4-2010

# Characterization of Glycation Sites on Human Serum Albumin using Mass Spectrometry

Omar S. Barnaby
*University of Nebraska-Lincoln*, obarnaby@huskers.unl.edu

# CHARACTERIZATION OF GLYCATION SITES ON HUMAN SERUM ALBUMIN USING MASS SPECTROMETRY

by

Omar St. Aubyn Barnaby

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor David S. Hage

Lincoln, Nebraska

April, 2010

# CHARACTERIZATION OF GLYCATION SITES ON HUMAN SERUM ALBUMIN USING MASS SPECTROMETRY

Omar S. Barnaby, Ph.D.

University of Nebraska, 2010

Advisor: David S. Hage

The modification of proteins by reducing sugars is a process that occurs naturally in the body. This process, which is known as glycation, has been linked to many of the chronic complications encountered during diabetes. Glycation has also been linked to changes in the binding of human serum albumin (HSA) to several drugs and small solutes in the body. While these effects are known, there is little information that explains why these changes in binding occur. The goal of this project was to obtain qualitative and quantitative information about glycation that occurs on HSA.

The first section of this dissertation examined methods that could be used to quantify and identify glycation that occurs on HSA. The extent of glycation that occurred on HSA was quantified using $^{18}$O-labeling mass spectrometry and the glycation sites were identified by observing the mass-to-charge (m/z) shifts that occurred in glycated HSA. This initial investigation revealed that $^{18}$O-labeling based quantitation can be improved over previous methods if a relative comparison is done with $^{18}$O-labeled peptides in a control HSA sample. Similarly, the process of making m/z shift-based

assignments could be improved if only the peptides that were unique to the glycated HSA samples were used with internal calibration.

These techniques were used in subsequent chapters for the assignment of early and late-stage glycation products on HSA. The regions on HSA that contained the highest amount of modification were identified, quantified, and ranked in order of their relative abundance. Of the commonly reported glycation sites, the *N*-terminus was found to have the highest extent of modification, followed by lysines 525, 199, and 439. The relative amount of modification on lysine 281, with respect to the aforementioned residues, varied with different degrees of glycation. The $^{18}O$ approach used for this analysis was novel because it allowed for the simultaneous quantification of all glycation-related modifications that were occurring on HSA. As such, several arginine residues were also found to have high amounts of modification on glycated HSA.

# Dedication

The work in this dissertation is dedicated to my mother, Paulette, my late-father, Christopher; and to my brothers, Tion, Olu, and Christopher. The love and support that they have given to me cannot be described in words. I love them dearly and appreciate the many sacrifices that they have made on my behalf. I also dedicate this work to the younger ones in my family. Let this be a constant reminder that hard work and dedication in anything that you pursue does not go unnoticed. Strive for the best in whatever you do and success will surely follow.

# Acknowledgements

I would like to thank my faculty advisor, Professor David S. Hage for his support and guidance during my time at UNL. His organization, drive for developing novel ideas, and approach to high quality research has always been an inspiration for me, and his willingness to support my ideas played a large role in my development as a scientist. I would also like to thank the professors my supervisory committee, Jody Redepenning, Anu Subramanian, Barry Cheung, and the late Adrian George. Their guidance played a major role in the development of my critical thinking skills and my ability to see the "big picture" when it comes to scientific research.

I would like to thank the researchers at the Nebraska Center for Mass Spectrometry, Ron Cerny and Kurt Wulser, for their support on several topics related to my graduate work. They were instrumental in teaching me the basics of mass spectrometry and providing advice in solving difficult problems that were related to my research. I am also grateful for my many colleagues in the Hage lab. It was a pleasure working with Corey, John, Rangan, Jeanethe, Ankit, Krina, David, Chad, Mia, Tong, Michelle, Abby, and Erica. I value the time that we spent toiling over classes and research, having random lunch time conversations on just about any topic, and having well needed comedic relief at times. I wish you all the best of success in graduate school and in any path that you choose in life.

I would also like to extend special thanks to additional family and friends. It was a pleasure collaborating with two of my friends, Ala Qadi and Ufuk Nalbantoglu, on the development of programs that were used for my research. The development of these tools

allowed me to process data at a rapid rate, and certainly allowed for the early completion of several projects. Special thanks are also in order for my many uncles and aunts, Paul, Annette, Pam, Janet, Madge, Vaughn, Val, Carla, Ruby, Brenton, Doreen, and Babeth; my godmother, Aprylle, and my grandmother, Sylvia. There is no doubt in my mind of their continued support over the years. I am grateful and love them dearly. There is simply not enough room to list the number of people that have not only supported me over the years but have made practical contributions to my success. To my family and friends, your support over the years has made this process an enjoyable one. Thank you!

# Table of Contents

## Chapter 1: General Introduction



## Chapter 2: Quantitative Analysis of Glycation Sites on Commercially Glycated Human Serum Albumin

## Chapter 3: Improved Techniques for Identifying Early and Advanced Glycation Adducts using Mass Spectrometry



## Chapter 4: Characterization of Glycation Sites on *In Vitro* Glycated Human Serum Albumin using Mass Spectrometry

## Chapter 7: Conclusion

# CHAPTER 1

# GENERAL INTRODUCTION

## 1.A. EFFECTS OF DIABETES AND PROTEIN GLYCATION

**1.A.1. Classification and Effects of Diabetes *in vivo*.** Diabetes is a disease state which is characterized by abnormally high blood sugar concentrations. The American Diabetes Association estimates that there are at least 24 million Americans that are currently living with diabetes (1). Of these 24 million people, older populations (i.e., above 60 years of age) and individuals of ethnic origin represent a large percent of the diabetic population. In healthy individuals, the blood glucose concentration is tightly regulated by insulin (2). After food ingestion, glucose is transported to the β-cells of the pancreas, initiating a reaction cascade that causes insulin to be released into the bloodstream. Insulin is subsequently able to interact with fat, liver, and muscle tissue to facilitate the storage of blood glucose as glycogen and triacylglycerides for later use. The high blood sugar concentration that is found during diabetes is related to either insufficient insulin production (i.e., type 1 diabetes) or resistance to insulin (i.e., type 2 diabetes) (3, 4).

