomarkers are being researched in two different ways. The first way is the traditional way of trying to analyze biomarkers based on the understood neuropathology of AD. While this approach is intuitive, it always has some amount of bias to it since the discovery process is limited to assumptions based on the currently available data regarding AD. The second way uses the technological advances in MS by calculating differences in protein/peptide masses in the samples of patients and controls for all detectable species. The advantage of this approach is that it’s an impartial and independent method that doesn’t take any preconceived notions into consideration and, therefore, gives an opportunity to find novel proteins and other biomolecules that are not necessarily associated with the amyloid and tau proteins. The disadvantage is the number of false positives that may arise due to the natural range of concentration for a protein, typically small numbers of specimens assayed and sampling bias (Lista, Faltraco et al. 2013). I have used the second method, described here, to find novel biomarkers.

1.4.3 Electrospray Ionization-Quadrupole-Time-Of-Flight Mass Spectrometer: Most Frequently Used Instrument for Biomarker Discovery

Mass analyzers such as quadrupole, time of flight, ion trap, or fourier transform ion cyclotron resonance and recently orbitrap instruments are used for biological mass spectrometry based biomarker discovery research. Mass spectrometry has become the preferred tool for biomarker discovery due to its increased resolution, sensitivity, high mass accuracy, and rapid tandem MS abilities, allowing for the identification of many biomarkers of interest (Alvarez 2013). I used
electrospray ionization (ESI) interfaced with quadrupole–time-of-flight (QqTOF) instrument for my experiments. It is described in the following sections.

William Stephens of the University of Pennsylvania first put forward the idea of a time-of-flight mass spectrometer, which was constructed in the late 1940s. In the mid 1950s, Wolfgang Paul of the University of Bonn detailed the concept of the quadrupole mass spectrometer (Paul and Steinwedel 1953). The quadrupole mass filter and time-of-flight mass analyzer were later coupled and the first commercial QqTOF was launched in 1996. The technique of QqTOF was modeled closely after the electrospray ionization time-of-flight (ESI-TOF) system (Chernushevich, Loboda et al. 2001). The tandem instrument provides for ion fragmentation with the second detector measuring the daughter ions and allowing for amino acid sequencing. Electrospray ionization is a soft ionization technique and is required to interface liquid chromatography systems with a tandem mass spectrometer set-up that involves converting liquid phase samples into gas phase ions.

The normal components of QqTOF instrument are a continuous ion source, ion transmitter/mass filter, ion optics, an ion modulator, a drift region, an ion reflectron mirror, a detector and data recording electronics. The ion source can be ESI, atmospheric pressure ionization, etc. and consist of a filtered stream of ions having mass-to-charge ratios in a specific range to which the quadrupole is tuned (Chernushevich, Ens et al. 1999). The ion transmitter or mass filter in such an instrument is a quadrupole. The ion modulator, drift region and ion reflectron mirror generally constitute the TOF analyzer. A schematic diagram of a QTOF instrument is shown in Figure 1.6.

As can be seen, ions pass through three quadrupoles (Q0, Q1 and Q2). Q0 and Q2 are radiofrequency (rf) only quadrupoles that act as collision dampeners for the beam coming in
The popularity of the QqTOF has been significantly advanced by the rapid growth of semi-automated instrument control and data processing (also a major factor in the commercial success of the quadrupole ion trap), and by continuing improvements in the core performance characteristics of mass resolution and sensitivity. In addition, the recent development of a matrix-assisted laser desorption/ionization (MALDI) ion source for the QqTOF has provided new capabilities in MS and MS/MS, which expand the applications in biological research. The recent explosion in the field of protein research (including commercial proteomics) also appears to be contributing significantly to its commercial success. It therefore seems timely to provide a snapshot of the technology as it currently exists, describing the operating principles as they relate to important performance features such as sensitivity, mass resolution and mass accuracy, with illustrations from some of the major fields of application. We will try to outline, in as practical a fashion as possible, the important aspects of tuning and operation, the strengths and the limitations of the technique (perhaps debunking some common perceptions while supporting others), the advantages provided by the ion sources that are most widely used and the power of instrument control and data processing software. It should be noted that the last aspect is a moving target; it is likely that by the time this paper is read, new automation capabilities will have been developed and exploited, and far more capabilities will have been proposed for the future.

**Figure 1.6. Quadrupole time-of-flight mass spectrometer.** Reprinted from (Chernushevich, Loboda et al. 2001) with permission. ©John Wiley & Sons, Inc., 2001.
through the ion source. This is required to narrow the spatial and velocity spreads of the beam resulting from the velocity distribution of the ions in the direction of injection. The r.f. field in the quadrupole acts as a potential well, which allows the ion beam to be radially constricted. For single MS operation, Q1 acts in the r.f.-only mode, i.e., as a transmitter, whereas for the MS/MS mode, Q1 acts as a mass filter, allowing only ions within a specific range of m/z to pass through to the analyzer. Q1 then accelerates the ions towards Q2 where the ion beam collides with the neutral gas molecules, such as argon undergoing collision-induced dissociation (Faria, Gonçalves et al. 2014) to generate fragment ions. These ions are again focused through the collision dampening process described above into the TOF analyzer (Chernushevich, Ens et al. 1999, Chernushevich, Loboda et al. 2001).

The focused beam is now passed to the ion modulator inside the TOF analyzer perpendicular to the z axis. This is called orthogonal modulation. The theory behind orthogonal injection stems from the fact that it is hard to feed a continuous beam of ions to the TOF analyzer. A simple but ineffective way is to open the gate at regular intervals and let ions through; however this drastically affects the throughput. Since TOF analyzers can tolerate larger spatial spread of ions, the ions can be fed parallel to the electrodes rather than through the electrodes. This is the most efficient way of injecting ions into the TOF without affecting throughput. Once the ions are in the TOF analyzer, they are repelled and accelerated by the electrodes into the field-free drift region and eventually into the reflector where the ions become reflected on to the detector. The extra shield is required for the inverted configuration to generate the field-free drift region in the TOF section. The mass spectra are measured using a time-to-digital converter (TDC) (Chernushevich, Ens et al. 1999, Chernushevich, Loboda et al. 2001).
1.4.3.1 Electrospray Ionization

Analysis of large biological molecules in mass spectrometry requires conversion of the molecules into gas-phase ions, which was not possible to achieve with conventional methods of ionization such as electron ionization, photon ionization and chemical ionization without significant pyrolysis and decomposition. That explains the focus over the last 45 years on developing soft ionization methods to ionize large molecules. One category of such methods is based on the working theory that when adequate energy is unleashed rapidly over the molecules, it will bring about evaporation before decomposition (Beuhler, Flanigan et al. 1974). The source of energy can differ based on the methods used. For example, in plasma desorption (PD), the molecules are bombarded with a radioactive isotope, usually californium-252. Fast ion bombardment (FIB) makes use of a beam of high-energy ions. Laser desorption (LD) makes use of photons as the source of energy. These methods are highly effective in extracting intact molecular ions from large molecules (Fenn, Mann et al. 1989).

The other category of ionization methods involves use of powerful electrostatic fields to obtain ions from molecules. Field desorption (FD) is an example of such a method where the molecules are applied to a fine wire covered with an array of sharp pointed needles that generate the strong electric field. Electrohydrodynamic ionization (EH) is another example that involves dissolution of the sample in a non-volatile, low vapor pressure liquid (like glycerol) passed into a high voltage chamber through a capillary tube. These methods have some commercial constraints such as costly magnetic sector analyzers, tiresome sample preparation processes, unavailability of low vapor pressure liquids, etc. (Fenn, Mann et al. 1989).

A more recent category of spray ionization techniques also involves use of electric fields to extract ions. The difference is that ambient bath gas is used instead of vacuum. There are three
methods currently, which use this technique: aerospray ionization (ASI), thermospray ionization (TSI) and ESI.

I used ESI to be able to observe multiply charged peptides of MW of $\leq 10000$ D. In ESI, the sample is continuously passed through a stainless steel or a quartz silica hypodermic needle that is maintained at a high voltage (2.5-6 kV) relative to the chamber around it. Because of the electric field generated around the tip of the needle, the surface of the liquid becomes charged with the same polarity as that of the needle and becomes dispersed into a very fine spray of charged droplets by Coulombic forces that move towards the end wall of the chamber. A concurrent flow of a high flowrate bath gas (such as nitrogen) in the chamber maintained at high temperature expedites the evaporation of the solvent from the droplets, thereby decreasing their radii progressively. Consequently, the charge density of the surface keeps on increasing until it reaches the Rayleigh limit at which the Coulombic repulsion is equivalent to that of the surface tension that eventually explodes the droplet apart, producing even smaller droplets. This cyclic process keeps repeating until the radius of the droplet becomes small enough to allow the surface charge density to be strong enough to extract the ions from the droplet and pass them into the ambient gas (Fenn, Mann et al. 1989, Ho, Lam et al. 2003). This explains ion evaporation model. According to the charge residue model the electrospray droplets undergo evaporation and fission cycles, subsequently forming smaller droplets containing on an average one one analyte ion.

The emitted ions are swept along with the flow of the bath gas into the glass capillary tube and emerge out of it as a free flowing jet of ions. These are then sampled through a skimmer cone and then passed on to the entrance of a mass spectrometer, such as a quadrupole TOF. The entire setup is shown in Figure 1.7.
Figure 1.7. Electrospray ionization setup. Reprinted from (Fenn, Mann et al. 1989) with permission.

©American Association for the Advancement of Science, 1989.
1.4.3.2 Quadrupole as Mass Filter

The quadrupole mass analyzer filters specific mass ions by applying an AC and DC potential on the electrodes. The effect of the potentials on the ions is the basis of the mass filtering property of the analyzer. Let me first take the effect of the AC potential on the ions traversing the X-Z plane in the absence of the DC potential. In this scenario, the electrodes spend half the cycle at a positive potential and half the cycle at a negative potential. In the positive cycle, the ions are pushed towards the central axis of the electrode structure, whereas in the negative cycle, the ions are pulled towards the negatively charged electrodes (Figure 1.8). The collision of these ions with the electrodes is dependent on the negative cycle duration, the magnitude of the negative cycle and the position, speed and mass/charge (m/e) ratio of the ion.

Next, let’s consider the case where a positive DC potential is applied to the electrodes on the X-axis along with an alternating AC potential of high frequency. Because of the high frequency of the AC voltage, the only ions that will be affected by the AC and DC voltage will be ions with low mass. Ions with heavy masses won’t ‘feel’ the negative effect of the AC voltage to a large extent since the frequency of oscillation between cycles is high. So, the amount of time the ions will be affected by the negative cycle of the AC voltage will be very low. Thus, the heavier ions tend to be affected more by the positive DC potential that pushes them towards the central Z-axis. In summary, ions with heavy masses tend to pass through, in this case, making the quadrupole a high-pass filter. The graph below [Figure 1.9 (a)] shows this behavior.

The potentials on the Y-Z plane will be opposite compared to those on the X-Z plane. That’s why there will be a negative DC potential along the Y-Z plane, and the AC potential will be 180 degrees out of phase compared to the X-Z plane. If the frequency of oscillations of the AC voltage is high, the only ions that will stay near the central axis will be ions with low mass, since
Figure 1.8. Operation of a quadrupole in the X-Z and Y-Z plane. Reprinted from (Ginsbach and Dunnivant 2008) with permission. ©American Chemical Society, 2008.
they will be repelled by the positive cycle of the AC potential. All other ions with heavier mass would be attracted towards the negatively charged DC potential. Therefore, ions with low mass would be those that would pass through the quadrupole, making it function as a low-pass filter in this case as seen in Figure 1.9 (b).

Because of the alternating behaviors of ions on the X-Z and Y-Z planes, the quadrupole acts as a band-pass filter where only ions with a certain mass are allowed to pass through. Thus, adjusting the AC and DC potentials can vary this band of m/z and allow ions only within a specific range to pass by (Miller and Denton 1986). The graph Figure 1.9 (c) combines the two graphs [Figure 1.9 (a) and 1.9 (b)] to show the band of masses that are allowed to pass.

1.4.3.3 Time-of-Flight Mass Spectrometer

A time-of-flight mass spectrometer consists of an electrode at a certain voltage and a grounded electrode separated by a certain distance. The ions, which come into the instrument get repelled by the higher potential electrode towards the grounded electrode thereby gaining velocity. The grounded electrode is basically a mesh, which is used to pass these ions through in a drift free region where they travel with constant velocity till they reach the detector. Based on the equations in the Appendix, the time required for an ion to reach the detector is directly proportional to the square root of its mass. This leads to ions of the same mass reaching the detector at the same time thereby differentiating them in the mass spectrum. This is the fundamental working principle of the time of flight mass spectrometer.

The resolution of the mass spectrometer decreases with increase in the mass of ions because of the square root factor. Since ions are not created at a unique point between the electrodes because of the width of the electron beam creating them, for any charge to mass ratio, there will
Figure 1.9. Quadrupole as a band pass filter. Reprinted from (Ginsbach and Dunnivant 2008) with permission. ©American Chemical Society, 2008.

a) The high pass mass filter in the X-Z plane allows heavy ions to be transmitted through the quadrupole and reach the detector.

b) The low pass mass filter in the Y-Z plane allows light ions to be transmitted through the quadrupole and reach the detector.

c) The combined effect of the dc and oscillating ac potential results in an area stability for a specific mass to charge ratio.
be a range of flight times which will make the mass peaks overlap each other and blur the spectrum, especially for high masses. Currently, there are two techniques to deal with this problem. The first is space focusing. The concept for this goes back to the fact that ions created near the repeller electrode will receive larger acceleration than ions created farther away. By adding an extra electrode for additional acceleration of the ions and choosing the distances and voltages with precision, it can be assured that ions with the same mass-to-charge ratio will reach the detector at the same time, thereby improving the resolution (Figure 1.10).

The second technique is time lag focusing. Molecules move at different speeds in a gas according to the Maxwell-Boltzmann velocity distribution. This leads to ions having varying initial velocities between the time they are created and the time they receive the acceleration from the electrodes. This also leads to blurring of the flight times, which affects the resolution.

To compensate for that, a time lag can be introduced to nullify the spread of the initial velocities of the gas molecules. The current way of achieving this effect is the use of a reflectron. A reflectron is basically a panel of electrodes with varying voltages having a constant potential difference between them starting from 0 V (ground). When ions with the same masses but varying velocities are directed towards the reflectron, they are repelled. However, ions having higher velocities go deeper into the reflectron before being repelled compared to the ions having lower velocities. The net result of this is that ions entering the reflectron with varying velocities will come out of it with similar velocities and reach the detector at the same time, thereby improving the resolution and making the mass peaks more distinct (Gross 2004, Whitaker 2012). The detector is an electron multiplier that amplifies the charge falling on it and converts it into a voltage signal, which is fed to an oscilloscope to generate a mass spectrum.
Figure 1.10. TOFMS with two electrodes for space focusing. Reprinted from (Yildirim, Sise et al. 2010) with permission. ©Elsevier, 2010.
1.5 Layout of the Dissertation

Chapters 2, 4, and 5 are different parts of the initial serum AD biomarker discovery study. Chapter 3 is a validation study that focused on validating the few most promising biomarkers that looked diagnostically different from the study described in Chapter 2. Chapter 3 also lays the foundation for future research. Chapter 6 concludes the dissertation by discussing potential pitfalls and future research.

Chapter 2 details work on the initial discovery of novel serum AD biomarkers. Chapter 3 describes the results of a study designed to validate the diagnostic biomarkers that are discussed in Chapter 2, particularly those that were found repeatedly in multi-marker panels. This chapter also details identification of the biomarkers. Chapter 4 describes the findings of novel biomarkers when genders were studied separately. Chapter 5 compares different stages of AD and describes attempts to find novel biomarkers that might be predictive of different stages, specifically earlier stage AD. Detailed equations of quadrupole and time of flight mass spectrometer are included in the appendix.
2.1 Abstract

Background: Deaths from AD increased by 68% in the first decade of the 21st century, while deaths from all other major diseases declined. This emphasizes the continued need for useful treatments, which in turn necessitates accurate serum biomarkers for AD. To address this need, we used an innovative serum proteomic approach to examine the low-molecular weight components in serum to find prospective biomarkers.

Methods: Serum samples from 58 cases with any clinical stage of AD and 55 matched controls from non-demented individuals were processed to deplete high abundance proteins. The processed samples were fractionated on a reversed phase capillary column and analyzed using an ESI-QTOF mass spectrometer (MS). Approximately 8000 MS features were observed. Data for species that appeared quantitatively different between case and control were evaluated statistically. Furthermore, using logistic regression analysis and a leave-one-out approach, combinations of statistically different or near different biomarkers were evaluated to determine if they provided improved diagnostic capability.

Results: We found 38 biomarkers with significant differences between cases and controls. Multi-marker panels were built that resulted in specificities and sensitivities greater than 80%.

Conclusion: The serum proteomics approach was successfully used to find serum biomarkers for AD.
2.2 Introduction

The rate of new Alzheimer’s disease (AD) continues to grow as the average length of life of the population increases (Association 2013). Likewise, the dramatic increase in diabetes appears to have added to the number of individuals developing AD. Currently, there are no treatment options for AD. The absence of medications that attenuate AD reflects the current inability to identify unambiguously those who have the disease and the unavailability of serous or radiologic changes that identify very early AD when an intervention could slow or interrupt disease progression. Drug studies in sporadic AD are not currently possible due to the lack of good diagnostic tests for earlier stage AD.

Clearly, having accurate and easily applied biomarkers specific for AD, especially early AD, is critical for developing pharmacologic interventions in AD and would be useful in clinical management of AD thereafter. Not surprisingly, there have been many attempts to find diagnostic biomarkers for AD over the years. Out of these several studies, there are a few cerebrospinal fluid (CSF) biomarkers that appear to have the most potential. Reduced levels of amyloid beta protein 1-42 (Aβ-42) and increased levels of microtubule associated protein tau (t-tau and phosphorylated tau) in the cerebrospinal fluid are frequently found in severe cases of AD patients and are believed to indicate the presence of significantly progressed disease and potentially earlier, developing disease. However, CSF levels of these two peptides have been found to be almost the same in patients with mild cognitive impairment compared with controls, thereby making use of these biomarkers very limited for early diagnosis of AD. Even though improvements have been made in the recent past to increase the sensitivities of these assays, there is still the drawback of actually getting CSF samples from the very large affected or at risk population. Getting cerebrospinal fluid for a test is invasive, slow and expensive and, hence, not
practical for testing the very large, aging population. All of these reasons support the need for finding biomarkers in blood (Selkoe 2000, Sunderland, Linker et al. 2003, Blennow 2004, Hansson, Zetterberg et al. 2006, Hu, Hosseini et al. 2007, Ringman, Coppola et al. 2012).

There is currently some evidence that the blood-brain barrier of AD patients is damaged, allowing Aβ to cross the barrier and enter the peripheral blood circulation. This may lead to changes in the presence or quantity of the proteins, peptides and other molecules in the blood, eventually opening the door to finding biomarkers to diagnose AD (Deane and Zlokovic 2007, Zipser, Johanson et al. 2007, Lue, Walker et al. 2009). The preponderance of blood-based biomarker research has focused on for Aβ-42, tau, and P-tau. However, the research shows major inconsistencies when these CSF markers are studied in blood. Studies have found an unusual expression pattern of amyloid-β protein precursor isoforms found in AD platelets, and altered gene expression profiles in AD lymphocytes, but these have not resulted in functional biomarkers (Mueller, Zhou et al. 2010, Watt, Perez et al. 2011).

In contrast to this, a very different approach to finding biomarkers has been gaining popularity. In this strategy, all the proteins that show up in a mass spectrum are considered and evaluated to find biomolecules in plasma or serum whose quantities differ between cases and controls (Zürbig and Jahn 2012, Ghidoni, Paterlini et al. 2013, Lista, Faltraco et al. 2013). Specifically, the approach involves sampling serum from symptomatic AD candidates and non-demented controls and performing two dimensional gel electrophoresis (2DGE) on the samples for differential comparison (Zhang, Barker et al. 2004, Lopez, Mikulskis et al. 2005, Hye, Lynham et al. 2006, Liu, Hu et al. 2006, German, Gurnani et al. 2007, Liao, Yu et al. 2007, Akuffo, Davis et al. 2008, Cutler, Akuffo et al. 2008, Thambisetty, Simmons et al. 2010, Henkel, Müller et al. 2012). The molecules that appear different between the cases and controls are then
removed and digested using trypsin and evaluated by MALDI-MS to identify the putative parent proteins. Most of the studies implementing this approach follow the identification process above with a second verification step where immunoassays are developed to the relevant proteins and are used to quantitate the proteins as part of a second study. This proteomic approach has some strong points, but it comes with a number of drawbacks (Henkel, Müller et al. 2012).

