IDENTIFICATION, CHARACTERIZATION, AND QUANTIFICATION OF DICARBONYL ADDUCTS IN THE PLASMA PROTEOME IN TYPE-2 DIABETES

by

Michael John Kimzey

Copyright © Michael John Kimzey 2011

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHARMACEUTICAL SCIENCES

In Partial Fulfillment of the Requirements For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2011
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Michael John Kimzey entitled “Identification, Characterization, and Quantification of Dicarbonyl Adducts in the Plasma Proteome in Type-2 Diabetes” and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Dr. Serrine Lau
Date: 3/30/11

Dr. Terrence Monks
Date: 3/30/11

Dr. Georg Wondrak
Date: 3/30/11

Dr. George Tsapraillis
Date: 3/30/11

Dr. Craig Stump
Date: 3/30/11

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Serrine S. Lau
Date: 3/30/11
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library. Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Michael John Kimzey
ACKNOWLEDGEMENT

I would like to express my deep and sincere gratitude to my advisor, Dr. Serrine Lau. Her vision, support, and guidance made this project possible. I am truly fortunate to have a committee comprised of respected scientists who helped shape this project. Dr. Terrence Monks has always challenged me to think about the deeper biological implications of my work. I would also like to thank Dr. George Tsapralis for his advice, knowledge, and expertise in mass spectrometry and proteomics. Dr. Georg Wondrak has guided me toward critical thinking about the roles of dicarboxyls. Dr. Craig Stump’s clinical perspective has been invaluable in directing the course of this research.

Dr. Hussein Yassine is a brilliant physician and I will always be fond of times spent together as budding scientists. I would also like to thank past members of the Lau lab for their help in assimilating me into the lab. I would also like to thank members of the Proteomics Core facility, especially Yelena Feinstein and Andrea Hunt for their advice and patience in training me on the MS instruments. I would also like to thank Dr. Dean Billheimer for his patience and statistical advice, as well as Dr. Vijay Gokhale for counsel in molecular modeling. I want to thank my friends and family, especially Raina for seeing me through tough times. Finally, I would like to dedicate this work to my mom, who let me make potions in the garage.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... 9

LIST OF ABBREVIATIONS .............................................................................................. 13

ABSTRACT ....................................................................................................................... 16

CHAPTER 1: INTRODUCTION ............................................................................................ 18

1.1. General Comments ................................................................................................. 18

1.2. Chemistry of glyco-oxidation and formation of AGEs ............................................ 19

1.2.1. The Maillard reaction ......................................................................................... 19

1.2.2. Glycation ........................................................................................................... 19

1.2.3. Glyco-oxidation ................................................................................................. 22

1.2.4. Glucosone .......................................................................................................... 24

1.2.5. 3DG .................................................................................................................... 25

1.2.6. MG .................................................................................................................... 25

1.2.7. AGEs derived from dicarbonyl intermediates ..................................................... 27

1.3. Advanced Glycation End Products and Diabetic Complications ............................ 31

1.3.1. AGEs and disease .............................................................................................. 31

1.3.2. AGEs and macrovascular diabetic complications ............................................. 32

1.3.3. AGEs and microvascular complications ......................................................... 34

1.3.4. AGEs and retinopathy ...................................................................................... 35

1.3.5. AGEs and neuropathy ...................................................................................... 35

1.3.6. AGEs and nephropathy .................................................................................... 36

1.4. Site-specific targets of glyco-oxidation ................................................................... 38

1.4.1. Specificity .......................................................................................................... 38

1.4.2. MG sites on proteins ........................................................................................ 38

1.5. Methods to analyze advanced glycation ................................................................ 40

1.5.1. AGE measurement by fluorescence ................................................................ 40

1.5.2. AGE measurement by immunoassay .............................................................. 41

1.5.3. Problems with measurement of free dicarbonyls .............................................. 42

1.5.4. Analysis of dicarbonyl adducts ........................................................................ 44

1.5.5. Site-specificity of MG adducts ......................................................................... 45

1.6. Analysis of peptides using mass spectrometry ...................................................... 46

1.6.1. Ionization sources ............................................................................................ 47

1.6.2. Mass analyzers .................................................................................................. 47

1.6.3. Peptide fragmentation ....................................................................................... 51

1.6.4. Bioinformatics .................................................................................................. 51

1.6.5. MRM quantification ........................................................................................ 52

1.7. Dissertation aims ................................................................................................... 54
# TABLE OF CONTENTS -- CONTINUED

## CHAPTER 2: IDENTIFICATION OF METHYLGLYOXAL MODIFICATIONS IN THE PLASMA PROTEOME

- 2.1. Introduction ............................................................................................................. 56
- 2.2. Methods .................................................................................................................... 60
  - 2.2.1. Materials ............................................................................................................. 60
  - 2.2.2. Subject selection ............................................................................................... 60
  - 2.2.3. Removal of abundant proteins ......................................................................... 60
  - 2.2.4. Enzymatic digestion ......................................................................................... 60
  - 2.2.5. MuDPIT (LC-LC-MS/MS) ............................................................................. 62
  - 2.2.6. *In vitro* MG modification ............................................................................. 65
  - 2.2.7. Database searching and validation ................................................................. 65
  - 2.2.8. Validation using accurate mass MS/MS ......................................................... 67
  - 2.2.9. Molecular modeling ......................................................................................... 68
- 2.3. Results .................................................................................................................... 69
  - 2.3.1. MG sites in the plasma proteome .................................................................. 69
  - 2.3.2. Microenvironment charged residue motif in MG sites .................................. 71
  - 2.3.3. Shotgun proteomics to search for *in vivo* modifications ............................ 83
  - 2.3.4. Determination of the site of MG modification in a model peptide ............. 86
- 2.4. Discussion ............................................................................................................... 89

## CHAPTER 3: VALIDATION OF DRUG SITE I AS A TARGET OF METHYLGLYOXAL ADDUCTION OF HUMAN SERUM ALBUMIN

- 3.1. Introduction ............................................................................................................. 95
- 3.2. Methods .................................................................................................................... 99
  - 3.2.1. Materials ............................................................................................................. 99
  - 3.2.2. 15-keto Prostaglandin E2 conversion assay ................................................... 99
  - 3.2.3. Prodan displacement assay .............................................................................. 100
  - 3.2.4. Affinity docking and molecular dynamics simulations ............................... 101
  - 3.2.5. Surface area calculations and molecular modeling .................................... 102
- 3.3. Results .................................................................................................................... 103
  - 3.3.1. MG-induced inhibition of PGE2 catalysis ....................................................... 103
  - 3.3.2. Displacement of prodan from HSA-prodan complex by MG ...................... 105
  - 3.3.3. Solvent accessibility of HSA side chains ......................................................... 109
  - 3.3.4. Molecular dynamics and affinity docking of MG-modified R257 .............. 112
- 3.4. Discussion ............................................................................................................... 115

## CHAPTER 4: DETERMINATION OF THE HIGH-AFFINITY SITES FOR OXIDATION AND DICARBONYL ADDUCTION OF HUMAN SERUM ALBUMIN

- 3.4. Discussion ............................................................................................................... 117
TABLE OF CONTENTS -- CONTINUED

4.1. Introduction.........................................................................................117
4.2. Methods.............................................................................................126
  4.2.1. Materials........................................................................................126
  4.2.2. Oxidation of peptides containing M, W, or C.................................126
  4.2.3. Modification of synthetic peptides with dicarboxyls....................128
  4.2.4. Transition optimization using synthetic peptides.........................130
  4.2.5. Preparation of in vitro oxidized and dicarbonyl modified HSA......130
  4.2.6. LC-MRM.....................................................................................132
4.3. Results...............................................................................................139
  4.3.1. Oxidation of peptides.....................................................................139
  4.3.2. Dicarbonyl modification of peptides.............................................141
  4.3.3. Transition optimization..............................................................145
  4.3.4. Oxidation hotspots in HSA..........................................................152
  4.3.5. Dicarbonyl hotspots in HSA.........................................................159
4.4. Discussion..........................................................................................166

CHAPTER 5: SITE-SPECIFIC QUANTIFICATION OF GLYCO-OXIDIZED HUMAN SERUM ALBUMIN FROM PLASMA.........................................................172

  5.1. Introduction........................................................................................172
  5.2. Methods............................................................................................173
    5.2.1. Materials......................................................................................173
    5.2.2. Subject selection..........................................................................173
    5.2.3. Blood sample handling and storage.............................................174
    5.2.4. Tryptic digestion of whole plasma.............................................174
    5.2.5. LC-MRM.....................................................................................177
    5.2.6. Data analysis................................................................................178
  5.3. Results...............................................................................................180
    5.3.1. Identification of oxidized and dicarbonyl-modified sites in vivo.....180
    5.3.2. Normalization of MRM data.......................................................181
    5.3.3. Clinical data and PCA.................................................................186
    5.3.4. Correlations between HSA modifications and clinical data........186
    5.3.5. ANOVA of discrete variables......................................................192
    5.3.6. The effect of metformin..............................................................196
  5.4. Discussion..........................................................................................201

CHAPTER 6: CONCLUDING REMARKS..........................................................207

  6.1. Introduction........................................................................................207
  6.2. Identification of adduct sites...........................................................208
  6.3. Functional consequences of dicarbonyl-induced protein damage........211
  6.4. Relative reactivity of sites for oxidation and dicarbonyl modification...212
TABLE OF CONTENTS – CONTINUED

6.5. Glyco-oxidation *in vivo* and metformin.................................................................213

CHAPTER 7: FUTURE DIRECTIONS..................................................................................216

7.1. Development of site-specific antibodies.................................................................216
7.2. Metformin intervention study....................................................................................217
7.3. Characterization of other reactive dicarbonyls.......................................................218

APPENDIX A: SUBJECT CONSENT FORMS AND IRB APPROVAL.........................220

A.1. Subject’s consent form.............................................................................................220
A.2. HSPP correspondence form......................................................................................225

REFERENCES..................................................................................................................226
LIST OF FIGURES

Figure 1.1. Overview of the formation of AGEs from early glycation and glyco-oxidation.................................................................................................21

Figure 1.2. Oxidative and metabolic formation of reactive dicarboxyls glucosone, 3-deoxyglucosone, and methylglyoxal..........................................................23

Figure 1.3. Metabolic sources of methylglyoxal........................................................................26

Figure 1.4. Lysine and arginine AGEs..................................................................................28

Figure 1.5. Dicarbonyl-derived hydroimidazolones................................................................30

Figure 1.6. Major pathways of diabetic complications..........................................................33

Figure 1.7. Electrospray ionization and triple quadrupole mass spectrometer..................48

Figure 1.8. Common ions from collision-induced dissociation of peptides......................50

Figure 1.9. Multiple reaction monitoring (MRM) scheme....................................................53

Figure 2.1. Chemistry of MG-arginine adduction.................................................................58

Figure 2.2. Approximate protein abundance in plasma.....................................................61

Figure 2.3. MuDPIT scheme...............................................................................................64

Figure 2.4. Bioinformatic pipeline from raw MS/MS data to protein identification and localization of modified arginines.................................................................66

Figure 2.5. Site-specific MG modifications in vitro................................................................70

Figure 2.6. Linear motifs adjacent to MG sites.....................................................................72

Figure 2.7. Methylglyoxal sites in human serum albumin.................................................74

Figure 2.8. Charged residues adjacent to MG sites in HSA...............................................75

Figure 2.9. Target arginines in drug site I and II pockets in HSA.......................................77

Figure 2.10. R61 MG site in apolipoprotein A1 is near a charged residue..........................79

Figure 2.11. Catalytic subunit of plasminogen.....................................................................80
LIST OF FIGURES– CONTINUED

Figure 2.12. Plasminogen MG sites as examples of the charged residue motif........81

Figure 2.13. Microenvironment of alpha 1 acid glycoprotein 1 and transferrin
MG sites..........................................................................................82

Figure 2.14. Unmodified, in vitro MG-modified, and in vivo MG-modified MS/MS
spectral profiles of 492HRHPDEAAFFDTASTGK508 peptide in
fibrinogen α chain.........................................................................84

Figure 2.15. Accurate mass MS/MS of the R257 MG-modified peptide indicates
arginine as the site of modification..............................................87

Figure 2.16. Accurate mass MS/MS of MG-modified peptide shows that the R+54
modification is not on the C-terminal lysine.................................88

Figure 2.17. Neighboring positive charge decreases the pKa and increases the
nucleophilicity of arginine..............................................................93

Figure 2.18. Base-catalyzed dehydration of dihydroimidazolidine by a neighboring
carboxylate.........................................................................................94

Figure 3.1. Conversion of 15-keto PGE\textsubscript{2} to 15-keto PGB\textsubscript{2} in drug site I of HSA......98

Figure 3.2. MG-modified HSA inhibits prostaglandin catalysis at drug site I..........104

Figure 3.3. Stoke’s shift of HSA-prodan complex yields fluorescent maxima at
465 nm..........................................................................................106

Figure 3.4. Prodan is displaced from drug site I by MG modification of HSA........108

Figure 3.5. Solvent accessibility of oxidizable (M,W, or C) and arginine (R)
HSA residues......................................................................................111

Figure 3.6. Methylglyoxal modification of R257 in drug site I changes
conformation of warfarin binding......................................................113

Figure 4.1. Hydroimidazolone adducts from dicarbonyl modification of arginine
and the corresponding mass changes..............................................118

Figure 4.2. Oxidation of methionine and corresponding mass changes.............120
LIST OF FIGURES– CONTINUED

Figure 4.3. Major oxidation products of tryptophan and the corresponding mass changes.................................................................122
Figure 4.4. Oxidation of cysteine...............................................................124
Figure 4.5. Tryptic HSA sequences containing major oxidizable residues........127
Figure 4.6. Tryptic HSA sequences containing arginine as a missed cleavage......129
Figure 4.7. MS of oxidized methionine and unmodified HSA peptide containing M548........................................................................134
Figure 4.8. MS of oxidized tryptophan and unmodified HSA peptide AWAVAL...135
Figure 4.9. MS of oxidized cysteine peptide containing C34.................................136
Figure 4.10. MS of MG and 3DG modification of HSA peptide containing R410.....137
Figure 4.11. MS of glucosone modification of HSA peptide containing R410........138
Figure 4.12. MS/MS of unmodified and oxidized methionine (M+16) peptides to select candidate ions to test for MRM..........................144
Figure 4.13. MRM collision energy (CE) optimization of unmodified and oxidized M548 peptides................................................................145
Figure 4.14. Validated transitions for unmodified HSA peptides.........................147
Figure 4.15. Validated transitions for MG-modified HSA peptides.....................148
Figure 4.16. Validated transitions for 3DG-modified HSA peptides....................149
Figure 4.17. Validated transitions for glucosone-modified HSA peptides..........150
Figure 4.18. Validated transitions for oxidized HSA peptides.............................151
Figure 4.19. Site-specific HSA methionine oxidation in vitro..............................154
Figure 4.20. Products of the oxidation of tryptophan at W214 in HSA...............156
Figure 4.21. Products of the oxidation of cysteine at C34 in HSA.........................158
LIST OF FIGURES– CONTINUED

Figure 4.22. Relative modification of arginine sites by methylglyoxal.......................161

Figure 4.23. Cooperativity and Ki of MG adduction to arginine sites is independent of solvent accessibility.................................................................163

Figure 4.24. Relative modification of arginine sites by 3-deoxyglucosone...............165

Figure 4.25. Relative modification of arginine sites by glucosone.............................167

Figure 4.26. Relative modification of HSA at R410 with three dicarbonyls.............168

Figure 5.1. Assay scheme for preparation of HSA peptides from whole plasma.....176

Figure 5.2. Example of a LC-MRM run.................................................................179

Figure 5.3. Scree plots for PCA of oxidized methionine, unmodified HSA, and glycation/glucose metrics.................................................................183

Figure 5.4. Hierarchical clustering of the AUC values for the three dicarbonyls at three R sites.................................................................185

Figure 5.5. Distribution fitting of variables to test for normality.............................188

Figure 5.6. Correlations between continuous clinical data and modified HSA.......191

Figure 5.7. Scatterplot of MG vs. PC glycation/glucose does not identify subjects with nephropathy.................................................................195

Figure 5.8. ANOVA of normalized MG and glucosones between normoglycemic subjects, diabetic subjects not taking metformin, and diabetic subjects taking metformin.................................................................198

Figure 5.9. ANOVA of PC glycation/glucose between normoglycemic subjects, diabetic subjects not taking metformin, and diabetic subjects taking metformin.................................................................200

Figure 6.1. Adjacent charge motif for arginine adduction by dicarbonyls............210
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>Advanced glycation endproduct</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>BUN</td>
<td>Blood urea nitrogen</td>
</tr>
<tr>
<td>C</td>
<td>Cysteine</td>
</tr>
<tr>
<td>CE</td>
<td>Collision energy</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-induced dissociation</td>
</tr>
<tr>
<td>CKD</td>
<td>Chronic kidney disease</td>
</tr>
<tr>
<td>CEL</td>
<td>Carboxyethyl lysine</td>
</tr>
<tr>
<td>CML</td>
<td>Carboxymethyl lysine</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>3-DG</td>
<td>3-deoxyglucosone</td>
</tr>
<tr>
<td>DN</td>
<td>Diabetic nephropathy</td>
</tr>
<tr>
<td>DP</td>
<td>Declustering potential</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>FL</td>
<td>Fructosyl-lysine</td>
</tr>
<tr>
<td>FN3K</td>
<td>Fructosamine 3-kinase</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated glomerular filtration rate</td>
</tr>
<tr>
<td>GSH</td>
<td>Glutathione</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>HbA1C</td>
<td>Glycated hemoglobin</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>Homeostatic model assessment-insulin resistance</td>
</tr>
<tr>
<td>HSA</td>
<td>Human serum albumin</td>
</tr>
<tr>
<td>Ki</td>
<td>Equilibrium dissociation constant</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-density lipoprotein</td>
</tr>
<tr>
<td>LSD</td>
<td>Least significant difference</td>
</tr>
<tr>
<td>M</td>
<td>Methionine</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser desorption ionization time-of-flight</td>
</tr>
<tr>
<td>MANCOVA</td>
<td>Multivariate analysis of covariance</td>
</tr>
<tr>
<td>MG</td>
<td>Methylglyoxal</td>
</tr>
<tr>
<td>MG-DH</td>
<td>Methylglyoxal-derived dihydroxyimidazolidine</td>
</tr>
<tr>
<td>MG-H</td>
<td>Methylglyoxal-derived hydroimidazolone</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple reaction monitoring</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MudPIT</td>
<td>Multidimensional protein identification technology</td>
</tr>
<tr>
<td>ND</td>
<td>Non-diabetic</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OGTT</td>
<td>Oral glucose tolerance test</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PC</td>
<td>Principal components</td>
</tr>
<tr>
<td>PGE$_2$</td>
<td>Prostaglandin E$_2$</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>pI</td>
<td>Isoelectric point</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>R</td>
<td>Arginine</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RP</td>
<td>Reverse phase</td>
</tr>
<tr>
<td>T2D</td>
<td>Type-2 diabetes</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl) phosphine</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>W</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>WC</td>
<td>Waist circumference</td>
</tr>
</tbody>
</table>
ABSTRACT

Glyco-oxidation is linked to the pathophysiology of diabetes and diabetic complications. The process of glyco-oxidation generates reactive dicarbonyls, which form adducts on arginine residues in distributions throughout the proteome that are site-specific depending on the protein microenvironment. Dicarbonyl adducts are thus markers for glyco-oxidative stress. Various approaches using mass spectrometry permits the identification, localization, and quantification of these dicarbonyl adducts. Using MG as a model dicarbonyl, a shotgun proteomics approach identified the sites for modification of major plasma proteins. Thirty five sites on seven abundant plasma proteins were found, and investigation into the microenvironment surrounding the target arginine sites revealed a neighboring charged residue motif where adjacent residues were either negatively or positively charged. One of the sites identified was R257 in HSA, which is located in the important drug binding site I. We validated drug site I as a target for MG modification by the adaptation of two assays to monitor the effect of MG modification. MG significantly decreases the rate of hydrolysis of PGE$_2$ in drug site I, and induces the displacement of prodan from drug site I. Molecular modeling of warfarin docking at drug site I with the MG-modified R257 resulted in significantly decreased binding and change in binding orientation. The oxidation products of susceptible residues methionine, tryptophan, and cysteine were evaluated using MRM of oxidized HSA peptides. Oxidation of methionine gave the M+16 single oxidized product, and M329 in HSA was the most responsive site. Oxidation of the sole W214 tryptophan produced the W+32 double oxidation product, and oxidation of C34 produced the C+48 triple
oxidation product. MG, 3DG, and glucosone were evaluated for propensity to modify 12 HSA sites based on MRM of dicarbonyl modified HSA. Dicarbonyl modification was independent of arginine solvent accessibility. In a clinical study using nephropathy as an endpoint, sites of oxidation and modification of HSA by MG, 3DG, and glucosone were quantified by MRM. The most important variable among diabetic subjects was metformin use, and subjects taking metformin had significantly reduced markers for glyco-oxidation. These findings may be useful in the development of new diabetes therapies that aim to ameliorate glyco-oxidative stress.
CHAPTER 1: INTRODUCTION

1.1. General Comments

Glycation and oxidation are major causes of biomolecular damage in physiological systems. These two processes operate in a collective mechanism called “glyco-oxidation”, which is strongly linked to the pathophysiology of many diseases, including diabetes and complications associated with diabetes. Advanced glycation end products (AGEs) are the stable end-stage adducts formed by the nonenzymatic glycation and oxidation of proteins, lipids, and nucleic acids. AGEs comprise a heterogeneous class of heterocycles, polymers, crosslinks, and saccharide derivatives. Due to the chemical diversity of such adducts there is no universally accepted method of detection or measurement.

This work focuses on protein AGEs that result from dicarbonyl intermediates of glycation and oxidation. The major target of dicarbonyl intermediates is arginine, which leads to a specific type of damage that is dependent on the site (protein microenvironment) and type of dicarbonyl. It is important to note that not all arginine residues exhibit identical reactivity; rather, particular arginine residues in certain proteins are considered “hotspots” in terms of dicarbonyl affinity. Just as not all arginine sites have equal reactivity, not all hotspots induce functional modifications upon adduction. One of the goals of this work is to identify these hotspots as markers for endogenous dicarbonyl exposure. This will be achieved by the construction of a panel of glyco-oxidative markers using human serum albumin as a model protein. The overarching aims of this work are to identify the biologically relevant AGE’s and determine the
relationship between these modifications and associated structural and functional consequences. This panel of markers will be compared against other clinical measures of glucose dysregulation, and it is envisioned that the sites that show the most significant clinical associations could be used as biomarkers for developing new diabetic therapies.

1.2. Chemistry of glyco-oxidation and formation of AGEs

1.2.1. The Maillard reaction

The basis for the formation of AGEs was discovered by LC Maillard in 1912 (Maillard 1912). Maillard proposed that the chemical reaction between amino groups (proteins and amino acids) and carbonyl groups on reducing sugars is the basis for browning and development of flavor in foods. This initial glycation reaction is followed by a cascade of complex chemical reactions collectively known as the “Maillard reaction” which yields a variety of advanced glycation end products (AGEs).

1.2.2. Glycation

Glycation begins with a condensation of the open-chain form of a sugar with an amine, resulting in a Schiff base. The rate of glycation has been shown to be a function of the anomerization rate of the sugar (Overend et al. 1961), so the greater the open-chain form the more reactive the aldehyde. Glucose is a six carbon sugar that exists in the ring-closed form proportionally more than any other six carbon sugar due to the stability of the equatorial hydroxyl groups in the ring conformation. It has been proposed that evolution has selected glucose as the universal metabolic fuel because glucose is the least reactive saccharide isomer (Bunn et al. 1981). Additionally, the reaction rate of sugars is
inversely proportional to the number of saccharide carbon atoms and is higher for phosphorylated sugars (Monnier 1990).

The Schiff base is a transient and reversible reaction that eventually undergoes rearrangement to the more stable Amadori product (Figure 1.1). The major Amadori product found in vivo is Nε-fructosyl-lysine (FL), formed from the reaction of glucose and the amino side chain of lysine. FL forms on proteins and free lysine in proportion to glucose concentration, and over time, these initial adducts can slowly undergo further
Figure 1.1. Overview of the formation of Advanced Glycation Endproducts (AGEs) from early glycation and glyco-oxidation.

Early glycation involves condensation of glucose with primary amines to give a Schiff base, which rearranges to an Amadori product. In the intermediate stage, the Amadori product is degraded into a series of reactive dicarboxyls, including 3-deoxyglucosone (3DG) through a series of oxidative mechanisms. These highly reactive α-oxoaldehydes are precursors for the formation of AGEs (Ahmed 2005).
rearrangements, dehydrations, fragmentations, cyclizations, and oxidations to yield a heterogeneous population of lysine and arginine protein adducts. Reactive α-oxoaldehydes (glyoxal, methylglyoxal (MG), 3-deoxyglucosone (3DG), glucosone, Figure 1.1) are important intermediates generated during the degradation of FL, and these reactive dicarbonyls are potent precursors of AGEs (Thornalley et al. 1999).

1.2.3. Glyco-oxidation

Formation of AGEs occurs by both oxidative and non-oxidative pathways (Figure 1.2). The term “glyco-oxidation” is used to describe glycation processes that also involve oxidation. Generally, this produces reactive dicarbonyl intermediates, but glyco-oxidation can also involve autoxidation of Amadori products to yield monolysine AGEs such as carboxymethyl lysine (CML) and carboxyethyl lysine (CEL) that are not formed through dicarbonyl intermediates. It is difficult to assess the degree to which AGE formation is the result of strictly oxidative or non-oxidative pathways, as these processes operate in parallel and physiological conditions are aerobic. In addition, experimental evidence with anaerobic glycation in vitro has demonstrated that non-oxidative glycation results in the production of hydrogen peroxide, a pro-oxidant (Jiang et al. 1990).

Glycation and oxidation are intertwined chemical mechanisms, and as such, this dissertation will refer to this combination as “glyco-oxidation”. It is generally accepted that the majority of glyco-oxidation operates via reactive α-oxoaldehyde (dicarbonyl) intermediates (Lapolla et al. 2005), primarily glyoxal, MG, 3DG, and glucosone. The other major source of reactive dicarbonyls other than through glyco-oxidation is through
Figure 1.2. Oxidative and metabolic formation of reactive dicarbonyls glucosone, 3-deoxyglucosone, and methylglyoxal.

FN3K: Fructosamine 3-kinase
glucose metabolism. Of the quantitatively important α-oxoaldehydes formed in vivo, glyoxal, MG, and 3DG are also formed through metabolic processes, whereas glucosone is primarily a product of glyco-oxidation.

1.2.4. Glucosone

Glucosone (2-ketoglucose) is the simplest, earliest and most upstream dicarbonyl product of glyco-oxidation. Open-chain glucose exists in equilibrium with its 1,2-enediol tautomer, which may be oxidized to the diol structure. Upon rearrangement, the ketone is formed at C2 and yields glucosone. Alternatively, the lysine/glucose Shiff base early glycation product can be oxidized at C2 to form a ketoimine, and reversal of the Schiff base (addition of water across the imine) yields glucosone. Reducing sugars in Schiff base configuration are more prone to transition metal catalyzed oxidation than free sugars, because the Schiff base traps the glycating agent in the open chain structure. The mechanism of glyco-oxidative formation of 3DG is not entirely clear, but most reports indicate that formation of the Amadori product is a necessary step (Niwa 1999). In the formation of 3DG, the Amadori product is activated by transition-metal catalyzed deprotonation of C2. Redistribution of electron density between C2 and C3 leads to dehydration at C3 forming the 2,3-enol, which upon tautomeric rearrangement and reversal of the Amadori rearrangement yields 3DG. 3DG is a precursor for MG, as 3DG can fragment and undergo a retro-aldol reaction which leads to MG and glyceraldehyde.
1.2.5. 3DG

A major source of 3DG other than glyco-oxidation is through a deglycation repair mechanism that attempts to reverse the initial glycation products FL and other fructosamines (Figure 1.2). Fructosamine 3-kinase (FN3K) phosphorylates fructosamines to fructosamine 3-phosphate, which spontaneously dephosphorylates to yield 3DG (Delpierre et al. 2000). Other minor sources of 3DG are from the aldose reductase (polyol) pathway of glucose metabolism, where fructose is generated. Fructose is also phosphorylated by FN3K to give fructose-3-phosphate, which also spontaneously dephosphorylates to 3DG. Once produced, “free” 3DG is able to react to form AGEs, or it is detoxified via reduction to 3-deoxyfructose or via oxidation to 3-deoxy-2-ketogluconic acid (Beisswenger, Howell et al. 2003).

1.2.6. MG

MG is the most well-studied and one of the most reactive dicarbonyls. Other than glyco-oxidative pathways, most MG is produced from spontaneous dephosphorylation of triose phosphate intermediates in the glycolytic pathway (Figure 1.3). As glucose is converted to pyruvate, it must be split into two three-carbon sugars, and this step is where most endogenous MG is generated. During intracellular glycolysis, glucose acquires two phosphate groups, resulting in fructose 1,6-diphosphate. Fructose 1,6-diphosphate is cleaved by aldolase into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate, both of which contain labile phosphate groups. Instability of these two triose phosphates results in a spontaneous degradation to inorganic phosphate (PO₄) and MG. It has also been demonstrated that during triose phosphate isomerase interconversion of these two
Figure 1.3. Metabolic sources of MG.

The primary source of endogenous MG is from spontaneous degradation of triose phosphates (green panel, left). MG can also be synthesized from enzymatic oxidation of aminoacetone (yellow) or acetol (blue).
triose phosphates, an ene-diol intermediate can leak from the enzyme active site to ultimately raise MG levels (Knowles 1991).

Other minor sources of MG include catabolism of threonine to aminoacetone, which can be metabolized to MG by semicarbazide-sensitive amine oxidase (SSAO) (Lyles et al. 1992). MG is also a byproduct of lipid metabolism. Acetone monooxygenase converts acetone from fat catabolism to acetal, and acetal monooxygenase (AMO) catalyzes acetal to MG (Casazza et al. 1984). Interestingly, MG is produced from pathways involved with the three major sources of dietary calories- fat, protein, and sugar. MG readily crosses cellular membranes and is found in both the cellular and plasma fraction of whole blood (Thornalley 1988). MG is detoxified by the glutathione (GSH)-dependent glyoxalase system. Glyoxalase I and Glyoxalase II are the main enzymes in this system. GSH will nonenzymatically react with MG to form a hemithioacetal, which is catalyzed by Glyoxalase I to S-D-lactoylglutathione. Glyoxalase II then hydrolyzes S-D-lactoylglutathione to D-lactate. Endogenous D-lactate levels have been identified as an indirect measurement of MG (Talasniemi et al. 2008).