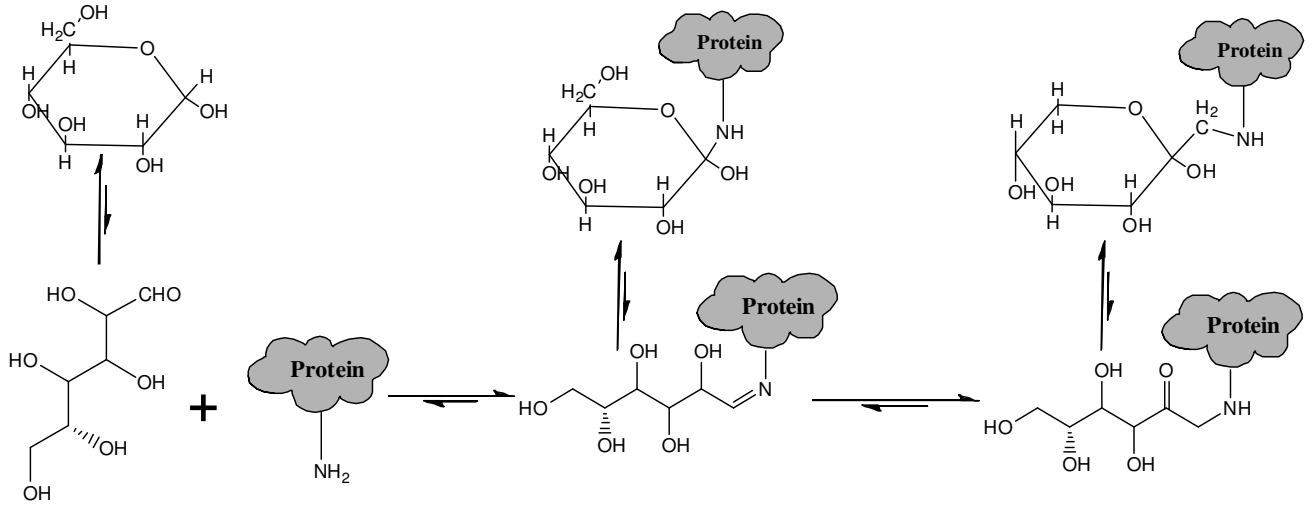
This increased blood glucose concentration has a number of effects in the body, which include an increased risk of heart disease (5), stroke (6), kidney disease (7), blindness (8), and amputations (9). Many of these complications are due to protein glycation and the formation of advanced glycation endproducts (AGEs) from blood

sugars (10). AGE-related complications in the body can be classified into two different areas. In one case, the interaction of AGEs with the receptor protein for AGEs (RAGE) is responsible for a wide range of inflammatory responses which eventually lead to tissue damage (11, 12). The second case involves direct modification of proteins by AGEs, leading to a loss of protein function (13) or tissue damage that results from protein crosslinking (14, 15).

**1.A.2. Glycation and Advanced Glycation End Products.** Glycation is a process that involves the non-enzymatic addition of reducing sugars to proteins. This reaction was first identified by Louis-Camille Maillard in 1911, who noticed a characteristic yellowing that occurred when a solution containing glycine was incubated with D-glucose (16). It was later determined that this characteristic coloring was due to complex reactions that eventually led to the formation of two different classes of compounds, reductones (i.e., dehydrated forms of sugars that are condensed on proteins) and melanoidins (i.e., nitrogen containing polymers initiated by late-stage glycation) (17). This process, which is now known as the Maillard reaction, was further organized into different stages by John E. Hodge in 1953 (18). The initial stage of glycation (i.e., early-stage glycation) involves the addition of reducing sugars to free primary amine groups that are present on a protein (see Figure 1.1) (19, 20). The predominant amino acids involved in glycation are therefore lysine residues and the *N*-terminus of a given protein. This initial step of this reaction results in the formation of a Schiff base. This reaction occurs fairly rapidly with the open chain form of a reducing sugar, where the equilibrium that is established favors the formation of a Schiff base under physiological conditions (21). The Schiff base, however, is unstable and subsequently rearranges over the course of about two

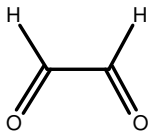
Figure 1.1 Diagram showing the initial steps of glycation. Reducing sugars condense on the primary amine groups (i.e., as present on lysine or the *N*-terminus) of proteins. This results in the formation of a Schiff base, which subsequently rearranges to form the more stable Amadori product. The α-oxaloaldehydes that result from prolonged glucose incubation are also shown.

## Early Stages of Glycation

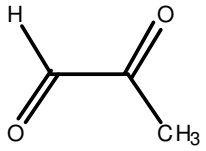



**D-Glucose** **Primary Amine** **Schiff Base** **Amadori Product (i.e., Fructosyl-lysine)**

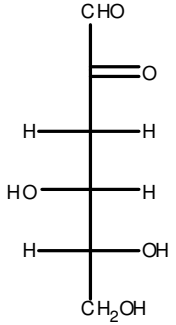
## α-Oxaloaldehydes



**Glyoxal**



**Methylglyoxal**



**3-Deoxyglucosone**

days to form a more stable class of compounds known as Amadori products (17, 22, 23).

While early glycation reactions in its own right is able to modify the function of many proteins, a more complex scenario unfolds when a protein is incubated with reducing sugars over extensive periods of time. This represents the intermediate-stage of the Maillard reaction, which is characterized by the formation of reactive dicarbonyl compounds (i.e., α-oxaloaldehydes) from either free sugars in solution (24-26). The rate of formation of these compounds varies with reaction conditions such as the temperature, pH, and the presence of buffers in the solution (10). While several α-oxaloaldehydes are involved in protein damage, the most significant and commonly cited compounds that result in protein damage *in vivo* are glyoxal, methylglyoxal, and 3-deoxyglucosone (see Figure 1.1) (27, 27, 28). Previous studies which looked at the formation of α-oxaloaldehydes during incubations of free glucose (Note: no proteins were added) over a three week period showed that the glyoxal was present at the highest concentrations, followed by 3-deoxyglucosone and methylglyoxal (10). These compounds are highly reactive and are able to bind with lysine, arginine, or the *N*-terminal regions of proteins, forming AGEs.