First, the insensitivity of the electrophoretic gel methods makes it difficult to reproduce the same results on a consistent basis (Zhang, Annan et al. 2010, Henkel, Müller et al. 2012). The current 2D-GE methods only allow abundant proteins to be measured. A study by Hye et al. found 15 different protein peaks that were AD diagnosis candidates, out of which 13 were components of immunoglobulin or serum albumin precursor (Hye, Lynham et al. 2006). Second, multiple peaks in 2D gels may point to the same protein that is changed by different post-translational modifications such as glycosylation or glycation. The low resolution of 2D gels makes it harder to comparatively and quantitatively differentiate one protein from the other. This is due to presence of multiple molecules in the same spot. Third, the non-specific stains (that often display increased intensity with an increase in the size of the protein) used in gel electrophoresis, like Coomassie blue, makes this method much less sensitive to smaller proteins and peptides that could include signaling and regulatory proteins. Fourth, posttranslational modifications could result in one pool of proteins having the same expression levels as another pool even when they appear different in quantity and vice versa. This is when two pools represent the same protein with different modifications. Finally, several of these studies use a single pooled specimen for cases and a single pooled specimen for controls. The ability to evaluate the uniformity of the observed change is often lost when the specimens are pooled together for experiments.
The research of blood biomarkers of AD using electrophoretic gel methods (2D-GE) until now has yielded ambiguous results with inconsistent markers (Lopez, Mikulskis et al. 2005, Hye, Lynham et al. 2006, Liu, Hu et al. 2006, German, Gurnani et al. 2007, Liao, Yu et al. 2007, Akuffo, Davis et al. 2008, Cutler, Akuffo et al. 2008, Thambisetty, Simmons et al. 2010, Henkel, Müller et al. 2012). Furthermore, with few exceptions, immunoassay studies on the biomarkers identified in the discovery phase have not found any quantitative difference between cases and controls in the validation phase. These results should not be interpreted as an absence of blood AD biomarkers, but rather a reflection of the insensitivity, non-specificity and poor reproducibility of 2D-GE protein observation and quantitation methods. More thorough reviews of current blood proteomic studies have been published (Zürbig and Jahn 2012, Ghidoni, Paterlini et al. 2013, Lista, Faltraco et al. 2013).

Alternative techniques have been tried to find new biomarkers for AD. Ray et al. used antibody arrays to evaluate plasma levels of 120 cell-signaling proteins in cases versus controls and found a combination of 18 proteins which diagnosed AD with 90% accuracy (Ray, Britschgi et al. 2007). Another approach was implemented by Reddy et al. who investigated a combinatorial library of immunoglobulins against serum to discover IgG molecules unique to AD (Reddy, Wilson et al. 2011). Some studies have focused on higher plasma levels of oxidized proteins or glycoproteins (Choi, Malakowsky et al. 2002, Yu, Chertkow et al. 2003), while other studies have not been able to reproduce the same results (Lista, Faltraco et al. 2013).

Newer proteomic methods have not been used to discover AD serum biomarkers. MUDPIT, a shotgun or global proteomics approach, is regularly used to investigate the set of proteins in a tissue or cell (Vaudel, Sickmann et al. 2012). However, present applications of MUDPIT don’t allow quantitative comparisons. My study makes use of an approach that focuses on the low
molecular weight components in serum by initially depleting high abundance proteins and then performing capillary liquid chromatography coupled to tandem mass spectrometry to study molecules in serum (Merrell, Southwick et al. 2004, Merrell 2009, Esplin, Merrell et al. 2011). This approach interrogates around 8000 spectral features (Alvarez, Shah et al. 2013) and has been applied successfully to proteomic biomarker discovery in other clinical indications (Esplin, Merrell et al. 2011).

I propose that this proteomics approach will identify biomarkers for AD in serum. This method, combined with biostatistical analysis should provide panels of serum biomarkers that effectively diagnosis AD.

2.3 Materials and Methods

2.3.1 Sample Collection and Study Population

Serum specimens were procured from the Knight Alzheimer’s Disease Research Center (ADRC) at the Washington University School of Medicine, St. Louis, MO (WUSTL). Formerly collected serum specimens from cases (n=58) and controls (n=55) were provided. The cases represented AD patients with different clinical dementia ratings (CDR), particularly: mild cognitive impairment or very mild AD (CDR= 0.5, n=7), mild AD (CDR= 1.0, n=4), moderate AD (CDR= 2.0 n=19) and severe AD (CDR= 3.0 n=28). The 58 cases and 55 controls did not have co-morbidities, including diabetes. All specimens were stored at −80 °C prior to processing and analysis.
2.3.2 Ethics Statement

Approval from Washington University Institutional Review Board (IRB) was in place at Knight ADRC, WUSTL, prior to the original studies being conducted, and IRB approval was likewise obtained at Washington University School of Medicine, Knight ADRC prior to the current specimens being provided to BYU for these studies. All participants signed a written informed consent form when the samples were originally collected, which also provided permission to use the sera for later laboratory research. Specimens were provided without any personal identifiers. Furthermore, the laboratory studies at BYU involving these specimens had approval from the Institutional Review Board for Human Subjects at BYU.

2.3.3 Sample Processing

Serum contains many highly abundant proteins like albumin. Highly abundant proteins lead to ion suppression of low abundance biomolecules when analyzed by MS. Hence, in the novel approach used here, these larger proteins were removed using an acetonitrile precipitation protocol (Merrell, Thulin et al. 2008, Merrell 2009, Alvarez, Shah et al. 2013). This substantially increases the number of biomolecules observable by MS. It also displaces many small molecules, particularly peptides, from large carrier proteins. Sera to be studied were thawed on ice, mixed and 200 μL aliquoted. To this, 400 μL of HPLC grade acetonitrile was added immediately, vortexed for 20 s and allowed to sit at room temperature for 30 min. Post centrifugation (10,000 rpm), the supernatant (~550 μL) was transferred to a clean tube and 300 μL of HPLC grade water was added. The samples were then evaporated to ~200 μL (the original volume of the aliquot) in a vacuum centrifuge (Labconco CentriVap Concentrator, Labconco Corporation, Kansas City, MO). This step helps remove the final traces of acetonitrile to prevent inconsistencies during LC
separation. Apparent protein concentration was determined by Bio-Rad microliter plate protein assay performed according to the manufacturer's instructions (Bio-Rad Laboratories, Hercules, CA). An aliquot containing an apparent protein concentration of 2 μg was transferred to a new microcentrifuge tube and brought down to <10 μL in the vacuum centrifuge. Samples were brought up to 10 μL with HPLC water and acidified with addition of 10 μL 88% formic acid (FA).

2.3.4 Analysis of Sera on a Capillary-LC-ESI-QTOF Mass Spectrometry System

The protein-depleted, acidified sample was loaded onto a capillary liquid chromatography (cLC) system capable of gradient elution to separate compounds according to their polarities prior to the column being interfaced with a tandem mass spectrometer. I randomized the samples to keep bias to a minimum and ran sample sets in as short a time as possible. The pump used was an LC Packings Ultimate Capillary HPLC pump system, with a FAMOS autosampler (Dionex Corporation, Sunnyvale, CA) maintained at 4°C. The quaternary pump delivers two different solutions: an aqueous solution (98% HPLC grade H₂O, 2% acetonitrile (ACN) and 0.1% FA) and an organic solution (2% H₂O, 98% ACN and 0.1% FA) at a specified composition via a mixing device situated prior to the pump. This helps in achieving more consistent chromatographic separation. A 1 mm (16.2 μL) microbore guard column from Upchurch Scientific, Oak Harbor, WA, and a 15 cm x 250 μm i.d. capillary column made in-house were utilized for chromatographic separation. While the capillary column was slurry packed; the guard column was dry-packed. Both columns used POROS R1 reversed-phase media from Applied Biosystems, Framingham, MA. Sample was introduced and retained on the guard column and then eluted from the capillary column using the following gradient: 3 min of 95% aqueous and
5% organic phase, followed by a linear rise in organic phase to 60% over the next 24 min. Thereafter, the solvent mix was increased linearly to 95% organic phase and 5% aqueous phase over the next 7 min, then held at 95% organic phase for 7 min and returned to 95% aqueous phase over 5 min. The column was allowed to re-equilibrate until the end time of the run (58 min). The flow rate used is 5.0 µL/min.

The cLC system was interfaced via an IonSpray source from Applied Biosystems to a QSTAR Pulsar I quadrupole/orthogonal time-of-flight mass spectrometer set at 4800 V. MS data were collected for m/z range of 500 to 2500 starting at 5 min and ending at 55 min of gradient elution. The positive ion mode was used with a scan rate of 1 spectrum/s. The Analyst QS® software package permitted effective data collection, preliminary formatting, MS spectral comparison, extraction of selected-ion chromatograms and analysis.

2.3.5 Time Normalization of the MS Data

There is a slight inherent variability in chromatographic elution times from one sample to another. A set of 10 endogenous molecular species, consistently present in all specimens, was selected as internal standards. The compounds in the series elute at approximately two-minute intervals from each another. This allowed for the same time windows to be considered for each specimen run. Data were organized in two-minute windows centroided on each marker or where there were gaps between two time markers. The time markers allowed for time normalization, providing uniformity in chromatographic elution windows over the important chromatographic region (~15 to ~35 min) as described in detail by Merrell et al. (Merrell, Thulin et al. 2008).
2.3.6 Selecting Candidate Biomarkers

MS spectra were compiled from 1 min before to 1 min after each endogenous time marker for each sample to generate a composite spectrum. In turn, spectra from AD cases and normal controls were color coded and overlaid to allow for direct analysis of each individual molecular species. Peaks that appeared quantitatively different between cases and controls were then quantitated by extracting the ion counts of the desired peak, specifying an m/z range to capture the peak envelope using the “XIC” function of the MS instrument software for each of the samples. The peak heights were then recorded for all samples in an Excel file for each of the potential AD biomarkers (Merrell, Thulin et al. 2008). The extracted data were then tested using the Student’s t-test to determine if they were statistically different. A p-value < 0.05 was considered statistically different.

2.3.7 Normalization of the Candidate Biomarkers

Non-biological variability can arise from several sources as part of proteomic studies. These include variability in specimen processing, specimen stability, ionization efficiency, and instrument sensitivity among others. To reduce variability, a previously tested approach was used. To accomplish this, a set of naturally present serum molecules was selected as internal references. The ideal reference was present in the same time window, had a similar m/z value, was consistently there, was well separated from other peaks, was easily measured and was quantitatively comparable between cases and controls. Each of these was then used to normalize a potential biomarker to compensate for the non-biologic variability between samples. The peak intensity of a candidate biomarker species was ratioed to the intensity of a co-eluting species. Normalization with the use of this endogenous control typically improved the statistical
difference between cases and controls for candidate peaks. Unnormalized raw and log
normalized values were calculated for each of the potential biomarker candidates and retested
statistically. Again, a p-value <0.05 was used to define candidate serum biomarkers.

2.3.8 Biostatistical Analysis

In addition to the Student’s t-test that was applied to the initial data screen, other statistical
tests were carried out. Logistic regression analysis was used to generate receiver operator
characteristic curves for each of the candidate biomarkers to find their sensitivities and
specificities. Sensitivity is defined as the fraction of cases correctly identified by a given
biomarker. Specificity reflects the fraction of controls correctly defined by a given biomarker.
Then, sets of candidate markers were modeled to obtain a combination that provided better
diagnostic discrimination. A p-value <0.05 was considered significant but all biomarker
candidates with p-values less than 0.10 were considered since there is likelihood that some of
these may be complementary to other candidate biomarkers even if they themselves do not
satisfy criteria for statistical significance. Logistic regression analysis was also used to evaluate
combinations of biomarkers to model the log-odds of developing AD. Correlation analysis
helped in studying the biomarker clusters. The biomarkers in the same subgroup within the
cluster provided similar discrimination between cases and controls, and may represent
components from the same parent compound or represent a common response or pathway, while
those that appeared to segregate provide complementary diagnostic information and may
represent differences in staging of disease or in the involvement of potential pathways. Finally, a
forward-selection statistical procedure was used. In this approach, the candidate markers were
added to the model one at a time, beginning with each individual marker. The inclusion of other
markers on a one by one basis was tested to see if the combination of the markers increased the area under the receiver operator characteristic curve (AUC, a measure of overall diagnostic utility) by at least 0.03. In this way it was possible to obtain combinations of biomarkers giving the highest possible AUCs. This defines the ‘leave-one out’ prediction approach. Several potentially useful panels of AD biomarkers were obtained this way. ROC curves were plotted once again to study the combined sensitivities and specificities of these multi-marker models.

2.4 Results

2.4.1 Evaluating Potential Batch Effects

Notwithstanding that a mixed set of samples representing cases and controls were run every day until they were all finished, there were batches of specimens run on different days. This can sometimes lead to bias. However, when a hierarchical clustering algorithm was applied to the data, no clusters or observable patterns associated with run date were observed.

2.4.2 Candidate Serum AD Diagnostic Biomarkers

When studying raw and un-normalized mass spectral data, we found 44 candidate serum AD biomarkers that were statistically significant in the all-stage AD set (CDR = 0.5, 1.0, 2.0, 3.0) compared with the matched, non-demented control set (CDR = 0). There were additionally 25 biomarkers that had p-values between 0.05 and 0.10. With log normalizing the data for each candidate to an endogenous molecular reference species (that was equivalently distributed between cases and controls), there remained 38 significantly different serum biomarkers and 21 candidates that were near significant (p = 0.05 to < 0.10). Many of these were significantly
Table 2.1. Biomarkers considered for biostatistical analysis with significant (< 0.05) or near significant (< 0.1) p values.

<table>
<thead>
<tr>
<th>Candidate number</th>
<th>Elution time window (min)</th>
<th>m/z of biomarker</th>
<th>Charge (Positive ion mode)</th>
<th>p value for ratio of potential biomarker to reference biomarker</th>
<th>p value for log of ratio of potential biomarker to reference biomarker</th>
<th>p value for unnormalized potential biomarker</th>
<th>p value for unnormalized ref. biomarker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8-15</td>
<td>921.4</td>
<td>1</td>
<td>0.0494</td>
<td>0.035</td>
<td>0.0214</td>
<td>0.4524</td>
</tr>
<tr>
<td>2</td>
<td>8-15</td>
<td>1091.4</td>
<td>1</td>
<td>0.0857</td>
<td>0.0192</td>
<td>0.0538</td>
<td>0.4472</td>
</tr>
<tr>
<td>3</td>
<td>8-15</td>
<td>752.3</td>
<td>2</td>
<td>0.1565</td>
<td>0.1189</td>
<td>0.0878</td>
<td>0.3806</td>
</tr>
<tr>
<td>4</td>
<td>11-15</td>
<td>531.3</td>
<td>1</td>
<td>0.0449</td>
<td>0.0414</td>
<td>0.0635</td>
<td>0.487</td>
</tr>
<tr>
<td>5</td>
<td>11-15</td>
<td>515.3</td>
<td>1</td>
<td>0.1249</td>
<td>0.0482</td>
<td>0.0052</td>
<td>0.487</td>
</tr>
<tr>
<td>6</td>
<td>11-15</td>
<td>1107.5</td>
<td>1</td>
<td>0.0731</td>
<td>0.0218</td>
<td>0.1535</td>
<td>0.487</td>
</tr>
<tr>
<td>7</td>
<td>11-15</td>
<td>561.2</td>
<td>1</td>
<td>0.0866</td>
<td>0.1615</td>
<td>0.2144</td>
<td>0.487</td>
</tr>
<tr>
<td>8</td>
<td>11-15</td>
<td>574.3</td>
<td>1</td>
<td>0.0778</td>
<td>0.3269</td>
<td>0.3238</td>
<td>0.487</td>
</tr>
<tr>
<td>9</td>
<td>11-15</td>
<td>597.8</td>
<td>2</td>
<td>0.386</td>
<td>0.3673</td>
<td>0.0703</td>
<td>0.487</td>
</tr>
<tr>
<td>10</td>
<td>11-15</td>
<td>602.3</td>
<td>1</td>
<td>0.0905</td>
<td>0.0787</td>
<td>0.1492</td>
<td>0.487</td>
</tr>
<tr>
<td>11</td>
<td>11-15</td>
<td>690.8</td>
<td>2</td>
<td>0.072</td>
<td>0.3121</td>
<td>0.3938</td>
<td>0.487</td>
</tr>
<tr>
<td>12</td>
<td>11-15</td>
<td>701.8</td>
<td>2</td>
<td>0.0865</td>
<td>0.0946</td>
<td>0.0684</td>
<td>0.487</td>
</tr>
<tr>
<td>13</td>
<td>11-15</td>
<td>708.3</td>
<td>1</td>
<td>0.0812</td>
<td>0.2063</td>
<td>0.0762</td>
<td>0.487</td>
</tr>
<tr>
<td>14</td>
<td>11-15</td>
<td>821.3</td>
<td>1</td>
<td>0.0606</td>
<td>0.045</td>
<td>0.1389</td>
<td>0.487</td>
</tr>
<tr>
<td>15</td>
<td>11-15</td>
<td>892.4</td>
<td>1</td>
<td>0.1487</td>
<td>0.217</td>
<td>0.0392</td>
<td>0.487</td>
</tr>
<tr>
<td>16</td>
<td>11-15</td>
<td>971.4</td>
<td>1</td>
<td>0.0836</td>
<td>0.0986</td>
<td>0.2517</td>
<td>0.487</td>
</tr>
<tr>
<td>17</td>
<td>11-15</td>
<td>1194.5</td>
<td>1</td>
<td>0.1015</td>
<td>0.0854</td>
<td>0.1503</td>
<td>0.487</td>
</tr>
<tr>
<td>18</td>
<td>11-17</td>
<td>1396.6</td>
<td>1</td>
<td>0.0506</td>
<td>0.0919</td>
<td>0.1442</td>
<td>0.2468</td>
</tr>
<tr>
<td>19</td>
<td>13-19</td>
<td>587.3</td>
<td>1</td>
<td>0.0918</td>
<td>0.014</td>
<td>0.0785</td>
<td>0.2468</td>
</tr>
<tr>
<td>20</td>
<td>13-19</td>
<td>989.3</td>
<td>4</td>
<td>0.0952</td>
<td>0.1129</td>
<td>0.094</td>
<td>0.4886</td>
</tr>
<tr>
<td>21</td>
<td>13-19</td>
<td>791.6</td>
<td>5</td>
<td>0.1144</td>
<td>0.1405</td>
<td>0.0837</td>
<td>0.4886</td>
</tr>
<tr>
<td>22</td>
<td>16-22</td>
<td>748.6</td>
<td>1</td>
<td>0.1808</td>
<td>0.1143</td>
<td>0.0309</td>
<td>0.3663</td>
</tr>
<tr>
<td>23</td>
<td>16-22</td>
<td>704.5</td>
<td>1</td>
<td>0.0831</td>
<td>0.04</td>
<td>0.0149</td>
<td>0.3663</td>
</tr>
<tr>
<td>24</td>
<td>16-22</td>
<td>778.6</td>
<td>1</td>
<td>0.0416</td>
<td>0.0214</td>
<td>0.0136</td>
<td>0.3663</td>
</tr>
<tr>
<td>25</td>
<td>16-22</td>
<td>583.3</td>
<td>1</td>
<td>0.0115</td>
<td>0.0379</td>
<td>0.0002</td>
<td>0.3663</td>
</tr>
<tr>
<td>26</td>
<td>16-22</td>
<td>660.5</td>
<td>1</td>
<td>0.1229</td>
<td>0.0533</td>
<td>0.0258</td>
<td>0.3663</td>
</tr>
</tbody>
</table>
different in both raw and normalized data sets. A summary of all 82 of these potential AD biomarkers is provided in Table 2.1. In addition, each candidate was submitted to receiver operator characteristic curve analysis. While some of the AD biomarkers were increased/up-regulated in quantity, others were decreased in subjects with AD.

2.4.3 Assessment of Diagnostic Similarities between Biomarkers

Correlation analyses of all biomarker pairs were applied to all of the biomarkers. This resulted in a heatmap, clustering dendogram and correlation matrix that identified markers that were closely correlated and, hence, provided similar diagnostic information, and those that showed little correlation and, hence, provided added and complementary diagnostic information.