1.2.7. AGEs derived from dicarbonyl intermediates

Dicarbonyls are potent precursors for AGEs. Several nucleophillic groups have been reported to adduct dicarbonyls, including protein nucleophiles (arginine, lysine, histidine, cysteine, N-terminus), DNA (N² of deoxyguanosine and N⁶ of deoxyadenosine) (Frischmann et al. 2005), and aminophospholipids such as phosphatidylethanolamine (Pamplona et al. 1998). The major AGE adduct formed, however, is hydroimidazolone.
Figure 1.4. Lysine and arginine AGEs.
resulting from arginine side chain glycation by reactive α-oxoaldehydes (Rabbani and Thornalley). Other arginine-dicarbonyl adducts are tetrahydropyrimidine (N⁸-[5-(2,3,4-trihydroxybutyl)-5-hydro-4-imidazolon-2-yl]ornithine), argpyrimidine (N⁵-(5-hydroxy-4,6-dimethyl-2-pyrimidinyl)ornithine), pentosidine ((2S)-2-Amino-6-[2-[((4S)-4-amino-5-hydroxy-5-oxopentyl]amino]-4-imidazo [4,5-b] pyridinyl] hexanoic acid), glucosepane, 3-deoxyglucosone-derived imidazolium cross link (DOGIC), arginine-lysine methylglyoxal crosslink (MODIC), carboxymethyl arginine, and carboxyethyl arginine.

Glyoxal, MG, and 3DG form hydroimidazolone on arginine residues as the primary AGE. Glyoxal, MG, and 3DG-derived hydroimidazolones are composed of three structural isomers- H1, H2, and H3, depending on which two of the three guanidino nitrogens the dicarbonyl adducts. It has been determined that the H1 isomer is the most abundant in vitro and in vivo (Thornalley et al. 2003), likely because the primary terminal nitrogens on the arginine side chain are more reactive than the secondary guanidino nitrogen. Hydroimidazolone formation is a multi-step mechanism that is initiated by the nucleophilic attack of the guanidino nitrogen from arginine to the aldehyde of the dicarbonyl, forming a carbinolamine intermediate. The carbinolamine is attacked again by the adjacent guanidino nitrogen from arginine to form the dihydroimidazolidine. The dihydroimidazolidine loses water to form the hydroimidazolone.
Figure 1.5. Dicarbonyl-derived hydroimidazolones and isomerism.

Glyoxal, MG, and 3DG form predominantly the H1 isomer (top). The H2 isomer forms on the guanidino side chain with the lactam carbonyl distal to the arginine side chain. The H3 isomer forms on the guanidino side chain with the lactam carbonyl proximal to the arginine side chain. H1, H2, and H3 are isobaric.
1.3. Advanced Glycation End Products and Diabetic Complications

1.3.1. AGEs and disease

Advanced glycation endproducts (AGEs) comprise a heterogeneous class of modified adducts that are implicated in disease states such as hyperglycemia, aging (Levine 2002), rheumatoid arthritis (de Groot et al.), atherosclerosis (Forbes et al. 2004), uremia (Agalou, Ahmed, Babaei-Jadidi et al. 2005), and neurodegenerative diseases such as alzheimer’s disease (Munch et al. 1997). AGEs are associated with and are implicated as causative factors for many inflammatory pathways and chronic diseases, yet the strongest evidence supports the link between increased AGEs, diabetes, and the development of diabetic complications.

Diabetes mellitus is a group of metabolic diseases characterized by high glucose levels in the blood and urine, and diabetes is generally caused by a lack of insulin and/or an inability to respond to insulin. The two main forms of diabetes are type 1 and type 2. Type 1 diabetes (previously referred to as insulin-dependent diabetes mellitus, IDDM) results from insufficient production of insulin, and is characterized by autoimmune destruction of pancreatic beta cells. Type 2 diabetes (previously referred to as non-insulin dependent diabetes mellitus, NIDDM) results from insulin resistance and a relative deficit of insulin secretion.

In 2001 a unifying theory of the molecular biology of diabetic complications proposed that excess superoxide partially inhibits GAPDH (Figure 1.6). This central mechanism is shared among the polyol pathway, the hexosamine pathway, the protein kinase C pathway, and the AGE pathway (Brownlee 2001). Hence, there appear to be
several mechanisms in which hyperglycemia induces diabetic complications, and all are linked to excess oxidative stress.

1.3.2. AGEs and macrovascular diabetic complications

Diabetic complications can generally be divided into two areas—microvascular and macrovascular. Macrovascular complications, such as coronary artery disease, peripheral arterial disease, and stroke share the central pathological mechanism of atherosclerosis. AGEs are directly linked to the development of atheromas. AGE-modification of LDL: 1) impairs LDL hepatic clearance; 2) increases the retention of LDL in the aortic wall; and 3) increases macrophage accumulation (Sobenin et al. 1993). This leads to increased localization of AGE-modified LDL in the subendothelium, which ultimately increases foam cell production and atheroma development. AGE’s are linked to premature atherosclerosis in diabetic patients (Lyons 1993) and increased AGE concentrations are found in atherosclerotic plaques located in coronary arteries of diabetic patients relative to normal controls (Nakamura et al. 1993). Arterial stiffness is also exhibited in diabetes. Impairment of vascular elasticity and vasoreactivity is due in part to the formation of AGEs, and AGE inhibitors have been shown to reduce arterial
Figure 1.6. Four major pathways of diabetic complications.

Excess mitochondrial superoxide activates four pathways of hyperglycemic damage. Superoxide partially inhibits GAPDH, which reduces upstream metabolism of glyceraldehyde-3-P (Brownlee 2001). Glyceraldehyde-3-P is a precursor for MG.
stiffness in rodent models of diabetes (Huijberts et al. 1993; Wolffenbuttel et al. 1998). In addition to the development of atherosclerotic plaques and arterial stiffening, AGEs contribute to macrovascular diabetic complications through many other thrombogenic mechanisms, namely, activation of platelet aggregation (Hangaishi et al. 1998), maintenance of a hypercoagulable state through stimulated expression of tissue factor (Khechai et al. 1997), and reduced fibrinolysis and clot dissolution (Murakami et al. 1990). Taken together, macrovascular complications due to AGE modification seem to impart structural modifications that harden and narrow the vasculature.

1.3.3. AGEs and microvascular complications

Microvascular complications are the result of the progressive pathology of the small arterioles, and susceptibility to this type of complication involves greater inter individual variation. The major microvascular diabetic complications target the eyes, nerves, and kidney, resulting in retinopathy, neuropathy, and nephropathy, respectively. There exists ample evidence that supports the involvement of AGEs in microvascular diabetic complications. In fact, many studies have demonstrated the predictive value of AGEs compared to other common risk factors, such as duration of diabetes and HbA1C (Wolffenbuttel et al. 1996; Genuth et al. 2005). In another clinical study, it was shown that AGE accumulation precedes and correlates with early manifestations of diabetic nephropathy and retinopathy (Beisswenger et al. 1995).

Microvessels are composed of pericytes and endothelial cells. One important function of pericytes is to regulate endothelial cell growth and regulate their antithrombogenic function. Apoptosis of pericytes is observed in diabetes and this
represents a key mechanism observed early in microvascular complications. Other central pathophysiological features of microvascular complications include increased leakage of proteins from the circulation, endothelial dysfunction, basement membrane thickening, activation of inflammatory cells, pericyte degeneration, and changes in the expression of vascular and neurotrophic factors (Sheetz et al. 2002).

1.3.4. AGEs and retinopathy

Diabetic retinopathy occurs in 75% of all persons with diabetes after more than 15 years of the disease (Klein et al. 1989). Diabetic retinopathy is the result of damage to the intraretinal blood vessels, which leads to retinal ischemia and closure of the retinal capillaries. AGE modification plays a major role in the pathogenesis of retinopathy, as demonstrated by evidence that AGEs accumulate in cultured bovine retinal pericytes (Chibber et al. 1999), AGEs accumulate proportionally to the severity of retinopathy in human diabetic neural retina from autopsies (Hammes et al. 1999), AGEs cause blood-retinal barrier breakdown in nondiabetic rats infused with AGE-modified proteins (Stitt et al. 2000), and AGEs induce abnormal proliferative signaling in retinal microvascular endothelial cells (Du et al. 2003). Perhaps the strongest clinical evidence that AGEs are linked to retinopathy is a trial that concluded that aminoguanidine, an AGE inhibitor and scavenger of dicarbonyls such as methyglyoxal (MG), significantly prevented a 3-step or more progression of retinopathy compared to placebo (Bolton et al. 2004).

1.3.5. AGEs and neuropathy

Neuropathy is the most enigmatic of the microvascular complications, and the most common form is a polyneuropathy characterized by the loss of peripheral sensation.
Decreased sensation combined with the aforementioned impaired vascular function leads to nonhealing ulcers, which are the leading cause of amputation in the United States (Shaw et al. 1997). Overall, neuropathy is thought to occur from glycation of the nerve parenchyma and ischemia from decreased neurovascular flow (Sugimoto et al. 2000). Murine models have shown that AGEs increase the severity of neuropathy by reducing sensorimotor conduction velocity and decreasing blood flow to peripheral nerves (Chen et al. 2004). AGEs likely damage proteins necessary for healthy neurologic function, such as the axonal cytoskeletal tubulin, neurofilament, ECM protein laminin, and actin (Negre-Salvayre et al. 2009). This results in axonal degeneration, impaired axonal transport, and segmental demyelination of peripheral nerves, which are prone to phagocytosis by macrophages (Sugimoto et al. 2008).

### 1.3.6. AGEs and nephropathy

Diabetic nephropathy is one of the most critical consequences of AGE-induced damage because the kidney is the major site of clearance of AGEs. Therefore, diabetic nephropathy increases circulating AGEs as a result of poor kidney function. Nephropathy is initially characterized by thickening of the glomerulus, specifically the basement membrane. This change is accompanied by microaneurysm formation, mesangial nodule formation and expansion, and glomerulosclerosis (Goh et al. 2008). Dysfunction of glomerular filtration is manifested by microalbumin, or leakage of serum protein into the urine. Expansion of the extracellular matrix in the mesangial area is associated with decreased glomerular filtration, as there is less surface area available for filtration. Loss of glomerular function is accompanied by reduced filtration capacity, which eventually
leads to complete renal failure. Clinical manifestations of diabetic nephropathy include leakage of albumin and other large serum proteins into the urine (proteinuria), and an increase in serum creatinine. These two complimentary measurements are used to assess chronic kidney disease (CKD).

Animal studies have clearly demonstrated that AGEs play a major role in hyperglycemia-induced diabetic nephropathy. Diabetic rats have significantly higher renal AGEs (Soulisliparota et al. 1995), and chronic infusion of nondiabetic animals with AGE-modified albumin resulted in glomerular hypertrophy, basement membrane thickening, mesangial extracellular matrix expansion, and albuminuria (Vlassara et al. 1992). Subsequent studies also determined that injection of AGE-modified albumin increased glomerular collagen IV and TGF-β expression in murine systems (Vlassara et al. 1994). These studies conclude that AGE modification of protein induce nephropathy.

In summary, experimental evidence links AGEs with vascular complications in four major areas: 1) Increased concentrations of AGEs localized to sites of development of complications; 2) Exposure to AGE precursors or AGE-modified proteins induce pathologies similar to diabetic vascular disease, 3) Vascular cells treated with AGE-modified proteins undergo changes in signaling mechanisms consistent with diabetic progression; and 4) Suppression of the development of vascular complications by AGE inhibitors in clinical trials and animal models of diabetes (Ahmed et al. 2007).
1.4. Site-Specific Targets of Glyco-Oxidation

1.4.1. Specificity

Glycation has similar damaging effects as point mutations, with profound effects on protein structure, stability, and function (Gomes et al. 2008). Advanced glycation is not a random reaction and various targets will be more or less susceptible to adduction based on a number of factors. The structural features of proteins are major determinants of the extent and distribution of electrophile adduction. Additionally, the nature of the electrophile itself will govern the site specificity of modification. Dicarbonyl modification of arginine is particularly damaging because arginine residues have the highest probability of any amino acid residue for location at functional sites of proteins (Gallet et al. 2000). A “hotspot” is defined as a site that has an extraordinarily high reactivity for modification. Despite current detailed knowledge of many protein structures, only a few dicarbonyl hotspots have been described in the literature.

1.4.2. MG sites on proteins

It has been reported that analysis MG modification of human serum albumin (HSA) revealed a hotspot at R410, with minor sites of modification at R114, R186, R218, and R428 (Ahmed, Dobler et al. 2005). It was concluded from this study that MG modification was not based on solvent accessibility; rather, it was determined by the microenvironment of the arginine sites. Specifically, an adjacent polar Y411 has a major influence over MG modification of R410. Other site-selective MG modifications were found on human hemoglobin, where out of 6 possible arginines, R104 of the β chain was modified preferentially (Gao et al. 2006). Glutathione peroxidase (GPx) was inactivated
by MG at millimolar concentrations, and it was determined that R184 and R185 are readily modified (Park et al. 2003). This is a critical site for GPx as both of the arginines are located in the glutathione binding site. β-casein contains a MG site at R183, which is in close proximity to D184 (Lima et al. 2009). R188 of heat shock protein 27 (Hsp27) is a hotspot for MG-argpyrimidine formation (Sakamoto et al. 2002), and plays a critical role in Hsp27 chaperone activity, oligomerization, and apoptosis. An interesting gain-of-function MG modification occurs at R21 and R103 of αA-Crystallin, in which MG induces enhanced chaperone activity as measured by the refolding of malate dehydrogenase (Biswas et al. 2006). In type IV collagen, MG-H1 residues were found at R390 (α1), R889 (α2), and R1452 (α2) (Dobler et al. 2006). These modified arginine residues are within α1 chain GFOGER and α2 chain RGD integrin binding sites, where O stands for hydroxyproline. MG-modification at these sites is implicated in endothelial cell apoptosis, detachment, and impairment of angiogenesis. In another study, MG modification of p300 at R354 decreases association of p300 with HIF-1α (Thangarajah et al. 2009). This MG hotspot is likely a primary modulator for impaired hypoxia-induced VEGF expression in diabetes. MG also may modify the B-chain of insulin at R22 and at the N-terminus (Jia et al. 2006). Bovine ribonuclease A has been used as a model protein to study the site-specificity of glyoxal and MG, and R39 was determined to be the primary site and R85 was a minor site for adduction of both dicarbonyls (Cotham et al. 2004; Brock et al. 2007). These two arginine residues are closest to the active site of RNase.
1.5. Methods to Analyze Advanced Glycation

The chemical diversity of AGEs and the complexity of the reactions involved in their formation makes AGE analysis a major technical challenge. Assays for AGEs utilize different detection methods such as fluorescent properties of certain AGEs, immunoassays that rely on nonspecific modifications, and mass spectrometry based assays. AGE measurement using fluorescence and immunoassays are nonspecific in identification of specific AGE structures, yet mass spectrometry has emerged as the method of choice for identification and quantification of AGEs.

1.5.1. AGE measurement by fluorescence

A few low abundant AGEs can be measured generically by fluorescence. Skin autofluorescence may be measured by illuminating it with an excitation light source of 300 - 420 nm and measuring the average light intensity in the 420 – 600 nm range (Meerwaldt et al. 2004). A similar measurement quantifies the fluorescence of the low-molecular weight fraction of serum in which de-proteinated serum is separated using HPLC and the total fluorescence (370/440 nm) of all the eluting peaks is reported (Januszewski et al. 2005). Both of these methods do not measure specific AGEs, and they are likely to recognize fluorescent adducts also derived from advanced lipoxidation end products (ALEs) in addition to trace pharmaceuticals, dietary nutrients, and other endogenous fluorophores. Not all AGEs are fluorescent, and not all endogenous fluorescence is attributed to AGEs. The spectral property of the skin autofluorescence measurement is loosely based on pentosidine, a low abundant arginine-lysine crosslink which is approximately 1000-fold less abundant than the hydroimidazolone adducts.
(Ahmed, Babaei-Jadidi et al. 2005). The HPLC assay to measure fluorescence is based on the excitation/emission maxima of trace amounts of crosslinked AGEs such as crosslines and vesperlysines (Tessier et al. 1999).

### 1.5.2. AGE measurement by immunoassay

Immunoassays are widely used in analysis of AGEs, but this approach also suffers from problems with AGE chemotype specificity. A common preparation method for generating antibodies to recognize AGEs is to incubate bovine serum albumin (BSA) with high concentrations of glucose for extended periods of time to yield AGE-modified BSA. This modified protein is used as an immunogen, and contains early glycation adducts such as FL as well as many various AGE structures. Horiuchi et. al. developed the 6D12 monoclonal antibody (Horiuchi et al. 1991) to recognize AGE-modified BSA over unmodified BSA. This clone is the most reported AGE antibody in the literature, but its specificity for individual AGE structures is not well established. CML and CEL are major epitopes for this AGE antibody (Ikeda et al. 1996), but hydromidazolones generated from dicarbonyls such as MG are also major epitopes of the 6D12 antibody. Monoclonal anti-CML (CM-10 clone) (Koito et al. 2004) and monoclonal anti-CEL (KNH-30 clone) (Nagai et al. 2008) were initially reported to claim specificity for CML and CEL, respectively, yet a subsequent study determined that these three monoclonal antibodies also cross react with HSA treated with glyoxal, MG, 3DG, glycolaldehyde, glyceraldehyde, and glucosone (Mera et al.). The high reactivity of dicarbonyl AGE precursors likely results in the generation of multiple epitopes upon protein modification, and the most abundant epitope produced is not necessarily the most antigenic. This case
is illustrated with the development of anti-MG, in which *in vitro* MG-modified HSA led to a monoclonal antibody (C3 clone) that preferentially recognizes the larger argpyrimidine epitope as opposed to the smaller hydroimidazolone epitope (Oya et al. 1999). This preference for argpyrimidine is contrasted with reports of endogenous protein-bound hydroimidazolone at 500-fold greater concentration *in vivo* than protein-bound argpyrimidine (Thornalley et al. 2003; Ahmed, Dobler et al. 2005).

Another type of immunoassay that is used to generically quantify protein AGE content is a carbonylation assay which utilizes hydrazine conjugated to a hapten such as dinitrophenol (DNP) or biotin. Hydrazine is a nucleophillic probe that will form stable adducts with carbonyl groups, and the DNP or biotin is recognized by anti-DNP or streptavidin conjugated to a reporter system. Hydrazine will attack all carbonyl groups, even carboxyls generated through oxidation and ALE reactions (Dalle-Donne et al. 2003).

**1.5.3. Problems with measurement of free dicarbonyls**

Analysis of free dicarbonyls is hampered by their high reactivity, low abundance “free” in solution, and instability throughout the assay procedure. Numerous platforms have been proposed that rely on derivitazation of the dicarbonyl (glyoxal, MG, 3DG, glucosone) prior to MS analysis. The process of derivitization creates another variable because both the ketone and aldehyde moieties must react to produce the analyte. An additional complication is that other endogenous electrophillic compounds are capable of derivitization, and may interfere with analysis. Dicarbonyls may be treated with aromatic 1,2-diamino compounds such as o-phenylene diamine, activated hydrazines such as 2,4-
dinitrophenylhydrazine, perfluorinated hydroxylamines such as o-(2,3,4,5,6-pentafluorobenzyl) hydroxylamine (PFBOA), aliphatic diamines such as 1,2-diamino propane, or nucleophillic enols such as 2-thiobarbituric acid. In acidic conditions, these various reactions yield compounds that are measured using spectrophotometric, spectrofluorimetric, or mass spectrometric techniques. Gas chromatography (GC) MS has been used to identify PFBOA derivatives, and the other derivatives largely rely on liquid chromatography (LC) coupled to electrospray ionization (ESI) for MS or MS/MS detection. These assays are not generally robust or reproducible, as evidenced by the fact that most of these studies published in the literature have not been repeated successfully. Moreover, as α-oxoaldehydes are highly reactive in biological matrices, most are protein bound and these assays must completely reverse the dicarbonyl from protein. An example of the degree of dicarbonyl that is bound is illustrated by an experiment in which human plasma was treated with 1μM [14C]methylglyoxal. This showed that all MG (>99%) was bound irreversibly after 24 h (Thornalley 2005).

Despite reports of new methodologies for analysis of free dicarbonyls in biological mixtures, no “gold standard” currently exists likely due to the limitations described above. These various derivatization platforms describe a wide range of dicarbonyl concentrations. For example, depending on the assay, normal plasma MG concentration was found to be 40nM- 4.5μM (Khuhawar et al. 2008). The difference in over two orders of magnitude between analytical platforms suggests that absolute quantitation based on any single assay may not be entirely accurate. Relative measurement of free dicarboxylics, however, has been demonstrated in side-by-side
comparisons to show that plasma free dicarbonyls are increased during the progression of diabetic complications (Nakayama et al. 2008).

1.5.4. Analysis of dicarbonyl adducts

Products of dicarbonyl adduction and advanced glycation are relatively stable and do not suffer from the same analytical problems as do free dicarbonyls. Hydroimidazolone adducts resulting from arginine-dicarbonyl condensation exist in two main forms, protein bound adducts and free adducts. After proteins are damaged by dicarbonyl modification, proteins containing site-directed hydroimidazolone adducts (protein bound) are targeted for proteolysis by the proteasome (Dudek et al. 2005). The proteasome then degrades the damaged protein into constituent amino acids, which also contain the hydroimidazolone free adducts. These free adducts are finally excreted in the urine. There is no known mechanism of de-glycation of protein hydroimidazolone ring structures other than urinary excretion of free adducts. These free adducts have been quantified in plasma, urine, cerebrospinal fluid, synovial fluid, and peritoneal dialysate (Thornalley et al. 2003; Ahmed et al. 2004; Agalou, Ahmed, Thornalley et al. 2005; Ahmed, Ahmed et al. 2005). Protein bound hydroimidazolone adducts have also been quantified in plasma protein, skeletal muscle, sciatic nerve, retina, and renal glomeruli of streptozotocin-induced diabetic rats (Thornalley et al. 2003). In addition, these protein bound adducts were analyzed in human blood cells such as red blood cells (RBC), neutrophils, and mononuclear leukocytes (Thornalley et al. 2003). Protein bound hydroimidazolone adducts are prepared for quantification by exhaustive digestion to amino acids using pronase. In amino acid form, both “free adducts” and “protein bound
adducts” may be quantified using stable isotope dilution tandem mass spectrometry (LC-MS/MS). This type of mass spectrometry utilizes a triple quadrupole instrument operating in multiple reaction monitoring (MRM) mode. MRM will be discussed in the next section.

1.5.5. Site-specificity of MG adducts

Analysis of free adducts and complete digests of protein bound adducts is informative in terms of overall degree of advanced glycation or oxidative stress in a system, but it does not indicate the site of the protein damage or extent of modification at individual sites. A few studies have quantitated the site-specific reactivity of model proteins toward dicarbonyls in vitro. Ribonuclease A has been used to determine arginine specificity of glyoxal (Cotham et al. 2004) and MG (Brock et al. 2007) using tryptic digestion of glyoxal or MG-treated protein. These peptides were analyzed by LC coupled to a full-scan MS. HSA has also been used to determine the site-specificity of MG using MALDI-TOF peptide mass mapping followed by full scan LC-MS. MG-treated hemoglobin was digested with Lys-C and analyzed using “semiquantitative” techniques such as LC-MS/MS in selective ion monitoring (SIM) mode (Gao et al. 2006). In this case, percent hydroimidazolone modification was calculated by integrating the peak areas extracted from the total ion chromatograms (TIC) based on an inclusion list of 6 parent arginine-containing peptides. These three examples using full scan MS and SIM are suitable for analysis of the relative reactivity of arginines in pure proteins treated with supraphysiological dicarbonyl concentrations. In order to quantify site-specific
modifications \textit{in vivo} in a complex matrix such as plasma, MS using MRM is a more suitable platform.

\section*{1.6. Analysis of peptides using mass spectrometry}

Mass spectrometry offers the most selective platform for analysis of AGEs. The masses of individual AGEs and fragmentation behavior using tandem mass spectrometry (MS/MS) are like fingerprints and these are powerful tools in identification and quantification of AGEs. Mass spectrometry has been employed to study both the reactive precursors for AGEs (dicarbonyls) and AGE adducts (free adducts and DNA/protein bound adducts).

Protein analysis by MS is generally approached in two ways. In the “top-down” approach, intact proteins are introduced into the mass spectrometer where they are mass analyzed based on the entire mass of the protein or protein complex. “Bottom-up” proteomics requires a fragmentation (proteolytic digestion) step prior to MS, and the resulting peptides may be further fragmented before another step of MS analysis. Masses of these fragments are often used to reconstruct the original peptide sequence, and this sequence information may be used for protein identification. Bottom up proteomics is the most widely used approach to obtain quantitative information about sites of modifications, and this dissertation will focus solely on this type of MS analysis.

There are many different types of mass spectrometers from different vendors, each suited for a particular type of analysis. Mass spectrometers are modular instruments, as no single combination of parts can encompass all types of MS-based analysis. In general, a mass spectrometer is composed of three parts: an ionization source that
introduces the sample into the gas phase where it obtains a charge, a mass analyzer which separates ions based on their mass-to-charge (m/z) ratio, and a detector which counts the ions.

1.6.1. Ionization sources

There are two main types of peptide or protein ionization methods, and both of these are considered “soft” ionization methods. “Soft” means that covalent bonds in the sample are not broken during the ionization process. The first is matrix-assisted laser desorption ionization (MALDI), where the sample is co-crystallized with organic matrix small molecules and dried into a thin layer. A laser excites the matrix and this process volatilizes the analyte into the gas phase, and at the same time, transfers a proton to the analyte in the case of positive mode for polypeptides. The second type of soft ionization is electrospray ionization (ESI) (Figure 1.7), where the sample is initially in liquid form. The solution flows through a capillary end with a high voltage applied, which creates a fine mist of charged droplets. As the solvent evaporates from the droplets, the sample ions remain in the gas phase and enter the source of the mass spectrometer.

1.6.2. Mass analyzers

Once the sample is charged and in the gas phase, mass analyzer(s) can separate or fragment the ions. In single-stage mass spectrometry, a single mass analyzer is used. In tandem mass spectrometry, two (MS/MS) or more (MS^n) mass analyzers are used. There are many types of mass analyzers, yet ion traps, quadrupoles, and time of flight (TOF) are three that are routinely used for proteomics. TOF mass analyzers are conceptually the
Figure 1.7. Electrospray ionization (ESI) and triple quadrupole (QqQ) mass spectrometer.

A) ESI generates charged fine droplets in the presence of an electric field to obtain ions.

B) QqQ comprising two transmission quadrupole mass spectrometers in series, with a non-resolving (RF-only) quadrupole between them to act as a collision cell.
simplest, whereby an electric field accelerates the ions in a vacuum and the time it takes to reach the detector is measured. Ions with a greater m/z take longer, and TOF analyzers are akin to velocity meters. Quadrupoles represent a different type of analysis, are also known as transmission quadrupoles or linear quadrupoles. Quadrupoles are essentially four parallel rods that use an oscillating electric field that acts as a mass selective filter for the path of ions. A common configuration is a linear series of three quadrupoles, termed triple quadrupole (QqQ). In this setup, the first (Q1) and third (Q3) quadrupoles act as mass filters and the second (q2) is a collision cell that fragments the ions. Q2 may induce collision-induced dissociation (CID) of the ions using a neutral gas such as nitrogen. Ion traps are another type of mass analyzer, and they are also known as quadrupole ion traps. Ion traps isolate packets of ions, which may then be selected, dissociated, and scanned all in the same component. A majority of MS presented in this dissertation was generated with a QTRAP 4000 (AB Sciex), a hybrid triple quadrupole linear ion trap system. In this configuration, Q3 in the QqQ was replaced with a linear ion trap for enhanced sensitivity.
Figure 1.8. Common ions from collision-induced dissociation of peptides.

CID generally generates b and y ions from fragmentation of the peptide bond. B ions result when the charge remains on the N-terminus of the peptide. Y ions result when the charge remains on the C-terminus of the peptide.
1.6.3. Peptide fragmentation

In gas phase CID, peptides will fragment along the peptide backbone at amide bonds to give mainly b and y ions (Wysocki et al. 2000). B-ions result from fragmentation that leave a charge (detectable m/z) on the N-terminus, and y-ions result from the charge remaining on the C-terminus. Fragmentation at different points along the peptide backbone produces different sized b- and y- ions (Figure 1.8). The difference in mass in a series of these ions is equal to the mass of the residue, hence, peptides may be sequenced from these fragmentation spectra. Protein modifications that are stable to CID can also be identified if the mass of the modification is known.

1.6.4. Bioinformatics

Once the MS/MS spectrum is acquired, it must be analyzed. A typical proteomics experiment can generate thousands of spectra which must either be matched to protein sequences in a database, or assigned a sequence based on only the masses of the ions (de novo sequencing). There are many software algorithms to do this. X!Tandem (Global Proteome Machine), SEQUEST (ThermoFisher Scientific), and Mascot (Matrix Science) use algorithms that compare each spectrum to all likely candidate peptides in a protein database. The PEAKS (Bioinformatics Solutions, Inc.) algorithm, however, performs sequencing directly from the MS/MS data and therefore does not rely on a protein database. All of these algorithms allow for the inclusion of post-translational modifications in the search, and this option is user-defined by the mass of any residue +/- any modified mass. For example, hydroimidazolone modification by methylglyoxal
increases the mass of arginine by 54. Knowing this, peptide search algorithms can include the possibility of R+54 in the search.

1.6.5. MRM quantification

MS can be used not only for identification of peptides and proteins, but also for quantification. MRM or selected reaction monitoring (SRM) in a QqQ instrument offers the highest sensitivity for quantitation of analytes in complex mixtures. The non-scanning use of quadrupoles translates into an increased sensitivity up to two orders of magnitude compared with conventional full scan techniques (Lange et al. 2008). In MRM, Q1 and Q3 act as filters to specifically select predefined m/z values corresponding to the parent peptide ions and daughter fragment ions, respectively. Q2 is used as a collision cell to produce the fragments that will be transmitted through Q3, and the collision energy is optimized for each parent/daughter pair. A transition is defined as the optimal m/z values for the parent/daughter as well as other optimal instrument-specific parameters that yield the maximum signal for each analyte. These may include collision energy (CE), which is the applied voltage in q2 to fragment ions; declustering potential (DP), which is the voltage potential between the orifice plate and ground, optimized to further strip off solvent molecules and reduce ion-ion aggregation; collision cell entrance potential (CEP); and collision cell exit potential (CXP).
Figure 1.9. Multiple Reaction Monitoring (MRM).

Specific product ions (Q3) corresponding to m/z from selected precursor ions (Q1) are recorded. (Kitteringham et al. 2009).
Proteomic analysis using MRM is uniquely suited for site-specific quantification of post-translational modifications (PTM’s). In contrast to a “shotgun proteomics” approach in which the analysis of a protein sample is unbiased, peptide MRM is a hypothesis-driven approach that has the capability to target the specific mass changes at a given protein site. In an MRM-based assay, a list of transitions is monitored as peptides are separated using LC such as reverse phase with C18 resin. As each transition is monitored in the mass spectrometer, a chromatogram for each transition is recorded. The area under the curve (AUC) value for each eluting peak indicates the abundance of the peptide analyte.