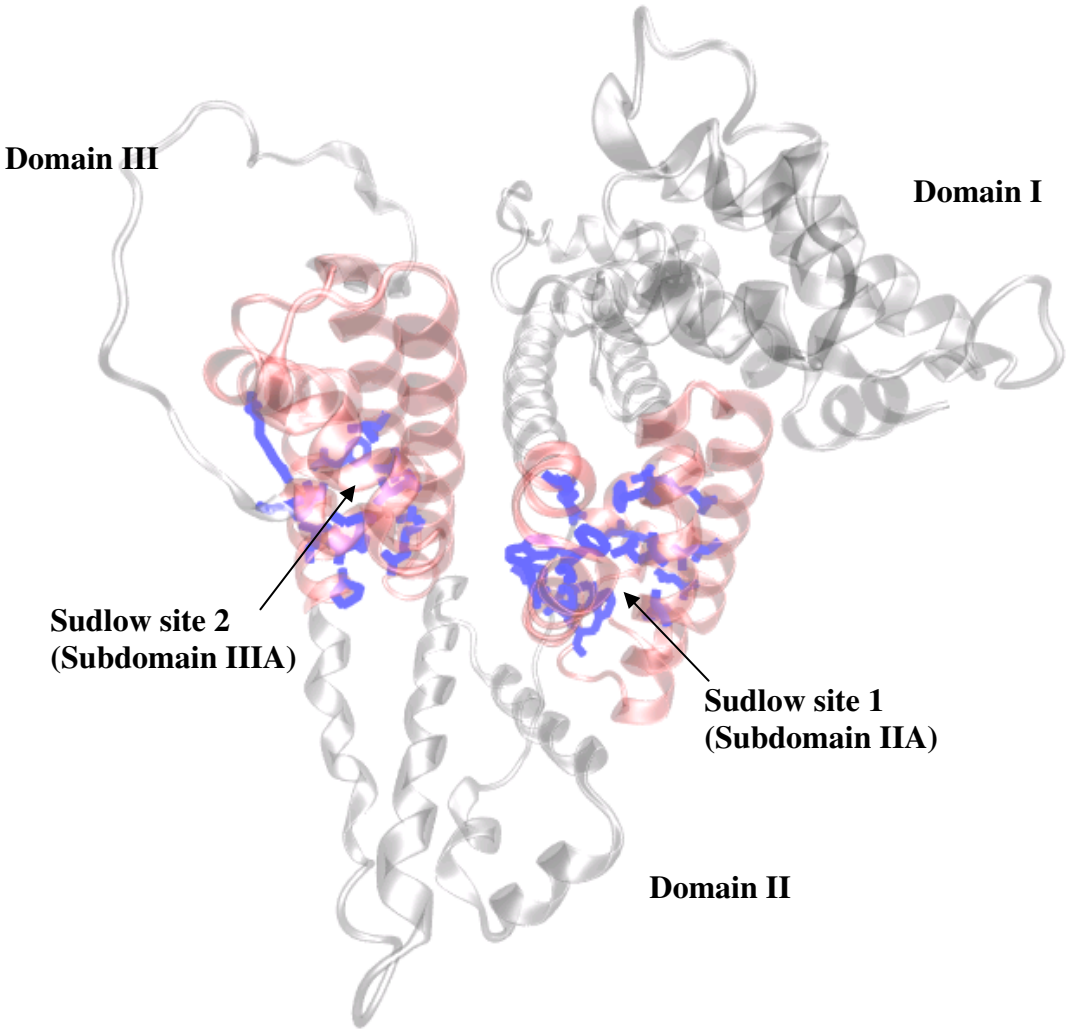
The study of AGEs is important because the loss of protein function and tissue damage that can occur *in vivo* is directly related to the formation of AGEs. This effect is fairly well studied for long lived proteins such as collagen (29-32) and lens crystallin (8, 14, 15), as well as some short lived proteins such as hemoglobin (33-35). However, there is little information that is available for the major serum transport protein, human serum albumin (HSA) (21). The research presented in this dissertation focuses on products that

result from the interaction of HSA with these reactive dicarbonyl compounds and D-glucose.

**1.A.3. Early and Late-Stage Glycation of HSA.** HSA is the most abundant protein in human serum, with concentrations that range anywhere from 30 – 50 g/L, representing approximately 50% of the total serum protein concentration (36, 37). As such, HSA exhibits large effects on a number of physiological processes, including the regulation of oncotic pressure (38) and blood pH (39), as well as the transport of fatty acids (40), drugs (41), and small solutes (42) in the body. HSA has 585 amino acids (43), which includes 59 lysines and 24 arginines that could potentially be modified to form glycation products. A crystal structure of HSA that shows the major drug binding sites on HSA is given in Figure 1.2. From this crystal structure, it can be seen that HSA has high helical content (43). There are also a number of drug interaction sites located in subdomains IIA (Sudlow site 1) and IIIA (Sudlow site 2) (44).

There are several reports suggesting that the glycation of HSA that occurs during diabetes may affect its drug-binding properties *in vivo* (41, 42, 45, 46). Additionally, the clearance of albumin from serum has been shown to be accelerated by approximately 2 fold in diabetes (47), which is thought to be due to interaction with RAGE receptors in the liver and kidney. While these effects are known to occur, the specific regions of HSA which are being modified by early or late-stage glycation products, which leads to these effects, have not yet been fully elucidated.

Figure 1.2 Crystal structure of HSA (48) that was developed using visual molecular dynamics (49). The majority of the structure of HSA (approximately 67%) is helical. Subdomains containing Sudlow sites 1 and 2, are shown in red and the amino acids that play a significant role in drug interactions at these sites are given in blue. The three main domains of HSA are also shown.