2.4.4 Evaluation of the Biomarkers in Multi-marker Panels

There is always the potential that changes related to AD may differ from one patient to the next. This can be due to differences in staging, gender, type of pathology present, and potentially even its cause. Use of a single biomarker may fail to identify all individuals with a given disease. Because of this, there is often improved diagnostic ability achieved with appropriate multi-marker sets. Hence, the combinations of my serum biomarkers were tested to see if they would provide better diagnostic discrimination for AD. As discussed in the Methods section, a forward selection approach was used to construct multiple marker panels. This considered all of the candidates, recognizing that even biomarkers that were only near significant might add complementary diagnostic information on patients with AD. All of the normalized biomarkers that were involved in making multi-marker panels with very high sensitivities and specificities were also significantly or near significantly different in the un-normalized results. Because
normalized data suggested fewer candidate biomarkers than raw data, we considered this set more conservative; in our constructing multi-marker panels, we focused on the normalized data. Multi-biomarker panels were created by including statistically different or near different candidates one by one until a combination of the panel yielded the best AUC possible. As mentioned earlier, each model was considered final when no marker improved the AUC by more than 0.03. With the construction of these several models, one for each biomarker as the starting point, it was also possible to determine which of those candidate markers showed up in more than one model. Markers occurring in several panels were considered to be more robust and promising among the biomarkers. There were 13 biomarkers that appeared in 10% or more of all the multi-marker models (See Table 2.2).

Further studies focused on these 13 biomarkers as a subset. Once again, the forward selection process was done with complementary correlation analysis, which included dendritic analysis and heat maps. ROC curves were plotted to study the performance of panels of these biomarkers in identifying AD. The best combinations of the 13 biomarkers provided AUCs of 0.908 and 0.912 (See Figure 2.1 and 2.2). The best combinations provided sensitivities of 88% and specificities of 87%. There were 6 multi-marker models that had an AUC of more than 0.80, where 11 of the 13 best biomarkers repeated themselves.

2.4.5 Cluster Analysis

Cluster analysis was done on the subset of 13 biomarkers. Some of these candidates correlated very well with others in the cluster. This might indicate a common single parent compound or perhaps represent similarities in biological staging, processes or pathways. The dendritic analysis suggested that there might be 4 groups (See Figure 2.3).
Table 2.2. Set of 13 biomarkers that appeared in 10% or more of all the multi-marker models.

<table>
<thead>
<tr>
<th>Candidate number</th>
<th>Elution time window (min)</th>
<th>m/z of biomarker</th>
<th>Charge (Positive ion mode)</th>
<th>p value for ratio of potential biomarker to reference biomarker</th>
<th>p value for log of ratio of potential biomarker to reference biomarker</th>
<th>p value for unnormalized potential biomarker</th>
<th>p value for unnormalized reference biomarker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11-15</td>
<td>531.3</td>
<td>1</td>
<td>0.0449</td>
<td>0.0414</td>
<td>0.0635</td>
<td>0.4870</td>
</tr>
<tr>
<td>2</td>
<td>11-15</td>
<td>602.3</td>
<td>1</td>
<td>0.0905</td>
<td>0.0787</td>
<td>0.1492</td>
<td>0.4870</td>
</tr>
<tr>
<td>3</td>
<td>11-15</td>
<td>701.8</td>
<td>2</td>
<td>0.0865</td>
<td>0.0946</td>
<td>0.0684</td>
<td>0.4870</td>
</tr>
<tr>
<td>4</td>
<td>11-15</td>
<td>708.3</td>
<td>1</td>
<td>0.0812</td>
<td>0.2063</td>
<td>0.0762</td>
<td>0.4870</td>
</tr>
<tr>
<td>5</td>
<td>11-15</td>
<td>892.4</td>
<td>1</td>
<td>0.1487</td>
<td>0.2170</td>
<td>0.0392</td>
<td>0.4870</td>
</tr>
<tr>
<td>6</td>
<td>13-19</td>
<td>989.3</td>
<td>4</td>
<td>0.0952</td>
<td>0.1129</td>
<td>0.0940</td>
<td>0.4886</td>
</tr>
<tr>
<td>7</td>
<td>16-22</td>
<td>583.3</td>
<td>1</td>
<td>0.0115</td>
<td>0.0379</td>
<td>0.0002</td>
<td>0.3663</td>
</tr>
<tr>
<td>8</td>
<td>17-25</td>
<td>804.6</td>
<td>1</td>
<td>0.0550</td>
<td>0.0188</td>
<td>0.0555</td>
<td>0.1461</td>
</tr>
<tr>
<td>9</td>
<td>22-32</td>
<td>660.4</td>
<td>1</td>
<td>0.0186</td>
<td>0.0113</td>
<td>0.0579</td>
<td>0.0984</td>
</tr>
<tr>
<td>10</td>
<td>29-37</td>
<td>874.6</td>
<td>1</td>
<td>0.1235</td>
<td>0.0044</td>
<td>0.0133</td>
<td>0.4480</td>
</tr>
<tr>
<td>11</td>
<td>31-38</td>
<td>1618.2</td>
<td>1</td>
<td>0.0086</td>
<td>0.0119</td>
<td>0.0161</td>
<td>0.1961</td>
</tr>
<tr>
<td>12</td>
<td>31-38</td>
<td>810.6</td>
<td>1</td>
<td>0.0044</td>
<td>0.0049</td>
<td>0.0022</td>
<td>0.1961</td>
</tr>
<tr>
<td>13</td>
<td>31-38</td>
<td>1568.2</td>
<td>1</td>
<td>0.0018</td>
<td>0.0020</td>
<td>0.0014</td>
<td>0.1961</td>
</tr>
</tbody>
</table>
Figure 2.1. Multimarker model (ROC curve) for biomarkers 531.3, 1568.2, 804.6, 602.3 and 708.3.
Figure 2.2. Multimarker model (ROC curve) for biomarkers 1618.2, 804.6, 531.3, 602.3 and 708.3.
Figure 2.3. Biomarker cluster analysis of 13 AD biomarkers subset.
2.5 Discussion

My goal was to find biomarkers for AD using a novel serum proteomics approach developed in my lab. This approach showed success in previous studies done on preterm birth, preeclampsia and endometriosis (Merrell 2009, Esplin, Merrell et al. 2011, Alvarez 2013). My approach enabled me to evaluate small proteins, peptides and other bio-molecular species that might show promise in diagnosing AD. My results suggested that no one molecule was sufficient to diagnose AD accurately. However, a number of molecules showed significant differences between cases and controls, and when selected candidates were combined, the best combination was able to diagnose AD with 88% sensitivity and 87% specificity. Hence, the approach worked and provided evidence that a serum biochemical diagnosis of AD is possible for most AD patients.

Studies conducted to find diagnostic biomarkers are dependent on the quality and number of samples available for analysis. I obtained the specimens used in these studies from the Knight Alzheimer’s Disease Research Center at Washington University School of Medicine. They are known to have stringent criteria for AD diagnosis and staging. The Knight ADRC ensured that cases and controls were age and gender matched along with the fraction of subjects with the APOE e4 allele (a genetic risk factor for AD). Care was taken that patients having co-morbidities such as diabetes were not included in the study. In this way, biomarkers that overlap with diseases commonly seen in AD patients (for instance diabetes) can be avoided. It also reduces the likelihood of discovering biomarkers for general brain disorders. Hence, the biomarkers I found appear to be specific to AD and not represent any other neurological disease. There is still the possibility that the changes observed here might occur in other neurologic conditions. Current understanding suggests that many neurologic diseases share some common features, e.g.,
increased levels of reactive oxygen species and inflammatory states. Clearly, verification studies of the biomarkers reported here would be required.

In my approach, I deliberately excluded highly abundant, larger molecular weight proteins that mask smaller and less abundant species. I recognize that potentially informative species may have been lost in this approach. Nevertheless, this approach allows for the survey of >8000 low molecular weight species, and although limited to this subset of serum molecules, it appears to provide many candidate biomarkers, which may lead to improvements in diagnosis of AD. This same serum proteomic method has also been previously successful in defining candidate biomarkers for medical indications, including preterm birth.

Both unnormalized (Abdullah, Paris et al. 2007) and normalized data were used for initial biostatistical analyses to generate as many candidates as possible. I found more potential markers using the raw data. This was important to rule out bias that may have arisen from the choice of corresponding reference markers. It can be difficult to find an ideal endogenous reference species that is close in m/z value to that of the candidate biomarker and whose quantitative values among cases and controls look equally distributed. The reference markers were not statistically different between cases and controls, and seemed to be uniformly distributed. Normalization removes some of the variability found from run to run. It is noted that all of the potential biomarkers found after normalization were represented in the set of unnormalized biomarker candidates. As mentioned, normalization reduced the number of candidates and may represent a more conservative approach; only those candidates that remained after normalization were considered in further biostatistical analyses.

Panels of biomarkers were developed using a ‘leave one out’ model, a forward selection approach. This provided several combinations of markers that had sensitivities and specificities
above those of single biomarkers. Most of the further analyses were restricted to the biomarkers that appeared in at least 10% of all the models. Note that this restriction eliminated candidates found only once in one model, making it less likely to have continued utility. All these considerations helped reduce the number of candidate markers from 44 significant and 25 near significant (found using raw MS data) to 38 significant and 21 near significant (found using normalized data), and of those, to 13 that appeared in at least 10% or more of all multi-biomarker panels. Using just the 13 candidate markers, I still had 6 panels with acceptable AUCs of 0.80 or better; of these, two had AUCs above 0.90, both of which provided ~85% sensitivity with ~85% specificity.

Cluster analysis suggested 4 different groups of related biomarkers. This could indicate: 1) the clustered biomarkers are fragments of a single parent compound; 2) different clusters may represent clinical subgroup differences (e.g., gender or disease stage); or 3) different clusters may possibly reflect different biological processes responsible for, or part of, AD.

In conclusion, the serum proteomic approach used here was able to find novel candidate biomarkers for AD. Some of these markers appear to be useful diagnostically when used independently. Combinations of candidate biomarkers allowed a larger portion of AD patients to be diagnosed. The better panels provided sensitivities and specificities greater than 80%. This far exceeds any current proposed AD marker or set of markers. Thus, I conclude from these data that the serum proteomics approach may be a useful adjunct in the serum diagnosis of AD and also in the study of the pathology of AD.
CHAPTER 3 - VALIDATION AND CHARACTERIZATION OF ALZHEIMER'S DISEASE BIOMARKERS

3.1 Abstract

Background: In all, 82 candidate biomarkers with significant and near significant p values, found in the first study, were further statistically analyzed to develop diagnostic multi-marker panels. Candidates seen in multiple panels were considered to be more promising. We found that 13 biomarkers appeared in 10% or more of all the multi-marker models. Multi-marker combinations limited to these biomarkers yielded specificities and sensitivities greater than 80%. A validation study of these 13 biomarkers was considered critical to determine if any of the biomarkers continued to hold promise. Further, if candidates were to be useful, chemical identification of specific biomarkers would be required. Hence, we have sought in this study to validate these 13 serum AD biomarkers in a blinded replication study analyzing additional patients and controls, and to find chemical identities of biomarkers with continued utility.

Methods: A blinded validation set consisting of 125 previously unstudied serum samples including different stages of AD (n = 68) and a similar number of non-demented matched controls (n = 57) were obtained, processed and analyzed by capillary liquid chromatography-tandem mass spectrometry as done in the initial study. The mass spectral data for the 13 AD biomarkers occurring in several panels as part of the initial study were compiled. Statistical analysis was performed on individual markers. Combinations of the validated biomarkers were statistically considered by logistic regression analysis. The validated biomarkers were further characterized to find their chemical identities using tandem mass spectrometry.
Results: We found 4 significant biomarkers that continued to be statistically different in the validation study. Combinations of these were more diagnostic than single biomarker performance. Also, there were enough very early stage AD patients (CDR = 0.5, n = 24) to consider them independently. Again, the same 4 biomarkers were significantly different between cases and controls. Additionally, we were also able to chemically identify these.

Conclusion: These data confirm that the serum proteomic approach has the ability to find novel serum biomarkers for AD, even in the earliest stage of AD.

3.2 Introduction

Every biomarker discovery study should be followed by a validation study to verify the usefulness of the candidate biomarkers. A valid biomarker is defined as “a biomarker that is measured in an analytical test system with well-established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the physiologic, toxicological, pharmacologic, or clinical significance of the test results” (Hunter, Losina et al. 2010). The validity of a biomarker is highly associated to how it can be implemented in diagnosis or understanding the disease pathology (Hunter, Losina et al. 2010).

Here, I proposed to use the serum proteomics approach to validate the novel, candidate serum biomarkers that appeared useful in diagnosing AD. I also focused on finding serum biomarkers that could aid in very early AD diagnosis. Further, chemical characterization of useful serum biomarkers was sought as part of these studies. Serum AD biomarkers diagnosing early stage AD should allow for evaluation of promising AD drugs much more efficiently and may be useful in personalized drug therapies.
3.3 Materials and Methods

3.3.1 Procurement and Storage of Samples

A validation study of the previously found biomarkers was proposed to the Knight ADRC at Washington University School of Medicine, and permission was granted for my procurement of a new set of serum samples. A set of 125 samples was provided. These comprised 68 cases constituting different stages of AD and 57 controls that were non-demented and matched for age, gender and ApoE 4 prevalence. Among 68 cases, I had 24 samples with CDR = 0.5. This made it possible for me to test the performance of my biomarkers between non-demented controls (CDR = 0) and individuals with mild cognitive impairment (CDR = 0.5). This comparison was not possible in the first study due to the very small number of samples with CDR = 0.5 in that set. An intermediary, aware of case/control status, created analysis sets that included both cases and controls randomized and submitted to my laboratory analysis in a blinded fashion. A total of 11 sets of samples, consisting each of 10-13 sera, were analyzed. The sets had approximately the same number of cases and controls as well as the same number of men and women to avoid bias due to sample processing, day to day changes in instrumental sensitivities, loading and ionization inefficiencies. All MS data were provided to the intermediary for statistical analysis independent of the laboratory investigators having any knowledge of case/control status. The samples were shipped on dry ice and stored at −80°C pre and post processing.

3.3.2 Validation of Biomarkers

Only one analysis set was processed at a time. Specimens were processed, subsequently loaded, and analyzed using as Agilent 6530 Accurate-Mass Q-TOF LC/MS mass spectrometer
system as part of a single day. All sets were analyzed in as short a time as possible to minimize chromatographic and instrumental variation. A 45-min wash was included between each sample to minimize any carry over effect while a more significant wash of 105 min was added at the end of every day. Great care was taken to maintain consistency in sample processing, quantitative loading, chromatographic elution profile and instrument calibration/sensitivity performance across all specimens and all days. Samples were processed using acetonitrile precipitation as described in Chapter 2 under the Methods section.

An Agilent 1260 Infinity Series HPLC system equipped with quaternary pump, online degasser, auto sampler, and thermostated column compartment (Agilent Technologies, Karlsruhe, Germany) was used to perform these validation experiments. Injection volumes of 5 µL, containing an apparent protein concentration of 5 µg were injected onto a 1 mm (16.2 µL) microbore guard column from Upchurch Scientific, Oak Harbor, WA, to retain specimen and eliminate salts and highly polar compounds. The guard column was then placed in series and sample loaded onto and fractionated by the analytical column (15 cm x 250 µm i.d. POROS R1 reversed-phase capillary column; media was obtained from Applied Biosystems, Framingham, MA, and columns were packed in-house) with a mobile phase flow rate of 5 µL/min. Mobile phase A was an aqueous solution water/acetonitrile/formic acid (98:2:0.1, v/v/v) and mobile phase B was primarily organic solution containing acetonitrile/water/formic acid (98:2:0.1, v/v/v). The gradient elution began at 100% A for 2 min followed by a 1 min transition from 100% A to 95% A. This in turn was succeeded by a linear increase in organic phase (B) to 60% over the next 24 min. Thereafter, the mobile phase mix was increased linearly to 95% organic phase B/5% aqueous phase A over the next 7 min, then held at 95% organic phase B for 7 min and returned to 95% aqueous phase A over 5 min. The column was allowed to re-equilibrate
until the end of the 58 min run.

In the cLC-MS system, the cLC was interfaced with an Agilent 6530 accurate-mass quadrupole time-of-flight mass spectrometer (QToF MS) equipped with an Agilent Dual ESI source. The Dual ESI source was operated in the positive ion mode, and instrument parameters were set as follows: sheath gas temperature, 300°C; sheath gas flow, 5 L/min; nebulizer, 15 psi; dry gas temperature, 300°C; dry gas flow, 5 L/min; and capillary entrance voltage (VCap), 3500 V. Fragmentor and Skimmer 1 were operated at 175 V and 65 V, respectively. The auto sampler temperature was maintained at 4 °C. The MS scan data were collected at a rate of 8 spectra/s over the range of m/z 400–3200. Data were acquired in profile mode. The instrument software MassHunter Data Acquisition B.05.01 (Agilent) and MassHunter Qualitative Analysis B.06.00 (Agilent) permitted effective data collection, extracting specific ion chromatograms and post hoc analysis.

3.3.3 Fragmentation and Chemical Identification Studies

Fragmentation and chemical identification of biomarkers were accomplished using an MS/MS approach with a QSTAR Pulsar I quadrupole orthogonal time-of-flight mass spectrometer and Agilent 6530 accurate-mass Q-TOF LC/MS. Primarily peptides were sequenced using a QSTAR Pulsar I quadrupole orthogonal time-of-flight mass spectrometer while lipid identifications were performed using both instruments.

3.3.3.1 Peptide Fragmentation Experiments Employing Applied Biosystem’s QSTAR Pulsar I Quadrupole Orthogonal Time-of-Flight Mass Spectrometer

The cLC system was interfaced to a QSTAR Pulsar I quadrupole orthogonal time-of-
flight mass spectrometer through an ESI IonSpray source (Applied Biosystems) set at 4800 V. All experiments were carried out in the positive ion mode and the instrument parameters were set as follows: sheath gas temperature, 25°C; nebulizer gas temperature, 25°C; nebulizer gas pressure, 20 psi; curtain gas pressure, 24 psi and electrospray voltage, 4800 V. The autosampler temperature was maintained at 4 °C. Either unit or low resolution was generally used. Samples run earlier were examined to find samples that naturally had a high concentration of the biomarkers of interest. Samples were processed using acetonitrile precipitation and loaded as described in Chapter 2. Prior to MS² runs, MS¹ runs were undertaken to determine the elution time. For MS² runs, an increased amount of protein-depleted serum was loaded to enable detection of the lower intensity fragment peaks clearly. Samples were manually injected and MS/MS fragmentation data were collected from m/z 50 up to 2000 depending on the initial m/z value and charge state of the precursor ion. The total run time was 55 min with a mass spectral scan rate of 1 spectrum/s. The delay time feature was used and fragmentation was only done during the relevant 2 min elution window. This 2 min window was begun 1 min before and ended 1 min after the biomarker peak. Either argon (Q-Star) or nitrogen (Agilent) was used as the collision gas. Typically, 2-4 runs at different fixed collision energies were performed to obtain the broadest set of fragments for the peaks of interest. Different spectra from different runs were first overlaid and then added using the Add Data feature in the Analyst QS 1.1 instrument software, yielding a summed MS² spectrum.

### 3.3.3.2 Peptide Sequencing via the Mascot Search Engine

I manually analyzed each mass spectrum and exported the peak m/z values and intensities to an Excel sheet. Here, any incorrectly assigned charge states were corrected by comparing these
peaks to the fragmentation data in the mass spectrum. All the peaks with charge state more than 1 in the adjusted data list were converted into their +1 m/z values using the formula: +1 mass = m/z value x charge – (charge – 1 H+) to simplify the database search. This amended list of ions was pasted into a Mascot generic format file and submitted to Mascot (www.jp1.chem.byu.edu/mascot - departmental mascot subscription). The “.mgf” file was uploaded, and the parameters used for Mascot searches were: no enzyme digestion, Uni_human database, no modifications, peptide tolerance of 1.2 Da, and MS/MS tolerance of 0.6 Da. Protein Blast search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was typically used in an effort to identify the parent protein compound for the peptides.