1.7. Dissertation Aims

This dissertation investigates the site-specificity of protein damage from oxidation and dicarbonyl adduction in vitro and in vivo. The second chapter will describe a shotgun proteomics approach used to identify sites of modification in the human plasma proteome. MG will be used as the model dicarbonyl/AGE precursor and the nature of these adducts on synthetic peptides and abundant plasma proteins will be characterized. In chapter 3, the functional consequences of MG modification of HSA will be described. This section will use molecular modeling approaches to simulate the site-specific structural modifications of dicarbonyl damage to HSA. Biochemical assays will be used to measure the functionality of HSA after dicarbonyl assault. Chapter 4 will describe the validation of a panel of glyco-oxidative modifications using MRM with HSA a model protein. Synthetic peptides representing 12 HSA arginines will be modified with
dicarbonyls (MG, 3-DG, and glucosone) at internal arginine sites, and an MRM transition list will be constructed and validated using *in vitro* dicarbonyl-modified protein. Similarly, the site-specificity of oxidation will be assessed using synthetic peptides containing the major residues that are susceptible for oxidation, such as methionine, tryptophan, and cysteine. Eight synthetic peptides were oxidized at internal (6M, 1W, 1C) residues, and an MRM transition list will also be constructed for this panel. The transition list were validated using MRM analysis of *in vitro* oxidized protein. Chapter 5 describes the clinical investigation and quantification of the site-specific glyco-oxidative targets for modification of HSA *in vivo*. Plasma was obtained from 120 subjects and HSA was analyzed for oxidation and dicarbonyl adduction using the validated MRM transition panel from chapter 4. These subjects represent the spectrum of diabetic nephropathy, from normal healthy controls to those needing dialysis treatment due to end-stage renal disease (ERSD). Endogenous modified HSA from either oxidation or dicarbonyl modification of hotspot arginines was quantified based on relative peptide MRM signal. The degree of oxidative or dicarbonyl modification was correlated to clinical measures of diabetic nephropathy, and differences between *in vitro* and *in vivo* modification will be addressed. Finally, the arginine and oxidative hotspots that show the most promising clinical significance will be discussed.
CHAPTER 2: IDENTIFICATION OF METHYLGLYOXAL MODIFICATIONS IN THE PLASMA PROTOEME

2.1. Introduction

Methylglyoxal (MG) is a toxic glucose metabolite formed by the spontaneous degradation of triosephosphates, oxidative metabolism of ketone bodies, and catabolism of threonine (Ahmed, Dobler et al. 2005). Endogenous MG is formed as a consequence of the combination of glycolytic flux and oxidative stress in a collective mechanism termed carbonyl stress (Wondrak et al. 2002). One of the critical pathogenic consequences of hyperglycemia in diabetes is a deficit in detoxication of reactive carbonyl compounds, such as MG. Many factors could trigger the accumulation of MG in vivo, including aging, hyperglycemia, inflammation, oxidative stress, and uremia (Ahmed et al. 2007).

MG is a reactive dicarbonyl electrophile that forms adducts with amino groups of proteins, nucleic acids, and phospholipids up to 20 X 10³ times more readily than glucose (Ahmed et al. 2007). The rate of MG formation has been calculated to be 120 µM/kg cell mass/day under normoglycemic conditions (Thornalley 1993). Moreover, the relative abundance of MG adducts has been estimated to be as high as one MG-derived adduct on 3-13% of all proteins, provided an average protein contains 26 arginines (Ahmed, Dobler et al. 2005). Depending on the protein, approximately 0.1% of arginines in plasma proteins bear the methylglyoxal-derived hydroimidazolone (MG-H1) modification, as determined by exhaustive digestion and quantification using MS/MS (Duran-Jimenez et al. 2009).
There is debate over the exact physiological concentration range of MG. Since MG is a reactive compound, the majority does not exist free in solution. Some estimates assert that endogenous MG is more than 99% protein bound (Dhar et al. 2009), so measurement of MG needs to correct for protein binding. Another technicality is that MG must be derivitized before analysis, and different analytical platforms to measure the MG-derivatives give different plasma concentrations, ranging from 40nM – 4.5μM (Khuhawar et al. 2008; Dhar et al. 2009). Nevertheless, these reports indicate that diabetic patients have higher concentrations of MG in side-by-side comparison with normoglycemic patients. Prolonged exposure to low amounts of endogenous MG throughout the half life of plasma proteins may have a profound effect in situations of poor glucose control and increased glyco-oxidative stress. It is often necessary to induce modifications with supraphysiological concentrations of MG to overcome nonspecific protein binding in vitro. It is reasonable to assume that the success rate for finding potential MG sites is increased with treatment of excess MG, and the relative reactivity of different arginines is similar whether they are exposed to low or high concentrations of MG.

One of the most quantitatively and functionally important MG adducts is the arginine MG-H1 adduct, producing a loss of positive charge via hydroimidazolone formation (Figure 2.1). MG will also adduct to the side chains on lysine and cysteine residues, but adducts formed on these residues are transient and exhibit faster off-rate kinetics (Lo et al. 1994). It is necessary to elucidate the site-specificity of MG damage, as critical residues involved in protein function could be susceptible to modification.
Figure 2.1. Chemistry of MG-Arginine Adduction.

MG primarily reacts with arginine residues to form neutral ring structures- MG-DH (dihydroxyimidazolidine, R+72) and MG-H1 (hydroimidazolone, R+54).
Moreover, particularly reactive sites could serve as biomarkers for MG exposure, similar to how glycated hemoglobin (HbA1C) is a marker for glucose exposure. These high-affinity sites for MG adduction may be a better gauge in vivo of the ability to detoxify reactive dicarbonyls, whereas HbA1C is an indirect measure of the history of glucose exposure.

Adduction by MG is likely to be of physiological importance because of the presence of a high amount of MG residues in cellular and extracellular proteins relative to other AGEs (Ahmed, Babaie-Jadidi et al. 2005). Some of these adducts are functionally important, as arginine residues have the highest frequency of occurrence in ligand and substrate recognition sites in receptors and enzyme active sites (Thornalley et al. 2003). Of all of the 20 amino acids, arginine has the highest probability of being located in active sites, with approximately 20% of active sites containing at least one arginine (Gallet et al. 2000). Receptor binding domain (RBD) analysis of 80,000 protein sequences estimate that this frequency is a 3.8-fold more frequent occurrence than a completely random distribution would predict (Gallet et al. 2000).

Due to the lack of commercial MG antibodies that recognize the MG-H1 epitope, mass spectrometry based analysis would be a preferred method to determine endogenously MG-modified human plasma in vivo. It is equally important to characterize the sites of modification so that specific antibodies can be designed for future high throughput analysis. Unbiased shotgun proteomics identified a MG target on plasma fibrinogen in vivo and several arginine modifications in the plasma proteome after MG incubation in vitro.
2.2. Methods

2.2.1. Materials

All materials were HPLC grade and were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. Synthetic peptides were custom ordered from New England Peptide (Gardner, MA). Sequencing grade trypsin was purchased from Promega (Fitchburg, WI).

2.2.2. Subject selection

All subjects provided informed consent (see Appendix A). Samples were obtained from the FDA-registered blood establishment ProMedDx (Norton, MA). ProMedDx ensures that all specimens are collected under IRB approved protocols according to 21 CFR, ICH/GCP guidelines and HIPAA Privacy Regulations. We obtained plasma from 3 middle-aged, obese persons, with type 2 diabetes mellitus (Sex: 2 M, 1 F, Age: 53 ± 6 years, BMI: 35 ± 5 Kg/m², duration of diabetes: 9 ± 5 years) all requiring insulin therapy, and 3 aged-matched healthy controls without a prior history of diabetes (Sex: 3 F, Age: 53 ± 8 years), and not on any medications.

2.2.3. Removal of Abundant Proteins

Plasma (50µl) from three diabetic and three non-diabetic patients was diluted to 600 µl with TBS pH 7.4 and centrifuged through a 0.2 µm pore size spin filter to remove particulates. This sample was applied to an IgY-12 (Genway Biotech, Inc.) affinity column using an AKTA FPLC instrument (GE Healthsciences) according to the manufacturer’s instructions. The flow-through and bound fractions were collected separately and, concentrated, and buffer exchanged into 100 mM ammonium bicarbonate
Figure 2.2. Approximate protein abundance in plasma.

About 90% of the total protein content of plasma is made up of only 10 proteins (left chart). The remaining 10% is largely made up of 12 proteins (right chart). These 22 proteins are considered to be high abundant and medium abundant proteins. (Schuchard)
pH 7.4 using Vivaspin centrifuge concentrators with a molecular weight cut-off of 3,000 Da. Protein samples were quantitated using DC Protein Assay (BioRad) prior to being frozen at -20˚C.

2.2.4. Enzymatic Digestion

  Immunodepleted plasma samples (60 μg) were reduced with DTT (20mM in 100 mM ammonium bicarbonate) for 30 minutes at 55°C and alkylated with iodoacetamide (55 mM in 100 mM ammonium bicarbonate) for 30 minutes at room temperature in the dark. Plasma protein was then digested with trypsin (protein to trypsin at 50:1 w/w ratio) overnight at 37°C. Peptides were desalted using Hypersep C18 columns (Thermo Scientific), lyophilized, and re-suspended in 10 μL of 1% TFA immediately prior to LC-LC-MS/MS.

2.2.5. MudPIT (LC-LC-MS/MS)

  Peptide digests were manually injected in technical replicates onto a microbore HPLC system (Paradigm MS4, Michrom, Auburn, CA) with two separate strong cation exchange (SCX) and reversed phase (RP) columns: a 100 μm I.D. capillary packed in-house with 7 cm of 5 μm Vydac C18 reversed phase resin and a separate 250 μm I.D. capillary packed in-house with 7 cm of 5 μm Partisphere strong cation exchanger resin (Whatman, Clifton, NJ). The twelve step MudPIT analyses (Figure 2.3) were as follows. Solutions: 10% methanol/0.1% formic acid, 0.01%TFA (buffer A), 95% methanol/0.1% formic acid, 0.01% TFA (buffer B), 10% methanol/0.1% formic acid, 0.01% TFA (buffer C) and 500 mM ammonium acetate/10% methanol/0.1% formic acid, 0.01% TFA (buffer D). Step 1 consisted of a 5 min equilibration step at 100% buffer A, followed by another
equilibration step for 5 min at 25% buffer B (75% buffer A), followed by a 40 min gradient from 25% buffer B to 65% buffer B, followed by a 10 min 65% buffer B and 10 min of 100% buffer A. Chromatography steps 2 to 12 followed the same pattern: 15 min of the appropriate % of buffers C & D followed by a 2 min 100% buffer C wash, a 5 min wash with 100% A, equilibration with 25% buffer B for 5 min, followed by a gradient from 25% buffer B to 65% buffer B in 40 min, followed by a 10 min 65% buffer B and 10 min of 100% buffer A. The buffer C/D percentages used were 95/5%, 90/10%, 85/15%, 80/20%, 70/30%, 60/40%, 40/60%, 20/80%, 0/100%, 0/100%, 0/100% respectively, for the 11 salt steps. Eluting peptides at 350 nL/min were sprayed into a ThermoFinnigan LCQ-Deca XP Plus ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) (Washburn et al. 2001) with a spray voltage of 1.6 kV. Spectra were scanned over the range 380-2000 mass units. Automated peak recognition, dynamic exclusion, and daughter ion scanning of the most intense ion were performed using the Xcalibur v1.3 software as described previously (Breci et al. 2005).
A protein mixture is digested and loaded onto a biphasic column. Peptides are bumped onto the reverse phase (RP) from the strong cation exchange (SCX) at intervals with a series of increasing salt concentrations. MS is performed, the top three most intense peaks are selected for MS/MS. This slow elution of peptides from the column allows time for MS and MS/MS scans.
2.2.6. *In Vitro* MG Modification

A 500 μg aliquot of the abundant (bound to IgY-12 column) and a 500 μg aliquot of the depleted (flow-through from IgY-12 column) human plasma protein from a healthy subject was buffer exchanged into 1X PBS pH 7.4 using Vivaspin centrifuge concentrators (MWCO 3K), treated with 5mM MG, and incubated at 37°C for 18 hours. The proteins were reduced, alkylated, digested, and desalted as previously described. A 60 μg aliquot of MG-modified plasma peptides from both the abundant and depleted fractions were subjected to LC-LC-MS/MS.

2.2.7. Database Searching and Validation

A total of 18 LC-LC-MS/MS runs from six patients in triplicate were categorized by normal vs. diabetic cohorts. The human Uniref50 version 8.0 (17,806 entries) database was searched using X!Tandem (Version: 2007.01.01.1, The Global Proteome Machine Organization), Mascot (Matrix Science), and Sequest (Version: 27, rev. 12, Thermo Finnigan) search engines. Search criteria included C+57, M+16, R+72, and R+54 as possible modifications, two missed cleavages were allowed, parent ion tolerance was set to 2.0 Da, and fragment tolerance was set to 1.0 Da. Database searching of peptide fragmentation spectra using the human Uniref50 database excluded protein sequences with >50% mutual sequence identity to speed search times and reduce redundancy. Spectra were considered for manual validation if Peptide Prophet probability >95%, and/or Mascot scores >40, and/or Sequest XCorr value greater than 1.9 or 2.2, or 3.7 for +1, +2 or +3 charge states, respectively. Spectral uniqueness was evaluated with Sequest DeltaCn >0.08 and X!Tandem –logE scores greater than 3. Spectral matches were
Figure 2.4. Bioinformatic pipeline from raw MS/MS data to protein identification and localization of modified arginines.

Database searching used Sequest, X!Tandem, and Mascot search engines to match MS/MS spectra to sequences in the human Uniprot database. Protein modification sites (R+54, R+72) were included as variable modifications. Peptide/Protein Prophet algorithms in the Scaffold (Proteome Software) platform assigned p-values to the search results and combined search results.
amalgamated into Scaffold (Version: Scaffold_2_05_02, Proteome Software), and probabilities of true matches were calculated as follows. Peptide probabilities were greater than 95% as determined by the Peptide Prophet (Figure 2.3) algorithm contained within Scaffold (Keller et al. 2002). Protein identifications were accepted if they could be established at greater than 95% probability as specified by the Protein Prophet algorithm (Nesvizhskii et al. 2003). Proteins that contained similar peptides and which could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. In order to manually validate new sites for adduction, we employ the following criteria: B and Y ion series must provide coverage at the site of modification; modifications that induce missed proteolytic (tryptic) cleavages are not scored for peptides matched with C-terminal modifications (false positives); and unmodified peptides from the protein must also be identified.

2.2.8. Validation using accurate mass MS/MS

A 10 mg sample of lyophilized synthetic peptide VHTECCHGDLECAADDRLAK (MW 2413 g/mol) was resuspended in deionized water to 5 mg/ml. Two 150 μL portions (0.31 μmol each) were incubated with 20 μL of 200 mM iodoacetamide (4 μmol) in 20 mM ammonium bicarbonate pH 7.4 in the dark at room temperature for 1 hr. This is a 4.3-fold excess of iodoacetamide over sulfhydryl content. MG (150 μL of 10 mM) in 2X PBS pH 7.4 was added to one of the peptide solutions (5-fold excess of MG to arginine), and 150 μL of 2X PBS was added to the blank reaction. The peptide reactions were incubated at 37°C for 2 hours. The reactions were stopped with the addition of formic acid (10 μl), and the modified peptides were
desalted with 50 mg Hypersep C18 RP cartridges (Thermo Scientific) and eluted with 1 mL 80% acetonitrile containing 0.1% TFA. The peptides were diluted thirty-fold with a solution of 0.5% formic acid in water prior to infusion. Peptide solutions were infused at 3 μL/min into an Apex-Ultra hybrid Qq-FTMS (Bruker). Parent ion window for MS/MS was set to 1.0 Da, and collision energy was 40eV for the +4 ions and 60eV for the +3 ions. Exact mass of the addition of hydroimidazolone is 54.011 amu (3C, 2H, 1O).

2.2.9. Molecular modeling

X-ray crystal structures were obtained from the Protein Data Bank and visualized using Discovery Studio Visualizer v2.5 (Accelrys Software Inc). PDB files are as follows: 1QRZ for human plasminogen catalytic subunit, 2HAU for human transferrin, 3KQ0 for alpha-1 acid glycoprotein I, 2A01 for human apolipoprotein A1, and 1A06 for human serum albumin.
2.3. Results

2.3.1. MG sites in the plasma proteome

MG primarily reacts with arginine residues to form relatively stable ring structures- dihydroxyimidazolidine (R+72) and hydroimidazolone (R+54). This adduction results in a net loss of positive charge from the arginine site, as arginine is positively charged at physiological conditions and the ring adducts are uncharged. In order for MG adducts to be detected using MS/MS, they must not be labile during ion activation. Upon treatment of plasma protein with MG, we identified several MG hotspots on plasma proteins in vitro that included candidate sites on albumin, fibrinogen, plasminogen, apolipoprotein A1, transferrin, α1 acid glycoprotein and haptoglobin (Figure 2.5). The proteins we found that harbor sites for MG modification are considered abundant proteins in the plasma proteome (Kuzyk et al. 2009). Identification of modifications on abundant proteins is the expected result of shotgun proteomics experiments because data-dependent MS/MS is unbiased and the most abundant parent peptide ions will be chosen for fragmentation. These peptides contained arginine (R) with atomic mass unit increases of R+54 and R+72 that fragment along the peptide backbone with minimal neutral loss of the modified arginine moiety. Our findings corroborate other studies that have determined that low-energy collision induced activation is appropriate for detection of MG hotspots (Gao et al. 2006; Brock et al. 2007).
Figure 2.5. Site-specific MG modifications in vitro.

Spectra were identified for the R+54 adduct using human plasma incubated with 5mM MG. Spectra containing the intermediate R+72 dihydroimidazoloidine adduct are indicated. All cysteines are carbamidomethylated (C+57) prior to digestion.
When adding potential modifications to database searches, false positives can result and it is necessary to manually validate the MG-modified spectra by using criteria specific for this modification. Since trypsin does not cleave at MG-modified arginine, any results that gave a sequence with a C-terminal modified arginine were considered false positives. The sequences of peptides where MG modification was identified with the intrapeptide arginine (R) modification site are listed in Figure 2.5.

2.3.2. Microenvironment charged residue motif in MG sites

Approximately 75% of the linear sequences of the MG-modified arginine sites appear to follow a motif (Figure 2.6) that involves a charged residue (DEKR) at either 3 or 4 residues upstream or downstream of the MG site. These charged residues may be negatively charged, such as aspartate (D) or glutamate (E), or may be positively charged, such as arginine (R) or lysine (K). Assuming alpha helix structure, amino acids will face the same side of the helix every 3.6 residues. Therefore, in a purely alpha helix conformation the third or fourth residues from the arginine sites exist in a position to interact with the arginine. However, not all of these sites may reside in a completely alpha helix structure, and it is likely that susceptibility to adduction has more to do with the overall three-dimensional microenvironment than simply the primary sequence. If the neighboring charged residues are positioned to interact close to the guanidino group, then the motifs in Figure 2.6 may be explained by neighboring acid/base chemistry.
<table>
<thead>
<tr>
<th>No apparent linear motif</th>
<th>(i+/-3) or 4 with positive/negative charge</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>CADDRADLA</td>
<td>ALLVRYTKK</td>
<td>HSA</td>
</tr>
<tr>
<td>LVNRRPCFS</td>
<td>VRVSRFVTW</td>
<td>Plasminogen</td>
</tr>
<tr>
<td>TSYNRGDST</td>
<td>LGCARPNKP</td>
<td>Plasminogen</td>
</tr>
<tr>
<td>TSDPRKQCS</td>
<td>EKSPRPSSY</td>
<td>Plasminogen</td>
</tr>
<tr>
<td>WWYNRCHAA</td>
<td>DFTNRINKL</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td>IPFNRLTIG</td>
<td>NKADRQYE</td>
<td>Transferrin</td>
</tr>
<tr>
<td>FSKLREQLG</td>
<td>DKELRTGKE</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td>TIFLREYQT</td>
<td>WAVARLSQR</td>
<td>HSA</td>
</tr>
<tr>
<td></td>
<td>GSCSRALAR</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>VEVSRNLGK</td>
<td>HSA</td>
</tr>
<tr>
<td></td>
<td>IEVSRFLFE</td>
<td>Plasminogen</td>
</tr>
<tr>
<td></td>
<td>LEDLRQGLL</td>
<td>Apo A1</td>
</tr>
<tr>
<td></td>
<td>GFRHRHPDE</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>EFPSRGKSS</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>NTYNRVSED</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>MEILRGDFS</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>EEIIRKGGE</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>DSSIRYLQE</td>
<td>Fibrinogen B</td>
</tr>
<tr>
<td></td>
<td>DMKRLFLED</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>ETNPRAGLE</td>
<td>Plasminogen</td>
</tr>
<tr>
<td></td>
<td>KCPGRVVGG</td>
<td>Plasminogen</td>
</tr>
<tr>
<td></td>
<td>KVCNRYEFL</td>
<td>Plasminogen</td>
</tr>
<tr>
<td></td>
<td>EGVMRNN(-)</td>
<td>Plasminogen</td>
</tr>
<tr>
<td></td>
<td>DLRSRIEVL</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td></td>
<td>ADKYRLTYA</td>
<td>Fibrinogen G</td>
</tr>
<tr>
<td></td>
<td>YYKLRTEGD</td>
<td>Haptoglobin</td>
</tr>
</tbody>
</table>

Figure 2.6. Linear motifs adjacent to MG sites.

Sequences from the 35 arginine sites found from \textit{in vitro} MG modification were categorized by Motif-X (Schwartz et al. 2009). Sites are plasma protein arginine hotspots from Figure 2.5. Central arginine \((i)\) is indicated in bold, and the 4 residues preceding and following the arginines show linear sequence adjacent to site. Left column lists the 8 sequences that do not contain a charged residue (DEKR) in either the \(i+/-3\) or \(i+/-4\) positions. Right column lists the sequences with the charge motif, and the DEKR residue(s) in \(i+/-3\) or \(i+/-4\) positions are underlined.
A glutamate/aspartate motif was observed within 3-4 residues of the arginine site, and in a strictly alpha helix secondary structure this would position the negatively charged glutamate adjacent to the positively charged arginine. A possible explanation for the role of glutamate/aspartate in MG-H1 formation is that immediately after MG adducts to arginine and generates the MG-DH (R+72) ring (Figure 2.1), the carboxylate conjugate base catalyzes the dehydration step to form MG-H1. In this proposed mechanism, the negative charge of glutamate/aspartate does not necessarily influence the pKa of arginine, but instead may promote ring formation by enhancing the elimination of water from the intermediate dihydroxyimidazolidine ring. A second lysine/arginine motif was observed within 3-4 residues of the arginine site. In addition to arginine, lysine also has a high incidence (12.5%) of location in receptor binding domains (Gallet et al. 2000). As both arginine and lysine have relatively long side chains, they may conceivably interact at a distance of three residues, but this would ultimately be determined by the three-dimensional structure. A closely positioned positively charged primary amino lysine side chain would effectively lower the pKa of arginine. Thus, the apparent lysine/arginine motif may be explained, at least in part, to the enhanced initial formation of MG-DH. Arginine has the highest pKa of all amino acid side chains (pKa~12), and a neighboring lysine or arginine would render the said arginine more nucleophilic in attacking dicarbonyl electrophiles. Another potential role for neighboring lysines involves the initial attack of dicarbonyls by lysine, and since the lysine-dicarbonyl intermediate is transient, the lysine is able to “pass off” or transfer the dicarbonyl to arginine.
Figure 2.7. Methylglyoxal sites in human serum albumin.

HSA crystal structure 1A06 was visualized using Discovery Studio Visualizer v2.5 (Accelrys Software Inc). The 5 arginines modified by MG are shown in purple. Drug site I contains R257 and R218, and drug site II contains R410. R428 is located in the cleft between the two major alpha-helical bundles in the heart-shaped protein. R485 is located near drug site II, but is not involved in ligand binding.
Figure 2.8. Charged residues adjacent to MG sites in HSA.

R428 (top) lies in a positively charged cleft in between two lysine residues- K432 and K519. Semi-opaque surface is colored blue for positive charge, red for negative charge, and white for neutral charge. R218 (bottom, left) is the “gatekeeper” residue for entry into drug site I. Distance between guanidino head groups is 4.65 Å. R485 (bottom, right) is buried between two negatively charged glutamate side chains, E383 and E450, forming a double salt bridge. Distances between R485 and E383 is 3.21 Å and R485 and E450 is 4.91 Å. Coordinates from PDB 1A06 were used for visualization.
Closer inspection of the arginine sites (Figures 2.8, 2.9, 2.10, 2.12, 2.13) shows that the proximity of neighboring residues in the microenvironment support the charged residue (DEKR) motif. Figure 2.8 illustrates the R428 MG site (top) in a region of positively charged lysines, K519 and K432. This positively charged cleft will be explored more in depth in later chapters because larger dicarbonyls such as glucosone and 3DG prefer to adduct at this site in vivo. The R218 MG site (bottom, left) is adjacent to another arginine (R222), yet this arginine was not identified in the MudPIT search. Since this is an i+4 distance on an alpha helix, the interaction would have been predicted from the linear sequence. The R485 MG site (bottom, right) is sandwiched between two glutamate side chains (E383, E450) in a double salt bridge. The contribution of two negatively charged carboxylates could aid in the formation of the R+54 adduct.

In terms of functionality of HSA and drug binding, R257 in drug site I and R410 in drug site II are critical arginines (Figure 2.9). R257 forms a salt bridge with E153 and likely is influenced by the negative charge to create a favorable MG site in this deep pocket. Drug site II is a shallower pocket, within which R410 is situated close to a positively charged K414 i+4 on an alpha helix. Taken together, the influences of neighboring charge on the two arginines in HSA drug sites likely influence not only dicarbonyl adduction, but also drug binding. The HSA drug binding sites are identified in Figure 2.7, which also shows the distribution of the MG sites.
Figure 2.9. Target arginines in drug site I and II pockets in HSA.

R257 (left) is adjacent to a negatively charged glutamate E153 at a distance of 3.30 Å in drug site I. This interaction forms a salt bridge between two alpha helices. R410 (right) is adjacent to a negatively charged lysine at a distance of 5.43 Å in drug site II. Coordinates from PDB 1A06 were used for visualization.
The question of how common neighboring charge motifs are encountered is illustrated by Apolipoprotein A1 in Figure 2.10. The R61 MG site is 6.06 Å from another arginine, R116, and the pair satisfy the close proximity requirements of the charged residue motif. The inset shown on the left depicts two blue clouds of positive charge relatively close together when compared with the entire ApoA1 protein (right). In other words, when assessing the entire ApoA1 charge distribution it appears that most of the charged residues (blue and red clouds) are spread out. Thus, the motif is not common for the rest of the protein.

Plasminogen has 8 MG sites and Figure 2.11 shows 7 of the sites on the catalytic subunit (residues 546-791). Figure 2.12 shows the MG sites that appear to follow the adjacent charged residue motif. Figure 2.13 A shows R63 of alpha 1 acid glycoprotein 1. It shows that R63 is next to a YYY motif rather than a charged residue. Figure 2.13 B shows R220 of transferrin next to a negatively charged aspartate D236.
Figure 2.10. R61 MG site in apolipoprotein A1 is near a charged residue compared to other charged residues.

Semi-opaque surface is colored blue for positive charge, red for negative charge, and white for neutral charge. R61 is a MG site and is adjacent to a positively charged R116. The proximity of R61 and R116 in the inset (left) is contrasted with the separation between positively charged surface residues over the entire ApoA1 protein (right, blue surfaces). R61 is remarkably close to R116 (6.06 Å) compared to other KR residues, so the motif does not occur at non-MG sites in this instance. Coordinates from PDB 2A01 were used for visualization.
Figure 2.11. Catalytic subunit of plasminogen.

Colors for secondary structure: red is alpha helix, green is β-turn, and blue is beta sheet. Seven of the eight MG sites in plasminogen are located in the catalytic subunit. R561 is located in the activation loop region, where tissue plasminogen activator cleaves the bond (yellow dashed line) between R561 and V562 in the conversion to plasmin, the active form. Coordinates from PDB 1QRZ were used for visualization.
Figure 2.12. Plasminogen MG sites as examples of the charged residue motif.

A. R637 is 5.16 Å from K651 on an antiparallel beta sheet. B. Salt bridge R789 is 3.22 Å from E785 on an alpha helix. C. R767 is 7.12 Å from K770 on a beta turn. D. R712 is 4.39 Å from E714 in a modified beta turn, and 13.15 Å from K708 in a partial alpha helix. E. R779 is 8.01 Å from R776 in an alpha helix. Coordinates from PDB 1QRZ were used for visualization.
Figure 2.13. Microenvironment of alpha 1 acid glycoprotein 1 and transferrin MG sites.

A. R63 is near three tyrosine residues on a beta sheet of alpha 1 acid glycoprotein 1. Distance to Y74 is 3.05 Å, to Y65 is 3.10 Å, and to Y50 is 5.25 Å. B. R220 of transferrin is located 4.44 Å from a negatively charged carboxylate side chain of aspartic acid D236. Coordinates from PDB 3KQ0 and PDB 2HAU were used for visualization of alpha 1 acid glycoprotein 1 and transferrin, respectively.
2.3.3. **Shotgun proteomics to search for in vivo modifications**

These experiments evolved as a result of a protein profiling study in search of plasma biomarkers for type 2 diabetes. This study was designed to analyze differences in medium abundant plasma proteins using spectral counting. In an effort to delve into the plasma proteome, we employed sub-proteome separation using immunodepletion with an IgY-12 affinity column. This column contains immobilized chicken antibodies that bind to the twelve most abundant plasma proteins: albumin, IgG, fibrinogen, transferrin, IgA, IgM, Apo A-1, Apo A-2, haptoglobin, alpha-1-antitrypsin, alpha-1-acid-glycoprotein, and alpha-2-macroglobulin. The column was efficient in removing albumin and immunoglobulins, but did not remove all of the fibrinogen. While we did not find significant differential protein expression between normoglycemic and diabetic subjects, a search of the spectra for R+54 modifications revealed fibrinogen α modified at R493, and this was validated via treatment of plasma protein with MG. We did not detect any modifications in normoglycemic subjects, and did not detect the R+72 dihydroxyimidazolidine (MG-DH) in any of the spectra.

A false-discovery rate (FDR) of 2.0% was calculated based on the total number of spectra from a search of a reversed (decoy) protein database (208 spectra) divided by the total number of spectra found from the search using the true database (10,385 spectra). The percent of false positive spectra that contained the R+54 modification was 21% (44 spectra).
Figure 2.14. Unmodified, *in vitro* MG-modified, and *in vivo* MG-modified MS/MS spectral profiles of $^{492}$HRHPDEAAFFDTASTGK$^{508}$ peptide in fibrinogen $\alpha$ chain.

(A) Product-ion spectra of the 944.3 [M+2H]$^{2+}$ ions from diabetic plasma. Matching peptide containing unmodified R493 with one missed tryptic cleavage site. (B) Product-ion spectra of the 971.2 [M+2H]$^{2+}$ ions mapping MG-modified at R493 after incubation of plasma protein with 5mM MG (*in vitro*) exhibits similar peptide fragmentation to (A), except with a mass increase of R+54. (C) Product-ion spectra of the 971.3 [M+2H]$^{2+}$ ions found *in vivo* in diabetic plasma matches MG-modified at R493 with the R+54 adduct. Bold “R” indicates the MG-H1 (R+54) adduct.
All patient samples contained fibrinogen α in its unmodified form, yet only one diabetic subject had two MS/MS spectra that matched MG-modified at R493. The MS/MS spectra for unmodified and R+54 fibrinogen α peptides in diabetic plasma are consistent with an increase of 54 mass units at the R493 site, and are shown in Figure 2.14. Unmodified and MG-modified spectra show that the fragment ions are similar, and the correlation of the tandem MS data with theoretical spectra generated from known protein sequences identified the modification at the R493 site. R493 is also a likely missed tryptic site, as the unmodified (Figure 2.14 A) spectra also contained a missed cleavage. This property is serendipitous, because it allows for vertical alignment of the MS/MS fragmentation pattern from the in vitro MG-modified plasma (Figure 2.14 B) and unmodified diabetic plasma (Figure 2.14 C) to visualize the mass shift of +54 at R.