Domain III
Domain I
Sudlow site 2
(Subdomain IIIA)
Sudlow site 1
(Subdomain IIA)
Domain II

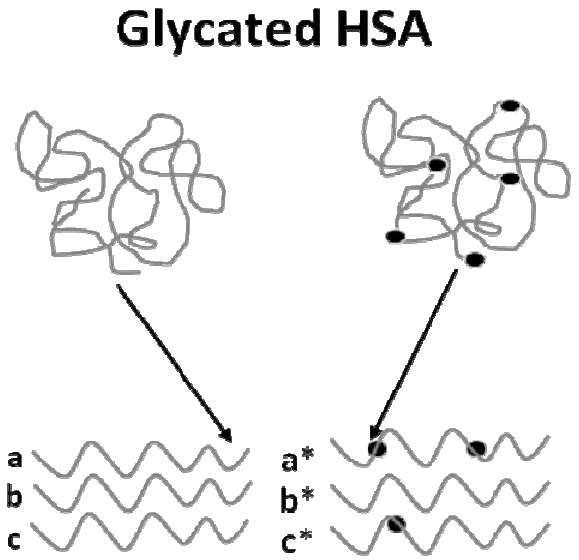
## 1.B. APPROACHES FOR THE ANALYSIS OF GLYCATED HSA

**1.B.1. Prior Attempts at Determining HSA Glycation.** Past studies of HSA glycation have involved total glycation analysis using spectroscopic techniques (50, 51) or the measurement of specific-glycation induced properties such as fluorescence (30) and interaction with affinity supports (52-54). The information obtained using these methods, however, provided no structural information. Studies using isotopically-labeled reducing sugars have also been used to quantify glycation sites and identify the location of these glycation sites (55, 56). These methods, however, have not given information about AGE-formation that occurs on HSA.

In an attempt to clarify the regions on HSA that are being modified during glycation, mass spectrometry offers the potential advantage of providing both structural and quantitative information. Mass spectrometry has an additional advantage, of being capable of global identification and quantification of modifications, when $^{18}$O-labeling is employed (57). The majority of reports describing the identification of modification sites on HSA have used peptide mapping for glycation product identification (19, 58-61). In peptide mapping, a protein is digested followed by mass spectrometry analysis, where shifts in the mass-to-charge (m/z) ratio can be matched to glycated peptides (see Figure 1.3). Typically, complex modification pattern can occur on HSA as a result of to glycation (see Table 1.1). This complexity, combined with the possibility that multiple lysine or arginine residues are present within a given peptide, often makes it difficult to pinpoint the exact location of modified residues on HSA.

Figure 1.3. Diagram showing the procedure that was used in this dissertation to identify glycation-related modification sites on HSA. HSA was first digested and the glycation sites were assigned based on the detected m/z shifts. The m/z shift depended on the mass of the glycation product and the number of glycation products that were present on any given peptide.

**Step 1**: Digestion of glycated HSA containing a mixture of modified and non-modified protein.




**Step 2**: Mass spectra are obtained, where the glycation sites are identified based on m/z shifts.

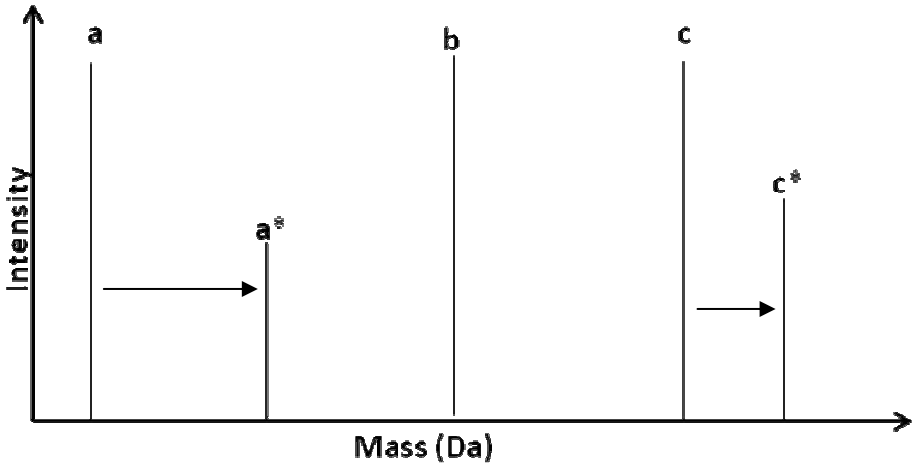


Table 1.1. Glycation products involving lysine and arginine residues

| **Glycation adducts involving lysine** | **Abbreviation** | **ΔM (Da)** |
|---|---|---|
| Fructosyl-lysine | FL | 162.0528 |
| Fructosyl-lysine-1$H_2O$ | FL-1$H_2O$ | 144.0423 |
| Fructosyl-lysine-2$H_2O$ | FL-2$H_2O$ | 126.0317 |
| $N_\varepsilon$-Carboxyethyl-lysine | CEL | 72.0211 |
| $N_\varepsilon$-Carboxymethyl-lysine | CML | 58.0055 |
| Pyrraline | Pyr | 108.0211 |

| **Glycation adducts involving arginine** | **Abbreviation** | **ΔM (Da)** |
|---|---|---|
| $N_\varepsilon$-[5-(2,3,4-Trihydroxybutyl)-5-hydro-4-imidazolon-2-yl]ornithine | 3-DG-H1 | 144.0423 |
| Tetrahydropyrimidine | THP | 144.0423 |
| Imidazolone B | IB | 142.0266 |
| Argpyrimidine | ArgP | 80.0262 |
| $N_\varepsilon$-(5-Hydro-5-methyl-4-imidazolon-2-yl)ornithine | MG-H1 | 54.0106 |
| $N_\varepsilon$-(5-Hydro-4-imidazolon-2-yl)ornithine | G-H1 | 39.9949 |

| **Glycation adducts involving lysine or arginine** | **Abbreviation** | **ΔM (Da)** |
|---|---|---|
| 1-Alkyl-2-formyl-3,4-glycosyl-pyrrole | AFGP | 270.0740 |

In recent publications, some authors have tried to circumvent this issue by using tandem mass spectrometry and collision-induced dissociation (CID). While this method can provide useful information, a limitation of this method is that the Amadori Product cannot be detected because water loss on the condensed sugar is favored over the cleavage of amide bonds during CID (19, 62). A common method for assigning the most likely regions of glycation, using mass spectrometry is to use known information about peptide chemistry such as the p$K_a$, fractional accessible surface area (FAS) (63, 64), and the location of certain residues that enhance glycation (56), to identify the most likely regions where modification could occur. The research presented in this dissertation uses this information as well as quantitative $^{18}$O-labeling data to identify which residues on HSA are most prone to glycation related modifications.