3.3.3.3 De Novo Sequencing of Peptides

On several occasions, Mascot has not generated conclusive results. This is often due to incomplete b- or y-ion series. In such cases, de novo sequencing was performed, where the fragmentation spectrum is manually evaluated to assign amino acid identities by studying the mass differences between fragment ions. There are typically both b- and y-ions observed when collision-induced dissociation is performed. Interpretation can be complicated when both incomplete representation of b and y ions are produced. First, I attempt to identify the b1 and/or y1 ions to specify the starting point of the peptide sequence. Immonium ions found in the low m/z range (typically < 200) often give indications about the amino acids present in the peptide sequence. Beginning with b1 or y1, mass differences between neighboring or near neighbor peaks in the MS² spectra were studied. When a mass difference corresponded with the mass of an amino acid residue, I identified the next amino acid. Then that peak was used as a reference, and the process was repeated to assign the next amino acid in the series starting from the b2 or
y2-ion fragment peak. This continued until a sequence of all amino acids contained in a peptide was ascertained. Several different partial combinations were sometimes obtained during the process. Often, with multiple collision energies and careful assignment, a tentative sequence was obtained. This could be compared to known amino acid sequences using a BLAST search for all proteins in the human using the NIH database. If the amino acid sequence was incorrect, no matches were found. With small peptides, the amino acid sequence may not be unique to a single parent protein. Usually, when something looked irresolvable by this approach, I checked for possible amino acid modifications. Occasionally, when a sequence could not be determined completely by de novo sequencing, the parent protein could still be determined using the protein BLAST search knowing that there must be a shared partial sequence and that the peptide’s total mass must still be consistent with the mass of possible extended peptides from any protein that contains that partial sequence (Alvarez 2013).

3.3.3.4 Lipid Identifications Employing Applied Biosystem’s QSTAR Pulsar I Quadrupole Orthogonal Time-of-Flight Mass Spectrometer

Fragmentation studies for lipids and their identification were carried out using the same instrument parameters as described in peptide identification studies above. The primary difference was that 1 or 2 collision energies were sufficient for fragmenting lipids. Fragmentation remained quite consistent across different collision energies and typically showed a very high intensity peak representing the head group.
3.3.3.5 Lipid Fragmentation Experiments Employing the Agilent 6530 Accurate-Mass Q-TOF LC/MS System

The Dual ESI source was operated in the positive ion mode and the instrument parameters were set as follows: sheath gas temperature, 300°C; sheath gas flow, 5 L/min; nebulizer, 15 psi; dry gas temperature, 300°C; dry gas flow, 5 L/min; and capillary entrance voltage, 3500 V. Fragmentor and Skimmer 1 were operated at 150 V and 65 V, respectively. The autosampler temperature was maintained at 4°C. The MS scan data were collected at a rate of 1 spectrum/s over the range of m/z 110–2000. Data were acquired in both profile and centroid modes. A reference solution having ions at m/z 121.0509 and 922.0098 was used during these experiments. Reference nebulizer pressure was maintained at 5 psi while auto calibration reference parameters were: detection window, 20 ppm, and minimum height, 1000 counts. The reference standards allowed me to correct unknown masses and also allowed me to determine the day-to-day drift in mass accuracy to ensure that no instrument adjustment needed to be made. In the case of lipids, where exact mass really matters, this was an important tool. The software, MassHunter Data Acquisition B.06.00 (Agilent) and MassHunter Qualitative Analysis B.05.01 (Agilent), permitted direct on-line data collection, extracting specific ion chromatograms and analysis.

Data-dependent Auto MS was most commonly used for fragmentation studies. However, if peak intensities were low, representing too few MS² scans, then data independent targeted MS was also used. The MS and the MS/MS information were collected from m/z 50. The ending m/z value depended on the mass of the parent lipid molecule, ensuring that the range would include all fragments of the parent lipid molecule. Typically, collision energies ranging from 15-30 eV were used for fragmentation. Not more than 3 different collision energies were used per run. Sometimes, a single collision energy was sufficient to obtain good fragmentation, while other
times additional energies were needed. The isolation width of the precursor ion was set to be narrow ($\sim 1.3 \text{ m/z}$). A narrow width precluded the possibility of fragmenting neighboring peaks. Typically, MS scan rates and MS/MS scan rates were 1 spectrum/s, although in a few cases, scan rates of 2 spectra/s were used. With a lower scan rate, more of the transients hit the detector; hence, there was an increase in the gain, which helped in obtaining higher signal/noise ratio of the fragment ions. Data were collected in both centroid and profile modes.

### 3.3.3.6 Identification of Lipids

While fragmenting some of our candidate biomarkers, a prominent peak at m/z 184.07 was observed. This is indicative of a phosphocholine head group, a distinctive feature of one class of glycerophospholipids. With lipids, there were not as many fragments as typically seen in peptide fragmentation spectra. Increasing the collision energies did not mean more fragmentation, but mostly just an alteration of peak intensities for the fragments that were already seen.

To further characterize lipid biomarkers, a search was performed on Lipidmaps (lipidmaps.org), the primary database of known lipids. The search involves using the mass of the neutral precursor ion ($([\text{M}+\text{H}^+] - \text{H}^+)$. Fragmentation data are not available and fragmentation spectra cannot be compared with archived data, because no fragmentation patterns exist. Consequently, the database would typically provide several compounds having similar masses from very close in m/z to others more unlikely. We focused only on the ones closest in m/z to the molecule of interest. Second, I only considered those classes of lipids that contained the observed head group. For instance, if a phosphocholine head group was observed in the spectrum, only phosphocholine lipids were considered. This typically limited the candidates dramatically. Then, more careful analysis of the MS$^2$ fragments was conducted, looking for fragments representing fatty acids that modified the glycerol backbone. This was often successful but did not allow
assignment of fatty acids to the sn1 or sn2 positions on the glycerol. In cases of lipid oxidation, it appeared to be unsaturated fatty acids that were modified; however, in the MS\textsuperscript{2} fragmentation, the oxygen containing fatty acid did not remain intact. Occasionally identification was achieved from published studies where known lipids had been studied by MS\textsuperscript{2} and fragmentation patterns were unambiguously the same as for the lipid biomarker. This was infrequent as there are very few reports of this nature.

Even after narrowing the field of possibilities as just described, there are almost always multiple lipids that share the same mass. The location of fatty acid substituents on the glycerol backbone cannot be determined by mass spectrometry, nor can the location of double bonds be determined by MS. There are double bond locations that are far more naturally frequent among fatty acids, but this cannot be determined. Given the very low abundance of these biomarkers, no other instrumental method, e.g., NMR, is possible. LIPID MAPS MS does provide a software tool (http://www.lipidmaps.org/tools/index.html) that predicts the fragments based on parent mass and sn1 and sn2 acyl chain information. However, this does not allow for the determination of sn1 or sn2 location or position of double bonds within fatty acid components. Standards of possible compounds are not typically available. Researchers have previously reported data regarding the positions where the fragments can break from a glycerophosphotydylcholine molecule (Byrdwell 2005). This information was used to predict various different fragments possible for the choices I had. Some of these would be ideally seen in the MS\textsuperscript{2} spectrum for confirmation (Alvarez 2013).

Some of my biomarkers were oxidized lipid molecules, and certain fragments indicative of loss of oxygen were observed. In case of hydroxylated species, most of the fragment ions have odd number m/z values and fragmentation resulted in neutral water loss. For others, a neutral
loss of 16 would represent a keto or epoxide modification. However, these attempts don’t typically provide the exact location(s) of the modifications or whether the fatty acid occupied the sn1 or sn2 position. These oxidized lipids represent the effects of reactive oxygen species generated by oxidative stress in subjects. Oxidized lipid molecules are not found in the Lipidmaps database. Their fragmentation spectra are also more complicated compared to other lipids, and very few publications include MS² data.

3.3.4 Statistical Analysis Including Construction of Multi-marker Model Using Validated Biomarkers

Simple statistics (Student’s t-test) were used to evaluate the performance of the 13 candidate markers in this replication study. A p-value < 0.05 was considered significant. The biomarker also had to show the same trend comparing cases to controls. Different combinations of the statistically significant candidate biomarkers were further studied to find out if they provided better diagnostic discrimination. Logistic regression analysis was used to consider combinations providing a receiver operator characteristic curve (plot of sensitivity, i.e., true positive versus 1-specificity, i.e., true negative) from which could be calculated an area under the curve and a sensitivity at any given specificity.

3.4 Results

3.4.1 Validation of Previously Found Biomarkers that Showed Significant Diagnostic Discrimination

I focused initially on how well the 13 biomarkers that had performed most consistently in the
initial serum AD biomarker study performed in this validation study. Of these 13, the biomarkers m/z 708.3 and 660.4 were not clearly seen in most of the current samples. In the case of m/z 708.3, in the validation set, there was a large, unrelated peak that likely led to ion suppression. Moreover, there was also in these specimens an overlapping peptide envelope, which made calculation of the target molecule impossible. For peak m/z 660.4, the intensity was too low to determine peak height or area with confidence. Hence, these two could not be analyzed.

The extracted MS data obtained for the remaining biomarkers were sent to the independent mediator who conducted the analyses and maintained the blind. The data were tested in four ways: 1) non-demented, matched controls (CDR = 0) vs. any stage AD cases, i.e., a CDR 0.5 or greater; 2) matched, non-demented controls (CDR = 0) vs. CDR > 0.5 (this excludes the earliest stage AD patients); 3) matched, non-demented controls (CDR = 0) compared with patients having moderate AD (CDR = 2); and finally 4) matched, non-demented controls (CDR = 0) vs. the earliest stage AD only (CDR = 0.5).

Among the remaining 11 biomarkers, m/z 602.3, 804.55 and 874.6 were significantly different between cases and controls for all 4 comparisons and followed the same trend as observed in the original study.

As I reviewed data from the initial study, I found that there were 2 candidate biomarkers with the same nominal m/z of 804.5, but offset somewhat in elution time from the 804.55 that was validated. To ensure that there was no misinterpretation of the data, both species were included. The second species had an m/z of 804.53, but eluted later between 39.5-41.5 min (compound with m/z 804.55 eluted between 33-35 min). This species had been a significant biomarker in the initial data, but was not part of the subset of 13. The candidate at m/z 804.53 in the later hydrophobic region also passed the validation. Cases had significantly higher quantities
compared with controls in the validation set as in the preliminary study. (See Table 3.1 for all
results of the validation of these biomarkers). Among the validated biomarkers, all were higher
in cases. This applied when I did comparisons between different stages of AD for these
biomarkers as well.

As part of the chemical identification studies, it was noted that the candidate with m/z 810.6
representing a protonated species was also present as m/z 832.6 representing a sodiated species
[M+Na⁺]. The peak m/z 832.6 was also a candidate biomarker in the initial study and belonged
to the list of significant and near significant biomarkers. Hence, I extracted data for the peak at
m/z 832.6 as well and included it. Also, I summed the data for the two ionic species m/z 810.6
and 832.6 as they represented the same biomolecule. Individually and collectively they did not
validate.

In a somewhat similar situation, as fragmentation studies were conducted the candidate m/z
989.3 was found to represent a peptide in its +4 charge state. The same peptide also occurred in
its +3 charge state (m/z = 1318.7) and, hence, both ion species were extracted and results of both
charge states were considered individually and combined. However, whether considered
individually or in combination, this peptide was no longer quantitatively different between cases
and controls in the replication study.

3.4.2 Identification of the Validated Biomarkers

The 4 biomarkers that validated in the replication set became the focus of efforts to
chemically identify them. Broadly, this was successful. Preliminary work demonstrated that 1 of
these was a peptide and 3 were glycerophosphocholine lipids. The structural features that can be
obtained from MS studies are summarized for each of the 4 biomarkers as follows.
Table 3.1. Student’s t-test results showing p values obtained for the unnormalized validation data. Different stages are compared and p values for them are listed in the table as well. Trends indicating whether values were higher in cases or controls were also studied. Observation was made if they followed the same or opposite trend. Finally, in the last column, indication is given as to whether or not the biomarker passed the validation study.

<table>
<thead>
<tr>
<th>Number/Comments</th>
<th>Biomarker (m/z)</th>
<th>Z</th>
<th>Retention Time (min)</th>
<th>CDR 0 vs CDR &gt;0</th>
<th>CDR 0 vs CDR &gt;0.5</th>
<th>CDR 0 vs CDR 2</th>
<th>CDR 0 vs CDR 0.5</th>
<th>Trend</th>
<th>Validation Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>531.3</td>
<td>1</td>
<td>19-21</td>
<td>0.5669</td>
<td>0.3645</td>
<td>0.6203</td>
<td>0.1037</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>2</td>
<td>602.3</td>
<td>1</td>
<td>19.5-22.5</td>
<td>0.0014</td>
<td>0.0584</td>
<td>0.0465</td>
<td>8.41x10^-5</td>
<td>Same</td>
<td>Passed</td>
</tr>
<tr>
<td>3</td>
<td>701.8</td>
<td>2</td>
<td>19-21</td>
<td>0.3251</td>
<td>0.2837</td>
<td>0.8336</td>
<td>0.6766</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>4</td>
<td>892.4</td>
<td>1</td>
<td>18.5-20.5</td>
<td>0.5819</td>
<td>0.6021</td>
<td>0.5393</td>
<td>0.7325</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>5</td>
<td>989.3</td>
<td>4</td>
<td>21.5-23.5</td>
<td>0.0586</td>
<td>0.0691</td>
<td>0.0787</td>
<td>0.2012</td>
<td>Opposite</td>
<td>Failed</td>
</tr>
<tr>
<td>Same molecule</td>
<td>1318.7</td>
<td>3</td>
<td>21.5-23.5</td>
<td>0.4336</td>
<td>0.3900</td>
<td>0.3264</td>
<td>0.6888</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>Values combined</td>
<td>989.3+1318.7</td>
<td></td>
<td></td>
<td>0.0709</td>
<td>0.0787</td>
<td>0.0877</td>
<td>0.2321</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>6</td>
<td>583.3</td>
<td>1</td>
<td>28-30</td>
<td>0.1010</td>
<td>0.2172</td>
<td>0.0200</td>
<td>0.0727</td>
<td>Opposite</td>
<td>Failed</td>
</tr>
<tr>
<td>7</td>
<td>804.6</td>
<td>1</td>
<td>33-35</td>
<td>3.60x10^-7</td>
<td>2.40x10^-7</td>
<td>2.00x10^-5</td>
<td>1.75x10^-8</td>
<td>Same</td>
<td>Passed</td>
</tr>
<tr>
<td>8</td>
<td>874.6</td>
<td>1</td>
<td>40-44</td>
<td>6.37x10^-5</td>
<td>0.0006</td>
<td>0.0025</td>
<td>3.00x10^-5</td>
<td>Same</td>
<td>Passed</td>
</tr>
<tr>
<td>9</td>
<td>810.6</td>
<td>1</td>
<td>45-48</td>
<td>0.4717</td>
<td>0.4593</td>
<td>0.9367</td>
<td>0.6800</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>Same molecule</td>
<td>832.6</td>
<td>1</td>
<td>43-46</td>
<td>0.2827</td>
<td>0.5921</td>
<td>0.4973</td>
<td>0.0489</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>Values combined</td>
<td>810.6+832.6</td>
<td></td>
<td></td>
<td>0.3948</td>
<td>0.4181</td>
<td>0.9792</td>
<td>0.5536</td>
<td>Not applicable</td>
<td>Failed</td>
</tr>
<tr>
<td>10</td>
<td>1568.2</td>
<td>1</td>
<td>44-47</td>
<td>1.66x10^-5</td>
<td>0.0001</td>
<td>0.0142</td>
<td>0.0115</td>
<td>Opposite but same for mild vs. severe AD</td>
<td>Failed</td>
</tr>
<tr>
<td>11</td>
<td>1618.2</td>
<td>1</td>
<td>44-47</td>
<td>0.9239</td>
<td>0.6117</td>
<td>0.0835</td>
<td>0.3607</td>
<td>Same</td>
<td>Failed</td>
</tr>
<tr>
<td>12</td>
<td>804.5</td>
<td>1</td>
<td>39.5-41.5</td>
<td>9.48x10^-7</td>
<td>3.68x10^-6</td>
<td>1.34x10^-5</td>
<td>1.36x10^-6</td>
<td>Same</td>
<td>Passed</td>
</tr>
</tbody>
</table>
Biomarker 602.31 (charge state = 1) is a peptide with sequence of amino acids L/IAENR. The parent protein cannot be unambiguously identified. This sequence is found in several proteins. (See Table 3.2.). The observed fragments can be seen in the fragmentation mass spectrum (See Figure 3.1).

Biomarker 804.55 (charge state = 1, 33-35 min) is a phosphatidylcholine with two fatty acid acyl side chains, one having 18 carbon atoms and 2 double bonds (18:2) and the other having 20 carbon atoms and 5 double bonds (20:5). The positions of the double bonds are not determinable. A high intensity fragment seen at m/z 542.3211 confirms the presence of a fatty acid (18:2), and a high intensity fragment observed at m/z 520.3342 confirms the presence of a fatty acyl chain (20:5) (See Figure 3.2). The molecular formula for the lipid molecule is \([C_{46}H_{78}NO_{12}P]^+H^+\).

Biomarker 874.59 (charge state = 1) is a peroxidated phosphocholine lipid molecule. In such cases, it is difficult to determine fatty acyl side chains that have been oxidized. However, reference to the literature provided a match, and from that I obtained a formula for this molecule, PC (38:4)-(OOH)₂. It is an oxidatively modified lipid representing the bis-peroxide the two fatty acids 18:0 and 20:4 in glycerophosphatidylcholine (810 + 64) with elemental composition of C₄₆H₈₄NO₁₂P (Ingram, Homer et al. 2003). Given the saturated nature of one fatty acid (stearic acid), it is almost certain that the oxidation is of arachidonic acid, containing 4 double bonds. Figure 3.3 shows the fragmentation spectrum of 874.59.

Biomarker 804.53 (charge state = 1 eluting at retention time 39.5-41.5 min) is an oxidized glycerophosphatidylcholine molecule of molecular composition [C₄₂H₇₈NO₁₁P+H⁺]. Again, saturated palmitic acid (16:0) is one of the fatty acids modifying the glycerol backbone and not subject to oxidation. The second fatty acid (18:2), containing double bonds, was oxidized (Pereira 2006). It is likely that the second chain contains one peroxide and one keto group.
3.4.3 Constructing Multi-marker Diagnostic Panels for AD

Logistic regression analysis was performed to evaluate the ability of multi-marker models to diagnose AD. They were attempted for any stage AD as well for the very earliest stage of AD (CDR = 0.5). This same statistical approach was also used in the original study, but was supplemented with other statistical comparisons. In this approach, all possible combinations of the 4 biomarkers were considered. Additionally, all 4 biomarkers were included in one model. The analysis provides a receiver operator characteristic plot of the performance of the model in correctly identifying cases (sensitivity) as a function of the false positive rate (1-true negative rate/specificity). When comparing controls (CDR = 0) vs. any stage AD cases (CDR = 0.5, 1.0, 2.0), the best model was obtained by using all four biomarkers as shown in Figure 3.5 with an AUC = 0.7865. When comparing controls (CDR = 0) vs. very mild stage AD (CDR = 0.5), the best multi-marker panel is the combination of biomarker m/z 804.6 and biomarker 602.3 as shown in Figure 3.6 with AUC = 0.8502. If all 4 biomarkers are included in the panel for early stage AD diagnosis, then a panel with AUC = 0.8494 is obtained as shown in Figure 3.7.

3.5 Discussion

It should be noted that for the validation study, an Agilent 6530 Accurate-Mass Q-TOF LC/MS system was used, which was different from the instrument used during the initial study. This was due to significant problems with the QStar instrument that occurred during the validation study. This made the instrument erratic and unusable for quantitative comparisons in the validation study. The Agilent instrument is, nonetheless, a capillary liquid chromatography
Table 3.2. Table of calculated and observed b and y ions, confirming the sequence L/IAENR for the biomarker peptide 602.3 (charge state 1).