R493 is located in fibrinogen α at the opposite end of the protein from the thrombin cleavage site, where thrombin cleaves the fibrinopeptides from fibrinogen to form fibrin. However, even though MG-modification of R493 may not effect thrombin-fibrinogen interactions, it is possible that R493 is involved with interactions between fibrin monomers, which are necessary for clotting. Disruption of this interaction could lead to increased or decreased clotting. Currently, no crystal structure of the sequence that covers R493 exists in the Protein Data Bank (PDB), so the structural consequence of MG modification at this site is not available for investigation by molecular modeling. The question of whether MG-modification at R493 is a functional modification, or simply an artifact discovered by chance is still unclear. In either case, R493 has potential as a site to monitor MG exposure as a high affinity site.
2.3.4. Determination of the site of MG modification in a model peptide

The synthetic peptide \(^{241}\)VHTECCHGDLLECADDRADLAK\(^{262}\) from HSA was selected for validation of MG modification because this site was found in the results of the MudPIT experiment. In addition, the arginine in this peptide is R257, located at a functional site in the drug site I binding pocket. The cysteines on this peptide were capped with iodoacetamide prior to MG modification to prevent thiohemiacetal formation. We determined that a 5-fold excess of MG was sufficient to induce the modification specifically at arginine, and higher concentrations formed adducts at the N-terminus and lysine side chain.

Figure 2.15 shows a close-up region of the MS/MS spectra from the +3 parent ions of the MG-modified and unmodified peptides. The y6 ion at m/z = 727.3960 is reasonably close to the experimental m/z of 727.3991. Therefore, the +54 modification is either at arginine or at a residue C-terminal to arginine. Figure 2.16 shows a close-up region of the MS/MS spectra from the +4 parent ions. The y3 ion identified in both spectra at m/z = 331.23353 (top, MG-modified) and 331.23369 (bottom, unmodified) closely match the y3 exact mass of 331.2340. Therefore, the +54 modification does not exist on the C-terminal lysine. Unfortunately, we were unable to obtain y4 (m/z = 446.2609) or y5 (m/z = 517.2980) ions from either the MG-modified or unmodified peptides.
Figure 2.15. Accurate mass MS/MS of the R257 MG-modified synthetic peptide indicates arginine as the site of modification.

Top spectra is MS/MS of the +3 charge state (m/z= 880.3) of the MG-treated synthetic peptide, and the arrow indicates the mass of the y6 ion (727.39601). Bottom spectra is MS/MS of the +3 charge state (m/z= 862.4) of the untreated synthetic peptide, and the arrow indicates the absence of the 727 peak.
Figure 2.16. Accurate mass MS/MS of MG-modified synthetic peptide shows that the +54 modification is not on the C-terminal lysine. Top spectra is MS/MS of the +4 charge state (m/z = 660.5) of MG-treated synthetic peptide, and the solid arrow indicates the unmodified y3 ion. Bottom spectra is MS/MS of the +4 charge state (m/z = 647.5) of the untreated peptide, and the solid arrow indicates the y3 fragment ion. Open arrow indicates the abundant fragment at 237.13 m/z. This corresponds to the b2 ion, and this fragment will be added to the transition list and monitored as discussed in chapters 4 and 5. The presence of this fragment in both spectra is evidence that the +54 modification does not reside on the N-terminus.
2.4. Discussion

This is the first study to detect multiple MG–modified proteins in the plasma proteome. We identified two spectra that matched MG-modified at R493 in the α chain fibrinogen in a subject with diabetes. Fibrinogen is involved in blood coagulation and as an acute phase reactant in inflammation. In the vascular system, diabetes mellitus encompasses a heterogeneous group of bleeding disorders that includes excessive clot formation and abnormal bleeding diathesis (Vague et al. 1997). In diabetic patients, various abnormalities in coagulation and fibrinolysis are found, i.e., higher fibrinogen concentrations, elongation of the thrombin clotting time, reduced fibrinogen survival, and elevated fibrinopeptide A and fibrin degradation products (Vague et al. 1997). Although diabetic subjects have an increased susceptibility to clot formation, they form abnormal clots (Vague et al. 1997). In poorly controlled diabetic patients, impaired alpha-chain cross-linking of fibrin monomers has been suggested to lead to a less stabilized clot (Lutjens et al. 1988), which is more susceptible to plasmin degradation. The propensity in diabetes toward increased clot formation is characterized by resistance to fibrinolysis by plasmin (Vague et al. 1997). We detected MG modified plasminogen in vitro, but not in vivo. Posttranslational modification of plasminogen by MG may contribute to the impaired fibrinolysis in diabetes (Lerant et al. 2000). Impaired clot formation and clot fibrinolysis characterize diabetes, and MG adduction may help explain these clotting defects.

The persistent hyperglycemia that occurs in diabetes is associated with enhanced fibrinogen glycation (Dunn et al. 2005), as all plasma proteins are exposed to the
deleterious effects of increased AGE precursors in diabetes. Poor glycemic control is, however, associated with complex metabolic derangements in addition to hyperglycemia, such as oxidative and carbonyl stress. Prior work using top-down proteomic analysis indicated an increase in the mean mass of individual $\beta$ and $\gamma$ chains of fibrinogen purified from diabetic subjects compared with those purified from controls (Dunn et al. 2006). Moreover, this increase was not an exact multiple of individual glucose molecules, but rather a mixture of their subsequent complex modifications by oxidation/reduction and dehydration.

Fibrinogen was not chosen to further study the functional consequences of MG modification because it has a short half-life (4 days), clotting is slightly delayed upon MG modification, and the structure of fibrinogen is not readily analyzed using molecular modeling. Inducible proteins with a short half life are less susceptible to the deleterious effects of biologically reactive intermediates because rapid protein turnover can nullify protein damage. Preliminary experiments also showed that MG-treated fibrinogen delayed clotting time (data not shown) only when using high concentrations of MG (250 $\mu$M). Therefore, MG-modification of fibrinogen is not likely a “lynchpin” that links diabetes with increased clot formation. It is entirely possible, however, that MG-modification of fibrinogen leads to clots that are resistant to fibrinolysis, and this would delay clot degradation and fit with the current understanding of diabetic coagulation abnormalities. Once crystal structures of fibrin-fibrin clots are available, we will be better equipped to determine if MG strengthens the interactions between fibrin monomers.
Although we did not find MG modifications in the spectra from normoglycemic subjects, MG modified proteins may exist in these subjects at lower abundance. It should be noted that in vivo modifications were found on fibrinogen, a protein that should have been depleted before MS/MS analysis with the immunoaffinity column. It is possible that glyco-oxidized fibrinogen is structurally altered so that the anti-fibrinogen antibodies do not bind modified protein. Hence, by serendipity, the IgY-12 column may enrich for certain modifications by pulling out abundant unmodified proteins from solution. This approach could be advantageous for the enrichment of modifications that do not yet have reliable antibodies directed against them, such as the MG-H1 modification. Another possible explanation is that the column was simply saturated with fibrinogen, so all forms (modified and unmodified) did not bind completely.

We found several sites that are susceptible to MG modification (Figure 2.5). Most of these arginines were also found with the +72 modification. The dihydroxyimidazolidine and open ring intermediates are isobaric and indistinguishable by mass spectrometry. However, the presence of the +72 adduct at these sites further validates these sites as real targets for MG, because arginine must form these structures before the subsequent loss of water to generate the +54 adduct. The 7 plasma proteins with MG modifications collectively contain 216 arginine residues, yet only 35 MG sites were found using the MudPIT approach. It is possible that the sites could be critical in terms of function for those particular proteins.

Several of these proteins were previously shown by others to have altered function after MG addition. MG-modified albumin at R410 in drug site II inhibited
albumin esterase activity and ketoprofen binding (Ahmed, Dobler et al. 2005). Another important albumin arginine that was identified was R257 in drug site I. This site was not identified by Ahmed et. al. (Ahmed, Dobler et al. 2005), possibly due to the interferences from multiple overlapping peaks with their LC-MS only approach. Tandem MS/MS is not affected by these interferences, because overlapping peaks are separated and individually fragmented. Using this approach we were also able to characterize the R485 MG site in HSA. Of the seven abundant plasma proteins, HSA was chosen for further study in terms of functional consequences of dicarbonyl adduction and oxidation, as well as site-specific quantification in human subjects.

The proposed protein motifs susceptible to MG modification are supported by the neighboring charged residue microenvironment (Figures 2.17 and 2.18). Based on observations of the structures analyzed herein, a negatively charged D or E located 5 Å or less or a positively charged R or K located 8 Å or less, from the arginine guanidinium reasonably approximates this motif. The ‘triple tyrosine’ surrounding the MG site in alpha 1 acid glycoprotein 1 may also exist in other proteins.

The formation of MG-modified proteins may help explain the development and pathology of diabetic complications. Chapters 4 and 5 of this dissertation involve quantitative analysis of these modifications in order to identify predictors of various diabetes complications, as these candidate biomarkers may serve as targets for new approaches of therapy that target both glucose and oxidative stress.
Figure 2.17. Neighboring positive charge decreases the pKa and increases the nucleophilicity of arginine.

Nucleophiles such as arginine and lysine are more reactive when they are in close proximity to another (top scheme). Neighboring nucleophiles such as lysine or arginine can increase the nucleophilicity of arginine by anchimeric assistance and/or neighboring group participation. The pKa of arginine is also decreased when two positive charges (bottom scheme) come in close contact. Due to the repulsive effects of two nearby positive charges, the likelihood of the two close side chains in sharing a single proton increases. This frees up arginine electrons to attack other electrophiles.
Figure 2.18. Base-catalyzed dehydration of dihydroimidazolidine by a neighboring carboxylate.

Dihydroimidazolidine rings have an acidic proton that can be deprotonated by carboxylates from nearby aspartate or glutamate. This leads to loss of water from the ring, and the enediol can tautomerize to the hydroimidazolone.
CHAPTER 3: VALIDATION OF DRUG SITE I AS A TARGET OF METHYLGlyOXAxl ADDUCTION OF HUMAN SERUM ALBUMIN

3.1. Introduction

Human serum albumin (HSA) is a multifunctional protein that exhibits important enzymatic functions in addition to playing a critical role in: ligand binding, maintenance of osmotic pressure, transport of biomolecules (fatty acids, hormones, calcium, bilirubin), and buffering of plasma pH. HSA is a negatively charged 585 residue single-chain protein with a molecular mass of 66,500 Da and an isoelectric point of 5.16 (Curry et al. 1998). HSA forms a heart-shaped protein with about 67% \( \alpha \)-helix and no \( \beta \) sheet secondary structure. HSA is not normally considered to be glycosylated, that is, sugar modification via enzymatic attachment of glycans (Carter et al. 1994). HSA contains 17 intramolecular disulfide bonds, which involve a total of 34 cysteine residues apart from a single unbonded cysteine.

Two major sites of ligand binding to HSA are at drug site I and drug site II. Drug site I is also known as the warfarin or phenylbutazone site, and drug site II is also called the benzodiazepine or ketoprofen site. Site II is smaller or shallower than site I, and site II is the less flexible pocket because binding at this site can be strongly affected by drug stereoselectivity (Kragh-Hansen et al. 2002). These two drug sites are highly adaptable binding cavities that are involved with charge neutralization and hydrogen bonding interactions with acidic or electronegative small molecule ligands (Ghuman et al. 2005). Drug site I contains basic residue side chains in the pocket, and does not contain acidic side chains. Broadly speaking, the small molecules that bind drug site I possess centrally
located anionic or electronegative features due to the presence of polar (negatively charged) patches in the middle of the pocket flanked by apolar regions (Ghuman et al. 2005). The integrity of drug site I is critical because warfarin has such a narrow therapeutic range (Ha et al. 2000). Approximately 99% of plasma warfarin is bound to HSA, and slight changes in the affinity of HSA for warfarin can have profound effects on bioavailability (Petersen et al. 2002). Drug site II, however, is generally selective for drugs with a terminal electronegative groups distal from a hydrophobic core (Ghuman et al. 2005).

A hotspot for MG adduction on HSA at R410 that impacts ketoprofen binding and esterase activity in drug binding site II has been described (Ahmed, Dobler et al. 2005). While this is a major site for MG modification in vitro, studies of human albumin mutants found that modification of R410 results in rapid elimination in mice (Iwao et al. 2006). MG modification at this hotspot may render HSA susceptible to elimination, and thus, may not accumulate to appreciable levels as a result of protein turnover. As we are unable to detect MG modification at R410 in vivo (Chapter 5), but are able to detect MG modification at drug site I (R257) in vivo, an investigation of the functional impact of MG modification at HSA drug site I is warranted.

In addition to esterase activity at drug binding site II, HSA plays a significant role in the breakdown of prostaglandins and stabilization of other eicosanoids in drug site I (Fitzpatrick et al. 1984; Yang et al. 2002). A spectrophotometric assay was developed to measure the catalytic dehydration of 15-keto PGE₂ to 15-keto PGA₂ during which the substrate undergoes keto-enol tautomerization and absorbs light at 505 nm (Figure 3.1)
(Yang et al. 2002). Site-directed mutagenesis and expression of fragments of HSA demonstrated that the R257M mutant and expression of albumin fragments other than domain II (containing drug binding site I) had the most profound decrease in formation of 15-keto PGA₂. It was determined that drug site I in HSA is an important regulator of prostaglandin metabolism, due especially to the high concentrations of albumin in blood.

The fluorescent probe prodan (6-propionyl-2–dimethylaminonaphthalene) binds with high affinity (~10 μM) to drug site I, accompanied by enhancement of fluorescence and a blue-shift of the emission maxima (Moreno, Cortijo et al. 1999). Displacement of prodan from drug site I results in decreased fluorescence when known site I compounds are added (warfarin, phenylbutazone) to the prodan-HSA complex, but not when drug site II compounds (e.g. diazepam) are added (Moreno, Cortijo et al. 1999).

Studies described in this chapter aim to validate drug site I (specifically R257) as a target for MG adduction. Two biochemical assays, one investigating prostaglandin catalysis and one which monitors the displacement of prodan, were utilized as functional assays to validate MG modification. Finally, in silico molecular modeling, dynamics, and docking simulations were employed to model the structure-function consequences of MG modification at R257 in terms of warfarin binding.
Figure 3.1. Conversion of 15-keto PGE\textsubscript{2} to 15-keto PGB\textsubscript{2} in drug site I of HSA.

Dehydration of 15-keto PGE\textsubscript{2} in drug site I of human serum albumin yields a keto-enol tautomer hybrid (15-keto PGA\textsubscript{2}) that absorbs light at 505nm. Further rearrangement of the chromophore leads to 15-keto PGB\textsubscript{2}, which does not absorb light at 505 nm.
3.2. Methods

3.2.1. Materials

All materials were HPLC grade and purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. Synthetic peptides were custom ordered from New England Peptide (Gardner, MA). Sequencing grade trypsin was purchased from Promega (Fitchburg, WI). 15-keto Prostaglandin E2 (PGE₂; catalog #14720) was purchased from Cayman Chemical (Ann Arbor, MI). Fatty acid-free human serum albumin (catalog # A3782) was purchased from Sigma-Aldrich. Lipidex-1000 was purchased from PerkinElmer (Waltham, MA). Prodan (6 - Propionyl - 2 – dimethylaminonaphthalene) was purchased from Anaspec Inc (Fremont, CA, catalog #88212, lot #64774).

3.2.2. 15-keto Prostaglandin E₂ conversion assay

Lipidex-1000 was buffer exchanged into 1 X PBS, pH 7.4. HSA (50 mg) was dissolved in 4.96 mL 1X PBS, pH 7.4 to a final concentration of 10 mg/mL. HSA was delipidated with Lipidex for 30 minutes. MG dilutions were created from 200 mM stock to concentrations of 2.0, 1.5, 1.0 and 0.5 mM. These dilutions were then used as stock for a 10 fold series dilution yielding 13 separate concentrations of MG in PBS. The concentrations used for the assay were 2, 1.5, 1, .5, .2, .15, .1, .05, .02, .015, .01, .005, .002 mM. 100 µL of HSA solution was incubated with 100 µL of MG solution, yielding protein concentration of 5 mg/mL and MG concentration half of those reported above, for 20 hours at 37°C. 5 mg 15-keto PGE₂ was dissolved in 1 mL ethanol. Aliquots of 20 µL (50, each containing 100 µg) were taken, the EtOH evaporated, and the aliquots stored at
-20°C. 15-keto PGE$_2$ was resuspended in 200 µL 1X PBS, pH 7.4. Following incubation, each of the 13 HSA/MG solutions was treated with 16 µL of PGE$_2$. The absorbance of each reaction at 505 nm ($A_{505}$) was measured immediately following the addition of PGE$_2$ giving a time point of $t=0$ using a blank containing 200 µL 1X PBS and 16 µL PGE$_2$. Measurements of $A_{505}$ were taken at time points $t=30, 60, 90, 120, 240$, and 300 minutes.

3.2.3. Prodan displacement assay

A 20 ml solution of HSA at 10 mg/ml in 1X PBS pH 7.4 was prepared. A 10 ml slurry of Lipidex-1000 in methanol was buffer exchanged into 1X PBS and the buffer was decanted from the aqueous slurry. The HSA solution was added to the slurry and the mixture was rotated at room temperature for 30 minutes. The Lipidex slurry was completely removed from HSA by plunging through a 0.22 µm filter using a 60 ml syringe. Prodan (25 mg) was dissolved in 5 ml of 40% ethanol, 60% methanol (final concentration is 5 mg/ml). A 10 ml portion of delipidated HSA at 10 mg/ml was incubated with 0.667 µL of the prodan solution (1:10 HSA:prodan), and this mixture was rotated at room temperature for 30 minutes. The HSA-prodan mixture was placed on dry ice and acetone precipitated by adding 40 mL of acetone at -20°C. This was vortexed and spun at 3000xg for 10 minutes to pellet the precipitated HSA-prodan. Two additional washes of cold acetone were used to remove any free prodan. The HSA pellet was brought to 10 mL with PBS and vortexed and resuspended.
A 2-fold serial dilution of MG in 1X PBS pH 7.4 was prepared starting from a stock solution of 10 mM. An equal volume of HSA-prodan and MG solution (200 μL each) was mixed in triplicate reactions for each MG concentration and incubated at 37°C for 30 minutes. A 300 μL aliquot of each reaction was read using a Gemini XPS Fluorescence Microplate Reader (Molecular Devices) in endpoint mode with 360 nm excitation, 420 nm filter cut-off, and 465 nm emission.

3.2.4. Affinity docking and molecular dynamics simulations

Our laboratory has previously reported the use of molecular modeling as a tool for simulating electrophile adduction (Fisher et al. 2007). The X-ray crystal structure coordinates for HSA complexed with warfarin (PDB code: 2bxd) were used as a starting model (Ghuman et al. 2005). MG-modified hydroimidazolone (MG-H1) was built using Insight II – builder module (Accelrys Inc., San Diego, CA). The charges were assigned using consistent valence force field (CVFF) parameters. Docking studies were carried out using Affinity docking program within Insight II 2005L modeling software. Twenty different conformations were generated for each ligand using Metropolis algorithm and distance dependent dielectric = 4.00 during docking. Best docking poses were refined using $10^5$ steps of Discover 3.0 minimization in case of methylglyoxal.

MG-modified adduct was formed on R257 of HSA. The modified structure was then subjected to $10^5$ steps of minimization using Discover 3.0. R-(+) warfarin and human serum albumin (native and MG-modified) complexes were then soaked with a 10 Å layers of TIP3P water molecules. This assembly was then subjected to dynamic
simulations for 250 ps. Trajectories were collected every 0.1 ps. The lowest-potential energy structure was selected and then minimized using $10^5$ steps of minimization. The final minimized structure was then used for the analysis.

### 3.2.6. Surface area calculations and molecular modeling

The GETAREA (v.1.0 beta) algorithm (Fraczkiewicz et al. 1998) was used to calculate solvent accessible surface area of side chains from the HSA crystal structure PDB 1AO6. The probe was set at 1.4 angstroms to simulate water, and the operations were performed without a gradient. The ratio of side-chain surface area to the "random coil" value is calculated per residue. The "random coil" value of a residue X is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations, where X is equal to M,W,C, or R. Residues are considered to be solvent exposed if the ratio value exceeds 50% and to be buried if the ratio is less than 20%. 
3.3 Results

3.3.1 MG-induced inhibition of PGE\textsubscript{2} catalysis

MG inhibits the HSA-mediated conversion of 15-keto PGE\textsubscript{2} to 15-keto PGA\textsubscript{2} (Figure 3.2). Absorbance vs. time was plotted and abs/min was calculated from the slope using linear regression. Dividing by the extinction coefficient (\(\varepsilon=35000\text{M}^{-1}\)) absorbance was transformed to concentration. The coefficient .000216 L (volume of the reaction in L) was applied to yield units of mol/min. Because 1 mg of HSA was used in the assay, and the mol unit can be transformed to nmol by application of a coefficient, the final units for the assay become nmol/min*mg HSA. These activities were plotted vs. log[MG].

The HSA concentration in this reaction was 75 \(\mu\text{M}\), and MG modification was found to inhibit the dehydration reaction significantly (p<0.05) at 100 \(\mu\text{M}\). This is near-equimolar inhibition (1.33:1 MG:albumin), suggesting that MG could play a significant role in prostaglandin metabolism \textit{in vivo}.
**Figure 3.2.** MG-modified HSA inhibits prostaglandin catalysis at drug site I.

Human serum albumin was modified with MG at various concentrations and 15-keto PGA$_2$ formation was monitored by absorbance at 505nm. Slope of the initial rate was converted to activity using a molar extinction coefficient of 35,000 M$^{-1}$ cm$^{-1}$. Means ± SD for three separate experiments are given.

*Significant (p<0.05) as compared to control (HSA without MG treatment)

***Significant (p<0.001) as compared to control (HSA without MG treatment)
3.3.2. Displacement of prodan from HSA-prodan complex by MG

The unique spectral properties of prodan make it an effective tool for the study of drug site I. When prodan is free in solution (Figure 3.3), it exhibits a fluorescent maximum at 520 nm when excited at 380 nm. When prodan is bound in drug site I of HSA, however, it undergoes a Stoke’s shift (blue shift) of 55 nm and this complex absorbs light at 380 nm and gives a fluorescent maximum at 465 nm (Figure 3.3). This phenomenon may be the result of radiationless energy transfer between W214 and bound prodan (Moreno and Gonzalez-Jimenez 1999). The prodan-HSA complex can be monitored spectrophotometrically because HSA alone does not exhibit autofluorescence in this range (Figure 3.3).

Using the fluorescent endpoint of 465 nm for the prodan-HSA complex allowed for the study of the impact of MG on drug site I. Any perturbations of this site can be indirectly measured by the displacement of prodan, which will ultimately decrease fluorescence at 465 nm. Therefore, maximum signal indicates unaffected prodan-HSA binding. Free prodan was purified from the HSA-prodan complex, and this complex was treated with a dilution of MG in triplicate reactions. Over time, prodan is naturally displaced from the HSA drug site I pocket, and the time point of 30 minutes was chosen because it permits sufficient time for MG to react with the protein complex. The “leakiness” of prodan from the complex does not permit analysis of sophisticated kinetic measurements, however, the measure of the complex can be determined relative to MG treatment at any given time point.
Figure 3.3. Stoke’s shift of HSA-prodan complex yields fluorescent maxima at 465 nm.

HSA was incubated with a 10-fold excess of prodan and allowed to equilibrate for 30 minutes. Unbound prodan was removed by acetone precipitation and the emission spectra of free prodan, HSA-prodan and HSA were obtained. Excitation $\lambda$ is 380 nm and upon binding HSA, prodan $\lambda_{\text{max}}$ blue-shifts from 520 nm to 465 nm. HSA without prodan (bottom spectra) shows negligible spectral emission. The endpoint values to monitor prodan displacement from the HSA-prodan complex is 380 nm excitation, 465 nm emission.
The effect of MG on the prodan-HSA complex is shown in Figure 3.4. Relative to the unmodified control, where no MG was added to prodan-HSA, MG treatment is significant in displacement of prodan at concentrations above 300 μM. The concentration of HSA (75 μM) in each reaction is the same as used in the prostaglandin assay. The data indicate that drug site I is targeted by MG in terms of binding to the prodan fluorescent probe. The difference in significant MG concentrations of this prodan assay compared to the prostaglandin assay may be due to the residues that are involved in prodan binding. Since the PG assay is more sensitive to MG than the prodan assay (100 μM vs. 300 μM), it is likely that prodan binding is affected less by arginine modification than PG hydrolysis at the same site. Currently, there is no available crystal structure of HSA bound to prodan, but from this data it seems that arginine side chains are not the primary means by which prodan interacts with drug site I. Nevertheless, eventual displacement of prodan from this site at higher MG concentrations is further evidence that drug site I is a target for MG adduction.
Figure 3.4. Prodan is displaced from drug site I by MG modification of HSA.

Prodan bound to HSA (prodan-HSA complex, 75 μM) was treated with a dilution of MG and the fluorescence (excitation 380 nm, emission 465 nm, filter cut-off 420 nm) was measured after 30 minutes. Means ± SD for three separate experiments are given. rfu: relative fluorescence units

*Significant (p<0.05) as compared to control (HSA without MG treatment)

**Significant (p<0.01) as compared to control (HSA without MG treatment)

***Significant (p<0.001) as compared to control (HSA without MG treatment)
3.3.3. Solvent accessibility of HSA side chains

Using the X-ray structure of HSA (PDB 1AO6), a relative value for each side chain was calculated. The GETAREA program rolls a probe over the surface of the crystal structure, then calculates the surface area for each side chain. This value is given relative to the solvent accessibility of that side chain in a random coil conformation represented by the tripeptide GXG, where X equals one of 20 different amino acid residues. The side chain in GXG is considered to be 100% solvent accessible.

The top chart in Figure 3.5 illustrates this relative solvent accessibility value for oxidizable residues methionine, tryptophan, and free cysteine of HSA. As these side chains are largely hydrophobic, it makes sense that they would be buried. In fact, all of these side chains are less than 50% solvent exposed in HSA relative to the random coil GXG peptide. There are 6 methionine residues in HSA, and M299 and M329 are the most solvent exposed (40-45%). This will become important in Chapter 4 because these two sites are the most responsive to oxidation. Solvent accessibility of tryptophan (W214, 20-25% relative solvent accessibility) is roughly halfway between the lowest and highest methionine solvent accessibilities. The sole free cysteine (C34, 0% relative solvent accessibility) is completely buried in this crystal structure.

HSA solvent accessibilities for all of the arginines are provided in the lower chart in Figure 3.5. Thirteen of the 24 arginines were analyzed for MG adducts in this dissertation, and this set provides a range of solvent accessibilities. At the low end, R485 is 4% solvent accessible, and at the high end, R81 is 80% solvent accessible. The remaining arginine side chains have solvent accessibilities that lie between these two
extremes. Drug site I contains R218, R222, and R257, all of which are relatively buried sites from 7-22% solvent accessible. R410 is the only arginine in drug site II, and this side chain is also relatively buried at 32% solvent accessibility.
Figure 3.5. Solvent accessibility of oxidizable (M, W, or C) or arginine (R) HSA residues.

Solvent accessible surface areas were calculated by using the PDB X-ray structure 1AO6 and GETAREA (v.1.0 beta). Radius was water (1.4 Å) and all side chains were relative to GXG side chain solvent accessibility. Top histogram are the methionine, tryptophan, and free cysteine side chain solvent accessibilities in the HSA crystal structure. Bottom histogram includes all arginines in HSA; R257 is in drug site I, and R410 is in drug site II.
3.3.4. Molecular dynamics and affinity docking of MG-modified R257

**Figure 3.6** presents an overlay showing the changes in drug site I pocket of unmodified R257 (yellow) to MG-modified R257 (purple), and the energy-minimized conformations of warfarin (dark yellow for unmodified R257, dark purple for MG-modified R257). Atomic distances consistent with hydrogen bonding interactions are indicated with dashed lines. Oxygen atoms that participate in hydrogen bonding interactions are designated in the warfarin structure (inset of **Figure 3.6**). In its unmodified form (yellow), R257 participates in hydrogen bonding with the aliphatic ketone (A) of warfarin. Another important interaction of warfarin and the drug site I pocket is between R222 and the lactone carbonyl (B). These two key arginine interactions (R257 and R222) help stabilize and orient warfarin in this drug binding site.
Figure 3.6. Methylglyoxal modification of R257 in drug site I changes conformation of warfarin binding.

Structure coordinates from PDB 2BXD were used as a starting model for molecular dynamics and energy minimization. Drug site I residues are overlaid for unmodified R257 (yellow) and MG-modified R257 (purple). Energy minimized warfarin conformation is colored dark yellow for unmodified R257, and dark purple for MG-modified R257. Dashed lines indicate hydrogen bonds. Inset shows warfarin structure labeled at three oxygen atoms (A), (B), and (C) that participate in hydrogen bonding.
MG-modified R257 (purple) is not able to form hydrogen bonding interactions with warfarin, and this causes a shift that frees warfarin to rotate and engage in additional interactions with H242 and K199. These two residues (H242, K199) do not form hydrogen bonding interactions with warfarin when R257 is in its unmodified, positively charged form. MG modification at R257 essentially pulls K199 into the binding pocket to replace this lost interaction, which is between the ketone (A) of warfarin and the ε-amino side chain of K199. H242 also picks up an interaction, and forms a hydrogen bond between the imidazole nitrogen of the histidine side chain and the keto-enol oxygen (C) of warfarin. The overall result of MG modification at R257 is that warfarin loses this arginine interaction, and this effect is compensated by a change in pocket conformation that pulls in H242 and K199, thus making a tighter pocket. One interaction that remains unchanged, however, is between R222 and the lactone ketone (B) of warfarin.

Although warfarin has a chiral center and is clinically available as a racemic mixture, the R-(+) and S-(−) enantiomers of warfarin bind HSA in similar positions as one another (Ghuman et al. 2005). Figure 3.6 illustrates the R-(+) enantiomer, and this isomer was chosen because the available crystal structure is based on this form. The hydroimidazolone formed from MG modification on arginine also has a chiral center, and the (R) enantiomer is shown. It is not known which hydroimidazolone enantiomer is formed predominantly in vivo, so we docked warfarin at both (R) and (S) isomers. Interaction energy values for R-(+) warfarin binding to drug site I were: R257 (unmodified) - 77.1 kcal/mol, R257-MG adduct (R isomer) - 60.6 kcal/mol, R257-MG adduct (S isomer) - 64.3 kcal/mol. The increase in free energy (+ΔG) as a result of
modification indicates that MG modification at R257 decreases HSA affinity for warfarin.

3.4. Discussion

Identification of the R257 MG modification was the starting point to test this modification in an assay that is sensitive to R257 function. It was determined through site-directed mutagenesis that R257 is critical for HSA-mediated prostaglandin catabolism (Yang et al. 2002). Although 15-keto PGE$_2$ and its derivatives are not found in nature, their spectral properties are useful in monitoring the integrity of drug site I. MG-modification at R257 may decrease the base-catalyzed elimination reaction, as the neutral hydromidazolone is not able to abstract a proton from carbon 10 of PGE$_2$. The potential link between MG adduction and prostaglandin metabolism is interesting, and certainly merits further investigation. Other enzymatic activities of drug site I include degradation of sulbenicillin, with the R-isomer being degraded much faster than the S-isomer (Fleury et al. 1997).