**1.B.2. Analysis of Glycated HSA using MALDI-TOF MS and $^{18}$O-Labeling.** Matrix-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) was the method used for obtaining the mass spectra that were used in this dissertation. For this analysis, a positive-ionization mode combined with delayed extraction and the use of a reflector were employed to enhance the mass resolution and signal quality that was obtained for a given spectra. The first step under this approach, involved the envelopment of a mixture of peptides in an acidic matrix that absorbs ultraviolet (UV) energy. The resulting sample was then crystallized on a MALDI plate. A pulsed UV laser was used to irradiate the sample, where the energy absorbed was used to both vaporize the protein/matrix mixture, and facilitate ionization of the peptides (65, 66). An acceleration voltage was then applied to the ionized sample after a preset delay time, causing these ions to drift to the detector at a velocity dependent on the mass of the peptide (67). The

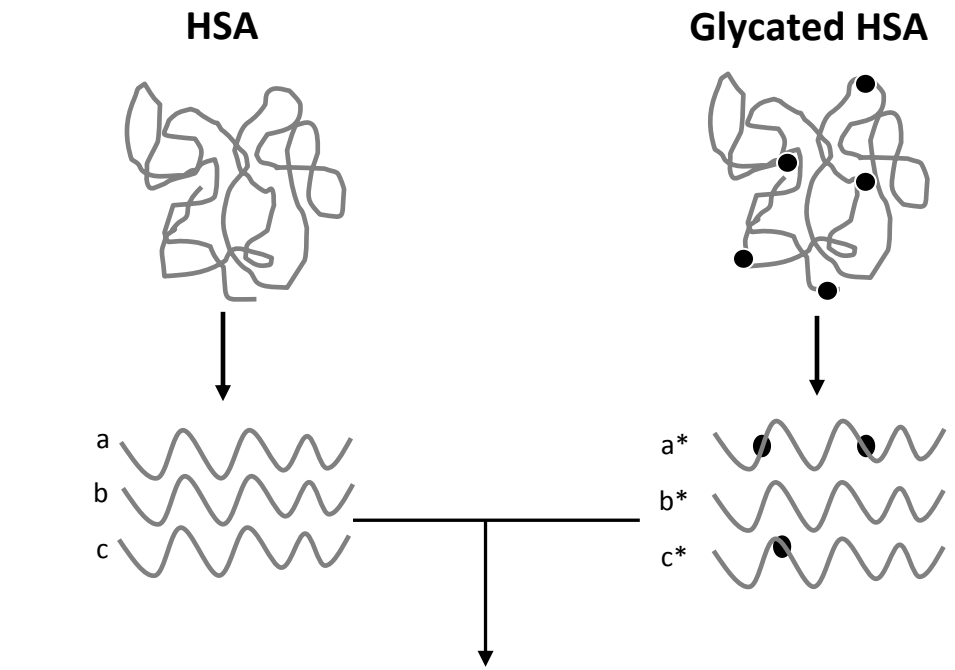
m/z of a peptide was then be determined from the time it took the peptide to reach the detector. The advantages of using MALDI-TOF MS for this analysis were that spectral interpretation was simple because of the low abundance of multiply charged species. MALDI-TOF MS also uses a soft ionization technique (66). This technique was also good for use with $^{18}$O-labeling, because interaction of $^{18}$O-labeled peptides with $^{16}$O was reduced in comparison to techniques based on electrospray ionization (68).

The $^{18}$O-labeling approach that was used for quantifying modification sites in this dissertation is shown in Figure 1.4. Under this scheme, HSA and glycated HSA were digested in $^{16}$O and $^{18}$O-labeled water, respectively. These samples were then mixed in a 1:1 (v/v) ratio, and the $^{16}$O/$^{18}$O ratios were obtained. Under this scheme, peptides that contained a large amount of glycation should exhibit large $^{16}$O/$^{18}$O ratios. In this technique, all of the peptides that result from an enzymatic digest can be labeled using in a single step (68, 69). This is an important feature for the quantification of modification sites, because of the sheer number of samples that were examined in this report.

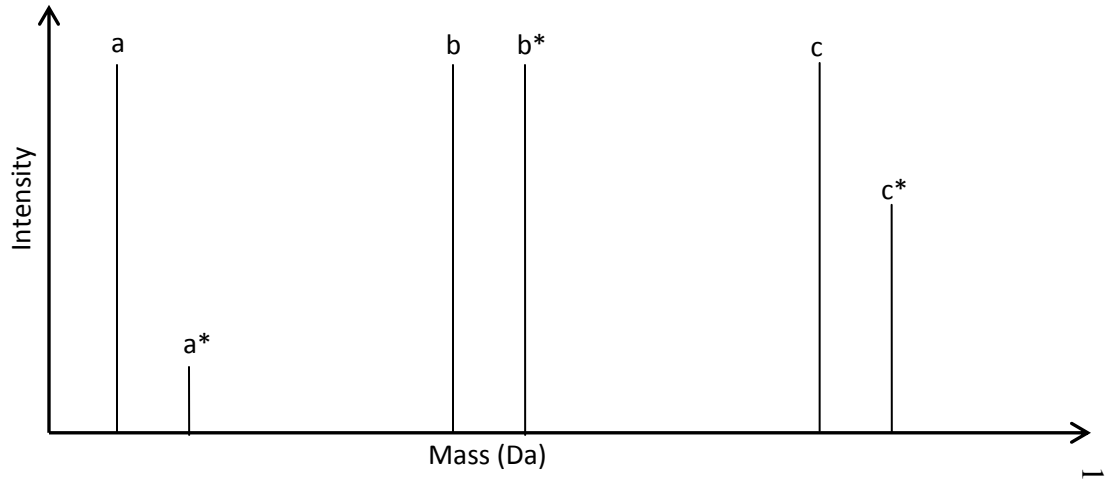
**1.B.3. Topics Covered in this Dissertation.** The first section of this dissertation examined methods that could be used for glycation site identification and quantification. More specifically, Chapter 2 of this dissertation described the use of $^{18}$O-labeling for quantifying the relative extent of modification that occurred on specific regions of HSA. The quantitative information obtained using this approach was compared to previous data that was obtained for the same type of HSA (63). Chapter 3 examined techniques that were used to improve the assignment of modification sites on HSA. This was done by

Figure 1.4. General procedure for obtaining quantitative estimates of glycation based on $^{18}O$-labeling MALDI-TOF MS. The non-modified peptides from HSA are represented by a-c, while a*-c* represent modified peptides from the same regions of glycated HSA.