<table>
<thead>
<tr>
<th>Amino acids</th>
<th>Calculated b ions</th>
<th>Calculated y ions</th>
<th>Observed b ions</th>
<th>Observed y ions</th>
<th>Immonium ions</th>
</tr>
</thead>
<tbody>
<tr>
<td>L/I</td>
<td>114.0919</td>
<td>602.3262</td>
<td>Absent</td>
<td>602.3079</td>
<td>86.0878</td>
</tr>
<tr>
<td>A</td>
<td>185.1291</td>
<td>489.2422</td>
<td>185.1247</td>
<td>489.2442</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>314.1716</td>
<td>418.2051</td>
<td>314.1593</td>
<td>418.1749</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>428.2146</td>
<td>289.1625</td>
<td>428.2129</td>
<td>289.1574</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>584.3157</td>
<td>175.1196</td>
<td>584.2684</td>
<td>175.1022</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.1 MS/MS spectrum revealing fragment ions of molecule m/z 602.31.
Figure 3.2 MS/MS spectra revealing fragment ions of molecule m/z 804.55 (second spectrum zoomed in to show fragments 520.3342 and 542.3211).
Figure 3.3 MS/MS spectrum revealing fragment ions of molecule m/z 874.59.
Figure 3.4 MS/MS spectrum revealing fragment ions of molecule m/z 804.53.
Figure 3.5. Receiver operator characteristic curve with AUC 0.7865 for diagnosis of AD using all 4 validated biomarkers 602.3, 804.6, 874.6, 804.5. The curve represents biomarkers that looked different between non-demented controls (CDR = 0) vs. any stage AD (CDR = 0.5, 1.0, 2.0). The numbers shown in the curve are observation numbers.
Figure 3.6. Receiver operator characteristic curve with AUC 0.8502 for early diagnosis of AD using biomarkers 602.3 and 804.6. The curve represents biomarkers that looked different between non-demented controls (CDR = 0) vs. very mild stage AD (CDR = 0.5). The numbers shown in the curve are observation numbers.
Figure 3.7. Receiver operator characteristic curve with AUC 0.8494 for early diagnosis of AD using all 4 biomarkers 602.3, 804.6, 874.6 and 804.5. The curve represents biomarkers that looked different between non-demented controls (CDR = 0) vs. very mild stage AD (CDR = 0.5). The numbers shown in the curve are observation numbers.
system with electrospray ionization introduction of specimen onto a quadrapole-time-of-flight instrument. Instrumental parameters and methods were made as close to the original as possible. My total ion chromatograms looked similar to what was observed in the original study. The Agilent system was new and has sub 1 ppm mass accuracy, mass resolution of >20000 and high femtogram sensitivity that is superior to the older QStar instrument used in the previous study (https://www.chem.agilent.com/Library/brochures/5989-9397EN_LO.pdf). The retention times and peak intensities appeared to be very consistent using this instrument compared with the previous instrument. If anything, these results should be more accurate and consistent.

The validation studies were done to verify the utility of previously found biomarkers. This study was performed in my lab in a blinded fashion where involved researchers did not know the case or control status of the subjects from whom the specimens were collected.

The validation results were successful and showed several candidate biomarkers that continued to show significant differences between cases and controls. While individual markers may be useful in AD diagnosis, additional diagnostic power was sought through the development of biomarker panels. Utilizing accepted statistical approaches, a diagnostically useful combination of biomarkers was constructed, the current optimum giving an AUC of 0.8502 for the diagnosis of the earliest stage of AD. It showed a sensitivity of ~83% for a specificity of ~78% or ~75% sensitivity at a specificity of ~85%. It would be interesting to evaluate the performance of these biomarkers taking into account other clinical or demographic features (for example, the effect of gender), but this would require much more substantial clinical information, which currently is not available. Nevertheless, these results are substantially better than any existing validated AD diagnostic approaches. Currently, there is no ability to biochemically or radiologically diagnose early stage AD. Also, it should be remembered that there were many
additional candidates in the original study that could also be explored in the validation set, which may add to the capability of diagnosing AD.

Tandem MS fragmentation studies to chemically identify biomarker candidates have allowed me to identify statistically significant peptide and lipid biomarkers to the extent that can be accomplished instrumentally. Of the 13 biomarkers considered here, 5 appeared to be peptides. Among the other 8 biomarkers, 7 appear to be lipids, which contain a phosphocholine group. Of the 13, 2 biomarkers could not be evaluated. An overlapping peptide envelope hid biomarker 708.3. Additionally, abundant neighboring peaks caused ion suppression and lowered the intensity of this biomarker in the majority of the samples. Biomarker 660.4 was of such low intensity in specimens in the validation set, that it could not be accurately quantified. Of the 4 validated markers, the peptide at m/z 602.3 has been sequenced, but is too small to assign a parent protein unambiguously. I have also characterized the lipid biomarkers as fully as can be accomplished with current instrumentation and methods. Work to better understand fragmentation of oxidized lipids will continue independently in my lab, making use of standards exposed to oxidative reagents. However, those are envisioned studies and not complete. Of interest, two of these validated biomarkers represented lipid oxidation products.

Research has shown that oxidative stress and lipid peroxidation may be contributing factors to AD pathology, leading to neuronal loss and mitochondrial dysfunction. These factors are predicted to be involved in the actual pathogenesis of AD during its very earliest stages (Jenner 1991, Beal 2003). One study done on a rat model of AD showed that brain tissue from these animals had a reduction in cardiolipin resulting from lipid peroxidation in the mitochondria exposed to free radical stress in the animal (Ruggiero, Cafagna et al. 1992). It has also been reported that the cardiolipin biosynthesis pathway may be selectively compromised by oxidative
stress, causing approximately 20% reduction and a change in composition of the cardiolipin content (Ellis, Murphy et al. 2005). Oxidized phospholipids are involved in other proinflammatory pathologies as well as in response to immunologic mechanisms. Although there has been tremendous research concerning oxidized phospholipids as a general phenomenon (Greig, Kennedy et al. 2012), individual phospholipids have not been studied. With more extensive research on these biomarkers, it might be possible to obtain more detailed information about the biology of these altered biomarkers and their potential role in the pathology of AD.

Recently, Mapstone et al. discovered and validated ten lipid peripheral blood biomarkers that predicted phenoconversion to either amnestic mild cognitive impairment or Alzheimer’s disease within a 2–3 year timeframe with over 90% accuracy (Mapstone, Cheema et al. 2014). Most of these would normally be considered small metabolites. However, there were no oxidized phospholipids among these, which might serve to be better predictors of early stage AD. My approach is unique and serves to find lipid as well as peptide biomarkers, which may provide greater detail about disease presence and staging. My results should at a minimum be considered complementary to Mapstone’s.

To conclude, my research verified the utility of several, novel biomarkers found to diagnose AD in the previous study. These markers have been as completely identified as can be accomplished scientifically by MS or by any other method given the nanomolar concentrations of these biomarkers. How or why these species are found in the circulation system is of interest, including exploration of where they originate in the body; however, such questions are beyond the scope of this research. Importantly, biomarkers specific to very early AD were found, consistent with the results reported in Chapter 5. Consideration of markers observed there as well as gender considerations as described in Chapter 4 may prove interesting in follow up studies. In
summary, this low molecular weight, serum proteomics approach discovered and validated novel diagnostic biomarkers for AD, even early stage AD.
CHAPTER 4 - IDENTIFICATION OF GENDER SPECIFIC BIOMARKERS FOR ALZHEIMER’S DISEASE

4.1 Abstract

Background: Approximately two-thirds of Americans with AD are females as reported by the Alzheimer’s Association. There is also anecdotal data that the course of AD in women is more aggressive. These observations suggest that there are likely gender differences in AD, but at present, no study has sought biomarkers that differ among men and women with AD or that are more specific to AD in women or in men. We predict there are biomarkers that are unique to males versus females. This could not only be important during diagnosis, but also when studying the pathways involved in AD. Hence, we attempted to restudy the biomarkers we found in all-stage AD cases versus matched controls, but consider potential differences based on gender.

Methods: Serum samples from 58 cases from individuals with any stage AD and 55 matched controls were depleted of high abundance, high molecular weight proteins, fractionated by capillary liquid chromatography and analyzed by ESI-MS. This has yielded 38 statistically significant biomarkers with another 21 near significant biomarkers. The data for males and females were separated into sets for these studies: 1) 29 male cases vs. 26 male controls, and 2) 29 female cases vs. 29 female controls. The biomarkers found in the initial study were reexamined in these sets to see if they were either male or female specific. Simple t-test statistical differences were determined as well as logistic regression analysis to test their diagnostic sensitivities and specificities.

Results: First, we found 31 significant (p < 0.05) or near significant (p = 0.05 to 0.099)
biomarker candidates for women and, second, we found 16 significant or near significant biomarkers for men. Additionally, 25 significant or near significant biomarkers were found to perform equivalently for women and men. Panels of markers specific for either women or men provided sensitivities of 60-70% with specificities >85% for diagnosis of AD.

Conclusion: This serum proteomics approach was successfully used to find gender-specific serum biomarkers for Alzheimer’s disease.

4.2 Introduction

Alzheimer’s disease is twice as prevalent in women as in men. As a function of age, the differences in AD incidence in women become even more substantial (Andersen, Launer et al. 1999, Lobo, Launer et al. 2000). Not only is there a higher probability of women getting AD, they also demonstrate greater cognitive declines compared with men. This even includes verbal skills that are stronger in healthy females compared with males (Hyde and Linn 1988, Weiss, Kemmler et al. 2003, Maylor, Reimers et al. 2007, Wallentin 2009).

In the Religious Orders Study (participants were older catholic nuns, priests and brothers), 141 brains were studied postmortem. It was reported that in women the association between the severity of AD pathology and clinical manifestations of AD were 20 times stronger compared to men (Barnes, Wilson et al. 2005). They concluded that AD tissue pathology would be diagnosed with clinical dementia much more frequently in women than in men. This conclusion concurs with the numerous reports stating the incidence of AD being higher in women than in men (Andersen, Launer et al. 1999), possibly because of longer life span (Hebert, Wilson et al. 2000).

Researchers compared both genders with equally severe AD using results of 18FDG-PET imaging. They found that women had significantly elevated glucose metabolism in the sections
principally associated with the pathological process of AD, i.e., the right inferior frontal, superior temporal and insular cortices, and the hippocampus, compared with men. This suggests again a gender difference in the disease process (Perneczky, Drzezga et al. 2007). In these studies, I sought to find serum biomarkers that confirmed a gender difference between AD women and AD men. This may allow me to define the biochemical pathways involved.

If there are gender differences in AD and in AD biomarkers, the most straightforward explanation would be hormone-driven differences. The metabolic changes caused by hormones may affect the pathology of AD. Estrogen is known to protect the brain. The decrease in the female hormones during menopause may reduce metabolism in the brain. This might increase risk for AD. Despite both sexes having receptors present for both estradiol and testosterone, their brains respond very differently to the hormones. Similarly, even though both sexes can synthesize estradiol in neurons, synaptic responses are sexually dimorphic in different regions of the brain. These studies demonstrate gender differences in brain biology and imply that the pathogenesis of AD will require deeper comprehension of the intricate interplay of genetics, hormonal regulation and environmental factors. The role of gender differences in the onset and progression of AD remains poorly understood, but is undoubtedly important (Carter, Resnick et al. 2012).

Longevity likely contributes to the higher fraction of women having AD compared to men beyond the age of 80 years. Longevity does not explain the higher number of AD cases among women between 60-80 years. There is currently effort to define potential pathogenic mechanisms to explain these findings. In one study, the authors noted that the mitochondria from young females are safeguarded against amyloid-β protein toxicity, produce fewer reactive oxygen species, and generate fewer apoptotic signals than males. But this was not the case for
mitochondria from older females (specifically during/post menopause). Clearly, these results suggest that estrogenic compounds and responses protect against the mitochondrial toxicity of amyloid-β protein (Viña and Lloret 2010). However, the issues are not simple. Studies of estrogen replacement therapies have not generally been successful in protecting against AD. Moreover, while all women eventually go through menopause, not all of them develop AD. Finding gender specific AD biomarkers holds open the possibility that more specific and more complete mechanisms involved in AD as a result of gender differences may be found.

Hence, I proposed to use the serum proteomics approach to look for biomarkers that might be useful to diagnose AD specifically among men or women. In turn, such gender specific molecular expression differences might identify gender differences in pathway involvement and may be useful in developing treatments for AD that are more effective in women or men.

4.3 Materials and Methods

As described in Chapter 2, serum specimens were obtained from the Knight ADRC at Washington University School of Medicine, and consisted of 58 cases and 55 controls. These were processed to deplete protein, separated by cLC and introduced onto the MS by electrospray ionization. These specimens came from AD patients and controls matched for gender, age and APO E4 status, and were devoid of any co-morbidity.

In the comparison of all stages of AD versus controls, 44 significant (p values less than 0.05) and 25 near significant (p values between 0.05 and 0.10) candidate biomarkers were found using raw MS data and 38 significant and 21 near significant candidate biomarkers using normalized data (Chapter 2). Almost all of the biomarker candidates found in the normalized data were present when using raw data. I looked for gender differences among all of the biomarkers found
in the first study (Chapter 2). For this purpose, I segregated the data into male cases (n = 29) vs. male controls (n = 26) and female cases (n = 29) vs. female controls (n = 29). I assessed the candidate marker’s ability to identify subjects having AD within each gender group. Several of the markers performed markedly better in one gender than the other. These were considered to be either male specific or female specific.

Combinations of candidate biomarkers were further modeled to find out if they provided better diagnostic discrimination. Note that candidates with p-values less than 0.10 were considered because some of them may provide complementary diagnostic information to other candidate biomarkers and give better diagnostic discrimination even if they themselves do not satisfy criteria for statistical significance. Logistic regression analysis and correlation analysis were performed to model the diagnostic capabilities of the candidates. Specifically, correlation analysis provided a grouping of biomarkers to identify clusters of candidates that appear to depict a common pathway and those that are more remotely connected, indicating complementary differences (potentially different pathways). Finally, a forward-selection method (leave-one-out approach) was used to build multi-marker diagnostic panels unique to males or to females. ROC curves were plotted for each gender and AUCs were calculated for the multi-biomarker combinations.

4.4. Results

4.4.1 Evaluation of Potential Male and Female Specific Biomarkers and Their Use in Multi-marker Panels

I observed that the p values of some of the biomarkers improved markedly (>2 fold better p-
value) when applied to just men or just women, even with fewer subjects in each gender comparison set. It is also true that if a specific marker showed improved diagnostic ability in females, it performed much less well in males (p > 0.15). There were other biomarkers that performed equivalently for either male or female cases versus controls. These markers showed p-values in both gender groups that were very similar to their original values, but somewhat less significant due to a smaller number of subjects in each of the subsets. Statistical analyses revealed 31 significant or near significant biomarker candidates for women, 16 for men, and 25 additional gender-neutral markers comparing any stage AD to controls. ROC curves were plotted for single candidate biomarkers.

**4.4.2 Evaluation of Biomarkers Specific to Males**

Multi-marker combinations were modeled for male-specific biomarkers. There were 10 biomarkers that appeared in 10% or more of all the multi-marker models (See Table 4.1). Multi-marker ROC curves were plotted. These provide measures of sensitivity and specificity in diagnosing AD in men. The best multi-marker panels gave AUCs of 0.820 (See Figure 4.1) and 0.784 (See Figure 4.2). These optimized combinations yielded sensitivities of ~65% with specificities of ~86%. Dendritic analysis, which relates one marker to another, showed 3-4 distinct groups (See Figure 4.3).

**4.4.3 Evaluation of Biomarkers Specific to Females**

Similarly, multi-marker combinations were constructed for female-specific biomarkers. The modeling gave 10 biomarkers that appeared in 10% or more of all female-specific, multi-marker panels (See Table 4.2). Correlation analysis, dendritic analysis, creation of heat maps and leave-
Table 4.1. Biomarkers that appeared in 10% of all male specific multi-marker models.

<table>
<thead>
<tr>
<th>Biomarker Number</th>
<th>Retention Time (min)</th>
<th>Biomarker (m/z ratio)</th>
<th>Charge State (Positive Ion Mode)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31-38</td>
<td>810.6</td>
<td>1</td>
<td>0.681</td>
</tr>
<tr>
<td>2</td>
<td>11-15</td>
<td>971.4</td>
<td>1</td>
<td>0.631</td>
</tr>
<tr>
<td>3</td>
<td>21-25</td>
<td>804.6</td>
<td>1</td>
<td>0.620</td>
</tr>
<tr>
<td>4</td>
<td>17-25</td>
<td>560.4</td>
<td>1</td>
<td>0.619</td>
</tr>
<tr>
<td>5</td>
<td>31-38</td>
<td>1618.2</td>
<td>1</td>
<td>0.613</td>
</tr>
<tr>
<td>6</td>
<td>17-25</td>
<td>1170.9</td>
<td>1</td>
<td>0.609</td>
</tr>
<tr>
<td>7</td>
<td>17-25</td>
<td>614.4</td>
<td>1</td>
<td>0.558</td>
</tr>
<tr>
<td>8</td>
<td>11-15</td>
<td>574.3</td>
<td>1</td>
<td>0.537</td>
</tr>
<tr>
<td>9</td>
<td>29-32</td>
<td>804.5</td>
<td>1</td>
<td>0.519</td>
</tr>
<tr>
<td>10</td>
<td>16-22</td>
<td>819.5</td>
<td>1</td>
<td>0.413</td>
</tr>
</tbody>
</table>
Figure 4.1. Multimarker model (ROC curve) for biomarkers 804.6 and 1618.2 for identification of AD in men.
Figure 4.2. Multimarker model (ROC curve) for biomarkers 574.3, 1170.0, 810.6 and 804.6 for identification of AD in men.
Figure 4.3. Biomarker cluster analysis of 10 best male specific AD biomarkers.
Table 4.2. Biomarkers that appeared in 10% of all female specific multi-marker models.

<table>
<thead>
<tr>
<th>Biomarker Number</th>
<th>Retention Time (min)</th>
<th>Biomarker (m/z ratio)</th>
<th>Charge State (Positive Ion Mode)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22-32</td>
<td>513.3</td>
<td>1</td>
<td>0.706</td>
</tr>
<tr>
<td>2</td>
<td>22-32</td>
<td>660.4</td>
<td>1</td>
<td>0.675</td>
</tr>
<tr>
<td>3</td>
<td>31-38</td>
<td>1576.2</td>
<td>1</td>
<td>0.662</td>
</tr>
<tr>
<td>4</td>
<td>31-38</td>
<td>675.6</td>
<td>1</td>
<td>0.662</td>
</tr>
<tr>
<td>5</td>
<td>29-37</td>
<td>1516.2</td>
<td>1</td>
<td>0.632</td>
</tr>
<tr>
<td>6</td>
<td>11-15</td>
<td>1107.5</td>
<td>1</td>
<td>0.618</td>
</tr>
<tr>
<td>7</td>
<td>11-15</td>
<td>531.3</td>
<td>1</td>
<td>0.580</td>
</tr>
<tr>
<td>8</td>
<td>25-34</td>
<td>790.6</td>
<td>1</td>
<td>0.547</td>
</tr>
<tr>
<td>9</td>
<td>11-15</td>
<td>515.3</td>
<td>1</td>
<td>0.545</td>
</tr>
<tr>
<td>10</td>
<td>8-15</td>
<td>921.4</td>
<td>1</td>
<td>0.545</td>
</tr>
</tbody>
</table>
one-out analysis were also executed. Multi-marker logistic regression analysis with receiver operator characteristic (ROC) curves demonstrated combinations that had AUCs of 1.000 (See Figure 4.4 and 4.5), 0.851 (See Figure 4.6), 0.837 (See Figure 4.7) and 0.828 (See Figure 4.8). The best of these displayed a sensitivity of 100% with a specificity of 100%. Dendritic analysis indicated as many as 4 different groups (See Figure 4.9).

4.5 Discussion

Disease prevalence, more so than incidence, provides a better assessment of disease risk. The older the population, the greater the risk of AD, and women in general live longer than men. Consequently, the incidence of new cases of AD is greater in women. However, AD also appears to affect a greater percentage of women than men, and studies suggest AD in women results in greater and more rapid declines in function (Musicco 2009).

Clinical studies find women to be more prone to develop AD and the progression and passage of their disease to be different than AD in men (Viña and Lloret 2010, Association 2013). AD pathology appears to be more extensive in women than in men when matched in other ways (Barnes, Wilson et al. 2005). All of this suggests differences in AD affecting women versus men. Consequently, I sought to know if my serum biomarkers would demonstrate gender-related performance differences. My results found that many biomarkers demonstrated a clearly better performance in women or in men. This suggests that my biomarkers will likely give very useful gender specific information in cases of AD.

While this study has primarily focused on diagnostically useful combinations of biomarkers that are more relevant to one gender, this same approach may be able to provide insights into the pathological differences of AD in women versus men. By identifying the several molecular
Figure 4.4. Multimarker model (ROC curve) for biomarkers 675.6, 531.3, 513.3 and 515.3 for identification of AD in women.
Figure 4.5. Multimarker model (ROC curve) for biomarkers 531.3, 513.3, 1576.2 and 515.3 for identification of AD in women.
Figure 4.6. Multimarker model (ROC curve) for biomarkers 675.6, 1516.2, 921.4 and 790.6 for identification of AD in women.
Figure 4.7. Multimarker model (ROC curve) for biomarkers 1107.5, 1516.2, 921.4 and 790.6 for identification of AD in women.
Figure 4.8. Multimarker model (ROC curve) for biomarkers 790.6, 1516.2 and 921.4 for identification of AD in women.
Figure 4.9. Biomarker cluster analysis of 10 best female specific AD biomarkers.
species that are altered in women or in men, by looking for connections (as suggested by the
dendritic maps) in these molecules, one may begin to explain clinical findings. However, this
would also require much more detailed information about the patients.