Warfarin has intrinsic fluorescence in solution, and this fluorescence increases nearly 10-fold when it binds to drug site I of HSA (Ha et al. 2000). However, the optimum spectra from warfarin (excitation 320 nm, emission 385 nm) are too close to the autofluorescence of aromatic side chains such as tryptophan, phenylalanine, and tyrosine to be used as a sensitive marker for displacement assays. This is why most of the studies that measure warfarin fluorescence do not use a standard plate reader, rather, they use a spectrafluorometer with custom filter sets to reduce spectral noise resulting from protein
autofluorescence. We wanted to establish an assay that was amenable to a standard plate reader, so we instead developed this assay based on prodan. The benefit of prodan is that it is much further red-shifted on the spectrum from the autofluorescence of HSA side chains. Another benefit is that, unlike warfarin, the emission peak of the bound prodan is quite shifted from the emission peak of the free prodan. This property is useful in detecting small changes in free vs. bound conformations. In addition to warfarin and prodan, other probes have been used to target drug site I, such as 5-dimethylaminonaphtalene-1-sulfonamide (DNSA), dansylamide, dansyl-L-glutamine, dansyl-L-asparagine, dansyl-L-lysine, n-butyl p-aminobenzoate, and phenol red (Kragh-Hansen et al. 2002). In a displacement assay to study MG, any one of these probes would likely exhibit similar behavior to prodan. Using two different assays we were able to determine that MG affects the integrity and function of drug site I in HSA.

Molecular dynamics simulations based on available crystal structures were useful in illustrating the impact of MG-modified R257. The interaction energy value estimate is based on computer algorithms used in *in silico* drug design, and is useful for ranking drug candidates based on docking interactions with an active site. Therefore, the approximate decrease of 15 kcal/mol in interaction energy as a result of R257 MG modification could result in a significant change in warfarin bioavailability. Decreased protein binding would increase the proportion of free drug, thus altering warfarin pharmacokinetics. Future studies will investigate the impact of MG modification on warfarin binding.
CHAPTER 4: DETERMINATION OF THE HIGH-AFFINITY SITES FOR OXIDATION AND DICARBONYL ADDUCTION OF HUMAN SERUM ALBUMIN

4.1. Introduction

Human serum albumin (HSA) is a good model protein with which to study protein damage by oxidation and reactive electrophiles, such as dicarbonyls. HSA is the most abundant protein in the plasma at ~0.6 mM (Suji et al. 2008), so quantification of the products of glyco-oxidation on HSA is important because these protein adducts are among the most abundant in vivo. Moreover, the HSA protein presents a diverse environment of nucleophiles and susceptible sites for oxidation. HSA may be used as a marker or index for glyco-oxidation, yet the relative reactivity of individual sites is largely unexplored. Site-specific details of HSA damage can be used to assess the impact of modification on the many functional roles of HSA. Such findings are also useful in applications to study other proteins that are also targets for glyco-oxidation. HSA contains 24 arginines, 1 tryptophan, 6 methionines, and 35 cysteines. Only 1 of the cysteines (C34) is not disulfide bonded, leaving 34 of the cysteines to participate in 17 intra-protein disulfide bonds. Dicarbonyl modifications mainly target the arginine residues to generate the corresponding hydroimidazolone (Figure 4.1). The mass changes associated with adduction of three dicarbonyls to internal peptide arginines will be investigated at the peptide level using 12 synthetic HSA sequences. These dicarbonyl-modified peptides will be used to generate an MRM transition list to validate in a protein. To date, no published studies have determined the site-specific reactivity of any protein-bound arginines toward either 3DG or glucosone.
Figure 4.1. Hydroimidazolone adducts from dicarbonyl modification of arginine and the corresponding mass changes.

Arrows represent arginine adduction. Dicarboxyls glucosone, 3DG, and MG will target arginine residues in proteins to yield corresponding mass increases of R+160, R+144, and R+54, respectively.
The effect of oxidation at MWC residues represents the most common target sites of physiological oxidative stress. However, other protein side chains are susceptible to oxidation, such as oxidation of histidine (H) to 2-oxo-histidine (Bridgewater et al. 2007), oxidation of proline (P) to hydroxyproline (Stadtman 1993), as well as oxidation of tyrosine (Y) to many different ring substituted products, such as 3-nitrotyrosine (Greenacre et al. 2001).

Oxidation of methionine (Figure 4.2) is one of the most common protein modifications (Wells-Knecht et al. 1997), and is likely the most abundant oxidation product of proteins. The sulfur group is readily oxidized to the sulfoxide (M+16), and under conditions of extreme oxidation the sulfone (M+32) can be generated (Khor et al. 2004). The susceptibility of methionine to oxidation is of concern because it is so sensitive to oxidation that even sample preparation may induce M+16 oxidation (Maier et al. 1995).
Figure 4.2. Oxidation of methionine (M) and corresponding masses.

The sulfur group is easily oxidized to form the sulfoxide (+16) and prolonged oxidation will produce sulfone (+32).
The indole ring of tryptophan (W) has a moderate reduction potential and is also susceptible to oxidation (Figure 4.3) (Davies et al. 1999). Initial oxidation generates the 2-hydroxytryptophan and oxindolyalanine tautomeric pair (W+16), although NMR studies of oxidized tryptophan have demonstrated that the oxindolyalanine structure predominates, despite the aromaticity of 2-hydroxytryptophan (Simat et al. 1998). Further oxidation breaks the C2-C3 bond and opens up the indole ring to give N-formylkynurenine, which may be identified as (W+32). Eventually, the N-formyl group is hydrolyzed, producing the advanced oxidation product kynurenine (W+4). Kynurenine is an example of a highly oxidized side chain with added mass that is not a multiple of oxygen.
Figure 4.3. Major oxidation products of tryptophan (W) and the corresponding mass changes.

The C2-C3 double bond of the indole ring on the side chain of tryptophan is initially oxidized to oxindolyalanine and its tautomer 2-hydroxytryptophan. Further oxidation cleaves the C2-C3 bond to yield N-formylkynurenine. Hydrolysis of the n-formyl group generates formic acid and the advanced oxidation product kynurenine.
Mass changes associated with oxidation of cysteine (C) (Figure 4.4) are similar to that of methionine, except the initial formation of sulfenic acid (C+16) is not considered a stable adduct. Cysteine sulfenic acid is even considered reactive and exhibits both nucleophillic and electrophillic properties (Carballal et al. 2003). Prolonged oxidation of the cysteine generates cysteine sulfone (C+32) and cysteine sulfonic acid (C+48). The C34 site in HSA represents the primary source for free thiol in the blood, and since glutathione (GSH) is absent from the blood, C34 of HSA acts as an important circulating antioxidant. Despite this, virtually no studies have been capable of quantifying the relative oxidized forms at the C34 site in vivo. There is a similar lack of understanding of the predominant oxidized forms of the sole tryptophan (W214) in HSA. Both of these sites, as well as the six methionine residues, could be useful as markers for oxidative stress in human blood. Similarly, of all the 24 arginine residues in HSA, none have been assessed in terms of their affinity for dicarbonyls.
Figure 4.4. Modification of cysteine (C).

During sample preparation of protein to peptides, cysteine is commonly treated with iodoacetamide to cap the reactive thiol (C+57, top scheme). Cysteine may be oxidized (left pathway) to the intermediate sulfenic acid (C+16) or to the sulfone (C+32) or to the sulfonic acid (C+48). Oxidized cysteines, such as sulfone and sulfonic acid are not reduced by DTT, so they are not capped by iodoacetamide during sample prep.
Synthetic peptides with oxidative modifications such as methionine sulfoxide and cysteic acid are commercially available from vendors that perform custom synthesis. However, some of these modifications are very expensive, so we obtained the relatively cheaper unmodified peptides, and oxidized the side chains post-synthesis. The overall goal is to obtain optimal transitions to be validated in \textit{in vitro} models and ultimately in human plasma. Chapter 4 attempts to identify the reactive sites for arginine-dicarbonyl (MG, 3DG, glucosone) adduction and oxidative modifications on cysteine, methionine, and tryptophan using HSA as a model protein. The approach will be targeted to the peptide sites based on the hypothesis that the mass changes associated with modification can be quantified using MRM of the digested proteins. Since the overall goal is to quantify these modifications in human plasma, an MRM transition list will need to be generated and validated \textit{in vitro}. This approach will first utilize MS/MS data generated from the shotgun approach in Chapter 2, as well as a theoretical digest to determine candidate tryptic HSA peptide sequences. Next, with synthetic peptides, a comprehensive transition list will be generated based on the parent/daughter ion pairs of the modified and unmodified peptides that give the maximum signal and lowest background. Finally, the putative transition list will be refined by analysis of digests of HSA protein that were modified (oxidation by H$_2$O$_2$/myeloperoxidase or “one pot” dicarbonyl adduction) in a dose-dependent manner at the \textit{protein} level.
4.2. Methods

4.2.1. Materials

All materials were HPLC grade and were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. Synthetic peptides were custom ordered from New England Peptide (Gardner, MA). Sequencing grade trypsin was purchased from Promega (Fitchburg, WI). Lipidex-1000 (Catalog # 6008301) was purchased from PerkinElmer (Waltham, MA). 2-keto-D-glucose (aka glucosone, Catalog #61793) was obtained from Sigma-Aldrich, and (4S,5R)-2-oxo-4,5,6-trihydroxyhexan-1-one (aka 3-deoxyglucosone, Catalog # OR3915T) was obtained from Apollo Scientific Limited.

4.2.2. Oxidation of synthetic peptides containing M, W, or C

HSA tryptic peptides (Figure 4.5) containing an internal M, W or C (C34 only) were synthesized, HPLC purified, lyophilized, and resuspended in 18Ω Milli-Q grade water to 2 mg/ml. A 150 μl aliquot was used for each reaction. In the methionine-containing peptides that also contained cysteine residues (i.e., peptides containing M87, M123, M299, M446), 200 mM iodoacetamide (10 μl) in 20 mM ammonium bicarbonate pH 7.4 was added and the solution was incubated in the dark at room temperature for 1 hour. Since unmodified HSA cysteine at C34 will be capped by iodoacetamide during protein processing, we generated the “unmodified,” or “not oxidized,” representative C34 peptide by treating it with iodoacetamide as described above. Each peptide solution was oxidized in 160 μl of mixture of 10% acetic acid/ 10% hydrogen peroxide for an overall concentration of 5% acetic acid/ 5% H₂O₂, and the reaction conditions for the three types
<table>
<thead>
<tr>
<th>Site</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M87</td>
<td>ETYGEMADCCAK</td>
</tr>
<tr>
<td>M123</td>
<td>LVRPEVDVLMCTAFHDNEETFLK</td>
</tr>
<tr>
<td>M299</td>
<td>SHCIAEVENDEMPADLPSLAADFVESK</td>
</tr>
<tr>
<td>M329</td>
<td>DVFLGMFLYEYAR</td>
</tr>
<tr>
<td>M446</td>
<td>MPCAEYLSWLNQLCVLHEK</td>
</tr>
<tr>
<td>M548</td>
<td>AVMDDFAAFVEK</td>
</tr>
<tr>
<td>W214</td>
<td>AWAVAR</td>
</tr>
<tr>
<td>C34</td>
<td>ALVLIAFAQYLQQCPFEDHVK</td>
</tr>
</tbody>
</table>

Figure 4.5. Tryptic HSA sequences containing major oxidizable residues.

Eight synthetic peptides were obtained that contain an internal residue capable of being oxidized.
of oxidizable residues are as follows. The six methionine-containing peptides (M87, M123, M299, M329, M446, M548) were incubated at room temperature for 30 minutes. The C34 cysteine-containing peptide was incubated at 37°C for 2 hours, and the W214 tryptophan was incubated overnight at 37°C. The peptides were diluted twenty-fold with a solution of 0.5% formic acid in 25% acetonitrile in water prior to MS analysis.

4.2.3. Modification of synthetic peptides with dicarboxyls

Tryptic peptides (Figure 4.6) containing an internal arginine (R) and a C-terminal lysine (K) were synthesized, HPLC purified, lyophilized, and resuspended in 18Ω Milli-Q grade water to 5 mg/ml. A 150 μl aliquot was used for each reaction. In peptides that also contained cysteine residues (ie., peptides containing R257, R485, R98, R81), 200 mM iodoacetamide (20 μl) in 20 mM ammonium bicarbonate pH 7.4 was added and the solution was incubated in the dark at room temperature for 1 hour. MG (150 μl of 10 mM) in 2X PBS pH 7.4 was added to each peptide solution, and the reactions were incubated at 37°C for 2 hours. Higher concentrations and longer incubations were required for the 3DG and glucosone modifications. 3DG (150 μl of 40 mM) or glucosone (150 μl of 200 mM) in 2X PBS pH 7.4 was added to each peptide solution, and the reactions were incubated at 37°C for 48 hours. The reactions were terminated by the addition of formic acid (10 μl), and the modified peptides were desalted with 50 mg Hypersep C18 RP cartridges (Thermo Scientific) and eluted with 80% acetonitrile containing 0.1% TFA. The peptides were diluted four-fold with a solution of 0.5% formic acid in water prior to MS analysis.
<table>
<thead>
<tr>
<th>arginine site</th>
<th>tryptic peptide_sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>R10</td>
<td>SEVAH**R(54)**FK</td>
</tr>
<tr>
<td>R81</td>
<td>LC(57)TVATL**R(54)**ETYGEMADC(57)C(57)AK</td>
</tr>
<tr>
<td>R98</td>
<td>QEPER**R(54)**NEC(57)FLQHK</td>
</tr>
<tr>
<td>R186</td>
<td>LDELR**R(54)**DEGK</td>
</tr>
<tr>
<td>R209</td>
<td>FGER**R(54)**AFK</td>
</tr>
<tr>
<td>R218</td>
<td>AWAVAR**R(54)**LSQR</td>
</tr>
<tr>
<td>R222</td>
<td>LSQR**R(54)**FPK</td>
</tr>
<tr>
<td>R257</td>
<td>VHTEC(57)C(57)HGDLLEC(57)ADDR**R(54)**ADLAK</td>
</tr>
<tr>
<td>R348</td>
<td>HPDYSVVLLLR**R(54)**LAK</td>
</tr>
<tr>
<td>R410</td>
<td>FQNALLVR**R(54)**YTK</td>
</tr>
<tr>
<td>R428</td>
<td>KVPQVSTPTLVEVS**R(54)**NLGK</td>
</tr>
<tr>
<td>R472</td>
<td>TPVSD**R(54)**VTK</td>
</tr>
<tr>
<td>R485</td>
<td>**R(54)**PC(57)FSALEVDETYPK</td>
</tr>
</tbody>
</table>

Figure 4.6. Tryptic HSA sequences containing R as a missed cleavage.
4.2.4. Transition optimization using synthetic peptides

Peptide solutions (either unmodified, oxidized, or dicarbonyl-adducted) from 4.2.3 were infused at 3 μl/min into a 4000 QTRAP (Applied Biosystems/ MDS Sciex) equipped with a Turbo Spray ion source and the manual transition optimization was performed as follows. The source temperature was set at 200˚C, source voltage was 5000 volts, GS1 and GS2 were set to 0 and 25 psi, respectively, and the declustering potential was set to 70 volts for all peptide parent ions.

Theoretical fragmentation spectra were obtained using the MS-Product function in ProteinProspector (Clauser et al. 1999), where R+54 was used to account for the MG modification, R+144 for 3DG, and R+160 for glucosone. The potential oxidized masses from Figures 4.2, 4.3, and 4.4 were considered for the eight sequences subjected to oxidation. While the individual peptide solutions were infused into the QTRAP, mass spectra of parent ions for each charge state (+2, +3, +4) were analyzed using enhanced MS (EMS) mode in Analyst v. 1.4 (AB Sciex). Parent ions from each charge state above were fragmented in enhanced product ion (EPI) mode (MS/MS) and a list of potential MRM transitions was generated. Out of all the possible intense ions from the MS/MS spectra, only b or y ions were considered that were in agreement with the site of modification and were selected for MRM optimization. In MRM mode, transitions were monitored as the collision energy was ramped from 5 to 100 eV and 1-5 candidate transitions were chosen for each modified or unmodified peptide.

4.2.5. Preparation of in vitro oxidized and dicarbonyl modified HSA
Delipidation of HSA. Lipidex-1000 slurry in methanol was buffer exchanged five times into an equal volume of 1X PBS, pH 7.4. Purified HSA (100mg, 96% fatty acid-free, Sigma-Aldrich catalog #A1887, lot #018K7540) was dissolved in 10 ml of 1X PBS pH 7.4 and was delipidated with aqueous Lipidex-1000 for 1 hour on a rotator at room temperature.

“One pot” dicarbonyl modification of HSA. A 200 μL stock solution containing 16.2 mM glucosone and 16.2 mM 3DG in 1 X PBS pH 7.4 was prepared. A 3-fold dilution series in 1 X PBS of the dicarbonyl solution was made by adding 100 μL of dicarbonyl solution to 200 μL 1 X PBS (5.4 mM, 1.8 mM, 600 μM, 200 μM). To the 300 μL aliquot of 200 μM dicarbonyl solution, 12 μL of 5 mM MG was added to make all dicarbonyls equal to 200 μM. A 2-fold dilution series from 200 μM was prepared (100, 50, 25, 12.5, 6.25, 3.13, 1.56, 0.780, 0.391 μM). The 13-point dicarbonyl series was added to an equal volume (100 μL each) of delipidated HSA at 10mg/ml for a final HSA concentration of 5 mg/ml (75 μM, total HSA in each is 500 μg). A control reaction was prepared without dicarbonyl. The reactions were incubated at 37˚C for 18 hours. Excess dicarbonyl and PBS were removed via centrifugation using Amicon Ultra 10K MWCO (Millipore) filters, and dicarbonyl-modified HSA was buffer exchanged into 50 mM ammonium bicarbonate pH 7.4.

Oxidation of HSA. A 10-fold dilution series of H₂O₂ in 1X PBS from a 0.3% stock solution was prepared. Each H₂O₂ dilution (20 μL) was added to 80 μL delipidated HSA at 10 mg /ml in PBS. Myeloperoxidase (2 μL of 570 nM) was added to each reaction, as well as a control reaction without H₂O₂. Another control reaction was prepared without
peroxide or MPO. All reactions were incubated overnight at 37°C. The reactions were quenched with 300 μL of 100 mM methionine. Excess methionine and PBS were removed via centrifugation using Amicon Ultra 10K MWCO (Millipore) filters, and oxidized HSA was buffer exchanged into 50 mM ammonium bicarbonate pH 7.4.

Reduction/alkylation/digestion/peptide purification. An aliquot (250 μg) of dicarbonyl-modified or oxidized (MPO-H₂O₂) HSA in 50 mM ammonium bicarbonate pH 7.4 from the section above was reduced with 100 μl of 20mM TCEP (tris(2-carboxyethyl)phosphine) in 0.4 M ammonium bicarbonate pH 7.4, and was incubated at 55°C for 30 min. Once the reactions reached room temperature, 100 μl of 25mM iodoacetamide in 50 mM ammonium bicarbonate pH 7.4 was added, and the reactions incubated for 30 minutes in the dark at room temperature. Trypsin was added at a 1:50 (w/w; trypsin: HSA) ratio, and the reactions incubated for 18 hours at 37°C. Formic acid (5 μl) was added to terminate the reactions, and the tryptic digests were desalted with 50 mg Hypersep C18 RP cartridges (Thermo Scientific) and eluted with 80% acetonitrile containing 0.1% TFA. The digests, in 80% acetonitrile, were frozen and lyophilized to dryness. The peptides were resuspended immediately prior to LC-MRM in 100 μl of a 1% formic acid solution containing the R218 stable isotope labeled peptide internal standard at 5 μg/ml prior to LC-MS/MS.

4.2.6. LC-MRM

Aliquots of oxidized, dicarbonyl-modified, or unmodified control HSA digest (25μg in 10 μl) were loaded onto a ZORBAX 300SB-C18 capillary column (5 μm, 0.5 x 150 mm, Agilent). Peptides were eluted from the column using an LC Packings Ultimate
II HPLC (Dionex) into a 4000 QTRAP using a flow rate of 40 μl/min of solvent A (0.01% TFA, 0.5% formic acid) with a 30 min linear gradient from 5% to 45% of solvent B (acetonitrile containing 0.01% TFA, 0.5% formic acid). The 4000 QTRAP was operated in MRM mode with the optimal transitions listed. Dwell time was set to 40 ms, Q1 resolution was set to low, and Q3 resolution was set to unit. Data analysis and peak integration was performed with Multiquant software (AB Sciex). Curve fitting of peak area ratio as a function of peroxide or dicarbonyl was performed using Prism (GraphPad Software Inc.) Analysis was nonlinear regression (fit Ki) assuming one binding site for each peptide. Slopes were determined from this best fit curve using the log(agonist) vs. response for variable slope analysis.
Figure 4.7. MS of oxidized methionine (M) and unmodified HSA peptide containing M548.

Upper is MS spectra of the +2 parent ions of M548 peptide after H₂O₂ treatment. Unmodified peptide is at m/z= 671.6, methionine sulfoxide (M+16) is at m/z =679.9. Lower spectra is MS of same peptide without H₂O₂ treatment.
Figure 4.8. MS of oxidized tryptophan (W) and unmodified HSA peptide AWAVAR.

Top is MS spectra of AWAVAR peptides after H$_2$O$_2$ treatment. Unmodified peptide is at m/z=683.8, kynurenine (W+4) is at m/z=687.8, oxindolylalanine (W+16) is at m/z=699.8, and N-formylkynurenine is at m/z=715.8. Lower spectra is MS of same peptide without H$_2$O$_2$ treatment.
Figure 4.9. MS of oxidized cysteine (C) peptide containing C34.

MS spectra of the +2 parent ions of C34 peptide after H$_2$O$_2$ treatment shows three oxidized forms in addition to the unmodified peptide. Unmodified peptide is at m/z = 1217.1, single oxidation to sulfenic acid (C+16) is at m/z = 1225.4, double oxidation to sulfone is at m/z = 1230.7, and triple oxidation to cysteic acid is at m/z = 1241.3.
Figure 4.10. MS of MG and 3DG modification of HSA peptide containing R410.

Upper MS is the +2 parent ions of the R410 peptide after treatment with MG. Grey arrow indicates the unmodified peptide at m/z=677, and the bold arrow indicates the +54 modification at m/z=703.9. Lower MS is the +2 parent ions of the R410 peptide after treatment with 3DG. Grey arrow indicates the unmodified peptide at m/z=677, and the bold arrow indicates the +144 modification at m/z=749.0.
Figure 4.11. MS of Glucosone modification of HSA peptide containing R410.

MS is the +2 parent ions of the R410 peptide after treatment with glucosone. Grey arrow indicates the unmodified peptide at m/z=677, and the bold arrow indicates the +160 modification at m/z=757.0.
4.3. Results

4.3.1. Oxidation of peptides

Despite the numerous methods to oxidize proteins, the most effective means to generate oxidized peptides was with mild oxidizing conditions using hydrogen peroxide. The rationale was to subject the peptide to just enough oxidation to get a reasonable distribution of all oxidized products. It is important to note that these reactions do not need to be driven to completion; however, they need to proceed sufficiently to induce the mass changes associated with modification so that those modified parent ions can be isolated and then fragmented to determine the best MRM transitions.

The reaction conditions of methionine versus cysteine or tryptophan-containing peptides reflect the susceptibility of these side chains to oxidation. Methionine containing peptides were immediately oxidized at room temperature, whereas the cysteine and tryptophan containing peptides needed incubation at 37°C for 2 days to generate the oxidized products. In all cases of oxidized peptides, the peptides were not purified prior to MS. Therefore, all oxidized species, including any “oxidation intermediates” will likely be present because the peroxide reaction mixture is directly infused into the mass spectrometer. This approach was chosen because it allows for the characterization of potential oxidation products that may be lost during traditional protein sample processing (i.e. reduction/alkylation). The underlying rationale was to show the difference between products of in situ oxidation versus product of oxidation that have been through reduction and alkylation (section 4.2.5).
The oxidation of the peptide containing M548 in Figure 4.7 is representative of the rapid and specific nature of oxidation of M to M+16. Under these mild conditions, the M+32 oxidized adduct did not manifest, so this oxidation product was not monitored. The 5 remaining peptides containing M (M87, M123, M299, M329, and M446) showed similar M+16 oxidized product formation as illustrated in Figure 4.7, and both the unmodified and modified were selected for MS/MS optimization in MRM mode.

Oxidation of W (Figure 4.8) in 5% H$_2$O$_2$ gave near-equimolar W+16 and W+32 oxidation products, assuming the ionization efficiency of these two peptides is equal. After 48 hours, the kynurenine W+4 peak began to emerge, and even though the MS peak is small, it was sufficient to isolate and fragment for MS/MS data. The strange “double peak” nature of the unmodified W214 peptide (bottom Figure 4.8) is likely an isotopic impurity and did not affect MRM optimization because only fragments that could be assigned to the proper theoretical MS/MS values were selected for MRM optimization.

Oxidation of C (Figure 4.9) produced the three expected oxidation products in multiples of 16 amu. In the same manner as the M and W peptides, the oxidized C34 peptide in the peroxide reaction mixture was directly infused for this analysis. The three modified forms were sufficient for MS/MS analysis, but not all of these oxidized forms exist after reduction/ alkylation. The proportion of these oxidized products is much different after oxidation of the entire HSA protein and subsequent processing. This will be discussed in section 4.3.4. Note that the “unmodified” peak (m/z= 1217.1) in Figure 4.9 will not be found in digests of HSA, because iodoacetamide will cap the free cysteine during sample preparation which increases the C mass by +57. Therefore, in order to
obtain MRM transitions for the unmodified C34 peptide, this peptide was treated with iodoacetamide in a separate reaction and infused for MRM analysis.

### 4.3.2 Dicarbonyl modification of peptides

Arginine-directed dicarbonyl adduction was used to modify a panel of HSA peptides. MG was more potent than 3DG or glucosone in terms of reactivity, as only a 2 hour incubation with peptides at 37°C was necessary. At MG concentrations higher than 5 mM or for longer incubation times, many reaction products were observed, and it was difficult to locate the exact site or nature of these modifications. At these high concentrations, modifications were in multiples of +54 and +72 of the parent ions, and MS/MS was not able to distinguish between adduction of the N-terminus, arginine or C-terminal lysine (K). At these high concentrations, arginine can be modified with two dicarbonyls, one for each guanidino terminal nitrogen. In general, the synthetic peptides (at ~2.5 mg/mL) that were modified by 5 mM MG gave peaks similar to the upper spectra in Figure 4.10. In this case, the +54 modification is prominent, with the +72 (at m/z= 720.3) present in lower abundance. As with the oxidized peptides, the purpose is to generate enough adducted peptide for MS/MS.

3DG and glucosone required higher concentrations and longer incubation times to generate their respective hydroimidazolone adducts. The modification of the R410 peptide is shown in the bottom spectra of Figure 4.10. The mass increase of R+144 of the parent ion is consistent with arginine modification. The modification of the R410 peptide with glucsone in Figure 4.11 shows the R+160 peak (m/z=757.0), which is also
consistent with arginine-glucosone hydroimidazolone. Also present in this spectra are several peaks of unknown identity.

Dicarbonyl modification of arginine induces a missed tryptic cleavage site. Of the 24 arginine sites in HSA, the 13 sites listed in Figure 4.6 were considered good candidates for MRM because they lie on tryptic peptide fragments, are of an appropriate length (5-26 residues), and because they contain a C-terminal lysine. The exception to this is the R218 peptide, which contains a C-terminal arginine. The synthetic tryptic peptide containing R218 was not chosen for dicarbonyl modification because it contains two arginines, and modification with dicarbonyl would likely target both arginines. Instead, LC-LC-MS/MS of a tryptic digest of MG modified HSA (section 2.3) was used to estimate the optimal transitions for the MG modified R218 peptide. For example, it was determined in this manner that the doubly charged peptide containing R218 at m/z=606.6 resulted in strong y6 and y8 fragments at 30% collision energy. Other sites identified from LC-LC-MS/MS of MG-treated HSA were R257, R410, R485, and R428. The spectra identifying the R428 peptide contained a missed tryptic cleavage at the N-terminus, and the synthetic peptide used for this study has a N-terminal lysine. The remaining 8 arginine containing peptides (R10, R81, R98, R186, R209, R222, R348, R472) were identified from a theoretical tryptic digest that assumed zero missed cleavages, other than the internal arginine.

One major difference between oxidized peptides and dicarbonyl-modified peptides is that due to the dicarbonyl missed tryptic cleavage, the unmodified arginine tryptic peptide actually represents two different peptides (Figure 4.6). The “unmodified”
arginine tryptic peptides will be chemically different than the longer modified peptide. In many cases, the missed tryptic cleavages cause these peptides to be over twice the length of the unmodified peptide, leading to not only increased mass, but also to major fundamental differences in ionization and fragmentation during MS/MS. Twelve synthetic peptides containing an internal arginine (all except R218) were modified with dicarbonyls and the transitions that gave the most intense fragment ions are reported in Figures 4.15-4.17.

Despite the differences between unmodified and modified arginine peptides, there are similarities regarding the fragmentation behavior of any one of these peptides modified with the different dicarbonyls. For example, in the R410 peptide, the y9 +2 ion is one of the most intense MRM fragments for MG modified peptide (Figure 4.15), for 3DG modified peptide (Figure 4.16), or for glucosone modified peptide (Figure 4.17). The R410 peptide thus is not majorly affected by the type of hydroimidazolone modification at this position in terms of favored ionization/fragmentation patterns of each dicarbonyl modification. By this logic, the same MRM transitions will be similar for any given peptide for the three different dicarbonyl modifications. This is the general trend, with a few exceptions such as R10. MG-modified R10 peptide prefers to fragment to the y5 ion from the +2 parent (Figure 4.15). 3DG-modified R10 peptide, however, gives the most intense signal when the triply charged parent fragments to the y4 +2 fragment (Figure 4.16). The glucosone-modified R10 peptide gave the most intense MRM signal when the triply charged parent fragmented to the y6 +2 ion (Figure 4.17).
Figure 4.12. MS/MS of unmodified (top) and oxidized methionine (M+16) peptides (bottom) to select candidate ions to test for MRM.

Upper is MS/MS spectra of the +2 parent ions (m/z=671.9) of unmodified M548 peptide. Lower is MS/MS of the +2 parent ions (m/z=679.9) of H₂O₂-treated M548 peptide. Based on a theoretical fragmentation of the peptides, only intense b and y ions consistent with site-specific modifications (in this case, M+16) are selected for MRM validation.
Putative MRM transitions were monitored as the CE in q2 was ramped from 5-50 eV. In both cases, the fragment at m/z=143.1 was one of the most intense peaks at CE above ~35. Not all of the MRM transitions are equal in relative signal between the two peptides, which suggests that M+16 oxidation at M548 effects ionization and/or gas phase fragmentation. For example, the intense fragment at m/z=1172 in the unmodified peptide in the top graph corresponds to the y10 ion, and has an optimal CE of 29. The y10 ion in the oxidized peptide (bottom graph) is lower in relative intensity (m/z=1188), however, the doubly charged y10 +2 ion at m/z=594.8 is the favored fragment of the oxidized peptide.
MG-modification of many peptides (R186, R209, R222, R257, R428, R472) resulted in parent ions from more than one charge state, producing fragments with intense signals, and monitoring transitions from more than one charge state provides further validation of transition specificity. In other cases of MG modified peptides (R10, R81, R98, R348, R485), a single transition was chosen as this was the sole dominant fragment and/or the other fragments were nonspecific. The topic of multiple charge states is related to the fact that these are longer (hence, larger and have more sites for gas-phase protonation) peptide sequences.