**Step 1:** HSA is digested in water and glycated HSA is digested in $^{18}O$-labeled water




**Step 2**: The HSA and glycated HSA digests are mixed in a 1:1 ratio; the $^{16}O/^{18}O$ ratios of the detected peptides are determined by mass spectrometry



performing HSA glycation *in vitro* and examining the effects of glycation-product heterogeneity, mass accuracy, and the level of selectivity on the identification of glycated peptides. In Chapters 4 and 5, the location and type of modifications that occurred on glycated HSA were examined. Chapter 4 focused on the identification of early-stage glycation products and Chapter 5 focused on the identification of AGEs in glycated HSA. The HSA that was used in both cases was glycated *in vitro*. The glycation products present at three stages of glycation were examined using the techniques that were optimized in Chapter 3.

Chapter 6 of the dissertation examined the use of $^{18}$O-labeling mass spectrometry for the quantification of glycation sites on HSA. Several improvements to the methods described in Chapter 2 were used here, such as the use of control HSA to obtain a set of reference $^{16}$O/$^{18}$O ratios. These reference ratios were directly compared to ratios that were obtained for the control and glycated HSA samples. Comparisons were made with the data obtained using commercially glycated HSA. The quantitative results were used to determine the most likely modification sites on glycated HSA from the data presented in Chapters 4 and 5.

## 1.C. REFERENCES

(1) American Diabetes Association. All about diabetes. http://www.diabetes.org/about-diabetes.jsp (accessed 9/2006, 2009).

(2) Nelson, D. L.; Cox, M. M., Eds.; In *Lehninger Principles of Biochemistry.* W. H. Freeman and Company: New York, 2005, pp 1119.

(3) Bennet, S. T.; Todd, J. A. Human type 1 diabetes and the insulin gene: Principles of mapping polygenes. *Annu. Rev. Genet.* **1996***, 3*0, 343-370.

(4) Turner, R. C.; Cull, C. A.; Frighi, V.; Holman, R. R. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus. *J. Am. Med. Assoc.* **1999***, 2*81, 2005-2012.

(5) Hartog, J. W. L.; Voors, A. A.; Bakker, S. J. L.; Smit, A. J.; Veldhuisen, D. J. V. Advanced glycation end-products (AGEs) and heart failure: Pathophysiology and clinical implications. *Eur. J. Heart Fail.* **2007***,* 9, 1146-1155.

(6) Olijhoeka, J. K.; Graafb, Y. V. D.; Bangaa, J. D.; Algrab, A.; Rabelinka, T. J.; Visserena, F. L. J. The metabolic syndrome is associated with advanced vascular damage in patients with coronary heart disease, stroke, peripheral arterial disease or abdominal aortic aneurysm. *Eur. Heart J.* **2004***, 2*5, 342-348.

(7) Perneger, T. V.; Brancati, F. L.; Whelton, P. K.; Klag, M. J. End-stage renal disease attributable to diabetes mellitus. *Ann. Intern. Med.* **1994***, 12*1, 912-918.

(8) Turk, Z.; Misur, I.; Turk, N. Temporal association between lens protein glycation and cataract development in diabetic rats. *Acta Diabetol.* **1997***, 3*4, 49-54.

(9) Bild, D. E.; Selby, J. V.; Sinnock, P.; Browner, W. S.; Braveman, P.; Showstack, J. A. Lower-extremity amputation in people with diabetes. Epidemiology and prevention. *Diabetes Care.* **1989***, 1*2, 24-31.

(10) Thornalley, P. J.; Langborg, A.; Minhas, H. S. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **1999***, 34*4, 109-116.

(11) Gebhardt, C.; Riehl, A.; Durchdewald, M.; Nemeth, J.; Furstenberger, G.; Muller-Decker, K.; Enk, A.; Arnold, B. RAGE signaling sustains inflammation and promotes tumor development. *J. Exp. Med.* **2008***, 20*5, 275-285.

(12) Stern, D. M.; Yan, S. D.; Yan, S. F.; Schmidt, A. M. Receptor for advanced glycation endproducts (RAGE) and the complications of diabetes. *Ageing Res. Rev.* **2002***,* 1, 1-15.

(13) Brownlee, M. Negative consequences of glycation. *Metab. Clin. Exp.* **2000***, 4*9, 9-13.

(14) Jansirani; Anathanaryanan, P. H. A comparitive study of lens protein glycation in various forms of cataracht. *Indian J. Clin. Biochem.* **2004***, 1*9, 110-112.

(15)Lewis, B. S.; Harding, J. J. The effects of aminoguanidine on the glycation (non-enzymatic glycosylation) of lens proteins. *Exp. Eye Res.* **1990***, 5*0, 463-467.

(16) Maillard, L. C. Action des acides amines sur les sucres; formation des melanoides par voie methodique. *C. R. Acad. Sci.* **1912***, 15*4, 66-68.

(17) Nursten, H. In *The Maillard Reaction;* Royal Society of Chemistry: Cambridge, UK, 2005, pp 214.

(18) Hodge, J. E. Chemistry of browning reactions in model systems. *J. Agric. Food Chem.* **1953***,* 1, 928-943.

(19) Lapolla, A.; Fedele, D.; Reitano, R.; Bonfante, L.; Guizzo, M.; Seraglia, R.; Tubaro, M.; Traldi, P. Mass spectrometric study of in vivo production of advanced glycation end-products/peptides. *J. Mass. Spectrom.* **2005***, 4*0, 969-972.

(20) Lapolla, A.; Fedele, D.; Seraglia, R.; Traldi, P. The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes: An update. *Mass. Spectrom. Rev.* **2006***, 25*, 775.

(21) Zhang, Q.; Ames, J. M.; Smith, R. D.; Baynes, J. W.; Metz, T. O. A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: Probing the pathogenesis of chronic disease. *J. Proteome Res.* **2009***,* 8, 754-769.