While my results, summarized here, are promising, I realize they are preliminary and require
additional studies to confirm the utility of these biomarkers. Such studies are currently underway
in my lab in a blinded fashion where involved researchers do not know the case or control status
of the subjects from whom the specimens were collected. These studies will take substantial
additional effort to complete. Moreover, there were other candidate biomarkers that were not
considered in the multi-marker panels that could and should be evaluated. Additionally, the
markers would require chemical identification before high-throughput, rapid assays could
possibly be developed. Chemical identification may explain the clustering of compounds seen in
the dendritic maps to know if this represents common pathways or disease processes, or some
other unifying factor.

In conclusion, research studies that have looked at differences in AD prevalence, progression,
and phenotype support differences between men and women. However, gender has not been
considered previously important in diagnostic approaches, but should be considered in both
future clinical and laboratory studies of AD. Very little research has been conducted to
potentially explain such gender differences in AD. Certainly no study of biomarkers for AD has
considered gender, and yet it is entirely possible that development of effective treatments may
differ by gender. My research hypothesis considered gender differences in subjects with AD, and
my data suggest strongly that such differences exist. My data also support gender-specific, serum
biomarkers that improve AD diagnosis in women or in men. My results may provide biomarkers
to detail AD mechanisms involved in the development and progression of AD in both males and
females. My results also underscore that our serum proteomics approach may be beneficial and complementary to existing clinical approaches to the diagnosis and, ultimately, treatment of AD in a gender appropriate approach in males and females.
5.1 Abstract

Background: Alzheimer’s disease (AD) is typically a slow, progressive disease. The changes in the brain are thought to start years, perhaps even decades, before the onset of clinical symptoms. Diagnosis currently relies on cognitive testing and is subjective. Typically, it requires years to diagnose AD with confidence. For these reasons, AD diagnosis often occurs after the disease has caused catastrophic and irreversible changes to the brain. Not surprisingly, drug studies that rely on cognitive testing for AD diagnosis have failed to demonstrate benefit. Early diagnosis of AD is not currently possible, and yet is paramount to the development of treatment options and to better understand the biochemical pathways that lead to AD. In an effort to find markers for early stage AD, we used a serum proteomics approach that allows us to evaluate thousands of low-molecular weight biomolecules in serum as potential biomarkers.

Methods: Serum samples from 19 individuals with moderate AD, CDR of 2.0, and 11 individuals with a very mild or mild clinical AD, CDR of 0.5 or 1.0, were previously analyzed by MS as part of an initial study to discover novel AD biomarkers for any stage AD. In turn this study focused on efforts to find AD biomarkers that differed between patients having mild AD compared with patients having moderately severe AD. Data were organized by CDR status and all peaks were overlaid and reviewed for possible quantitative differences. Initially, peak heights for peaks that appeared quantitatively different were obtained, combined by group and assessed statistically using the Student’s t-test. Multi-marker panels that can diagnose AD among patients
at different stages of the disease were also modeled.

Results: We found 44 significant and near significant biomarkers that were quantitatively different between mild and moderately severe AD. The best combination resulted in an ROC curve having an AUC of 0.892 and providing a sensitivity and specificity ≥90%.

Conclusion: The serum proteomics approach appears able to find serum biomarkers that are different at different stages of Alzheimer’s disease. This can potentially not only help in diagnosis, but also give clues about the biochemical pathways involved in the disease.

5.2 Introduction

5.2.1 Onset of Alzheimer’s Disease

AD is an immedicable and relentless brain disease. It is known to steadily diminish the ability to remember and think, and to perform simple daily activities/chores. Despite its prevalence and devastating consequences, we do not understand precisely how the disease process starts. Current research indicates that detrimental changes in the brain begin at least a decade before the onset of dementia and other clinical symptoms (Foundation).

5.2.2 Review of Current Biomarkers for Early Diagnosis

There have been attempts at finding early stage biomarkers. These are reviewed in the following paragraphs.

5.2.2.1 CSF biomarkers

Currently, increased CSF tTau or pTau protein levels and decreased Aβ-42 levels are the
only biomarkers with demonstrated sensitivity and specificity to differentiate between AD and other dementias in their early stages (Rösler N 2002, Sunderland, Linker et al. 2003, Galasko 2005, Wiltfang, Lewczuk et al. 2005, Hampel and Buerger 2006, Hansson, Zetterberg et al. 2006, Parnetti, Lanari et al. 2006, Bailey 2007, Finehout, Franck et al. 2007, Engelborghs, De Vreese et al. 2008). They are possible indicators of disease severity and may correlate with increasing numbers of AD symptoms (Hampel, Mitchell et al. 2004, Lewczuk, Esselmann et al. 2004, Vanderstichele, De Meyer et al. 2005, Jellinger 2006, Stefani, Martorana et al. 2006, Wallin, Blennow et al. 2006, Speth, Sindic et al. 2007). Increased values of tTau CSF concentrations were documented between early and advanced stages of AD (Andersson, Blennow et al. 2008). However, changes in soluble Aβ-42 have not been found to parallel the degree of cognitive decline (Stefani, Martorana et al. 2006). In one study, the Aβ-42/40 ratio was associated with changes in pTau levels in patients with low and high CSF Aβ-40 levels, but was not associated with Aβ-42 levels alone (Wiltfang, Esselmann et al. 2007).

Studies have shown that a marked increase of beta-secretase 1 (BACE 1) CSF levels and activity were correlated with mild cognitive impairment (MCI). Also, increases in other neuroimmune markers appear to precede clinically severe AD in some studies (Gruden, Davidova et al. 2007, Zhong, Ewers et al. 2007).

Longitudinal cohort studies done recently found increased levels of CSF Aβ-42 and tTau (but not pTau phosphorylated at threonine 181) in a group of patients in a memory clinic with similar levels in all diagnostic groups. In that study, the cross-sectional variation between diagnostic groups was substantially greater than the longitudinal differences observed within individuals. This demonstrates that these markers were not biomarkers of disease advancement (Bouwman, van der Flier et al. 2007). A problem has been that most studies evaluating these biomarkers have
used clinical diagnoses alone rather than including pathological changes (Jellinger, Janetzky et al. 2008). This may or may not explain the general poor performance of these biomarkers in staging AD.

5.2.2.2 Plasma biomarkers

In plasma, total amyloid precursor protein or Aβ-42 was found increased in cases of familial AD (Schupf, Patel et al. 2001, Ertekin-Taner, Younkin et al. 2008), but has failed as diagnostic biomarkers in cross-sectional studies of sporadic AD (Tamaoka, Fukushima et al. 1996, Mehta, Pirttilä et al. 2000, Vanderstichele, Kerschaver et al. 2000, Fukumoto, Tennis et al. 2003, Assini, Cammarata et al. 2004, Crystal and Davies 2008, Ertekin-Taner, Younkin et al. 2008). In some studies, higher levels of Aβ-42, decreased levels of Aβ-40 and a reduced Aβ-42/Aβ-40 ratio have been found in the plasma of cases compared with controls and levels changed with disease progression from cognitively normal to MCI to AD (van Oijen, Hofman et al. 2006, Graff-Radford, Crook et al. 2007, Sundelöf, Giedraitis et al. 2008). In direct contradiction, other studies concluded that lower levels of Aβ-42 plasma might indicate the transition from MCI to AD (Song, Mook-Jung et al. 2007). Perhaps owing to high intra- or inter-personal variability of serum and plasma Aβ levels (Abdullah, Paris et al. 2007), plasma levels of Aβ-40 and Aβ-42 have failed to correlate with histologically or biochemically assessed amyloid plaques in the brain (Freeman, Raju et al. 2007). Currently, it is accepted that plasma Aβ-42 levels alone are not reliable biomarkers for MCI and AD (Mehta, Pirttilä et al. 2000, Brettschneider, Morgenthaler et al. 2005, Blasko, Jellinger et al. 2008).

Inflammatory molecules found in blood, such as C reactive protein (CRP), interleukin-6, and others, have been proposed as biomarkers for vascular dementia (Bibl, Esselmann et al. 2007,
Ravaglia, Forti et al. 2007) or appear to be increased prior to the clinical onset of symptoms in both AD and vascular dementia (Engelhart, Geerlings et al. 2004). However, these markers are not specific for AD, but are increased in multiple diseases. Plasma Aβ is a risk factor for AD in several longitudinal cohort studies with changes in Aβ levels observed 5–20 years before the patient dies (Mayeux, Honig et al. 2003, van Oijen, Hofman et al. 2006).

Proteomics approaches used on plasma have revealed increased concentrations of α-1-antitrypsin and apolipoprotein J in individuals with AD (Liao, Yu et al. 2007). Plasma homocysteine has been shown to be directly related to Aβ-40 levels, while the association with Aβ-42 was not significant, suggesting that homocysteine is related to aging but not specifically to AD, although it may interact to affect AD risk and cognition in Parkinson’s disease (Mayeux, Honig et al. 2003, Irizarry, Gurol et al. 2005, Luchsinger, Tang et al. 2007). Thus, the ineffectiveness of Aβ as an AD marker in plasma/serum as well as its non-specificity and the absence of other AD markers currently require further work to find diagnostic serum biomarkers for AD, and especially biomarkers for early stage AD (Jellinger, Janetzky et al. 2008).

5.2.2.3 Platelet biomarkers

Studies of platelets in AD suggest that changes in amyloid precursor protein predict conversion of MCI into dementia (Borroni, Colciaghi et al. 2003, Tang, Hynan et al. 2006). Other platelet studies have found an association of membrane fluidity with cognitive decline (Zainaghi, Forlenza et al. 2007). Higher BACE-1 activity in platelets from AD patients compared with controls (Johnston, Liu et al. 2008), and an increase in monoamine oxidase B expression have been reported in the platelets and brain tissue of demented individuals having AD or PD. This has been replicated by others (Adolfsson, Gottfries et al. 1980, Jarman, Glover et al. 1993,
Jellinger, Janetzky et al. 2008). Recently, increased platelet phospholipase A2 activity has been reported in patients with AD, vascular dementia or ischemic stroke (Krzystanek, Krzystanek et al. 2007, Jellinger, Janetzky et al. 2008). Although some of these may have appeared changed in response to disease, they lack specificity.

5.2.2.4 Structural biomarkers

The usefulness of magnetic resonance imaging (Blasko, Jellinger et al. 2008) in observing structural changes in the diagnosis and staging of AD has been reviewed (Barnes, Foster et al. 2007, Jellinger, Janetzky et al. 2008, Vemuri, Gunter et al. 2008). Medial temporal lobe atrophy (MTA) as assessed by MRI is sensitive to primary degenerative hippocampal atrophy in elderly seniors. However, a moderate MTA score has not been linked with dementia (Barkhof, Polvikoski et al. 2007), but has been linked to MCI (Mevel, Desgranges et al. 2007). Nevertheless, these changes are not specific for AD. In contrast, the amount of increased positron emission tomography (PET) measured amyloid-binding ligand (11C) PIB uptake in the brain of subjects with possible MCI/mild AD appears indicative of early AD (Klunk, Engler et al. 2004, Kemppainen, Aalto et al. 2007, Rowe, Ng et al. 2007, Jellinger, Janetzky et al. 2008). Examination of CBF-SPECT, CMRgl-PET (glucose metabolism), proton spectroscopy (H-1 MRS), high-field strength functional MRI, voxel-based morphometry, increased activation of the mediobasal temporal lobe detected by fMRI, (R)-[(11)C]PK11195 PET to detect microglia, combined PiB imaging and structural MRI studies appear to be predictive for conversion of MCI into early AD (Apostolova, Steiner et al. 2007, Bracco, Bessi et al. 2007, Caroli, Testa et al. 2007, Hämäläinen, Tervo et al. 2007, Huang, Eidelberg et al. 2007, Kantarci, Weigand et al. 2007, Kircher, Weis et al. 2007, Krogholler, Boellaard et al. 2007, Petrell, Wang et al. 2007,
Rami, Gomez-Anson et al. 2007, Jack, Lowe et al. 2008, Prince, Woo et al. 2008). Similarly accelerated rates of hippocampal atrophy and ApoE ε4 prevalence (Van de Pol, van Der Flier et al. 2007), FDG-PET (Samuraki, Matsunari et al. 2007), detection of subcortical hyperintensities (Debette, Bombois et al. 2007, Huang, Friedland et al. 2007, Smith, Egorova et al. 2008) and MRI patterns of grey matter atrophy (Whitwell, Shiung et al. 2008) are also indicative of AD type pathology and conversion of MCI into early AD. Some argue that relatively normal metabolism of glucose in the presence of high frontal amyloid load suggests that amyloid plaque formation may not be directly responsible for neuronal dysfunction in this disorder (Edison, Archer et al. 2007). Cortical and hippocampal atrophy in aging and dementia are complicated by the many processes acting on brain structures that can mediate cognitive deterioration (Jagust, Zheng et al. 2008).

5.2.2.5 Alzheimer Associated protein (ALZAS)

Alzheimer Associated protein is an Aβ protein expressed in aging patients with the likely diagnosis of AD. The gene for the protein was discovered on chromosome 21 within the APP gene region. This protein, has a 79 amino acid sequence and contains the Aβ-42 fragment, the amyloid protein transmembrane signal and a unique 12 amino acid c-terminal, which is absent in all known allelic forms of the amyloid precursor protein (Kienzl, Jellinger et al. 2002, Kienzl, Jellinger et al. 2006). Reverse transcription PCR studies revealed that the mRNA expression of this protein was found in cortical and hippocampal regions of the brain and in lymphocytes of older individuals with AD (Kienzl, Jellinger et al. 2002, Kienzl, Jellinger et al. 2006, Kienzl, Jellinger et al. 2007, Jellinger, Janetzky et al. 2008).

In summary, despite all of these efforts to identify biochemical biomarkers or radiologic
measured changes in brain architecture, there are no clinically approved early stage biomarkers for AD. Moreover, the great majority of studies focusing on AD serum biomarkers have attempted to demonstrate differences in the Aβ or tau proteins without success. My approach is completely novel and has ability to find biomarkers not previously considered.

5.2.3 Alzheimer’s Disease is Multifactorial

Recently, Iqbal et al. proposed that AD is a heterogeneous disease (Iqbal and Grundke-Iqbal 2010). They suggested that AD is caused by a number of factors including multiple genetic, environmental, or lifestyle factors, or from a combination of all three. Less than 1% of AD cases are a result of genetic mutations. Known mutations include transmembrane proteins like amyloid precursor protein, presenilin 1, and presenilin 2 (Bird 2008). Sporadic AD that is not associated with any known genetic mutations represents greater than 99% of all AD. This classification has been further subdivided into AELO (AD cases with low Aβ1–42, high incidence of APOE ε4, and late onset), ATEO (AD cases with low Aβ1–42, high tau, and early onset), LEBALO (AD cases with high incidence of Lewy bodies, low Aβ1–42, and late onset), HARO (AD with high Aβ1–42, and recent onset) and ATURO (AD with low Aβ1–42, high tau, high ubiquitin, and recent onset), each presumably due to several different etiopathogenic mechanisms as shown in Figure 5.1. Also, if an individual carries one or two copies of the APOE ε4 allele instead of APOE ε2 or APOE ε3, the risk for the development of sporadic AD is increased by many fold (Corder, Saunders et al. 1993).

Even though many plaques and neurofibrillary tangles are observed in AD pathology, the two lesions occur in different proportions among different cases of AD. This is the basis for separating the plaque- and tangle-dominant AD subgroups (Katzman, Terry et al. 1988,
Figure 5.1. Possible subgroups of Alzheimer's disease based on the risk factors. Reprinted from (Iqbal and Grundke-Iqbal 2010) with permission. ©Elsevier, 2010.
The Amyloid Cascade Hypothesis (Hardy and Higgins 1992, Hardy and Selkoe 2002) states that the main cause of neurofibrillary degradation and dementia is the β-amyloid precursor protein (β APP) and its fragment, Aβ. However, the absence of any correlation between the number of plaques and the number of tangles in AD, the existence of various phenotypes of Aβ plaques, together with the lack of neurofibrillary degeneration in normal elderly humans is contradictory to the hypothesis. On the other hand, the hypothesis is reinforced by the findings of a study that showed that introduction of Aβ1–42 in P301L tau transgenic mice (Götz, Chen et al. 2001) and P301L tau transgenic mice with APPSWE (APP with Swedish mutation) (Lewis, McGowan et al. 2000) worsened tau neurofibrillary pathology. This worsening of tau pathology may be attributed to activation of the stress-activated protein kinases, known to phosphorylate tau at the various proline-directed sites. However, none of these tau mutations have been observed in actual AD patients. Moreover, abundant Aβ plaques have been observed in the neocortex of healthy, aging people as well as abundant Aβ content found in genetic cerebral hemorrhage with amyloidosis, but without any neurofibrillary pathology present (Levy, Carman et al. 1990, Dickson, Crystal et al. 1992). All of this makes the Amyloid Cascade Hypothesis untenable for AD and its attendant changes in the human brain. Thus, we do not as yet understand the origins of neurofibrillary tangles and Aβ plaques. This situation suggests strongly the need for more global evaluation of molecules in AD that might be useful as biomarkers and that additionally might shed light on the progressive mechanisms leading to last stage AD.

5.2.4 Unknown Mechanisms Involved

While the cause of AD is unknown, the genetic and pathological studies done over the past
20 years still suggest that aggregation of amyloid-beta contributes to the pathogenesis of AD (Hardy and Selkoe 2002). Aβ plaque formation is likely to contribute to several later biochemical changes including neuroinflammation, synaptic dysfunction, tauopathy, and apoptotic events leading to cell death. This concept is called “amyloid cascade hypothesis.” Notwithstanding this theory, β-amyloidosis and Aβ plaque burden do not correlate with the onset or level of cognitive decline in the transition from MCI to dementia to later stages of AD. This clearly implies that other factors are mediating or contributing to the advancement of the disease. New research data show that approximately 30% of all individuals over the age of 75 have the characteristic neuropathological features of AD when their brains were studied by autopsy without functional evidence of AD. These individuals were considered clinically normal at the time of death. Hence, significant AD lesions evident in these individuals’ brains without disease symptoms could be interpreted as no direct causal link between Aβ pathology and AD cognitive function, or to the possibility of yet-to-be-specified protective mechanisms that preserve cognitive function in the face of pathologic processes (Iacono, O’Brien et al. 2008, Iacono, Markesbery et al. 2009, Alkadhi and Eriksen 2011, Driscoll and Troncoso 2011). To identify such compensating mechanisms, many approaches may be needed, including proteomic studies that might reveal novel pathways. Such approaches, especially comparing changes as a function of AD disease stage can potentially map out the one or more mechanisms involved in this disease. While identifying these potentially new pathways was not a direct goal of our studies, demonstrating differences in the complement of peptides and proteins present in serum at different stages of AD progression opens the door to that possibility.

5.3 Materials and Methods

This study relied on the same set of specimens as described in Chapter 2 and included sera
from 58 cases with AD and sera from 55 non-demented controls obtained from Knight ADRC, Washington University School of Medicine. These were processed as previously described and analyzed by cLC-ESI-MS/MS. Of these, samples of 19 cases with a moderate clinical dementia rating of 2.0 and 11 cases with very mild or mild clinical dementia ratings of 0.5 (n = 7) or 1.0 (n = 4) were compared. All peaks were reviewed in a de novo effort to locate biomolecules that were differentially expressed at different stages of AD. All aspects of data analysis, statistical analysis and testing of multi-marker sets to find AD stage-specific serum biomarkers in this study were performed again on these data sets using the procedures outlined in Chapter 2 in the Methods section.

5.4 Results

5.4.1 Discovery of Candidate Serum Biomarkers for Early Stage AD

On examining raw, un-normalized MS data, 4 statistically significant AD stage dependent serum biomarkers were found comparing patients having very mild and mild stage AD (CDR = 0.5, 1.0) with patients having later stage, moderate AD (CDR = 2.0). Additionally, 16 near-significant biomarkers that had p-values between 0.05 and 0.10 were found. Upon log normalizing the mass spectral data for each candidate to an internal molecular species from the same time window (that was equivalently distributed between cases and controls), I found 13 significantly different serum biomarkers along with 19 candidates that were near significant (p = 0.05 to < 0.10). When un-normalized or normalized data were combined, 20 unique potential biomarkers were found (p values < 0.05) with another 24 near-significant biomarkers (p values between 0.05 and 0.10) (See Table 5.1). These were further statistically assessed by receiver
operator characteristic curve analysis applied to all 44 candidate biomarkers. Some of the AD biomarkers were higher in cases and others in controls.