4.3.3. Transition optimization

MS/MS was used to confirm the location of the oxidized or dicarbonyl-modified residues for all modified peptides. After confirmation of sequence specificity of the modification, a list of all intense ions from the MS/MS spectra was generated for each modified peptide. Each peptide (modified or unmodified) may generate tens of putative transitions that must be tested because the intensities of individual fragments derived from a single parent ion differ considerably. Only the intense ions that matched a theoretical fragmentation as determined by ProteinProspector were considered for further MRM optimization. These putative transitions were tabulated and monitored as the CE was ramped from 5 to 50 eV. It is not feasible to show all of the spectra in this dissertation. However, Figures 4.12 and 4.13 are representative of the process of both screening putative MRM transitions and selection of optimized transitions for all (unmodified and modified) peptides. Hundreds of potential transitions were narrowed down to the top three for validation of each peptide at the protein level.
<table>
<thead>
<tr>
<th>parent</th>
<th>daughter</th>
<th>CE</th>
<th>charge</th>
<th>ion</th>
<th>peptide description</th>
<th>peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>471.3</td>
<td>596.4</td>
<td>30</td>
<td>2</td>
<td></td>
<td>internal control</td>
<td>DDNPNLPR</td>
</tr>
<tr>
<td>481.4</td>
<td>685.4</td>
<td>30</td>
<td>2</td>
<td></td>
<td>R410t unmodified</td>
<td>FONALLVR</td>
</tr>
<tr>
<td>575.3</td>
<td>937.5</td>
<td>25</td>
<td>2</td>
<td></td>
<td>internal control</td>
<td>LVNEVTEFAK</td>
</tr>
<tr>
<td>637.9</td>
<td>851.3</td>
<td>25</td>
<td>3</td>
<td>y7</td>
<td>R485 unmodified</td>
<td>RPC(57)FSALEVDETYVPK</td>
</tr>
<tr>
<td>637.9</td>
<td>961.4</td>
<td>30</td>
<td>3</td>
<td>b8</td>
<td>R485 unmodified</td>
<td>RPC(57)FSALEVDETYVPK</td>
</tr>
<tr>
<td>637.9</td>
<td>950.4</td>
<td>26</td>
<td>3</td>
<td>y8</td>
<td>R485 unmodified</td>
<td>RPC(57)FSALEVDETYVPK</td>
</tr>
<tr>
<td>696.5</td>
<td>925.8</td>
<td>34</td>
<td>3</td>
<td>y15</td>
<td>R257 unmodified</td>
<td>VHTEC(57)C(57)HGDLLEC(57)ADDR</td>
</tr>
<tr>
<td>992.2</td>
<td>780.4</td>
<td>40</td>
<td>3</td>
<td>y15</td>
<td>M299 unmod</td>
<td>SHC(57)AEVENEMPADLPSLAADFVESK</td>
</tr>
<tr>
<td>992.2</td>
<td>1163.6</td>
<td>40</td>
<td>3</td>
<td>y11</td>
<td>M299 unmod</td>
<td>SHC(57)AEVENEMPADLPSLAADFVESK</td>
</tr>
<tr>
<td>840.5</td>
<td>676.9</td>
<td>30</td>
<td>3</td>
<td>y11</td>
<td>M446 unmod</td>
<td>MPC(57)AEDYLSVVLNQLC(57)VLHEK</td>
</tr>
<tr>
<td>671.9</td>
<td>1172.6</td>
<td>29</td>
<td>2</td>
<td>y10</td>
<td>M548 unmod</td>
<td>AVMDFAAFVEK</td>
</tr>
<tr>
<td>671.9</td>
<td>143.1</td>
<td>35</td>
<td>2</td>
<td>a2</td>
<td>M548 unmod</td>
<td>AVMDFAAFVEK</td>
</tr>
<tr>
<td>717.8</td>
<td>724.2</td>
<td>36</td>
<td>2</td>
<td>y6</td>
<td>M87 unmod</td>
<td>ETYGEMADC(57)C(57)AK</td>
</tr>
<tr>
<td>717.8</td>
<td>135.8</td>
<td>50</td>
<td>2</td>
<td>y1</td>
<td>M87 unmod</td>
<td>ETYGEMADC(57)C(57)AK</td>
</tr>
<tr>
<td>337.2</td>
<td>416.2</td>
<td>18</td>
<td>2</td>
<td>y4</td>
<td>W214 unmod</td>
<td>AWAVAR</td>
</tr>
<tr>
<td>337.2</td>
<td>258.1</td>
<td>16</td>
<td>2</td>
<td>y2</td>
<td>W214 unmod</td>
<td>AWAVAR</td>
</tr>
<tr>
<td>830.8</td>
<td>991</td>
<td>31</td>
<td>3</td>
<td>y16</td>
<td>C34 unmod</td>
<td>ALVLIAFAQYLQQC(57)PFEDHVK</td>
</tr>
<tr>
<td>830.8</td>
<td>1047.5</td>
<td>31</td>
<td>3</td>
<td>y17</td>
<td>C34 unmod</td>
<td>ALVLIAFAQYLQQC(57)PFEDHVK</td>
</tr>
<tr>
<td>663.4</td>
<td>809.5</td>
<td>37</td>
<td>4</td>
<td>b7</td>
<td>M123 unmod</td>
<td>LVRFVEDVMC(57)TAHFNEETFVK</td>
</tr>
<tr>
<td>663.4</td>
<td>806.4</td>
<td>24</td>
<td>4</td>
<td>y13</td>
<td>M123 unmod</td>
<td>LVRFVEDVMC(57)TAHFNEETFVK</td>
</tr>
<tr>
<td>812.4</td>
<td>1149.6</td>
<td>38</td>
<td>2</td>
<td>y9</td>
<td>M329 unmod</td>
<td>DVFLGFMFLYEYAR</td>
</tr>
<tr>
<td>542</td>
<td>701.4</td>
<td>26</td>
<td>3</td>
<td>y5</td>
<td>M329 unmod</td>
<td>DVFLGFMFLYEYAR</td>
</tr>
<tr>
<td>840.1</td>
<td>526.3</td>
<td>50</td>
<td>3</td>
<td>y4</td>
<td>M446 unmod</td>
<td>MPC(57)AEDYLSVVLNQLC(57)VLHEK</td>
</tr>
<tr>
<td>840.1</td>
<td>413.3</td>
<td>53</td>
<td>3</td>
<td>y3</td>
<td>M446 unmod</td>
<td>MPC(57)AEDYLSVVLNQLC(57)VLHEK</td>
</tr>
</tbody>
</table>

**Figure 4.14. Validated transitions for unmodified HSA peptides.**

The first 5 peptide sequences listed do not contain MW or free C residues. The remainder of the sequences are used as controls to monitor the unmodified M, W, or C containing peptides. R410, R485, and R257 unmodified sequences contain the respective arginine and represent the N-terminal portion of the dicarbonyl modified sequence. In the case of multiple transitions for each sequence, the first transition listed was used for quantification.
Figure 4.15. Validated transitions for MG-modified HSA peptides.

Twelve sequences containing an internal arginine-MG modification of +54 were synthesized and the transitions with the highest signal and lowest background are listed here. Transitions for the R218 MG peptide were based on MS/MS spectra from MG-modified HSA from section 2.3. In the case of multiple transitions for each sequence, the first transition listed was used for quantification.
Figure 4.16. Validated transitions for 3DG-modified HSA peptides.

Twelve sequences containing an internal arginine-3DG modification of +144 were synthesized and the transitions with the highest signal and lowest background are listed above. Transitions for the R218 3DG peptide is based on intense ions corresponding MS/MS spectra from MG-modified HSA from section 2.3, then the theoretical mass of +144 was added to the internal arginine. In the case of multiple transitions for each sequence, the first transition listed was used for quantification.
<table>
<thead>
<tr>
<th>parent</th>
<th>daughter</th>
<th>CE</th>
<th>charge</th>
<th>ion</th>
<th>peptide description</th>
<th>peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>505.1</td>
<td>120.0</td>
<td>35</td>
<td>3</td>
<td>b1</td>
<td>R410-Glucosone modified FQNALLVR(160)YTK</td>
<td></td>
</tr>
<tr>
<td>505.1</td>
<td>619.4</td>
<td>23</td>
<td>3</td>
<td>y9</td>
<td>+2 R410-Glucosone modified FQNALLVR(160)YTK</td>
<td></td>
</tr>
<tr>
<td>505.1</td>
<td>276.2</td>
<td>25</td>
<td>3</td>
<td>b2</td>
<td>R410-Glucosone modified FQNALLVR(160)YTK</td>
<td></td>
</tr>
<tr>
<td>687.1</td>
<td>109.9</td>
<td>80</td>
<td>4</td>
<td>y2</td>
<td>+2 R257-Glucosone modified VHTEC(57)C(57)HGDLLEC(57)ADDR(160)ADLAK</td>
<td></td>
</tr>
<tr>
<td>549.9</td>
<td>109.9</td>
<td>80</td>
<td>5</td>
<td>y2</td>
<td>+2 R257-Glucosone modified VHTEC(57)C(57)HGDLLEC(57)ADDR(160)ADLAK</td>
<td></td>
</tr>
<tr>
<td>549.9</td>
<td>567.9</td>
<td>23</td>
<td>5</td>
<td>y9</td>
<td>+2 R257-Glucosone modified VHTEC(57)C(57)HGDLLEC(57)ADDR(160)ADLAK</td>
<td></td>
</tr>
<tr>
<td>549.9</td>
<td>662.4</td>
<td>23</td>
<td>5</td>
<td>y21</td>
<td>+4 R257-Glucosone modified VHTEC(57)C(57)HGDLLEC(57)ADDR(160)ADLAK</td>
<td></td>
</tr>
<tr>
<td>412.3</td>
<td>561.3</td>
<td>20</td>
<td>3</td>
<td>y8</td>
<td>+2 R186-Glucosone modified LDELR(160)DEGK</td>
<td></td>
</tr>
<tr>
<td>617.9</td>
<td>200.1</td>
<td>50</td>
<td>2</td>
<td>a2</td>
<td>R186-Glucosone modified LDELR(160)DEGK</td>
<td></td>
</tr>
<tr>
<td>378.6</td>
<td>459.3</td>
<td>18</td>
<td>3</td>
<td>y6</td>
<td>+2 R10-Glucosone modified SEVAHR(160)FK</td>
<td></td>
</tr>
<tr>
<td>567.4</td>
<td>216.8</td>
<td>40</td>
<td>2</td>
<td>b2</td>
<td>R10-Glucosone modified SEVAHR(160)FK</td>
<td></td>
</tr>
<tr>
<td>837.1</td>
<td>538.2</td>
<td>50</td>
<td>3</td>
<td>y4</td>
<td>R81-Glucosone modified LC(57)TVATLR(160)ETYGEMADC(57)C(57)AK</td>
<td></td>
</tr>
<tr>
<td>628.1</td>
<td>724.2</td>
<td>32</td>
<td>4</td>
<td>y6</td>
<td>R81-gluicosone modified LC(57)TVATLR(160)ETYGEMADC(57)C(57)AK</td>
<td></td>
</tr>
<tr>
<td>738.2</td>
<td>662.5</td>
<td>34</td>
<td>3</td>
<td>y17</td>
<td>+3 R428-gluicosone modified KVPQVSPTTLVEVSR(160)NLGK</td>
<td></td>
</tr>
<tr>
<td>553.9</td>
<td>737.0</td>
<td>21</td>
<td>4</td>
<td>y12</td>
<td>+2 R428-gluicosone modified KVPQVSPTTLVEVSR(160)NLGK</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.17. Validated transitions for glucosone-modified HSA peptides.**

Six sequences containing an internal arginine-glucosone modification of +160 were synthesized and the transitions with the highest signal and lowest background are listed above. In the case of multiple transitions for each sequence, the first transition listed was used for quantification.
<table>
<thead>
<tr>
<th>parent</th>
<th>daughter</th>
<th>CE</th>
<th>charge</th>
<th>ion</th>
<th>peptide description</th>
<th>peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>997.5</td>
<td>780.4</td>
<td>40</td>
<td>3</td>
<td>y15</td>
<td>+2 M299- sulfoxide [O]</td>
<td>SHC(57)AEVENDEM(16)PADLPSLAADFVESK</td>
</tr>
<tr>
<td>997.5</td>
<td>1163.6</td>
<td>40</td>
<td>3</td>
<td>y11</td>
<td>M299- sulfoxide [O]</td>
<td>SHC(57)AEVENDEM(16)PADLPSLAADFVESK</td>
</tr>
<tr>
<td>820.4</td>
<td>1165.6</td>
<td>34</td>
<td>2</td>
<td>y9</td>
<td>M329- sulfoxide [O]</td>
<td>DVFLGM(16)FLYEYAR</td>
</tr>
<tr>
<td>820.4</td>
<td>1101.4</td>
<td>45</td>
<td>2</td>
<td>b9</td>
<td>M329- sulfoxide [O]</td>
<td>DVFLGM(16)FLYEYAR</td>
</tr>
<tr>
<td>820.4</td>
<td>713.4</td>
<td>32</td>
<td>2</td>
<td>y11</td>
<td>+2 M329- sulfoxide [O]</td>
<td>DVFLGM(16)FLYEYAR</td>
</tr>
<tr>
<td>845.4</td>
<td>526.3</td>
<td>50</td>
<td>3</td>
<td>y4</td>
<td>M446- ox albumin</td>
<td>DVFLGM(16)FLYEYAR</td>
</tr>
<tr>
<td>845.4</td>
<td>413.3</td>
<td>34</td>
<td>2</td>
<td>y9</td>
<td>M446- ox albumin</td>
<td>DVFLGM(16)FLYEYAR</td>
</tr>
<tr>
<td>845.4</td>
<td>1140.6</td>
<td>41</td>
<td>3</td>
<td>y11</td>
<td>M446- ox albumin</td>
<td>DVFLGM(16)FLYEYAR</td>
</tr>
<tr>
<td>667.4</td>
<td>809.5</td>
<td>37</td>
<td>4</td>
<td>b7</td>
<td>M123- ox albumin</td>
<td>LVRPEVDVM(16)C(57)TAFHDEETFLK</td>
</tr>
<tr>
<td>667.4</td>
<td>806.4</td>
<td>24</td>
<td>4</td>
<td>y13</td>
<td>+2 M123- ox albumin</td>
<td>LVRPEVDVM(16)C(57)TAFHDEETFLK</td>
</tr>
<tr>
<td>667.4</td>
<td>879.9</td>
<td>26</td>
<td>4</td>
<td>y14</td>
<td>+2 M123- ox albumin</td>
<td>LVRPEVDVM(16)C(57)TAFHDEETFLK</td>
</tr>
<tr>
<td>879.9</td>
<td>498.3</td>
<td>45</td>
<td>3</td>
<td>y4</td>
<td>albumin</td>
<td>ALVIQAQLQY(48)PFEDHVK</td>
</tr>
<tr>
<td>879.9</td>
<td>510.4</td>
<td>27</td>
<td>3</td>
<td>b5</td>
<td>albumin</td>
<td>ALVIQAQLQY(48)PFEDHVK</td>
</tr>
<tr>
<td>679.9</td>
<td>594.8</td>
<td>30</td>
<td>2</td>
<td>y10</td>
<td>+2 M548- sulfoxide [O]</td>
<td>AVM(16)DDFAAFVEK</td>
</tr>
<tr>
<td>679.9</td>
<td>143.1</td>
<td>37</td>
<td>2</td>
<td>a2</td>
<td>M548- sulfoxide [O]</td>
<td>AVM(16)DDFAAFVEK</td>
</tr>
<tr>
<td>725.8</td>
<td>610.7</td>
<td>34</td>
<td>2</td>
<td>y10</td>
<td>+2 M87- sulfoxide [O]</td>
<td>EYGE(16)A(57)C(57)AK</td>
</tr>
<tr>
<td>725.8</td>
<td>358.8</td>
<td>50</td>
<td>2</td>
<td>Y</td>
<td>M87- sulfoxide [O]</td>
<td>EYGE(16)A(57)C(57)AK</td>
</tr>
<tr>
<td>345.2</td>
<td>175.1</td>
<td>30</td>
<td>2</td>
<td>y1</td>
<td>W214- oxindolylalanine [O]</td>
<td>AW(16)AVAR</td>
</tr>
<tr>
<td>345.2</td>
<td>416.2</td>
<td>20</td>
<td>2</td>
<td>y4</td>
<td>W214- oxindolylalanine [O]</td>
<td>AW(16)AVAR</td>
</tr>
<tr>
<td>353.2</td>
<td>416.2</td>
<td>21</td>
<td>2</td>
<td>y4</td>
<td>W214- n-formylkynurenine [O]</td>
<td>AW(32)AVAR</td>
</tr>
<tr>
<td>353.2</td>
<td>345.2</td>
<td>20</td>
<td>2</td>
<td>y3</td>
<td>W214- n-formylkynurenine [O]</td>
<td>AW(32)AVAR</td>
</tr>
<tr>
<td>353.2</td>
<td>175.1</td>
<td>32</td>
<td>2</td>
<td>y1</td>
<td>2[O]</td>
<td>AW(32)AVAR</td>
</tr>
<tr>
<td>339.2</td>
<td>416.2</td>
<td>20</td>
<td>2</td>
<td>y4</td>
<td>W214- kynurenine</td>
<td>AW(4)AVAR</td>
</tr>
</tbody>
</table>

Figure 4.18. Validated transitions for oxidized HSA peptides.

Eight tryptic sequences were oxidized with 5% H$_2$O$_2$ and optimal transitions for the different oxidized species are listed. For M peptides, only M+16 was monitored. In the case of multiple transitions for each sequence, the first transition listed was used for quantification.
4.3.4. Oxidation hotspots in HSA

Transitions for each modified peptide were considered valid if they had overlapping peaks for multiple transitions at identical retention times. A list of the transitions that showed increases in peak area as a function of HSA oxidation is given in Figure 4.18. Unmodified HSA peptides were used as controls to monitor assay conditions and are shown in Figure 4.14. For every modification, the area under the curve (AUC) for the corresponding unmodified (not oxidized) peptide was used as an internal control to account for assay and measurement variation. It is reasonable to assume that the ionization/fragmentation properties of the unmodified and oxidized peptides that have the same sequence are similar. Therefore, the ratio of the AUC’s (modified peptide AUC/ unmodified peptide AUC) for the oxidized peptide set is related to % modification. If this ratio equals 1, then this approximates a 50% modification if the ionization/fragmentation efficiencies are identical and the fragment ion being monitored is the same in both cases. As this ratio approaches zero, the abundance of modified peptide relative to the total amount of peptide becomes negligible. At these relative abundances, the peak area ratio equals or is approximately equal to % modification given the above assumptions. For example, at a peak area ratio (modified AUC / unmodified AUC) of 0.01 the % modification is approximately 1%.

The dose-response curve for site-specific MPO/H₂O₂ oxidation of methionine is shown in Figure 4.19. This log-log plot shows the response of M329, the most solvent accessible methionine residue in HSA. The second most solvent accessible methionine is M299, and this is also the second most susceptible methionine to oxidation. The order of
reactivity for these methionines to oxidation is M329 > M299 > M548 > M123 > M87 > M446 based on the peak area ratio (modified/unmodified) at the highest peroxide concentration. Higher concentrations of peroxide caused problems with the tryptic digestions, so these were not used for this analysis. Moreover, the basal levels of oxidation in this HSA batch from Sigma is different than what is observed in fresh human plasma, which will be explored in Chapter 5. In this HSA preparation, M548 has the highest basal level of oxidation (low peroxide), but in human plasma M329 has the highest level.
Figure 4.1. Site-specific HSA methionine oxidation \textit{in vitro}.

HSA was treated with myeloperoxidase in the presence of increasing concentrations of peroxide. HSA was digested and analyzed by MRM using optimal transitions. Peak area ratio is area under the curve of transition for the oxidized peptide divided by the area under the curve for the transition for the unmodified peptide.
Oxidation of tryptophan (Figure 4.20) at the protein level shows that the W+32 double oxidation is clearly the favored oxidation product. The singly oxidized W+16 and advanced oxidation product W+4 appear to increase slightly in signal at the highest peroxide concentrations. These findings are in contrast to the oxidation of the W214 synthetic peptide, in which approximately an equal signal came from the +16 and +32 mass additions. The difference is that the protein microenvironment may play a role in which species is formed. Another possible explanation is that protein sample processing (reduction/alkylation) affects the W+16 oxidation, but not the W+32 oxidation. The kynurenine W+4 is the least abundant W oxidation product in the context of both peptide and protein-level peroxide reactions. The absolute value of W oxidation is much lower compared with M. The peak area ratios for W oxidation range from 0.002-0.0025, whereas for M oxidation the peak area ratios start at 0.1 and go as high as 60. This is a 50-fold difference in signal between basal levels of oxidized M.
Figure 4.20. Products of the oxidation of tryptophan at W214 in HSA.

HSA was treated with myeloperoxidase in the presence of increasing concentrations of peroxide. HSA was digested and analyzed by MRM using optimal transitions. The oxidized forms of tryptophan were monitored at the peptide level, and the double oxidation (W+32, N-formylkynurenine) appears to be the abundant W oxidation product. Orange hook indicates the y₄ ion is monitored for all peptides.
Oxidation of C34 in HSA showed a similar discordance with the nature of the oxidized product in the protein versus peptide oxidation. Figure 4.21 reveals that the triply oxidized cysteic acid (C+48) is the dominant oxidation product at C34. The next most abundant C oxidation product is the doubly oxidized cysteine sulfone (C+32). The singly oxidized sulfenic acid (C+16) does not appear to increase as the amount of peroxide is increased. This is likely due to the reduction conditions that eliminate oxidation intermediates such as the unstable sulfenic acid group. The effect of the C+16 modification during reduction/alkylation is that TCEP reduces this intermediate oxidized sulfur, which is then capped during the iodoacetamide step. These findings are consistent with studies that have suggested that sulfenic acids are reversed during mild reduction with reducing agents such as DTT (Kettenhofen et al. 2011). The low levels of signal from the monitoring of the C+16 oxidation may be an artifact in the protein digest. The C+16 modification also may be formed at the ESI source as a byproduct of high voltage under atmospheric pressure.

Validation of these transitions at the protein level indicates the relative abundance of oxidative modifications at these sites based on the values of the peak area ratios. Methionine is the most reactive toward oxidation, cysteine the next most reactive (at least at C34), and tryptophan has the lowest proportion of oxidized product.
HSA was treated with myeloperoxidase in the presence of increasing concentrations of peroxide. HSA was digested and analyzed by MRM using optimal transitions. Peak AUC’s were normalized to the C34 unmodified peptide. The oxidized forms of cysteine were monitored at the peptide level (C34 only) and the triple oxidation product (C+48, cysteic acid) appears to be the abundant C oxidation product. Y-axis is peak area ratio to unmodified. Orange hook indicates b_{13} ion is monitored for all peptides.

Figure 4.21. Products of the oxidation of cysteine at C34 in HSA.
4.3.5. Dicarbonyl hotspots in HSA

The fact that dicarbonyl modification induces a missed tryptic site poses a unique challenge in terms of assay normalization. In all cases, the unmodified peptide is significantly different in that it is composed of a shorter sequence and the likely site of gas phase protonation has changed from an arginine on a short unmodified peptide to a lysine on a longer modified peptide. These changes affect both ionization and fragmentation. The smaller, unmodified peptide will prefer to adopt a lower charge state during ionization, while the larger, modified peptide will prefer to adopt a higher charge state during ionization. Fragmentation mechanisms are also altered because the longer modified peptide will generate more fragments. Therefore, MRM-based intensities of unmodified versions of the dicarbonyl-modified peptides are not an accurate measure of the proportion of the peptide that is modified. However, peptide MRM signals must be normalized to an internal control to account for assay (i.e. digestion and purification) and MS variation. We propose that as a means to compare different R sites, a HSA peptide that does not overlap with any of the R sites should be used for normalization.

The 42LVNEVTEFAK51 peptide in HSA was validated for MRM analysis in a previous study (Kuzyk et al. 2009) and this peptide does not overlap with any of the arginines monitored in our study. Four unmodified R sites were also monitored: R218, R257, R410, and R485. Unmodified R257 and R485 peptides contain cysteines and these peptides were used as controls for reduction and cysteine alkylation.

The “one pot” method to adduct arginine sites with three different dicarbonyls has the advantage of multiplexed competition which is likely more relevant to physiological
conditions. MG was by far the most reactive of the three dicarbonyls, and it was not added at concentrations higher than 100 μM because this caused significantly decreased tryptic digestion.

Figure 4.22 shows the binding curves for MG sites in HSA. In the top plot, the AUC of the modified peptide is divided by the AUC of the internal control \(^{42}\text{LVNEVTEFAK}^{51}\) peptide and this is plotted as a function of MG concentration. It is tempting to ascribe the values of the peptide ratio to abundance of the modification, but this is incorrect because it does not account for differences in ionization/fragmentation. In order to do this, these values were further normalized to their respective minima and maxima. The lower plot in Figure 4.22 shows these normalized values, which may be used to calculate relative arginine reactivity.
Figure 4.22. Relative modification of arginine sites by methylglyoxal.

Human serum albumin was treated with a 2-fold dilution series (0.39-100 μM) of MG in 1X PBS pH 7.4. MG-modified albumin was reduced, alkylated, digested, and analyzed by LC-MS/MS in MRM mode using the transitions in Figures 4.14 and 4.15. Peak AUC’s were expressed relative to the unmodified tryptic albumin peptide $^{42}$LVNEVTEFAK$^{51}$ (top plot), and those values were then normalized to adjust for differences in ionization/fragmentation efficiency (bottom plot).
The two measures that will be used to gauge site reactivity are slope and Ki of the binding curves. The slope (aka Hill slope) is a measure of binding cooperativity. A slope of 1 is considered non-cooperative, a slope less than 1 is negative cooperativity, and a slope greater than 1 is positive cooperativity. Cooperative binding measures how the affinity for ligands change as initial binding occurs. The Ki (equilibrium dissociation constant) is another measure of affinity, and it is the concentration of dicarbonyl at which half of the binding occurs. In order for binding curve analysis to characterize this MRM-based data, some assumptions have to be made. First, the initial abundance of modification is near zero. Second, the upper limit of MRM signal as expressed by the asymptote of the peak area ratio is near binding/modification saturation. Third, normalization to the lower and upper limit of this curve eliminates all effects of differential peptide ionization and/or fragmentation.

*MG site reactivity.* MG modification of R98 and R348 containing peptides was not observed at any of the MG concentrations. It is likely that the R348 peptide is too hydrophobic for the assay conditions, as the synthetic peptide was not immediately soluble. It is possible that trypsin does not produce the R98 peptide that we predicted from the theoretical digest. If there are missed cleavages at the N or C terminus of this peptide, then all of the masses would shift. Another possible explanation for the lack of signal at R98 is that glutamine (Q) at the N-terminus may undergo gas phase deamination (Q-17), which would also shift the masses.
Figure 4.23. Cooperativity or Ki of MG adduction to arginine sites is independent of arginine solvent accessibility.

Two measures of arginine reactivity are the slope of the binding curve (upper plot) and the concentration required to get a 50% binding response, or Ki (lower plot). Correlation between hill slope and solvent accessibility is $r^2=0.005$. Correlation between Ki and solvent accessibility is $r^2=0.32$. 
Of the remaining 11 R sites that were responsive to MG, modification is not correlated to side chain solvent accessibility in the two measures (slope, Ki) of reactivity. This is depicted in the two plots in Figure 4.23, where the upper scatterplot is solvent accessibility vs. Hill slope and the lower scatterplot is solvent accessibility vs. Ki. The points in the upper plot appear to form a normal distribution centered on a slope of 1. The three sites that have a slope > 1 are R257, R472, and R485. These sites are relatively buried in HSA. Consistent with the concept of the importance of neighboring microenvironment in dicarbonyl-arginine modification is the fact that the most solvent accessible side chain, R81, has a slope of 1. This is to be expected, as a “free” side chain should not exhibit the effects of cooperativity. The lower scatterplot of Figure 4.23 shows the lack of correlation between solvent accessibility and Ki. The curve fitting software within Prism was unable to calculate a Ki for R218 and R485. With the exception of R81, most of the R sites have relatively close Ki values. The R81 outlier in this case also underscores the importance of microenvironment for site reactivity; the most solvent exposed residue has the highest Ki.
Figure 4.24. Relative modification of arginine sites by 3-deoxyglucosone.

Human serum albumin was treated with a dilution series (0.39-8100 μM) of 3DG in 1X PBS pH 7.4. 3DG-modified albumin was reduced, alkylated, digested, and analyzed by LC-MS/MS in MRM mode using the transitions in Figures 4.14 and 4.16. Peak AUC’s were expressed relative to the unmodified tryptic albumin peptide 42LVNEVTEFAK51 (top plot), and those values were normalized to adjust for differences in ionization/fragmentation efficiency (bottom plot).
3DG site reactivity. 3DG is less reactive than MG and did not modify as many sites. Of the 13 sites monitored for 3DG modification, only R186, R410 and R428 had detectable peak area increases as a function of 3DG (Figure 4.24). Ki values for 3DG modified arginines are 23.8 μM for R186, 39.5 μM for R410, and 23.4 μM for R428. Hill slope values for 3DG modified arginines are 1.05 for R186, 1.78 for R410, and 0.77 for R428.

Glucosone site reactivity. Glucosone is also less reactive than MG and did not modify as many sites. The same sites as 3DG were found for glucosone: R186, R410 and R428 (Figure 4.25). Ki values for glucosone modified arginines are 94.7 μM for R186, 30.0 μM for R410, and 179 μM for R428. Hill slope values for glucosone modified arginines are 1.01 for R186, 0.51 for R410, and 1.00 for R428.

R410 adduction of three dicarbonyls. The shift in Ki of the three dicarbonyls in Figure 4.26 shows that MG is more reactive than the larger dicarbonyls, 3DG and glucosone. The Ki of MG at this site is 6.71 μM, which is 5.9-fold lower than 3DG (39.5 μM) and 4.5-fold lower than glucosone (30.0 μM).

4.4. Discussion

The generation of a comprehensive panel of MRM transitions for glyco-oxidative modifications on HSA is an important first step in the site-specific bottom-up quantification of said modifications. This transition list is empirical data, in that other researchers who want to study these modifications simply need to set up their instruments to include the transitions listed herein, assuming they are using QqQ MS.
Figure 4.25. Relative modification of arginine sites by glucosone.

Human serum albumin was treated with a dilution series (0.39-8100 μM) of glucosone in 1X PBS pH 7.4. Glucosone-modified albumin was reduced, alkylated, digested, and analyzed by LC-MS/MS in MRM mode using the transitions in Figures 4.14 and 4.17. Peak AUC’s were expressed relative to the unmodified tryptic albumin peptide \(^{42}\text{LVNEVTEFAK}^{51}\) (top plot), and those values were normalized to adjust for differences in ionization/fragmentation efficiency (bottom plot).
Those transitions that showed a dose-dependent increase in AUC.

Figure 4.26. Relative modification of HSA at R410 with three dicarbonyls.