(22) Hayashi, T.; Mase, S.; Namiki, M. Formation of three-carbon sugar fragment at an early stage of the browning reaction of sugar with amines or amino acids. *Agric. Biol. Chem.* **1986***, 5*0, 1959-1964.

(23) Monnier, V. M. Nonenzymatic glycosylation, the Maillard reaction and the aging process. *J. Gerontol.* **1990***, 45*, B105-B111.

(24) Stitt, A. W.; Curtis, T. M. Advanced glycation and retinal pathology during diabetes. *Pharmacol. Rep.* **2005***, 5*7, 156-168.

(25) Ahmed, N.; Argirov, O. K.; Minhas, H. S.; Cordeiro, C. A. A.; Thornalley, P. J. Assay of advanced glycation endproducts (AGEs): Surveying AGEs by chromatographic assay with derivitization by 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and application to $N_{\varepsilon}$-carboxymethyl-lysine- and $N_{\varepsilon}$-(1-corboxyethyl)lysine-modified albumin. *Biochem. J.* **2002***, 36*4, 1-14.

(26) Thornalley, P. J. Dicarbonyl intermediates in the Maillard reaction. *Ann. N. Y. Acad. Sci.* **2005***, 104*3, 111-117.

(27) Takeuchi, M.; Kikuchi, S.; Sasaki, N.; Suzuki, T.; Watai, T.; Iwaki, M.; Bucala, R. l. Y., S. Involvement of advanced glycation end-products (AGEs) in Alzheimer's disease. *Curr. Alzheimer Res.* **2004***,* 1, 39-46.

(28) Ahmed, N.; Thornalley, P. J. Chromatographic assay of glycation adducts in human serum albumin glycated in vitro by derivitization with 6-aminoquinlyl-N-hydroxysuccinimidyl-carbamate and intrinsic fluorescence. *Biochem. J.* **2002***, 36*4, 15-24.

(29) Saito, M.; Marumo, K. Collagen cross-links as a determinant of bone quality: A possible explanation for bone fragility in aging, osteoporosis, and diabetes mellitus. *Osteoporos. Int.* **2010***, 2*1, 195-214.

(30) Sell, D. R.; Nemet, I.; Monnier, V. M. Partial characterization of the molecular nature of collagen-linked flourescence: Role of diabetes and end-stage renal disease. *Arch. Biochem. Biophys.* **2010**, *49*3, 192-206.

(31) Bank, R. A.; Baylis, M. T.; Lafeber, F. P.; Maroudas, A.; Tekoppele, J. M. Ageing and zonal variation in post-translational modification of collagen in normal human articuar cartilage. The age-related increase in non-enzymatic glycation affects iomechanical properties of cartilage. *Biochem J.* **1998**, *33*0, 345-351.

(32) Monnier, V. M.; Kohn, R. R.; Cerami, A. Accelerated age related browning of human collagen in diabetes mellitus. *Proc. Natl. Acad. Sci.* **1984**, *8*1, 583-587.

(33) Biroccio, A.; Urbani, A.; Massoud, R.; di Ilio, C.; Sacchetta, P.; Bernardini, S.; Cortese, C.; Federici, G. A quantitative method for the analysis of glycated and glutathionylated hemoglobin by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Anal. Biochem.* **2005**, *33*6, 279-288.

(34) Nakanishi, T.; Miyazaki, A.; Kishikawa, M.; Yasuda, M. Quantification of glycated hemoglobin by electrospray ionization mass spectrometry. *J. Mass. Spectrom.* **1997**, *3*2, 773-778.

(35) Chandalia, H. B.; Krishnaswamy, P. R. Glycated hemoglobin. *Curr. Sci.* **2002**, *8*3, 1522-1532.

(36) Rodkey, F. L. Direct spectrophotometric determination of albumin in human serum. *Clin. Chem.* **1965**, *1*1, 478-487.

(37) Putignano, P.; Kaltsas, G. A.; Korbonits, M.; Jenkins, P. J.; Monson, J. P.; Besser, G. M.; Grossman, A. B. Alterations in serum protein levels in patients with Cushing's syndrome before and after successful treatment. *J. Clin. Endocrinol. Metab.* **2000***, 8*5, 3309-3312.

(38) Shaklai, N.; Garlick, R. L.; Bunn, H. F. Nonenzymatic glycosylation of human serum albumin alters its confirmation and function. *J. Biol. Chem.* **1984***, 25*9, 3812-3817.

(39) Peters, T. In *All About Albumin: Biochemistry, Genetics, and Medical Applications.* Academic Press: San Diego, 1996, pp 102-126.

(40) Curry, S.; Mandelkow, H.; Brick, P.; Franks, N. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nat. Struct. Biol.* **1998***,* 5, 827-835.

(41) Koyama, H.; Sugioka, N.; Uno, A.; Mori, S.; Nakajima, K. Effects of glycosylation of hypoglycemic drug binding to serum albumin. *Biopharm. Drug. Dispos.* **1997***, 1*8, 791-801.

(42) Vorum, H.; Fisker, K.; Otagiri, M.; Pedersen, A. O.; Hansen, U. K. Calcium ion binding to clinically relevant chemical modifications of human serum albumin. *Clin. Chem.* **1995***, 4*1, 1654-1661.

(43) He, X. M.; Carter, D. C. Atomic structure and chemistry of human serum albumin. *Nature.* **1992***, 35*8, 209-215.

(44) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattachara, A. A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **2005***, 35*3, 38-52.

(45) Ahmed, N.; Dobler, D.; Dean, M.; Thornalley, P. J. Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. *J. Biol. Chem.* **2005***, 28*0, 5724-5732.

(46) Doucet, J.; Fresel, J.; Hue, G.; Moore, N. Protein binding of digitoxin, valproate and phenytoin in sera from diabetics. *Eur. J. Clin. Pharmacol.* **1993***, 4*5, 577-579.

(47) Nakajou, K.; Watanabe, H.; Kragh-Hansen, U.; Maruyama, T.; Otagiri, M. The effect of glycation on the structure, function and biological fate of human serum albumin as revealed by recombinant mutant. *Biochim. Biphys. Acta.* **2003***, 162*3, 88-97.