5.4.2 Assessment of Diagnostic Similarities between Biomarkers

First, all the data was clustered and it was observed that the markers could distinguish mild stage AD versus moderate stage AD very well. Second, correlation analyses of all biomarker pairs were applied to all of the biomarkers. During the process, a heatmap, clustering dendogram and correlation matrix that identified biomarkers that were closely correlated were produced. The markers that closely correlated provided similar diagnostic information, while those that showed little correlation provided added and complementary diagnostic information.

5.4.3 Evaluation of the Biomarkers in Multi-marker Panels

Further, the combinations of these serum biomarkers were tested to see if they would provide better diagnostic discrimination between different stages of AD. Again, the forward selection approach was used to construct multi-biomarker models. Significant as well as near significant biomarkers were considered, owing to the fact that near significant biomarkers might potentially add complementary diagnostic information on subjects with earlier stage AD as opposed to the ones with more severe AD. Log normalized data was used to build multi-marker panels. Multi-biomarker panels were created by including statistically different or near different candidates one by one until a combination of the panel yielded the best AUC possible. There were 13 biomarkers that appeared in 9% or more of all the multi-marker models (See Table 5.2). These were considered to be more robust and likely more valuable.

Additional statistical analyses centered on the 13 biomarkers, on which again a forward
Table 5.1. List of 20 significant (p value < 0.05) and 24 near-significant (p value between 0.05 and 0.1) biomarkers when raw and normalized data were considered. [Letters in the brackets indicate position of the biomarker within the time range (A = aligned, L = left, R = right, RO = right, outside)]. These are indicated, in cases, where there are multiple marker peaks with the same m/z ratio in the same time window.]

<table>
<thead>
<tr>
<th>Peptide number</th>
<th>Approximate elution time window (min)</th>
<th>m/z of biomarker</th>
<th>Charge (Positive ion mode)</th>
<th>p value for ratio of potential biomarker to reference biomarker</th>
<th>p value for log of ratio of potential biomarker to reference biomarker</th>
<th>p value for unnormalized potential biomarker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8-15</td>
<td>744.35</td>
<td>2</td>
<td>0.1102</td>
<td>0.0339</td>
<td>0.2475</td>
</tr>
<tr>
<td>2</td>
<td>8-15</td>
<td>808.80</td>
<td>2</td>
<td>0.0755</td>
<td>0.0699</td>
<td>0.1483</td>
</tr>
<tr>
<td>3</td>
<td>8-15</td>
<td>675.80</td>
<td>2</td>
<td>0.0926</td>
<td>0.0813</td>
<td>0.2356</td>
</tr>
<tr>
<td>4</td>
<td>11-17</td>
<td>1446.50</td>
<td>1</td>
<td>0.1910</td>
<td>0.0637</td>
<td>0.0688</td>
</tr>
<tr>
<td>5</td>
<td>11-17</td>
<td>1073.46</td>
<td>1</td>
<td>0.0585</td>
<td>0.0790</td>
<td>0.2467</td>
</tr>
<tr>
<td>6</td>
<td>11-17</td>
<td>1307.50</td>
<td>1</td>
<td>0.0217</td>
<td>0.0485</td>
<td>0.1362</td>
</tr>
<tr>
<td>7</td>
<td>11-17</td>
<td>1379.50</td>
<td>1</td>
<td>0.0190</td>
<td>0.0224</td>
<td>0.1064</td>
</tr>
<tr>
<td>8</td>
<td>13-19</td>
<td>566.30</td>
<td>1</td>
<td>0.4779</td>
<td>0.4849</td>
<td>0.0978</td>
</tr>
<tr>
<td>9</td>
<td>13-19</td>
<td>1238.50</td>
<td>1</td>
<td>0.0986</td>
<td>0.1242</td>
<td>0.3367</td>
</tr>
<tr>
<td>10</td>
<td>13-19</td>
<td>603.30</td>
<td>1</td>
<td>0.0716</td>
<td>0.1606</td>
<td>0.3844</td>
</tr>
<tr>
<td>11</td>
<td>13-19</td>
<td>589.30</td>
<td>1</td>
<td>0.2988</td>
<td>0.3484</td>
<td>0.0480</td>
</tr>
<tr>
<td>12</td>
<td>13-19</td>
<td>571.36</td>
<td>1</td>
<td>0.0355</td>
<td>0.0487</td>
<td>0.2834</td>
</tr>
<tr>
<td>13</td>
<td>13-19</td>
<td>593.35</td>
<td>1</td>
<td>0.0217</td>
<td>0.0463</td>
<td>0.2766</td>
</tr>
<tr>
<td>14</td>
<td>13-19</td>
<td>1260.52</td>
<td>1</td>
<td>0.0118</td>
<td>0.0541</td>
<td>0.3631</td>
</tr>
<tr>
<td>15</td>
<td>13-19</td>
<td>764.50</td>
<td>1</td>
<td>0.0113</td>
<td>0.0504</td>
<td>0.3281</td>
</tr>
<tr>
<td>16</td>
<td>13-19</td>
<td>620.30</td>
<td>2</td>
<td>0.1220</td>
<td>0.0366</td>
<td>0.0938</td>
</tr>
<tr>
<td>17</td>
<td>16-22</td>
<td>552.30</td>
<td>1</td>
<td>0.0444</td>
<td>0.0070</td>
<td>0.0517</td>
</tr>
<tr>
<td>18</td>
<td>17-25</td>
<td>536.30</td>
<td>1</td>
<td>0.0975</td>
<td>0.1494</td>
<td>0.1996</td>
</tr>
<tr>
<td>19</td>
<td>17-25</td>
<td>538.30</td>
<td>1</td>
<td>0.0776</td>
<td>0.1695</td>
<td>0.1882</td>
</tr>
<tr>
<td>20</td>
<td>17-25</td>
<td>529.30</td>
<td>1</td>
<td>0.1328</td>
<td>0.0839</td>
<td>0.0784</td>
</tr>
<tr>
<td>21</td>
<td>17-25</td>
<td>558.3(L)</td>
<td>1</td>
<td>0.0801</td>
<td>0.0637</td>
<td>0.1830</td>
</tr>
<tr>
<td>22</td>
<td>17-25</td>
<td>534.30</td>
<td>1</td>
<td>0.0480</td>
<td>0.0672</td>
<td>0.0798</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>23</td>
<td>17-25</td>
<td>518.30</td>
<td>1</td>
<td>0.0147</td>
<td>0.0270</td>
<td>0.0758</td>
</tr>
<tr>
<td>24</td>
<td>17-25</td>
<td>558.3(R)</td>
<td>1</td>
<td>0.0280</td>
<td>0.0366</td>
<td>0.2001</td>
</tr>
<tr>
<td>25</td>
<td>20-27</td>
<td>531.39(R)</td>
<td>1</td>
<td>0.1121</td>
<td>0.1009</td>
<td>0.0662</td>
</tr>
<tr>
<td>26</td>
<td>20-27</td>
<td>504.30</td>
<td>1</td>
<td>0.1104</td>
<td>0.0916</td>
<td>0.1654</td>
</tr>
<tr>
<td>27</td>
<td>20-27</td>
<td>583.35</td>
<td>1</td>
<td>0.0777</td>
<td>0.1625</td>
<td>0.0847</td>
</tr>
<tr>
<td>28</td>
<td>20-27</td>
<td>546.40</td>
<td>1</td>
<td>0.0644</td>
<td>0.0963</td>
<td>0.2014</td>
</tr>
<tr>
<td>29</td>
<td>20-27</td>
<td>531.39(A)</td>
<td>1</td>
<td>0.1195</td>
<td>0.0608</td>
<td>0.0404</td>
</tr>
<tr>
<td>30</td>
<td>20-27</td>
<td>1067.70</td>
<td>1</td>
<td>0.0317</td>
<td>0.0314</td>
<td>0.0716</td>
</tr>
<tr>
<td>31</td>
<td>20-27</td>
<td>1091.70</td>
<td>1</td>
<td>0.0195</td>
<td>0.0224</td>
<td>0.0644</td>
</tr>
<tr>
<td>32</td>
<td>22-32</td>
<td>692.50</td>
<td>1</td>
<td>0.1777</td>
<td>0.2178</td>
<td>0.0904</td>
</tr>
<tr>
<td>33</td>
<td>22-32</td>
<td>513.35</td>
<td>1</td>
<td>0.1064</td>
<td>0.1824</td>
<td>0.0623</td>
</tr>
<tr>
<td>34</td>
<td>22-32</td>
<td>1019.70</td>
<td>1</td>
<td>0.0995</td>
<td>0.1317</td>
<td>0.1400</td>
</tr>
<tr>
<td>35</td>
<td>22-32</td>
<td>1196.80</td>
<td>1</td>
<td>0.0972</td>
<td>0.1366</td>
<td>0.3047</td>
</tr>
<tr>
<td>36</td>
<td>22-32</td>
<td>652.40</td>
<td>1</td>
<td>0.0856</td>
<td>0.0704</td>
<td>0.0991</td>
</tr>
<tr>
<td>37</td>
<td>22-32</td>
<td>567.30</td>
<td>1</td>
<td>0.0688</td>
<td>0.0558</td>
<td>0.1340</td>
</tr>
<tr>
<td>38</td>
<td>22-32</td>
<td>662.40</td>
<td>1</td>
<td>0.0557</td>
<td>0.0546</td>
<td>0.0685</td>
</tr>
<tr>
<td>39</td>
<td>22-32</td>
<td>1123.70</td>
<td>1</td>
<td>0.0442</td>
<td>0.0306</td>
<td>0.1575</td>
</tr>
<tr>
<td>40</td>
<td>22-32</td>
<td>617.4(R)</td>
<td>1</td>
<td>0.0417</td>
<td>0.0681</td>
<td>0.0332</td>
</tr>
<tr>
<td>41</td>
<td>22-32</td>
<td>617.4(RO)</td>
<td>1</td>
<td>0.0428</td>
<td>0.0716</td>
<td>0.0179</td>
</tr>
<tr>
<td>42</td>
<td>27-36</td>
<td>922.60</td>
<td>1</td>
<td>0.0796</td>
<td>0.0762</td>
<td>0.3618</td>
</tr>
<tr>
<td>43</td>
<td>27-36</td>
<td>924.60</td>
<td>1</td>
<td>0.0096</td>
<td>0.0117</td>
<td>0.2433</td>
</tr>
<tr>
<td>44</td>
<td>29-37</td>
<td>766.60</td>
<td>1</td>
<td>0.0737</td>
<td>0.0788</td>
<td>0.0946</td>
</tr>
</tbody>
</table>
selection approach was repeated with complementary correlation analysis. This resulted in dendritic analysis and heat maps. ROC curves were plotted to study the performance of panels of these biomarkers in identifying early stage AD. The best combinations using two or more of the 13 biomarkers provided AUCs of 0.892 and 0.882 (See Figure 5.2 and 5.3). The best combination resulted in an ROC curve with an AUC of 0.892 and provided a sensitivity and specificity $\geq 90\%$, while the second best combination with an AUC of 0.882 provided a sensitivity of 100\% at a specificity of $\sim 75\%$. There were 6 multi-marker models that had an AUC of greater than 0.80 and another 2 with an AUC of $\sim 0.80$.

5.4.4. Cluster Analysis

Cluster analysis was done on the subset of 13 biomarkers. The dendritic analysis suggested that there might be 4-5 groups (See Figure 5.4).

5.5 Discussion

Chemical diagnosis of AD is not currently possible. Biochemical staging of AD using serum biomarkers, however important, likewise does not exist. My hypothesis was that a more global, proteomics approach could be successfully applied to serum to find AD disease stage specific differences in the molecular expression levels in serum. My results suggest strongly that the approach is feasible and the hypothesis correct. Comparison of very mild/mild stage AD (CDR = 0.5, 1.0) samples to moderate stage AD (CDR = 2.0) provided 44 candidate biomarkers that looked promising. The best multi-marker combination was able to diagnose early stage AD with $\sim 90\%$ sensitivity and $>90\%$ specificity. Validation and identification of the biomarkers that appear to be diagnostically useful will be performed in the future.
Table 5.2. Set of 13 biomarkers that appeared in 9% or more of all the multi-marker models.

<table>
<thead>
<tr>
<th></th>
<th>Elution time window (min)</th>
<th>m/z of biomarker</th>
<th>Charge (Positive ion mode)</th>
<th>p value for ratio of potential biomarker to reference biomarker</th>
<th>p value for log of ratio of potential biomarker to reference biomarker</th>
<th>p value for unnormalized potential biomarker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8-15</td>
<td>744.35</td>
<td>2</td>
<td>0.1102</td>
<td>0.0339</td>
<td>0.2475</td>
</tr>
<tr>
<td>2</td>
<td>8-15</td>
<td>808.80</td>
<td>2</td>
<td>0.0755</td>
<td>0.0699</td>
<td>0.1483</td>
</tr>
<tr>
<td>3</td>
<td>11-17</td>
<td>1073.46</td>
<td>1</td>
<td>0.0585</td>
<td>0.0790</td>
<td>0.2467</td>
</tr>
<tr>
<td>4</td>
<td>13-19</td>
<td>571.36</td>
<td>1</td>
<td>0.0355</td>
<td>0.0487</td>
<td>0.2834</td>
</tr>
<tr>
<td>5</td>
<td>16-22</td>
<td>552.30</td>
<td>1</td>
<td>0.0444</td>
<td>0.0070</td>
<td>0.0517</td>
</tr>
<tr>
<td>6</td>
<td>20-27</td>
<td>504.30</td>
<td>1</td>
<td>0.1104</td>
<td>0.0916</td>
<td>0.1654</td>
</tr>
<tr>
<td>7</td>
<td>20-27</td>
<td>1067.70</td>
<td>1</td>
<td>0.0317</td>
<td>0.0314</td>
<td>0.0716</td>
</tr>
<tr>
<td>8</td>
<td>22-32</td>
<td>692.50</td>
<td>1</td>
<td>0.1777</td>
<td>0.2178</td>
<td>0.0904</td>
</tr>
<tr>
<td>9</td>
<td>22-32</td>
<td>662.40</td>
<td>1</td>
<td>0.0557</td>
<td>0.0546</td>
<td>0.0685</td>
</tr>
<tr>
<td>10</td>
<td>22-32</td>
<td>617.4(RO)</td>
<td>1</td>
<td>0.0428</td>
<td>0.0716</td>
<td>0.0179</td>
</tr>
<tr>
<td>11</td>
<td>27-36</td>
<td>922.60</td>
<td>1</td>
<td>0.0796</td>
<td>0.0762</td>
<td>0.3618</td>
</tr>
<tr>
<td>12</td>
<td>27-36</td>
<td>924.60</td>
<td>1</td>
<td>0.0096</td>
<td>0.0117</td>
<td>0.2433</td>
</tr>
<tr>
<td>13</td>
<td>29-37</td>
<td>766.60</td>
<td>1</td>
<td>0.0737</td>
<td>0.0788</td>
<td>0.0946</td>
</tr>
</tbody>
</table>
Figure 5.2. Multimarker model (ROC curve) for biomarkers 552.3, 744.4 and 662.4.

AUC: 0.892
Figure 5.3. Multimarker model (ROC curve) for biomarkers 808.8, 552.3, and 571.4.
Figure 5.4. Biomarker cluster analysis of 13 AD biomarkers subset.
Studies conducted to find diagnostic early stage biomarkers heavily rely on the quality and number of samples available for analysis. My samples were obtained from the Knight ADRC, Washington University School of Medicine. They are known to have stringent criteria for AD diagnosis and staging. The cases and controls were age and gender matched along with the fraction of subjects with the APOE e4 allele. Patients having co-morbidities such as diabetes were not included in the study. This helps increase the likelihood of higher true positive diagnostic rate among the individuals with early stage AD as opposed to other diseases or dementias.

We had fewer samples for comparison between different stages. However, this method has worked very well in all the studies described in the previous chapters in this dissertation. The thoroughness and complexity of the serum proteomic approach has enabled us to observe greater than 8000 spectral peaks and obtain promising results.

Normalized data were used for initial biostatistical analyses to remove any variability if present from run to run.

Panels of biomarkers were developed using a ‘leave one out’ model to provide several panels of markers that had sensitivities and specificities greater than those of individual biomarkers. All further analyses were limited to the 13 biomarkers that appeared in more than 9% of all the models. Using just these 13 candidate markers, I still had 8 panels with acceptable AUCs of ≥0.80. Cluster analysis suggested 4-5 different groups of related biomarkers. The markers with strong correlation might represent a common parent compound or represent similarities in biological staging, processes or pathways.

In conclusion, the serum proteomic approach used here was able to find novel candidate biomarkers for staging mild AD compared with moderate AD.
6.1 Summary

In conclusion, my research can be summarized as follows.

A sizeable initial study of serum from 58 AD cases and 55 age and gender-matched controls was carried out using a novel serum proteomics approach, focusing on lower molecular weight species. This yielded 38 statistically significant biomarkers and another 21 near significant biomarkers. Combinations of these provided panels of markers that diagnose AD correctly ~90% of the time. Currently there are no approved diagnostic biomarkers for AD in blood or cerebrospinal fluid.

In that same data set, I found several biomarkers that were significantly different only in women or only in men. While there is evidence for gender differences in AD, my findings have implications for diagnosis, disease mechanism(s) and the development of effective therapeutics.

Analyzing data from the initial AD study, 44 additional significant and near significant candidate serum biomarkers were found that were quantitatively different between mild and severe AD.

Subsequently, a follow up, blinded validation study with 125 additional, previously unstudied subjects was performed, and data were extracted for the 13 most promising diagnostic biomarkers from the initial study. Of these, 4 of the original markers have replicated successfully, and I have identified these as completely as is currently analytically possible. Sets of these provide accurate diagnosis of ~83% of very early AD with a specificity of ~78%.
6.2 Limitations

As with every research study, potential pitfalls exist. These are detailed as follows.

When I began my discovery study to find serum biomarkers for Alzheimer’s disease using a proteomics approach, I anticipated finding primarily peptide biomarkers. However, I found that the latter half of our useful chromatographic region contained predominantly lipid molecules. Of these, most detected in the positive ion mode were phosphatidylcholines and oxidation products of phosphocholines, because most other lipid classes are not easily positively ionized without forming adducts. Another possible problem is that lipids are very sticky using reversed phase chromatographic phases and may, if one is not careful, get carried over into the next run. For this reason, I conducted a wash run with high organic phase followed by adequate equilibration time after every sample during the replication study, but not during the initial study.

Lipids with few exceptions are stable under the conditions used here and this would make them attractive as biomarkers. Oxidized lipids are often an indication of a diseased state, and they point to mechanisms involved in a disease. However, my procedure used a bicinchoninic acid (BCA) protein assay to quantitate 5 μg of apparent protein in each sample to be loaded on the cLC-MS/MS system. While this might have quantitated the proteins correctly, I am unsure if lipid loading was quantitatively accurate by this method. A possible downside could be that too little specimen was loaded and potential lipid biomarkers might be lost. As is clear from a review of lipidomics research, there are no assays used to quantitate lipids. Most of the lipidomics researchers just use equal volumes for quantitating their samples. These issues would lead to greater random variability and decrease statistical significance, reducing the number of candidates. The ability for results to be replicated then suggests that differences are real and not artifacts of processing. My research has replicated 3 lipid biomarkers and their p-values were
Another challenge is that lipid identification is not nearly as straightforward as protein/peptide identification for which we have search engines, easily recognizable fragments representing b- and y-series ions, etc. I have used the Lipidmaps database and found some possible matches for lipid molecules. However, this database relates to accurate masses and does not provide fragmentation data or allow for fragmentation data to be matched to archived MS data. Neither does it contain oxidation products of lipids in its database. Smaller amounts of serum samples and the exceptionally low concentrations present in those specimens rule out options to characterize them as organic chemists do. Research on lipids is in its infancy and is limited compared to proteins, but occasionally answers can be found by surveying the literature. Even so, this does not always guarantee an exact identity. It is practically impossible to locate the positions of double bonds. If available, lipid standards of expected molecules may be bought and fragmented to see if MS data match the fragmentation spectrum of the biomarker of interest. The selection of commercially available lipids is very limited and contains no oxidized species.