The R site that gives the highest relative signal is R410 for modification by MG, 3DG, and glucosone. Plotted against each other, the curves show the response of MG compared to the less reactive 3DG and glucosone.
There exists a debate that oxidation (especially oxidation of methionine) can occur during sample preparation and/or atmospheric ESI (Perdivara et al.) as chemical artifacts as well as important post-translational modification markers. Oxidation likely occurs during all of these processes, so it is important to minimize sample-to-sample variation so that the relative oxidation can be measured. This concept proposes that if the samples are handled minimally and treated identically, then the differences in oxidation between them are due to biological (intrinsic) differences.

The oxidation profiles of proteins are simplified as to the location and extent of modification if we first take an approach toward a general understanding of how each type of residue is oxidized. In other words, M+16, W+32 and C+48 are the most common oxidized forms for these residues. This framework is a starting point in deconstructing the stoichiometry (multiples of 16) associated with analysis of oxidation in top-down proteomics. This information, when combined with the quantitative results, show that M oxidation is about an order of magnitude more abundant than C oxidation, which is in itself an order of magnitude more abundant than W oxidation. The degree of solvent exposure of M is likely an important factor in determining which M sites are the most susceptible to oxidation. This could be confirmed with other proteins, and might apply to W and C residues in proteins with multiple W or C sites that can be monitored on the same protein.

The peak area ratios for the dicarbonyl modification set can be misleading. Unlike the oxidative modifications, the dicarbonyl adducted peptides needed to be further normalized to account for differences in ionization and fragmentation. This produced
binding curves where Ki could be determined. MG sites had relatively similar Ki values, except for the most solvent accessible R81. This data supports the findings that neighboring residues in the microenvironment help stabilize and generate the hydroimidazolone adducts. 3DG and glucosone only appeared to target three sites- R186, R410, and R428. This is likely due to the fact that these dicarboxylics are larger (Figure 4.1), are more saccharide-like (poly-hydroxylated), and may not fit into many different R sites like MG can. Head-to- head comparisons of the three dicarboxylics at R410 show that MG is roughly 5-fold more potent than the larger 3DG and glucosone. 3DG and glucosone are “glucose-like” and probably exist primarily in a non-reactive closed ring hemiacetal because both of these structures contain a C5 hydroxyl and C1 aldehyde.

The steepness of the slopes in the binding curves indicates cooperativity. Cooperativity is an allosteric phenomenon, so at low levels of binding, the ligands bind at allosteric sites and influence further binding. Since the dicarboxyl reactions were run in the same mixture, it is difficult to assess which dicarboxyl was responsible for either positive or negative cooperativity.

Commercially available HSA used in this study may not reflect the true nature of modifications at the lower limit of detection because this protein has been through unknown manufacturing processes. This processing could introduce artifacts like oxidation and glycation. The trade-off for this is traceability and reproducibility as this protein has a catalog number and lot number should other researchers want to reproduce this data. As will be described in Chapter 5, modifications found in human plasma have a
different adduction profile than those found *in vitro*. The difference between modification *in vitro* and *in vivo* is a major theme in this dissertation.
CHAPTER 5: SITE-SPECIFIC QUANTIFICATION OF GLYCO-OXIDIZED HUMAN SERUM ALBUMIN FROM PLASMA

5.1. Introduction

Increasing evidence identifies the formation of advanced glycation end products (AGEs) as a major pathogenic link between hyperglycemia and diabetes complications, such as retinopathy, neuropathy, nephropathy and cardiovascular disease (Ahmed et al. 2007). Current clinical trials in diabetes highlight unmet challenges in treating diabetic complications, where treatment aimed exclusively at decreasing glucose levels does not translate into fewer cardiovascular complications (Gerstein et al. 2008; Patel et al. 2008; Duckworth et al. 2009). A therapeutic approach that targets carbonyl stress, long before the appearance of overt oxidative stress and damage has been recommended (Ahmed et al. 2007). However, novel targets that combine the risk of glucose and oxidative stress have mostly been unexplored. Since the discovery of insulin, the treatment of diabetes has focused on the development of glucose lowering agents. We propose that markers that capture glyco-oxidative risk are needed to guide development of emerging therapies.

We performed a cross-sectional study of 120 subjects with varying degrees of type-2 diabetes to assess oxidative and dicarbonyl markers on HSA in a broad continuum of this complex disease. Specifically, we examined these markers with respect to other clinical measures of glucose dysregulation in order to evaluate the utility of these modifications.
5.2. Methods

5.2.1. Materials

All materials were HPLC grade and were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. Synthetic peptides were custom ordered from New England Peptide (Gardner, MA). Sequencing grade trypsin was purchased from Promega (Fitchburg, WI). Lipidex-1000 (Catalog # 6008301) was purchased from PerkinElmer (Waltham, MA).

5.2.2. Subject selection

All subjects provided informed consent. The study approval was granted by the Human Subjects Protection Program at The University of Arizona under the project title: Proteomic and Metabolomic Biomarker Investigation of Type 2 Diabetes. Subjects were recruited from University Medical Center, Southern Arizona VA Health System, UPH-Kino, and El Rio diabetes and primary care clinics. We collected blood and urine samples from 120 subjects who were instructed to fast overnight. Subjects were assigned to 3 cohorts (controls, type 2 diabetic subjects, and type 2 diabetic subjects with diabetic nephropathy) matched for age, sex, and racial/ethnic background over 1 year. There are 46 subjects with normal or impaired glucose tolerance (pre-T2D), and 73 subjects with T2D. Non diabetes is defined as fasting glucose below 100 mg/dL; pre-type 2 diabetes is fasting glucose between 100 and 125 mg/dL; diabetes is fasting glucose above 125 mg/dL. New diabetes diagnosis was based on 2-hour glucose greater than 200 mg/dl. Twenty one subjects had diabetic nephropathy based on a spot urine microalbumin > 30
mg/g of creatinine (52 T2D without nephropathy). Twenty subjects (all T2D) were diagnosed with cardiovascular disease (CVD). Subjects were tested for multiple metabolic markers that include fasting insulin, fasting lipid profile, serum creatinine, uric acid, in addition to the oral glucose tolerance test (OGTT) with a 75 gram glucose solution. Volunteers were also assessed for body mass index, waist circumference, blood pressure, family history of diabetes, medication history and an activity diary.

5.2.3. Sample handling and storage

Blood was collected into heparin coated vacutainer tubes and immediately placed on ice. Blood samples were centrifuged at 4°C and plasma was aspirated and stored at -80°C in 200 μL aliquots. The total time between blood collection and sample storage was less than one hour.

5.2.4. Tryptic digestion of whole plasma

Plasma samples (200 μL aliquots) from 120 of the subjects were snap thawed by immediately placing the frozen tubes into a 37°C water bath for 10 minutes. Plasma samples were then centrifuged at 14,000 x g for 5 minutes to pellet any plasma precipitates. Plasma (5 μL) from the top layer was added to 100 μL of Lipidex-1000 slurry that was buffer exchanged into an equal volume of 100 mM ammonium bicarbonate (Ambic) pH 7.4. Samples were rotated at room temperature for 30 minutes to allow for slurry mixing and delipidation of plasma. Plasma proteins were separated from slurry by centrifuging through a 0.22 μm centrifugal filter units (Cat# UFC30GVNB, Millipore) for 5 minutes at 12,000 x g. A 100 μL aliquot of 20mM tris(2-
carboxyethyl)phosphine (TCEP) pH 7.4 was added to the filtered protein samples, and these solutions were incubated at 55°C for 30 minutes. To further denature the protein and cool the samples, the samples were then sonicated at room temperature for 10 minutes. A 100 μL aliquot of 25 mM iodoacetamide in 100 mM Ambic pH 7.4 was added to the cooled solutions and the samples were incubated in the dark at room temperature for 30 minutes. Sequencing-grade trypsin at 0.1 μg/μL in 100 mM Ambic pH 7.4 was added to each sample (50 μL, 5 μg of total trypsin) and the samples were incubated at 37°C for 16 hours. Formic acid (5 μL) was added to stop the reactions, and the peptide samples were C18 purified using 50 mg Hypersep C18 RP cartridges (Thermo Scientific) as follows. The cartridges were washed with 5 mL acetonitrile and then equilibrated with 3 X 5 mL of 1% formic acid in water. The acidified peptide solutions were slowly plunged through the C18 resin at a rate of about 1 drop per second. The cartridges were washed 2 X 5 mL of 1% formic acid and eluted with 1 mL of 80% acetonitrile containing 1% formic acid. The peptide solutions were frozen at -80°C, lyophilized to dryness, and stored at -80°C until LC-MRM.
5μl plasma
Delipidate, reduce, alkylate, trypsin (1:70)
C18 purify, lyophilize

1/10 (0.5μl plasma) per LC-MRM

Technical duplicates

n = 120 subjects

73: diabetic  47: normal

21: nephropathy

20: CVD

Figure 5.1. Assay scheme for preparation of HSA peptides from whole plasma.
5.2.5. LC-MRM

A solution of 5 μg/mL synthetic stable isotope-labeled peptide AWAVAR(+10) (corresponding to unmodified R218) peptide was freshly prepared in 1% formic acid and 2% acetonitrile just before MS analysis. This peptide solution was used as a diluent to reconstitute the lyophilized peptide samples. The peptide diluent (100 μL) was added to each lyophilized sample, which were then vortexed for 30 seconds. Samples were centrifuged at 15,000 X g for 5 minutes, and the resuspended peptide solutions were transferred to 300 μL autosampler vial inserts.

Peptide solutions were loaded onto a ZORBAX 300SB-C18 capillary column (5 μm, 0.5 x 150 mm, Agilent) using a Famos autosampler (LC Packings) with a 10 μL injection loop. Peptides were loaded and eluted from the column using an LC Packings Ultimate II HPLC (Dionex) into a 4000 QTRAP using a flow rate of 40 μl/min of solvent A (0.01% TFA, 0.5% formic acid) with a 30 min linear gradient from 5% to 45% of solvent B (acetonitrile containing 0.01% TFA, 0.5% formic acid). The 4000 QTRAP was operated in MRM mode with the optimal transitions listed in Chapter 4. Dwell time was set to 40 ms, Q1 resolution was set to low, and Q3 resolution was set to unit. The source temperature was set at 200˚C, source voltage was 5000 volts, GS1 and GS2 were set to 0 and 25 psi, respectively, and the declustering potential (DP) was set to 70 volts for all peptide parent ions. There were two technical replicates per sample. An example of a typical MRM spectrum is shown in Figure 5.2.
5.2.6. Data analysis

Data analysis and MRM peak integration was performed with Multiquant software (AB Sciex). All dicarbonyl-modified and oxidized peptide peaks were manually inspected for correct assignment and peak integration. The default peak integration parameters for unmodified peptides are as follows. MQL was used as the default integration algorithm, the smoothing width was set to 1.0 points, and the peak splitting factor was set to 4 points. The area under the curve (AUC) was then calculated for modified and unmodified peptides.

Statistical analysis was performed using Statistica v.7.0 (Stat Soft Inc., Tulsa, OK) with the Multivariate Exploratory Techniques module. For Principal Components Analysis (PCA), analysis was based on correlations, variances were computed based on SS/(n-1), and MD deletion was based on mean substitution. Post-hoc tests for analysis of variance (ANOVA) were performed using the Fisher LSD test. Cluster analysis was based on single linkage and the distance measure was 1-pearson r.
Figure 5.2. Example of a LC-MRM run.

Left spectrum shows the full profile of all HSA MRM transitions, and the visible peaks are mostly unmodified peptides. Peptide elution from the column is between 15 and 35 minutes. Right spectrum is a zoomed region to show the dynamic range of the MRM set, with well defined peaks for the modified peptides. Each color represents a different MRM transition.
5.3. Results

5.3.1. Identification of oxidized and dicarbonyl-modified sites in vivo

There were 13 unmodified HSA peptides monitored for each sample, 6 of which contain methionine and were used as controls for oxidation of methionine, 1 contained tryptophan and was used as a control for oxidized tryptophan, and 1 contained free cysteine (C34, not disulfide bonded) and was used as a control for oxidized cysteine. This leaves 5 peptides that were monitored that did not contain MWC oxidizable residues. Of all the 13 unmodified peptides, all 120 subjects had MRM signals for each, which exclude the presence of any HSA polymorphisms or mutations in these subjects that would shift the peptide masses at these monitored sites.

Oxidized methionine containing peptides gave relatively strong signals with peak area ratios between 0.01 and 0.1 to the corresponding unmodified peptides. All oxidized methionine peptides eluted earlier than the unmodified peptides, indicating the increased polarity/hydrophilicity of M+16 oxidation. Tryptophan (W+16, W+32, W+4) and cysteine (C+16, C+32, C+48) oxidations, however, were lower in abundance and were in most cases indistinguishable from other interfering peaks. The W and C oxidized modifications can be identified at higher sample input (the equivalent of 2 μL of plasma per LC run, as opposed to 0.5 μL of plasma per LC run), but problems with assay linearity in terms of detecting hydroimidazolone was observed.

The three sites that consistently gave strong MRM signals for arginine modification by dicarbonyls were R186, R257, and R428. R186 was shown to be modified by all three (MG, 3DG, glucosone) dicarbonyls, yet R257 was only shown to be
modified by MG, and R428 was only modified by 3DG and glucosone. This is consistent with the in vitro modified HSA data, and suggests that the low solvent accessibility of R257 means that the smaller MG can fit at this site, and larger 3DG and glucosone do not fit into this deep pocket of drug site I. R428 was found to be adducted by the larger dicarbonyls 3DG and glucosone, and this site may be favored due to the interactions of the R428 site and the hydroxyl groups of these two carbohydrates. Modification of R410 by dicarboxyls was not observed in any of the subjects, despite having strong MRM signals when HSA was modified in vitro.

5.3.2 Normalization of MRM data

Inherent assay and MS variation must be normalized before reporting the relative values of these modifications. The most accurate means to accomplish this is to measure the relative modification based on the MRM signals for the modified sites compared to the MRM signals for the unmodified peptides. In doing this, variation in assay (initial protein concentration, delipidation non-specific binding, reduction, alkylation, digestion, purification, lyophilization, and reconstitution) and LC-MS/MS (autosampler, chromatography, instrument drift) should be accounted for by the unmodified peptides. Once the ratio of modified to unmodified has been determined, these ratios are log transformed to adjust for extremes at the low and high ends.

Oxidized methionine containing peptides were normalized as follows. The raw AUC of the oxidized peptide was divided by the raw AUC for the corresponding unmodified peptide, whereby each transition monitored the same parent and fragment
ion. This value was log transformed, and the log transformed peak area ratios were averaged between technical replicates. The averaged log(peak area ratio) was calculated for each M site. Therefore, each subject had six values for oxidized methionine, one value for every M site. It was apparent that the trends for oxidized methionine were similar for all sites, so principal components analysis was performed to reduce these six variables into one variable or “principal component” (PC). PC analysis of the six methionine sites showed that the first principal component captured 63% of the variation of this data set. Figure 5.3 shows the scree plot of these six variables to illustrate the fraction of total variance represented by each PC. The eigenvalue number on the X axis is a ranking of the PC in decreasing order. Since the first PC of this data set was able to reduce so much variation in the data, the value of the first PC coordinate was used for correlation with clinical data. This value will be referred to as “PC oxidation.”
Figure 5.3. Scree plots for PCA of oxidized methionine, unmodified HSA, and glycation/glucose metrics.

The averaged log(modified AUC/unmodified AUC) values for 6 methionine sites (top left) were analyzed for principal components, and the first principal component (Eigenvalue #1) captured 63% of the variation of the set. This value is referred to as: “PC oxidation”.

The log AUC values for 13 unmodified HSA peptides (top right) were analyzed for principal components, and the first principal component captured 60% of the variation of the set. This value is referred to as: “PC unmodified”.

Log transformed fasting glucose, OGTT, and HbA1C values for each subject were analyzed for principal components (bottom left). The first PC captured 80% of the variation of this data set. This coordinate value is hereafter referred to as “PC glycation/glucose”.
As was discussed in Chapter 4, the unmodified arginine containing peptides are quite different than the corresponding dicarboxyl modified versions due to the missed tryptic cleavage. Therefore, a means for normalization based on a composite of the AUC of MRM signals from unmodified peptides was sought. In theory, any single unmodified HSA peptide can be used for normalization, but a representative value that encompasses all of the unmodified sites should be a more accurate common denominator for normalization. To identify this single value that represents “unmodified HSA signal”, PCA was performed on the log AUC values for the 13 unmodified peptides of all 120 subjects in duplicate. Surprisingly, the first PC captured a significant majority of the variation (>60%), as illustrated in the scree plot in Figure 5.3.

PC1 (first principal component) of the 13 unmodified HSA peptides was used to normalize the MRM signals for arginine containing peptides modified by dicarboxyls. As with the oxidized methionine peptides, the final form that will be used for clinical analysis is: log(modified/unmodified). Since the PC coordinates are already log transformed, this can be rearranged by taking the difference of the log transformed values: log(AUC dicarbonyl modified)-PC1. The PC for each sample was subtracted from the log(modified AUC) for each dicarbonyl, and the technical duplicates were averaged. Hierarchical clustering (Figure 5.4.) shows that the MG sites cluster together and the larger 3DG and glucosone sites also cluster together. Normalized MG values from two sites (R186 and R257) were averaged to give a single “normalized MG” value for each subject. Normalized 3DG and glucosone from two sites (R186 and R428) were averaged (4 values total) to give a single “normalized glucosones” value for each subject.
Figure 5.4. Hierarchical clustering of the AUC values for the three dicarbonyls at three R sites.

Log transformed MRM AUC values from each subject in technical duplicates were clustered. MG sites R186 and R257 clustered together, and the values were averaged between the two sites for each subject. 3DG and glucosone sites R186 and R428 clustered together, and these four values were averaged to give a single “normalized glucosones” value for each subject.
5.3.3. Clinical data and PCA

In an effort to simplify the clinical values representing glucose dysregulation, PCA was performed on three major measures of diabetes—fasting glucose, oral glucose tolerance test (OGTT) and glycated hemoglobin (HbA1C). These three values were log transformed and analyzed using PC. Figure 5.3. shows the first principal component for these three variables reduced the variation by 80%. Coordinates for this first PC will be referred to as “PC glycation/glucose” and it will serve as a benchmark to indicate the severity of diabetes. All of the other continuous variables (eGFR, WC, etc.) were log transformed prior to statistical analysis.

5.3.4. Correlations between HSA modifications and clinical data

The PC for glycation/glucose was used as a benchmark to evaluate the oxidized HSA (PC oxidation) and dicarbonyl modified HSA (normalized MG, normalized glucosones) as markers in the context of the other clinical data. There will be two general ways to evaluate this clinical data. The first approach is to analyze the continuous variables against each other in a correlation matrix. The second approach is to group the subjects according to clinical diagnoses or other discrete variables. Discrete variables will be discussed in section 5.3.5. One of the first decisions to be made with this analysis is whether to apply parametric or non-parametric methods in the statistics. There is no simple way to determine which method to use, however, if the data is normally distributed, then parametric methods offer more statistical power. Moreover, non-parametric methods are useful for small samples (n<50), and parametric methods are
used for larger sample numbers. In studies with larger samples, it makes little sense to
use non-parametric analysis. The sample size for this study is n=120.

In order to test for data normality of the HSA modifications, a Shapiro-Wilks’ W
test was performed in order to assess whether the data are accurately modeled by a
normal distribution. In this test, if the W statistic is significant (p<0.05), then the
hypothesis that the respective distribution is normal should be rejected. For all three HSA
modification variables (PC oxidation, normalized MG, normalized glucosones) the W
statistic was found to be insignificant (p>0.05), therefore, the data fits a normal
distribution (Figure 5.5). It is for this reason, and the fact that the sample number is
relatively large, that parametric statistics were employed for this analysis.
Figure 5.5. Distribution fitting of variables to test for normality.

The first principal component (PC) values for methylglyoxal (red), glucosones (pink), and oxidation of methionine (green) were tested for normality using a Shapiro Wilks’ W test. The W statistic for all of these variables was not significant, therefore, the null hypothesis that the respective distributions are not normal should be rejected.
Continuous variables. A summary of the correlations between the continuous variables are tabulated in **Figure 5.6**. The correlation coefficient R values were calculated for the log transformed continuous variables. Urine albumin levels that were undetectable were assigned a value of 1 μg of urinary albumin/mg creatinine for log transformation and subsequent statistical analysis purposes. In most cases, PC1 for glycation/glucose was highly correlated with other clinical measures. Lipids (HDL, LDL, total cholesterol, triglycerides) were not significantly correlated with either oxidized or dicarbonyl modified HSA, and are not shown in **Figure 5.6**. Triglycerides (TG) and HDL, however, had significant correlations with PC glycation/glucose (TG: \( r = 0.468, p < 0.001 \); HDL: \( r = -0.350, p < 0.001 \)).

The first continuous clinical variable to be compared in **Figure 5.6.** is the estimated glomerular filtration rate (eGFR). This was calculated using the simplified Modification of Diet in Renal Disease (MDRD) formula (Levey et al. 1999; Stevens et al. 2006) where 
\[
eGFR (\text{mL/min/1.73 m}^2) = 175 \times ([\text{serum creatinine (\( \mu \text{M} \))/88.4}]^{-1.154}) \times (\text{age in years})^{-0.203} \times (0.742 \text{ if female}) \times (1.210 \text{ if African American}).
\]
EgFR is used to assess kidney function and is also used to stage chronic kidney disease (CKD). Negative correlations with the PC for glycation/glucose, PC oxidation, MG, and glucosones (**Figure 5.6.**) are expected as the presence of these markers should increase as kidney function (filtration rate) decreases. Correlations (\( r = -0.252 \)) and p values (\( p = 0.006 \)) using PC glycation/glucose were closely matched for eGFR compared to the HSA modifications. Positive correlations with age are observed with the PC glycation/glucose
narrowly beating the PC oxidation in terms of correlation (r=0.392) and significance (p<0.001).

Waist circumference (WC) was also significantly correlated with PC glycation/glucose, and interestingly, WC was also significantly correlated with MG and glucosones. Blood urea nitrogen (BUN) was significantly correlated to PC glycation/glucose, as was body mass index (BMI) and homeostatic model assessment-insulin resistance (HOMA-IR). The HOMA-IR value is calculated by multiplying fasting glucose by fasting insulin and dividing by 22.8, so it makes sense statistically that these values are correlated because fasting glucose is a common factor in both metrics.
Correlation coefficient R values were calculated for the three groups of modified HSA (PC oxidation, PC methylglyoxal, and PC glucosones) and clinical variables. The PC for glycation/glucose was used as a benchmark to assess severity of diabetes. P values <0.05 for significant correlations are indicated in bold.

eGFR: estimated glomerular filtration rate, WC: waist circumference, BUN: blood urea nitrogen, BMI: body-mass index, HOMA-IR: Homeostatic model assessment-insulin resistance
5.3.5 ANOVA of discrete variables

Two major clinical endpoints for this study are cardiovascular disease (CVD) and diabetic nephropathy (DN). CVD is defined by history of a major cardiovascular event, such as stroke or myocardial infarction. DN is defined by the presence of protein in the urine. The cutoff for this is \( \geq 30 \mu g \) of urinary albumin/mg creatinine. Other discrete variables are sex, medications (e.g. metformin), and diagnosis of type 2 diabetes (T2D). T2D is defined by having a fasting blood glucose \( \geq 126 \) mg/dL. Subjects grouped into these cohorts are evaluated using analysis of variance (ANOVA) to identify markers or variables that are best able to distinguish the different groups. Post hoc analysis of the ANOVA data assigns significance to each variable in terms of being able to discriminate the different groups.

There were no significant differences in the clinical or HSA modification data between the sexes except for HDL (mean F: 56.0 mg/dL, mean M: 45.9 mg/dL, p=0.001). The next step was to evaluate whether pre-type 2 diabetes (pre-T2D) should be considered as a separate cohort for this analysis. Pre-T2D is defined as having a fasting blood glucose level 100-125 mg/dL. ANOVA was unable to discriminate between non-diabetic subjects (ND) and pre-T2D for the continuous variables except for PC glycation/glucose (mean ND: -1.786, mean pre T2D: -1.046, p=0.048) and age (mean ND: 43.0, mean pre-T2D: 50.3, p=0.039). Therefore, these two groups were considered as one group: ND. The ND group serves as a control group that does not have subjects with CVD, DN, or subjects taking metformin.
The data shows major differences between the ND group (n=46) and the T2D group (n=74) as a whole, when including all complications and metformin use. The variables that have p < 0.05 for ANOVA of these two cohorts are: PC glycation/glucose, PC oxidation, normalized MG, normalized glucosones, eGFR, triglycerides (TG), age, WC, BUN, BMI, HOMA-IR, and urine albumin. Most of the significant differences between the ND and T2D groups in terms of HSA modifications were attributed to the differences between the ND group and the T2D cohort not taking metformin. The effect of metformin will be discussed in the next section. Because there are major differences between the ND cohort and the T2D cohort regardless of any diabetic complications, the differences between ND and T2D without complications vs. ND and T2D with complications will not be discussed. An important question that will be addressed, however, is whether or not these modifications are markers to identify complications within the T2D group. We investigated whether the HSA modifications were significantly different in the CVD or DN cohorts compared to T2D without these complications, and compared that ANOVA to the PC glycation/glucose metric. In T2D subjects with CVD, the post hoc p values for one-way ANOVA are all greater than 0.05 (PC glycation/glucose p= 0.596, MG p=0.665, glucosones p=0.603), yet the PC oxidation approached significance at p=0.089. The DN group also had p >0.05 for these variables within the T2D cohort (PC glycation/glucose p=0.195, PC oxidation p=0.437), but the dicarbonyl modifications had ANOVA p values that were the lowest with MG p=0.128 and glucosones p=0.112. Even though the p values are >0.05 for HSA modifications, the trends are approaching significance for PC oxidation (in CVD) and dicarboxylic (in DN).
In both cases, the $p$ values are less than the $p$ value for the benchmark PC glycation/glucose. A scatter plot of MG vs. PC glycation/glucose for three cohorts (ND, T2D, and T2D with DN) is shown in Figure 5.7.
Figure 5.7. Scatterplot of MG vs. PC glycation/glucose does not identify subjects with nephropathy.

$r=0.214$, $p=0.019$

5.3.6. The effect of metformin

It is clear in the analysis of this data that subjects taking metformin have significantly lower values for dicarbonyl and oxidative modifications compared to T2D subjects not on metformin therapy. Subject selection for this study was not initially designed (or statistically powered) to study this effect. It was even written in the original study plan to have the clinicians’ advise subjects who are taking metformin to stop metformin use two weeks before the blood draws, yet many of these subjects did not stop taking metformin. However, those subjects who withheld or did not withhold metformin were recorded. Of the 73 diabetic subjects in this study, 26 were taking metformin and 47 were not taking metformin. Metformin is associated with decreased oxidative and dicarbonyl stress, and clinical studies have demonstrated decreased levels of free dicarbonyls, dicarbonyl adducts, and oxidized adducts in diabetic subjects taking metformin relative to matched diabetic subjects not on metformin (Beisswenger et al. 1999; Rabbani, Chittari et al.). Our findings corroborate the results of these other clinical studies. Figure 5.8. shows the differences between the groups ND, T2D on metformin, and T2D not on metformin for the dicarbonyl modifications. In the top plot in Figure 5.8., ANOVA of normalized MG means and standard errors for each of these groups is shown. On the y-axis is normalized MG, and the x-axis is grouped according to subjects +/- metformin use. In blue are ND subjects, and none are taking metformin. In red are T2D subjects, the left data range (in red color) is for T2D not taking metformin, and the right data range (in red color) is the mean and standard error of T2D on metformin. Post hoc (Fisher LSD) tests were performed against all groups, and the most significant
differences (p<0.001) were found between the ND group and the T2D group not on metformin. One important note is that there are also significant differences (p=0.031) between T2D subjects not taking (left) and taking (right) metformin. There is a substantial and significant decrease in MG modifications in T2D subjects taking metformin, and this population is not statistically significant (p=0.342) from the ND population.

This effect of metformin is consistent for other HSA modifications like glucosones and oxidations. In the bottom plot of Figure 5.8, the means for the normalized glucosones are plotted in the same fashion as the normalized MG (top plot). For glucosones, statistical significance between the groups is generally the same as MG. Glucosones are highest in T2D not taking metformin when compared to the ND group (p=0.001) or the T2D group on metformin (p=0.059). Again, the ND group is not statistically different (p=0.29) than the T2D group on metformin for glucosone modification. The values for PC oxidation between the three groups (data not shown) is similar in terms of trends and significance to that of the glucosone modifications. Oxidation is the greatest in the T2D group not taking metformin, which is significantly higher (p<0.001) than the ND group. The mean of the PC oxidation for the T2D cohort taking metformin is also between that of the ND group (p=0.111) and the T2D cohort not on metformin (p=0.157).
Figure 5.8. ANOVA of normalized MG and glucosones values between non-diabetic subjects, diabetic subjects not taking metformin, and diabetic subjects taking metformin.

Post hoc test (Fisher LSD) p values are indicated between the groups. Values are means of the groups +/- standard error.

ND: not diabetic, n=46  T2D taking metformin, n=26  T2D not taking metformin, n=47
We wanted to examine the differences between the T2D subjects taking or not taking metformin in order to rule out any effects of covariation. A simplified (T2D subjects only) multivariate analysis of covariance (MANCOVA) was performed as follows. The dependent variables analyzed for covariance were normalized MG, normalized glucosones, and PC oxidation. The categorical predictor was metformin use, and the continuous predictors were eGFR, lipids (TG, HDL, LDL, TC), age, sex, WC, BUN, BMI, HOMA-IR, and urine albumin. None of the modified HSA dependent variables (MG, glucosones, oxidation) shared any significant (p<0.05) F values for any of the candidate covariables, so covariance was ruled out as a means to describe the effect of metformin on the HSA modifications.

One of the most interesting aspects of the effect of metformin is that the clinical glucose measurements (fasting glucose, HbA1C, OGTT) as represented by PC glycation/glucose are not markers for whether or not subjects are taking metformin in this study (Figure 5.9). In fact, the glucose values for subjects taking metformin are slightly higher on average than in T2D subjects not taking metformin. ANOVA is unable to distinguish these two groups (p=0.098), however, the differences between ND and either T2D group is significant (p<0.001).
Figure 5.9. ANOVA of PC glycation/glucose values between non diabetics, diabetics not taking metformin, and diabetics taking metformin.

Post hoc test p values are indicated between the groups. Values are means of the groups +/- standard error.

ND: not diabetic, n=46   T2D taking metformin, n=26   T2D not taking metformin, n=47
5.4. Discussion

Sample preparation. This study used methods that involved minimal sample preparation in order to generate tryptic HSA peptides for LC-MRM. We tried a variety of HSA enrichment and isolation methods, including Cibracon blue binding and immobilized anti-HSA methods. However, it is not known whether or not these affinity-based methods will bind to the modified (oxidized, dicarbonyl modified) versions of HSA. Furthermore, an additional enrichment step could introduce variation into the system. Whole plasma was therefore used to measure the relative level of HSA modifications because HSA is the major constituent of plasma at about 50% of the total plasma protein content by mass (Prinsen et al. 2004). In a different study that used hydrazide labeling of abundant plasma proteins, of four major plasma proteins (albumin, fibrinogen, immunoglobulins, and transferrin), albumin was the major plasma protein target of oxidant stress in uremia (Himmelfarb et al. 2001).