Experience tells me that there should be hydrophobic peptides eluting in some hydrophobic regions where lipids are seen. I have observed this to be true. However, these peptide biomarkers undergo ion suppression due to the presence of more prominent lipid peaks. This reduces the likelihood of finding many hydrophobic peptide biomarkers. I have discovered a charge state +5 peptide biomarker in the hydrophobic region that was highly significant in the validation study. It has isotopic peaks overlapping with a validated lipid biomarker in that region. My efforts to manually analyze all of my data, although very time consuming, makes this possible. Hence, I continued to use this approach in the absence of software that can work with the complexity of our data.
Another thing I have noticed is the presence of lipid dimers in our MS² spectra data. These are potentially formed in the ESI step. Some of these performed well enough to be considered candidates in the initial study. However, given their potential artifactual production, it is not surprising that they did not replicate well. Even so, they are not useful biomarkers if they are not formed inside the body.

Peak overlaps are common for complex biological samples. Using high-resolution instruments helps but does not completely eliminate this problem.

Use of several different software programs to compare group data have been tried and have been disappointing. Manual analysis with all of its problems has frequently helped me find biomarkers that were missed by automated software programs. Each of my files, which were from 1 to 5 GB in size for 2 min windows, limited comparison of 60 cases and 60 controls. My method demanded much time, manual power and high speed computing resources. Such an approach prolonged the analysis time substantially.

I overlaid MS spectra during the initial study to visually look for peaks that appeared quantitatively different between cases and controls. This might not be the best procedure to look for potential biomarkers, given the desire to compare normalized samples. Normalization of the MS data is intended to reduce variability due to sample processing, sample loading, ionization inefficiencies and changes in instrument sensitivity. Normalization then should improve statistical differences, but this did not always occur, suggesting that the endogenous reference marker was not adequate or appropriate for that specific biomarker candidate. A significant concern was that I might have lost good biomarkers due to this technique, both without normalization and with normalization. Unfortunately, finding reliable reference peaks was difficult, especially for lipids, did not improve p-values or improved them negligibly and was not
attempted in the replication study.

Some of the issues with reference peaks included peaks too large or too small, distant in their elution time, markedly different in m/z value, having complicated isotope envelopes, uncertainty as to their quantitative neutrality between cases and controls, their absence in some specimens and more. I have strived to find the best reference markers possible.

Another concern in using internal reference markers was that a lipid molecule was used to normalize a peptide biomarker. While this is not ideal, I have attempted, where possible, to make reference and marker chemically similar.

Ideally, I could use a range of m/z values that takes in just one monoisotopic peak when recording extracted ion chromatogram intensities. However, this usually led to very low extracted ion chromatogram peak intensities for several biomarkers, so low that it could affect the accuracy due to low signal to noise. I overcame this problem by using slightly larger ranges such that they covered two isotopic peaks. This came with a potential complication of its own, risking the quantitation of neighboring peaks having some level of overlap with our molecule of interest.

While there are no FDA approved drugs given to AD patients, many of the subjects may be on some type of medication. Subjects had no major intercurrent diseases, but I cannot rule out some medication use among subjects, and this could modify the quantities of biomarkers. Certainly, the biomarkers I have found are not drugs or drug metabolites. Availability of more patient data would be reassuring.

I have studied only a cross section of people representative of AD and matched, non-demented controls. However, in a larger cohort or individuals from another part of the world, there is always a possibility of these biomarkers working differently. Remember, there is
evidence that AD is not uniform in its cause, presentation or progression, leaving open the possibility that there may be environmental factors. Longitudinal studies, while currently almost impossible to carry out, would be very important.

Also, patients with AD have several other diseases since they are typically old. Other diseases may affect levels of AD biomarkers found in these studies. My study used cases and controls without co-morbidities, and assumed that the levels will continue to be different in individuals with AD and comorbidities.

In my initial sample set, I had several samples with CDR 2.0 and 3.0, which are representative of severe stage disease, and very few with CDR 0.5 and 1.0 that are indicative of mild stage AD. In the replication study, I had many more very mild and mild AD subjects and no CDR 3 subjects. All the stages were studied in the initial set. This might have reduced replication of some biomarkers.

6.3 Future Research

This was and continues to be a vast project. Although I finished preliminary studies and assessed validation of the 13 most promising biomarkers found in the initial study, there is much more that can be done. A few possible future research studies are as follows:

The validation study should be extended to the 38 statistically significant biomarkers and the additional 21 near significant biomarkers found in the initial study. Since two of the biomarkers, 1568.2 and 1618.2, among the 13 were lipid dimers and two others, 708.3 and 660.4, could not be quantified with confidence, I have really studied only a small portion of the potential biomarkers. The fact that 804.55 found between 39.5–41.5 min was validated successfully, but did not belong to the list of 13 biomarkers, proves that more useful results can be obtained by
evaluating the other biomarkers from the initial study. Further biostatistics involving regression and correlation studies could be added. Multi-marker panels should be reconsidered after studying all of the biomarkers from the initial study.

Finding more biomarkers among the currently unevaluated candidates found in the initial study would then require more chemical characterization and identification studies. Hence, more fragmentation studies are anticipated in the future. Additionally, I don't have absolute identities of our oxidized biomarkers. These could be investigated further until complete identities are obtained.

Data can be further statistically evaluated to build multi-marker panels for comparisons between very mild and mild AD (CDR = 0.5, 1.0) and moderate AD (CDR = 2.0) for biomarkers found in the initial study when these stages were compared (see Chapter 5).

Further, biomarkers that were found while studying different stages have not been validated. A study could be undertaken to validate and identify these.

Also, biomarkers that were male and female specific have not yet been validated. Replication studies with identification of these in the same data set that I used can be done in the future as well. Also, the validated biomarkers have not been tested as a function of gender.

It would be interesting to obtain serial specimens from AD patients and look for biomarkers that predict those who have a slow time course for their disease and those that demonstrate rapid decline in their cognitive function. Such markers would be valuable in designing drug studies.

Once all validation and identification studies are done, disease mechanisms can be investigated.

For diagnostic purposes, if possible, there could be studies to build more quantitative assays using some of the biomarkers for diagnosing AD.
Other labs could test these biomarkers in a clinical setting and check for further replication and reproducibility.

Potential treatments could be found, once useful pathways are found, using these biomarkers to guide enrollment into those investigations.

Additionally, it would be important to develop programs or use new software to allow faster and more effective data analysis and statistical comparisons. Likewise, better search engines to facilitate peptide identification would be valuable, especially those that take more sophisticated approaches to peptide sequencing. My sample preparation method uses acetonitrile precipitation eliminating most proteins and hence, most peptide identifications rely on sequencing and can be complex and not typically successful on current search engines. Additionally, most online search engines cannot handle post-translational modifications or oxidations that are encountered in my studies. This has required me to do manual de novo sequencing routinely even for short peptides, and the process is very time consuming. When there are many peptides to be sequenced for a project, a good de novo sequencing program would be very useful.

With the computational proteomics expertise of Dr. Prince and programming guidance from my husband (Jigar B. Shah), I wrote a program, which employed de novo sequencing, to overcome weaknesses of the online search engines like MASCOT. With additional work, this could help in better and faster identification of peptides. However, better peak normalization procedures would be required to allow such a program to succeed as well as the ability to account for more complex post-translational modifications. Efforts could be made in this direction in the future as well as in developing programs that can pick out peaks, compare them and calculate the differences between cases and controls statistically.
REFERENCES


Henkel, A. W., K. Müller, P. Lewczuk, T. Müller, K. Marcus, J. Kornhuber and J. Wiltfang (2012). "Multidimensional plasma protein separation technique for identification of


7.1 Quadrupole Mass Spectrometer

7.1.1. Equations of Motion

The voltage distribution (potential) experienced by ions as they enter the mass analyzer hinges on two factors: the shape of the electrodes and the time dependent nature of the voltage applied to the electrodes, which can be illustrated as follows:

\[ \phi = (U_{DC} + V_{RF} \cos \omega t) \frac{x^2 - y^2}{2r_0^2} \]

where \( U_{DC} \) is the magnitude of the DC voltage applied, \( V_{RF} \) is the magnitude of the RF voltage, \( \omega \) is the angular frequency of the RF voltage \( (2\pi f) \), \( r_0 \) is the distance from the center axis \( (z \text{ axis}) \) to any electrode and \( x \) and \( y \) are distances along the given coordinate axes.

The electric field at any point of time within the electrodes can be calculated by taking the partial derivative of the potential defined above as a function of the distance along any of the coordinate axes.

\[ E_x = -\frac{d\Phi}{dx} = -(U_{DC} + V_{RF} \cos \omega t) \frac{x}{r_0^2} \]
\[ E_y = -\frac{d\Phi}{dy} = (U_{DC} + V_{RF} \cos \omega t) \frac{y}{r_0^2} \]
\[ E_z = -\frac{d\Phi}{dz} = 0 \]

Acceleration of the ion along the coordinate axes can be defined as the rate of change of velocity (which itself is the rate of change of displacement of the ion along the axes) of the ion.
along the coordinate axes. That’s why it can be defined as:

\[ a_x = \frac{d}{dt} \left( \frac{dx}{dt} \right) = \frac{d^2 x}{dt^2} \]

\[ a_y = \frac{d}{dt} \left( \frac{dy}{dt} \right) = \frac{d^2 y}{dt^2} \]

By Newton’s law of motion,

\[ F = ma \]

\[ F_x = ma_x = m \frac{d^2 x}{dt^2} \]

\[ F_y = ma_y = m \frac{d^2 y}{dt^2} \]

The force exerted on the ions can also be defined as the product of the charge of the ion (e) and the magnitude of the electric field (E).

\[ F = Ee \]

\[ F_x = -(U_{DC} + V_{RF} \cos \omega t) \frac{ex}{r_0^2} \]

\[ F_y = (U_{DC} + V_{RF} \cos \omega t) \frac{ey}{r_0^2} \]

\[ F_z = 0 \]

\[ m \frac{d^2 x}{dt^2} = -(U_{DC} + V_{RF} \cos \omega t) \frac{ex}{r_0^2} \]

\[ \frac{d^2 x}{dt^2} = -(U_{DC} + V_{RF} \cos \omega t) \frac{ex}{mr_0^2} \]

\[ \frac{d^2 x}{dt^2} + (U_{DC} + V_{RF} \cos \omega t) \frac{ex}{mr_0^2} = 0 \]

Similarly,

\[ \frac{d^2 y}{dt^2} - (U_{DC} + V_{RF} \cos \omega t) \frac{ey}{mr_0^2} = 0 \]
\[ \frac{d^2 z}{dt^2} = 0 \]

The solutions to the differential equations above give us a complete description of the trajectory of any ion in terms of each ion’s initial conditions. As can be seen, the voltages of the electrodes do not affect the acceleration of the ion along the z-axis. If we define parameters a and q such that

\[ a = \frac{4eU_{DC}}{\omega^2 mr_0^2} \]

\[ q = \frac{2eV_{RF}}{\omega^2 mr_0^2} \]

the differential equations can be restated now as

\[ \frac{d^2 x}{dt^2} + \frac{\omega^2}{4} (a + 2q \cos \omega t)x = 0 \]

\[ \frac{d^2 y}{dt^2} - \frac{\omega^2}{4} (a + 2q \cos \omega t)y = 0 \]

These equations can be transformed into the accepted form of Mathieu’s differential equation by doing the following change of variables:

\[ \xi = \frac{\omega t}{2} \]

Mathieu’s equation consists of a second derivative of the distances along the coordinate axes as a function of \( \xi \) that can be calculated using the chain rule:

\[ \frac{du}{d\xi} = \frac{du}{dt} \frac{dt}{d\xi} \]  (where \( u \) is \( x \) or \( y \))

\[ \frac{d^2 u}{d\xi^2} = \frac{d}{d\xi} \left( \frac{du}{dt} \frac{dt}{d\xi} \right) \]

By the product rule of derivatives,
\[ \frac{d^2u}{d\xi^2} = \frac{d^2u}{dt^2} \left(\frac{du}{dt}\right) + \frac{d}{d\xi} \left(\frac{du}{dt}\right) \frac{d\xi}{dt} \]

Based on the definition of \( \xi \),

\[ \xi = \frac{\omega t}{2} \]

\[ t = \frac{2\xi}{\omega} \]

\[ \frac{dt}{d\xi} = \frac{2}{\omega} \]

\[ \frac{d^2t}{d\xi^2} = 0 \]

Therefore, the equation now becomes

\[ \frac{d^2u}{d\xi^2} = \frac{2}{\omega} \frac{d}{d\xi} \left(\frac{du}{dt}\right) \]

We can calculate the right part of the equation above as follows:

\[ \frac{du}{dt} = \frac{du}{d\xi} \frac{d\xi}{dt} \]

\[ \frac{d}{d\xi} \left(\frac{du}{dt}\right) = \frac{d}{d\xi} \left(\frac{du}{d\xi} \frac{d\xi}{dt}\right) \]

By quotient rule of derivatives,

\[ \frac{d}{d\xi} \left(\frac{du}{dt}\right) = \frac{\frac{dt}{d\xi} \frac{d}{d\xi} \left(\frac{du}{dt}\right) - \frac{du}{d\xi} \frac{d}{d\xi} \left(\frac{dt}{d\xi}\right)}{\left(\frac{dt}{d\xi}\right)^2} \]

\[ \frac{d^2}{d\xi^2} \left(\frac{du}{dt}\right) = \frac{\frac{dt}{d\xi} \frac{d^2u}{d\xi^2} - \frac{du}{d\xi} \frac{d^2t}{d\xi^2}}{\left(\frac{dt}{d\xi}\right)^2} \]
Since \( \frac{d^2 t}{d \xi^2} = 0 \)

\[
\frac{d}{d \xi} \left( \frac{d u}{d t} \right) = \frac{d t}{d \xi} \frac{d^2 u}{d t^2} \frac{d^2 u}{d \xi^2} \left( \frac{d t}{d \xi} \right)^2
\]

\[
\frac{d}{d \xi} \left( \frac{d u}{d t} \right) = \frac{d^2 u}{d \xi^2} \frac{d t}{d t} \frac{d^2 u}{d \xi^2} \frac{d t}{d \xi}
\]

\[
\frac{d}{d \xi} \left( \frac{d u}{d t} \right) = \frac{d^2 u}{d t^2} \frac{d^2 u}{d r^2} \frac{d t}{d \xi}
\]

\[
\frac{d}{d \xi} \left( \frac{d u}{d t} \right) = \frac{d^2 u}{d t^2} \frac{d t}{d \xi}
\]

\[
\frac{d}{d \xi} \left( \frac{d u}{d t} \right) = \frac{2 d^2 u}{\omega d t^2}
\]

Therefore, the final equation now becomes

\[
\frac{d^2 u}{d \xi^2} = 4 \frac{d^2 u}{\omega^2 d t^2}
\]

\[
\frac{d^2 u}{d \xi^2} = 4 \frac{d^2 u}{\omega^2} \left( - \frac{\omega^2}{4} \right) \left( a + 2q \cos \omega t \right) u
\]

\[
\frac{d^2 u}{d \xi^2} + \left( a + 2q \cos \omega t \right) u = 0
\]

which is the canonical form of Mathieu’s equation (Miller and Denton 1986).

### 7.1.2 Stability Diagram

An examination of the solutions of Mathieu’s equations indicates that stable oscillatory trajectories can be established with certain values of \( a \) and \( q \), which will allow ions to safely pass
through the quadrupole mass analyzer. All other values will lead to unstable trajectories that cause the ions to be filtered out. Since Mathieu’s equations are represented in terms of two variables, a simple graph can be plotted showing stable solutions of the equation for values of a and q, making it easier to visualize. Figure 7.1 shows the stable solutions of Mathieu’s equations in the X direction when a and q are positive. This basically shows the values of a and q for which the ions will have a stable path through the analyzer in the X-Z plane. Figure 7.2 shows the solutions in the Y direction for negative values of a and q, which indicates the safe passage of ions along the Y-Z plane. As we can see, both the graphs are mirror images of each other.

Overlaying the two graphs gives us solutions that would allow safe trajectories for ions in all directions. Summarily, the values of a and q that lie at the intersection of the two graphs (Figure 7.3) will give us solutions that would let the ions pass through the filter safely.

Since quadrupole mass analyzers are used at low voltages, I am more interested in the lower values of a and q in the stable regions in Figure 7.3. This figure shows a magnified image of the stability region of the overlaid graph. Every mass has its own a and q values under a set of operating conditions. Thus, masses can be plotted on the stability region of a and q. This forms the basis of the mass filtering mechanism, since values of a and q can be chosen such that only ions within a range of masses are allowed stable trajectories across the analyzer. Consider two ions with masses $m_0$ and $m_1$. If conditions are chosen such that for $m_0$ the values are $a_0$ and $q_0$, $m_1$ can be calculated taking the ratio of $a_s$ and $q_s$ as shown in the following equations.

$$\frac{a_1}{a_0} = \frac{8eU/m_1 \text{ } w^2 r_0^2}{8eU/m_0 \text{ } w^2 r_0^2} = \frac{m_0}{m_1} \rightarrow a_1 = \left(\frac{m_0}{m_1}\right) a_0$$

$$\frac{q_1}{q_0} = \frac{4eV/m_1 \text{ } w^2 r_0^2}{4eV/m_0 \text{ } w^2 r_0^2} = \frac{m_0}{m_1} \rightarrow q_1 = \left(\frac{m_0}{m_1}\right) q_0$$
Figure 7.1. Stability diagram showing stable solutions of Mathieu’s equation in the X direction.

Figure 7.2. Stability diagram showing stable solutions of Mathieu’s equation in the Y direction.

Figure 7.3. Overlaid stability diagram showing the stable regions at the intersection points.

This means that all masses can be represented by a straight line starting from (0,0) in the a and q graph. The slope of that line will indicate the filtering of masses of ions that would be allowed to pass. Figure 7.4 shows a magnified image of the stability region in the a and q graph with three different lines indicating different values of DC and AC potentials applied to the quadrupole. As one can see, the top line would not allow any ions through, since none of the masses lie within the green stability region. The second and third lines would allow masses to pass through that lie in the green stability region.

7.2 Time-of-Flight Mass Spectrometer

The time of flight mass spectrometer’s entry point consists of two conducting plates. Let us assume that they are separated by distance d. If voltage is applied to one of the plates, an electric field will be created. Now, if we pass ions through this electric field, the potential energy of an ion is equal to the charge of the ion (q) times the strength of the electric field at the point at which it was created (U).

\[ E_{PE} = qU \]

The charge of an ion is equal to the product of an integer number z of electron charges e.

\[ q = ez \]

Therefore, the potential energy of the ion now becomes

\[ E_{PE} = ezU \]

Because there is a field gradient between the two plates and like charges repel each other, the ion will experience a push towards the grounded plate. Due to this push, the potential energy of the ion is converted into kinetic energy that is defined by

\[ E_{KE} = \frac{1}{2}mv^2 \]
where \( m \) is the mass of the ion and \( v \) is the velocity imparted on the ion. Because potential energy of the ion is converted into kinetic energy, I can equate both equations and find the value for the time-of-flight of the ion:

\[
E_{PE} = E_{KE}
\]

\[
ezU = \frac{1}{2} mv^2
\]

\[
v^2 = \frac{2ezU}{m}
\]

\[
v = \sqrt{\frac{2ezU}{m}}
\]

\[
v = \sqrt{\frac{2eU}{m}}
\]

Thus, it is seen that the velocity imparted to the ion because of the voltage applied is inversely proportional to the square root of the mass of the ion. If the grounded plate is a wire mesh or has a slit in it for the ion to pass through, the ion will pass through and achieve a velocity \( v \) defined above. After passing through the grounded plate, the ion is in a field-free zone and drifts at a constant velocity. Therefore, the time required for the ion to travel a distance \( D \) post acceleration can be calculated as follows:

\[
t_{TOF} = \frac{D}{v}
\]

\[
t_{TOF} = \frac{D}{\sqrt{2eU} \sqrt{\frac{z}{m}}}
\]

\[
t_{TOF} = \frac{D}{\sqrt{2eU} \sqrt{\frac{m}{z}}}
\]

Thus, the time required for an ion to travel in a field-free region is directly proportional to the square root of the mass-to-charge ratio of the ion (Figure 7.5).
Figure 7.4. Magnified image of the stability region. Reprinted from (Group 1998) with permission.

©University of Richmond, 1998.
Figure 7.5. Concept of time-of-flight mass spectrometer.