The caveat to the use of whole plasma is that there are many interfering substances in plasma that cause downstream problems for LC and MS. Specifically, hydrophobic compounds such as lipids and drugs can clog capillary separations, shift retention times during chromatography, dramatically alter ionization of peptides, and can even saturate the MS detector. Our success in using the hydrophobic resin Lipidex (Perkin Elmer) in removing trace amounts of prostaglandins (Chapter 3) was also applied to using this resin as a novel means to delipidate plasma prior to protein processing. Initial experiments (data not shown) demonstrated that trypsin digestion was more efficient in reducing sample-to-sample variation if the plasma was delipidated using
this resin. Although a rigorous evaluation of this product was not established here, it is anticipated that this simple technique will receive widespread implementation as a robust method for improving sample preparation for MS-based analyses of complex mixtures.

**Data reduction and normalization.** The major trends, correlations, and significant variables described in this study hold true whether or not PC was used to reduce the number of variables. For the clinical manifestation of diabetes represented by PC glycation/glucose, any single constituent variable (e.g. fasting glucose) could have been used as a standalone variable and the major outcomes and significant associations would have been the same. By consolidating these overlapping variables into one, it not only eliminates gaps in the data for which no value was recorded (e.g. missing clinical data), it also offers a more powerful means to simplify redundant variables. In using many variables at once in the analysis, it reduces any bias that any single variable alone might create.

The signals from the unmodified peptides did not have to be consolidated either, as any single unmodified HSA peptide could have been used to adjust for assay conditions. Using PC to identify a value that represents something akin to an average of all 13 unmodified peptides is a better choice because it reduces any bias that exists in choosing a single unmodified peptide. In order to streamline the analysis of the individual dicarbonyl modifications at the different sites, an average of 2 sites was chosen for MG, and an average of 2 sites and 2 dicarboxyls (glucosone,3DG) was chosen for the glucosones based on hierarchical clustering (Figure 5.4). While these individual sites were evaluated for statistical correlations, no single site outperformed or was more
significant than the other. It was because of this that the sites were averaged based on clustering similarity, which was in good agreement with the chemical properties (dicarbonyl size, reactivity) of the two major types (MG and glucosones) of dicarboxyals. Site specificity was not as important as the nature of the adduct for both oxidation of M and dicarbonyl adduction of R. In other words, there was not a single site that was more associated with diabetes for any of the three groups of modifications.

Clinical significance. Two major aspects arise in the analysis of this data: 1. how well these markers (oxidation, dicarbonyl adducts) capture glyco-oxidative stress, and 2. how well glyco-oxidative stress associates with major clinical endpoints compared to traditional measures of glucose dysregulation. In other words, are these markers of glyco-oxidation and how good are they? To answer how well these markers capture glyco-oxidative stress, it should be noted that these markers do not measure the reactive species, rather, they are the products of glyco-oxidative damage. In this sense, they are analogous to HbA1C in that they measure the history of glyco-oxidation in the plasma over a period of time. As the average half-life of HSA is 19 days, these markers should capture plasma free dicarbonyls and ROS over the lifetime of the protein. The day-to-day fluctuations of free dicarbonyls and ROS is not well studied. The degree of glyco-oxidized protein modification should account for major swings in concentrations of reactive species to average this potential variability.

In this study, these markers do not offer significantly better statistical power than standard glucose metrics in identifying the two major clinical endpoints, CVD and DN. This is partially due to the covariation of diabetic subjects taking metformin (discussed
It is tempting to use the lower p values for oxidation of M to describe how oxidation is a better marker for CVD than glucose measurements. It is also tempting to claim that the lower p values for MG and glucosone markers for DN are more powerful than glucose measurements. These are extraordinary claims, and extraordinary claims require extraordinary evidence. In both cases, lower p values point to trends that could be significant, but another study would need to validate this. This study would need to be larger, and it would need to exclude subjects taking metformin. Ideally, it would focus on one endpoint, either CVD or DN.

The use of metformin has a remarkable effect on glyco-oxidative adduct levels. The results of this study were in close agreement with a study that quantified the levels of MG-H1, 3DG-H, and methionine sulfoxide from apoB100 in T2D subjects +/- metformin (Rabbani, Chittari et al.). This apoB study isolated apoB100 from LDL and performed exhaustive digestion and quantification of glyco-oxidized adducts by LC-MRM. While they do not report mean values, the median values, ranges, and significance between cohorts (ND n=21, T2D –met n=19, T2D+met n=13) are in close agreement with our data. T2D subjects not taking metformin had significantly higher MG, 3DG, and methionine sulfoxide levels than T2D taking metformin. Metformin is a weak carbonyl scavenger, and its ability to decrease dicarbonyls is not thought to be primarily through direct binding of reactive carbonyls (Engelen et al.). Metformin will, however, produce the cyclic triazpinone adduct when treated with MG (Ruggiero-Lopez et al. 1999), and these adducts in the urine are inversely proportional to plasma free MG levels (Beisswenger and Ruggiero-Lopez 2003). Metformin does not directly scavenge free
radicals (Khouri et al. 2004). The primary mechanisms underlying how metformin decreases dicarbonyl and oxidative stress have surprisingly not been thoroughly researched. Some promising work has looked into the upregulation of plasma and erythrocyte superoxide dismutases, catalase, and glutathione levels as a result of metformin treatment (Faure et al. 1999; Pavlovic et al. 2000). The molecular target(s) for these activities could be related to actions on the generally agreed upon target of metformin, AMP-activated protein kinase. It is widely believed, however, that the benefits of metformin are multifaceted, and not strictly attributed to the actions on a single target (Scarpello et al. 2008). The large doses (1-3 g daily) also increase the likelihood of off-target effects. Metformin is particularly interesting because of the cardiovascular benefit in addition to the glucose-lowering effects. This could be the true utility for glyco-oxidative markers, as they are able to measure an effect of metformin that traditional measures of glucose do not. The question of whether or not these markers are able to measure cardiovascular risk is certainly worth investigating. In other words, is the metformin effect on glyco-oxidative markers related to CVD risk reduction? An example where this would be useful is in the development of new anti-diabetic drugs. In the wake of Avandia-related (rosiglitazone) cardiovascular problems, the FDA requires evidence of CVD benefit for all new diabetic therapies. These glyco-oxidative markers could be useful in determining if new diabetic therapies are similar to metformin in their ability to reduce protein damage by glyco-oxidation.
An ideal marker would be one that is able to predict the development of diabetic complications with greater accuracy than currently available clinical assays. This is, however, quite a lofty goal. It is more likely that markers described herein can serve as part of a panel containing multiple markers that can help to guide and optimize therapeutic regimens. Thus, early treatment of the disease would have beneficial outcome in improving the quality of life in diabetic patients. Further evaluation and final validation of these biomarkers may not only help predict patient populations that are susceptible to develop diabetes and diabetic complications, they may also serve to optimize pharmaceutical treatment during the course of therapy. These markers might also be useful in the development of new therapies that aim to reduce blood sugar as well as oxidative and carbonyl stress.
CHAPTER 6: CONCLUDING REMARKS

6.1. Introduction

A major byproduct of glyco-oxidation is the generation of reactive dicarbonyl intermediates. These electrophillic intermediates are toxic because they damage biomolecules, such as proteins, leading to an altered function primarily as a result of neutralization of protein positive charges. As a class of saccharide metabolites, dicarboxyls represent a different aspect of metabolic dysregulation than just glucose levels alone. Inefficient metabolism of phosphorylated glycolytic metabolites is considered a major source of MG and 3DG. Spontaneous dephosphorylation of triose phosphates and hexose phosphates, respectively, are the major metabolic mechanisms for endogenous formation of these dicarboxyls. Glucosone has not been well studied, however, it is likely that the majority of glucosone is formed without a metabolic component like the sister dicarboxyls MG and 3DG. The majority of endogenous glucosone probably results from a standard 2 electron oxidation of glucose. To a lesser extent, MG and 3DG can be formed directly from oxidation of sugars such as glucose. These mechanisms likely involve transition metal (Fe$^{2+}$, Cu$^{2+}$) catalyzed oxidation reactions in vivo.

The products of dicarboxyl modification and oxidation are more robust measures of glyco-oxidative stress than measurement of the reactive intermediates.Dicarboxyls will primarily adduct arginine residues on proteins to generate hydroimidazolone end stage products. The abundance of MG and 3DG hydroimidazolones in plasma protein is approximately 0.1% of total arginine content each. However, the distribution of these
sites throughout the proteome has not been previously established. We selected HSA as a model protein to investigate the distribution of dicarbonyl modifications and oxidation *in vitro* and *in vivo*.

### 6.2. Identification of adduct sites

Using MG as a model dicarbonyl, plasma protein from a non diabetic subject was treated with MG to induce modifications. The MG-modified plasma proteins were digested into peptides and analyzed using shotgun proteomics to identify the sites for adduction. As expected, the modifications were found on abundant plasma proteins. Each site and every mass spectrum that identified the sites were manually inspected. The processes for manual validation are dependent on a number of criteria that are not included in the database search/matching algorithms. These criteria are dependent on the nature of the modification as well as the assay procedures used to generate the peptides. The process of identifying real sites of protein modification from MS/MS spectra is to eliminate all of the false positive matches. Since this process is dependent on statistics, it is helpful to separate the true positives from false positives based on what is the most likely explanation for the combinations of peptide fragments. A good starting point is to assume that all of modifications listed in the results of the database searches are false positives, then to work backwards to identify the true modified spectra. In this way, the true positives are separated from the false positives because the only possible explanation for the combination of fragments is that the peptide is modified. There were many false positive matches that were eliminated from consideration as true matches.
In order to gain a better understanding of likely and unlikely spectral matches for modified peptides, there are three areas that provide information about the likelihood of identification of peptide modification by shotgun MS. The first, and one of the best places to develop an understanding for protein modification, is to closely examine the unmodified spectral matches from the same experiment (same data set). Since protein modification in itself is an unlikely event, only modified peptides that match to proteins that were also identified from (several) unmodified spectra should be considered. In other words, it is likely that the match to that modification is real only if other peptides from that particular protein are also identified. The unmodified spectra from the same data set also indicate how well trypsin cleaved at arginine and lysine (missed cleavages should match, except if the modification induces a missed cleavage), as well as efficiency of cysteine capping (uncapped C’s are unlikely). Two other ways to test the assignment of false positives are in: 1) MS/MS spectra sets that do not contain the modification, and 2) reverse sequence databases in which the sequences for each protein is in the reverse order. Both of these approaches can be used to calculate a false discovery rate (FDR) for each modification.

The list of MG sites (Figure 2.4) on abundant plasma proteins were examined in terms of the linear sequence. In most cases, a negatively or positively charged residue was within 3-4 residues of the hotspot arginine. While this does not mean that these charged residues are necessarily interacting with the arginine site, an examination of these arginine sites from their X-ray crystal structures indicate that an adjacent charge motif exists (Figure 6.1). Generally speaking, the negatively charged residues (D,E) were
Figure 6.1. Adjacent charge motif for arginine adduction by dicarbonyls.

Positive charge- Lysine or Arginine
- Decrease pKa of target arginine
- Initial binding / transfer of dicarbonyl

Negative charge- Glutamate or Aspartate
- Facilitate hydroimidazolone adduct formation by base catalysis of dehydration step
found relatively close (3-5 Å) to the hotspot arginines compared to the positively charged (K,R) adjacent residues (4-9 Å).

6.3. Functional consequences of dicarbonyl-induced protein damage

Using HSA as a model protein with which to validate dicarbonyl modification, we chose to focus on drug site I (R257) as a target for modification. This was based on the search results that identified R257 as a target for in vitro MG modification as well as in vivo modification from MRM of HSA from the clinical subjects. There are two assays that we adapted to study this specificity of binding at drug site I. The general concept is that if MG is binding to drug site I then these assays will be perturbed. The first assay measures a relatively esoteric property of HSA, which involves hydrolysis of a model prostaglandin. The second assay measures the displacement of a fluorescent probe from the drug site I pocket. There was a difference in the concentration of MG required to significantly alter these assays, as the prostaglandin conversion assay was significantly inhibited at approximately 100 μM MG and the displacement of prodan was significant at approximately 300 μM MG. The difference between the two assay results probably relates to the specific residues that prodan and prostaglandin interact with in the pocket of drug site I. Molecular modeling of the hydroimidazolone adduct on R257 reduced the docking affinity (12.8-16.5 kcal/mol) of warfarin in drug site I.
6.4. Relative reactivity of sites for oxidation and dicarbonyl modification

In the field of protein-electrophile adduction research, there is an emerging trend toward not only understanding the protein sites that are susceptible to modification, but also in determining the relative site-specific reactivity of these covalent modifications. With advances in mass spectrometry and other analytical technologies, the concomitant direction of research will shift from site identification toward site quantification. Based on the reactive species under investigation, this means that every site should have a value for the degree of modification. As the drug industry is moving toward the widespread implementation of biopharmaceuticals, they must be able to accurately quantify all of the low level PTM’s on their engineered proteins. It is likely that in the future the FDA will require methods such as the ones described in this dissertation to accomplish this comprehensive characterization.

The application of binding curve analysis to determine the reactivity of HSA sites toward MG based on peptide MRM signals is a new concept. It must be stressed, however, that intensity of MS does not exactly equal abundance, and these intensities must be normalized to account for differences in ionization and/or fragmentation. There must be some assumptions made before saturation binding curves can be fit to the data. We propose that a value such as the equilibrium binding constant (Ki) can be used to compare site reactivity. This is simply the value (in concentration of reactive species) at which 50% of the binding response is achieved. Sites in HSA that were monitored for MG modification and calculated for Ki gave a narrow range (10-35 μM) for 8 arginine sites. These sites were relatively buried compared to the most solvent exposed arginine,
R81, which had a Ki of 123 μM. This data suggests that the reactive sites need to be somewhat buried so that the microenvironment can facilitate adduct formation.

The relative abundance of oxidation was studied using some of the major known protein targets of oxidation, namely, the side chains of methionine, cysteine, and tryptophan. Synthetic HSA peptides were oxidized in order to generate a comprehensive MRM transition list to quantify each oxidized species. HSA was treated with a dilution of hydrogen peroxide and oxidized peptides were normalized to the unoxidized versions. Methionine was the most abundant oxidized side chain and the M+16 adduct was the only detectable oxidized form. The next most abundant oxidation product was cysteine, and C+48 was the predominant oxidation product. Finally, tryptophan was the least abundant, and W+32 was found to be the major oxidized species. Of the 6 M sites in HSA, M329 was the most reactive. It is also the most solvent accessible, and it is by the most abundant oxidized methionine on HSA in vivo.

6.5. Glyco-oxidation in vivo and metformin

Despite finding many sites for dicarbonyl modification in vitro, only three (R186, R257, and R428) were found in vivo. This could be due to differential protein turnover, where certain modifications are recognized and targeted for degradation, leaving other modifications to remain in the plasma. It is known that albumin is predominantly synthesized by the liver, yet the factors that regulate albumin degradation are not clearly established (Prinsen et al. 2004). Muscle, liver, and kidney are the main contributors to albumin catabolism, and these three tissues comprise 40-60% of albumin degradation in
rabbit models (Yedgar et al. 1983). The HSA peptide that had the best MS signal upon modification by MG was the R410 containing peptide, yet this was not found in any of the subjects. Modification of HSA at this site may be rapidly eliminated, as at least one study has shown that the R410M HSA mutant is rapidly degraded in mice (Iwao et al. 2006). Therefore, anything other than a positive charge at this site (i.e. methionine or hydroimidazolone) could be targeted for degradation.

In MRM analysis of the dicarbonyls, we decided to average the MG sites (R186 and R257) and the glucosones (R186 and R428 for both 3DG and glucosone) because they clustered well together. The glucose metric “PC glycation/glucose” is analogous to a score for severity of diabetes. In most cases, this value correlates well with other measures of glucose intolerance, such as age, waist circumference, blood urea nitrogen, body mass index, HOMA-IR, and lipids. This metric is the benchmark used in this study. The oxidation metric “PC oxidation”, however, has a slightly better correlation with urine albumin concentration than the glucose metric (p=0.015 vs. p=0.027).

Dicarbonyls and oxidation predict CVD or DN with slightly lower p values than the PC glycation/glucose metric. The p values for PC oxidation are lower than PC glycation/glucose in predicting CVD in diabetic subjects (p=0.089 vs. p=0.596). In addition the p values for MG and glucosones are lower than PC glycation/glucose in predicting DN (p=0.128 for MG, p=0.112 for glucosones, and p=0.195 for PC glycation/glucose). The major findings in the other cohort was the subjects taking metformin had significantly lower MG, glucosones, and oxidations than the diabetic
subjects not on metformin. The PC glycation/glucose metric did not reflect this major
difference.

These findings corroborate the results of a recently published study (Rabbani,
Chittari et al.) that found that diabetic subjects had significantly (p< 0.01 for MG-H1,
p<0.05 for 3DG-H) lower levels of the MG-H1 and 3DG-H adducts in apoB100 if they
were taking metformin. This apoB100 study also found significantly lower (p<0.01)
levels of the methionine sulfoxide (M+16) adduct in apoB100 protein for diabetic
subjects on metformin compared to diabetic subjects not taking metformin. The study
presented in this dissertation and the apoB100 study are in close agreement in terms of
similar findings. This is despite the analysis of two different plasma proteins (HSA,
apoB100) and two different methodologies (peptide-specific quantification vs.
quantification of exhaustive digests to amino acids). Data management for this clinical
study involved tens of thousands of data points from MRM transitions and clinical data.
CHAPTER 7: FUTURE DIRECTIONS

7.1. Development of site-specific antibodies

One of the major outcomes of this work is to elucidate the differences between electrophile modification patterns in vitro and in vivo. Modification sites that were found in vitro were not necessarily found in vivo, likely because of differential recognition and degradation of modified proteins with different modification sites. Some MG sites, such as R410 on HSA, are modified as readily and are probably just as reactive in vivo as in vitro. However, this site is not detectable for in vivo modification, despite having the best ionization/fragmentation properties of all the MG-modified HSA peptides monitored. Therefore, it is likely preferentially recognized and eliminated over other modified HSA proteins that have altered sites (e.g. R257) that persist in the plasma. Without analyzing protein modifications that exist in vivo, the assumption that an antigen such as MG-modified HSA by in vitro methods is a representative of the modification pattern in vivo is wrong. Therefore, a site-based approach to generating antibodies is sought.

Of all of the dicarbonyl sites that demonstrated dose-dependent modification at HSA arginines in vitro, only a few were found reproducibly in vivo. R186, R257, and R428 were all found to be modified by dicarbonyls in vivo, however, R186 is the “goldilocks site” as all three dicarbonyl (MG, 3DG, and glucosone) modifications were found here. Generation of a plasma biomarker for dicarbonyls could be based on this site to raise antibodies that recognize dicarbonyl modifications here. An outline for the production of antibodies that recognize dicarbonyl modifications at R186 is as follows. The tryptic peptide that contains the internal R186 site would be modified with the
appropriate dicarbonyl and purified from the N-terminal and lysine adducted impurities. This purified modified peptide would be conjugated to an immunogenic protein such as KLH using carbodiimide chemistry. This protein conjugate would be injected into an animal, such as a rabbit, and the antisera titers that recognize the modified peptide over the unmodified peptide would be selected for validation. It is anticipated that site-specific antibodies that recognize individual AGE structures will greatly accelerate research into carbonyl and oxidative stress related mechanisms of toxicity. This will provide tools for both basic science and clinical assays for high-throughput analysis of the dicarbonyl proteome.

### 7.2. Metformin intervention study

To date, no study has performed an intervention of metformin to study the before and after effects of bisguanides on dicarbonyl content and oxidative stress. There have been several studies that have analyzed different groups of T2D +/- metformin (Beisswenger et al. 1999; Rabbani, Chittari et al. 2010; Engelen et al. 2011), but to our knowledge none have analyzed adducts/dicarbonyls before and after metformin treatment. This would be useful in determining the degree to which metformin (and other diabetic medication) is able to decrease levels of glyco-oxidation. This type of study would also be statistically more useful because of the elimination of covariation that can exist in analyzing two separate cohorts.

In addition to the clinical future studies, a more fundamental understanding of the anti-oxidant and anti-glycation properties of metformin is warranted. This could involve a
dose-dependent treatment of cells with metformin followed by measurement of gene expression changes. It is likely that metformin signals anti-oxidant response elements, perhaps mediated in part by NRF-2. These studies could also help pinpoint the effect of protein environment on dicarbonyl modification. In other words, does modification of plasma proteins take place in the plasma or does it occur in the cells that produce the proteins? Modification is likely to take place in both environments because cellular concentrations of dicarboxylics such as MG are higher than the plasma.

7.3. Characterization of other reactive dicarboxylics

It has been accepted that the majority of physiologic dicarboxylics are glyoxal, MG, and 3DG (Thornalley et al. 1999). This initial study attempted to identify major adducts of N-acetyl arginine as it was incubated for extended periods of time with glucose. However, not all of the modified arginines in this study were characterized. These three dicarboxylics also were studied in terms of formation from other biochemical pathways, and many researchers in the field have focused primarily on these three. Nearly a decade had passed before this experiment was repeated (Usui et al. 2007), and using 2,3-diaminonaphthalene as a trapping agent, many more abundant α-dicarboxylics were identified. These include tetroside, triosone, and 3-deoxytetrosone. Two years later, another study (Gobert et al. 2009) used o-phenylenediamine as a trapping agent and identified glucosone, 3-deoxypentosone, 1-deoxypentosone, 1-deoxytetrosone, and 1-deoxyglucosone under both aerobic and anerobic glycation conditions. Clearly, there are
many different dicarboxyls to be studied. We anticipate that R186 in HSA is a good place to start looking for these endogenous dicarboxyls.
APPENDIX A: SUBJECT CONSENT FORMS AND IRB APPROVAL

SUBJECT’S CONSENT FORM

Project Title: Population-Based Proteomic Investigation of Type 2 Diabetes

You are being asked to read the following material to ensure that you are informed of the nature of this research study and of how you will participate in it, if you consent to do so. Signing this form will indicate that you have been so informed and that you give your consent. Federal regulations require written informed consent prior to participation in this research study so that you can know the nature and risks of your participation and can decide to participate or not participate in a free and informed manner. If you choose not to participate, your refusal will involve no penalty or loss of benefits which you would normally experience.

PURPOSE

You are being invited to participate voluntarily in the above-titled research project at University Medical Center and University Physicians Healthcare (UPH) hospitals. The main purpose of this project is to discover new predictors that can accurately detect individuals at high risk for developing type 2 diabetes. You may be asked to participate in the additional study of the effect of diabetes on West Nile virus immunity and genetic markers.

SELECTION CRITERIA

The Principal Investigator or a member of his/her study staff will discuss the requirements for participation in this study with you. To be eligible to participate, you must be older than 12 years old. Approximately 300 individuals will be enrolled in this study. Participation is voluntary and you may withdraw from the study at any time.

PROCEDURE(S)

The following information describes your participation during sample collection which will last up to four and a half hours.

1. You will be asked to report to the research center at University Medical Center or the primary care and diabetes clinics at UPH-Kino Hospital in the morning after an 8-hour overnight fast.
2. Your body weight, height, and waist circumference will be measured.
3. Up to 50 ml of blood will be collected from your arm vein for measurement of blood sugar, cholesterol, proteins, blood cells and genetic markers.
4. A urine sample will be collected in a small container—on day of testing.
5. If you have never been diagnosed with Diabetes before you will be given 75 g of oral glucose (sugar) solution and a repeat blood test will be performed 2 hours later, or fatty meal that consists of a cup of heavy whipped cream and a repeat blood test will be performed 4 hours later. If you have lactose intolerance, we do not recommend drinking this meal as it may lead to gastrointestinal discomfort.
6. You may be asked to return to a follow up visit

Version Date: 05/25/2010
Would you like to be contacted for this study in the future?

If YES, please initial: ________________

Would you like to be contacted for future diabetes studies?

If YES, please initial: ________________

**RISKS**

When your blood is collected, you may experience minor pain, bruising or swelling of the skin where the needle is inserted. The amount of blood withdrawn, the overnight fasting for 8 hours and ingesting the oral glucose solution have no known serious health risks. There is a potential impact of a new diagnosis of diabetes on your psychological, social, employability, financial or insurability status, but this will be reduced by creating a safe link to your data that is only accessible to the study investigators.

The amount of blood that is withdrawn is less than ¼ of a typical blood donation.

**BENEFITS**

You will be given information on effective lifestyle interventions to prevent the development of diabetes, the metabolic syndrome and related health risks. If you are diagnosed with diabetes or assessed to be at high risk for diabetes you will be informed and we will encourage you to communicate this to your primary physician.

Identification of a useful marker obtained from a simple blood test or urine sample that might predict which people are at risk for developing diabetes will be extremely useful in preventing the development of related complications. The prevention and treatment of type 2 diabetes in large numbers of people could have immense benefits to public health and healthcare costs.

**CONFIDENTIALITY**

Representatives of regulatory agencies (including the University of Arizona Human Subjects Protection Program) may access your records to ensure quality of data and study conduct. Personal identifying of data will remain confidential. Samples will only be used to measure predictors of type 2 diabetes and study the effect of diabetes on blood cells. Results from genetic measurements will not be connected to your name or other identifying information. These or any other laboratory results will not be released to anyone outside of the investigative team. The files will be stored at a secure location in the Medical Research Building. Your identifying information linked to samples and data will only be accessed by the principal investigator Craig Stump and designated personnel approved by the IRB committee. We will destroy the paper records 10 years after the study completion.
Would you like to be contacted for this study in the future?

If YES, please initial: __________

Would you like to be contacted for future diabetes studies?

If YES, please initial: __________

RISKS

When your blood is collected, you may experience minor pain, bruising or swelling of the skin where the needle is inserted. The amount of blood withdrawn, the overnight fasting for 8 hours and ingesting the oral glucose solution have no known serious health risks. There is a potential impact of a new diagnosis of diabetes on your psychological, social, employability, financial or insurability status, but this will be reduced by creating a safe link to your data that is only accessible to the study investigators.

The amount of blood that is withdrawn is less than ¼ of a typical blood donation

BENEFITS

You will be given information on effective lifestyle interventions to prevent the development of diabetes, the metabolic syndrome and related health risks. If you are diagnosed with diabetes or assessed to be at high risk for diabetes you will be informed and we will encourage you to communicate this to your primary physician.

Identification of a useful marker obtained from a simple blood test or urine sample that might predict which people are at risk for developing diabetes will be extremely useful in preventing the development of related complications. The prevention and treatment of type 2 diabetes in large numbers of people could have immense benefits to public health and healthcare costs.

CONFIDENTIALITY

Representatives of regulatory agencies (including the University of Arizona Human Subjects Protection Program) may access your records to ensure quality of data and study conduct. Personal identifying of data will remain confidential. Samples will only be used to measure predictors of type 2 diabetes and study the effect of diabetes on blood cells. Results from genetic measurements will not be connected to your name or other identifying information. These or any other laboratory results will not be released to anyone outside of the investigative team. The files will be stored at a secure location in the Medical Research Building. Your identifying information linked to samples and data will only be accessed by the principal investigator Craig Stump and designated personnel approved by the IRB committee. We will destroy the paper records 10 years after the study completion.
PARTICIPATION COSTS AND SUBJECT COMPENSATION

Participation in this study is cost free. You will be provided $20 as compensation for your time.

CONTACTS

You can obtain further information about the research or voice concerns or complaints about the research by calling the Principal Investigator Craig Stump, M.D., Ph.D at 520-626-4317. If you have questions concerning your rights as a research participant, have questions, complaints, or concerns about the research and cannot reach the Principal Investigator, or want to talk to someone other than the Investigator, you may call the University of Arizona Human Subjects Protection Program office at (520) 626-3515. If you would like to contact the Human Subjects Protection Program via the web, please visit the following website http://ocrp.vpr.arizona.edu/irb/contact.

LIABILITY

Side effects or harm are possible in any research program despite the use of high standards of care and could occur through no fault of yours or the investigator involved. Known side effects have been described in this consent form. However, unforeseeable harm also may occur and require care. You do not give up any of your legal rights by signing this form. In the event that you require or are billed for medical care that you feel has been caused by the research, you should contact the principal Craig Stump, M.D., Ph.D at (520) 626-4317.

AUTHORIZATION

Before giving my consent by signing this form, the methods, inconveniences, risks, and benefits have been explained to me and my questions have been answered. I may ask questions at any time and I am free to withdraw from the project at any time without causing bad feelings or affecting my medical care. My participation in this project may be ended by the investigator or by the sponsor for reasons that would be explained. New information developed during the course of this study which may affect my willingness to continue in this research project will be given to me as it becomes available. This consent form will be filed in an area designated by the Human Subjects Protection Program with access restricted by the principal investigator Craig Stump, M.D., Ph.D at 520-626-4317, or authorized representative of the Endocrinology Department. I do not give up any of my legal rights by signing this form. A copy of this signed consent form will be given to me.

____________________________  ____________________
Subject’s Signature            Date

____________________________
Subject’s printed name

Version Date: 05/25/2010      Page 3 of 4
Parent/Legal Guardian (if necessary)  Date

Witness (if necessary)  Date

INVESTIGATOR'S AFFIDAVIT:
Either I have or my agent has carefully explained to the subject the nature of the above project, hereby certify that to the best of my knowledge the person who signed this consent form was informed of the nature, demands, benefits, and risks involved in his/her participation.

Signature of Representative  Date
**HSPP Correspondence Form**

**Date:** 06/23/09  
**Investigator:** Craig Stump, M.D., Ph.D.  
**Department:** Med  
**Project No./Title:** 07-0812-01 Proteomic and Metabolomic Biomarker Investigation of Type 2 Diabetes  
**Current Period of Approval:** 10/14/08 - 10/08/09  

### IRB Committee Information

- IRB1 - IRB00000291
- IRB2 - IRB00001751
- IRB3 - IRB00003012
- IRB4 - IRB00005448
- Administrative Action

**FWA Number:** FWA00004218

### Nature of Submission

- New Project  
- Amendment  
- Unanticipated Problem Involving Risks to Subjects or Others  
- Response to IRB Committee  
- Other (define):  
- Continuing Review  
- Protocol Deviation/Violation/Waiver  
- Non-Compliance  
- Not Applicable

### Documents Reviewed Concurrently

- Request for Amendment Form – VOTF (dated 06/22/09)  
- VOTF (received 06/22/09)

### Description of Submission

*Personnel changes [adding Armstrong, Bharara, Sen, Ingram; removing Peterson].*

### Committee/Chair Determination

- Approved as submitted

### Additional Determination(s)

- Not Applicable

---

*Wendy R. Tate, PSM*

Wendy R. Tate, PSM, CIP  
Chair Designee, IRB3 Committee  
UA Institutional Review Board  
WRT/ap

---

*Reminder:* Continuing Review materials should be submitted 30-45 days prior to the expiration date to obtain project re-approval.

- Projects may be concluded or withdrawn at any time using the forms available at [www.irb.arizona.edu](http://www.irb.arizona.edu).
- No changes to a project may be made prior to IRB approval except to eliminate apparent immediate hazard to subjects.
- Original signed consent forms must be stored in the designated departmental location determined by the Department Head.
REFERENCES


Products Precedes and Correlates with Early Manifestations of Renal and Retinal Disease in Diabetes." Diabetes 44(7): 824-829.


Schuchard, M. "Comparison of Precipitation Methods Following Depletion of Twenty High Abundance Proteins from Human Plasma."