(48) Sugio, S.; Kashima, A.; Mochizuki, S.; Kobayashi, K. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* **1999***, 1*2, 439-446.

(49) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996***, 1*4, 33-38.

(50) Baker, J. R.; Zyzak, D. V.; Thorpe, S. R.; Baynes, J. W. Chemistry of the fructosamine assay: D-Glucosone is the product of oxidation of Amadori compounds. *Clin. Chem.* **1994***, 4*0, 1950-1955.

(51) Armbuster, D. A. Fructosaminase: Structure, analysis and clinical usefulness. *Clin. Chem.* **1987***, 3*3, 2153-2163.

(52) Lui, X.; Scouten, W. H. In *Boronate affinity chromatography.* Hage, D. S., Cazes, J., Eds.; Handbook of Affinity Chromatography; CRC: Boca Raton, FL, 2006; pp 215-230.

(53) Koyama, T.; Terauchi, K. Synthesis and application of boronic acid-immobilized porous polymer particles: A novel packing for high-performance liquid affinity chromatography. *J. Chrom. B.* **1996***, 67*9, 31-40.

(54) Bailon, P.; Ehrlich, G. K.; Fung, W. J.; Berthold, W., Eds.; In *Methods in Molecular Biology;* Lui, X. C., Scouten, W. H., Eds.; Affinity Chromatography: Methods and Protocols. Humana Press: NY, 2000; Vol. 147, pp 119-128.

(55) Capote, F. P.; Scherl, A.; Mu''ller, M.; Waridel, P.; Lisacek, F.; Sanchez, J. C. Glycation isotopic labeling (GIL) with 13C-reducing sugars for quantitative analysis of glycated proteins in human plasma. *Mol. Cell. Proteomics.* **2009***, in press*.

(56) Iberg, N.; Fluckiger, R. Nonenzymatic glycosylation of albumin in vivo. *J. Biol. Chem.* **1986***, 261*, 13542-13545.

(57) Yan, W.; Chen, S. S. Mass spectrometry-based quantitative proteomic profiling. *Briefings Funct. Genom. Proteomics.* **2005***, 4*, 27-38.

(58) Bidasee, K. R.; Nallani, K.; Yu, Y.; Cocklin, R. R.; Zhang, Y.; Wang, M.; Dincer, U. D.; Besch, H. R. Chronic diabetes increases advanced glycation end products on cardiac ryanodine receptors/calcium-release channels. *Diabetes.* **2003***, 52*, 1825-1836.

(59) Capote, F. P.; Sanches, J. C. Strategies for proteomic analysis of non-enzymatically glycated proteins. *Mass Spectrom. Rev.* **2008***, 28*, 135-146.

(60) Lapolla, A.; Fedele, D.; Reitano, R.; Arico, N. C. Enzymatic digestion and mass spectrometry in the studies of advanced glycation end products/peptides. *J. Am. Soc. Mass. Spectrom.* **2004***, 15*, 496-509.

(61) Zhang, X.; Medzihradskzy, K. F.; Cunningham, J.; Lee, P. D. K.; Rognerud, C. L.; Ou, C. N.; Harmatz, P.; Witkowska, H. E. Characterization of glycated hemoglobin in diabetic patients: Usefulness of electrospray mass spectrometry in monitoring the extent and distribution of glycation. *J. Chrom. B.* **2001***, 759*, 1-15.

(62) Zhang, Q.; Frolov, A.; Tang, N.; Hoffmann, R.; Goor, T. V. D.; Metz, T. O.; Smith, R. D. Application of electron transfer dissociation mass spectrometry in analysis of non-enzymatically glycated peptides. *Rapid Comm. Mass. Spec.* **2007***, 21*, 661-666.

(63) Wa, C.; Cerny, R. L.; Clarke, W. A.; Hage, D. S. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta.* **2007***, 385*, 48-60.

(64) Wa, C.; Cerny, R. L.; Hage, D. S. Identification and quantitative studies of protein immobilization sites by stable isotope labeling and mass spectrometry. *Anal. Chem.* **2006***, 78*, 7967-7977.

(65) Fitzgerald, M. C.; Parr, G. R.; Smith, L. M. Basic matrixes for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides. *Anal. Chem.* **1993***, 65*, 3204-3211.

(66) Gusev, A. I. Interfacing matrix-assisted laser desorption/ionization mass spectrometry with column and planar separations. *Fresen. J. Anal. Chem.* **2000***, 366*, 691-700.

(67) Vestal, M. L.; Juhasz, P.; Martin, S. A. Delayed extraction matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid. Comm. Mass. Spec.* **2005***, 9*, 1044-1050.

(68) Stewart, I. I.; Thomson, T.; Figeys, D. $^{18}O$ labeling: A tool for proteomics. *Rapid. Comm. Mass. Spec.* **2001***, 15*, 2456-2465.

(69) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **2007***, 389*, 1017-1031.

# CHAPTER 2

# QUANTITATIVE ANALYSIS OF GLYCATION SITES ON COMMERCIALLY GLYCATED HUMAN SERUM ALBUMIN

## 2.A. INTRODUCTION

This chapter evaluates the use of $^{18}$O-labeling quantitative proteomics as a means of quantifying modifications that occur on commercially glycated human serum albumin (HSA). The glycation of HSA has been shown to affect its drug binding kinetics (1-3) as well as its structure (4). These functional effects have not yet been fully characterized and better information regarding the glycation pattern that occurs on HSA, the extent of modification that occurs on various parts of HSA, and whether the modifications are consistent in both *in vitro* and *in vivo* glycated HSA are needed to get a better understanding of why these functional changes are occurring. Some information about the glycation pattern that occurs on HSA has previously been obtained (5-8); however, quantitative information about the degree of modification that occurs on different regions of HSA is an area where little work has been done (9-11). Similar trends have been noted in other proteins where qualitative data is available (12-14), however very little quantitative information is available.

In this study, isotopic labeling and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) will be employed as tools for

quantitative studies by comparing HSA and minimally glycated HSA that have been digested in the presence of $^{16}O$- or $^{18}O$-enriched water (see general approach in Figure 1.4) (15-21). The resulting digests will be mixed in a fixed ratio, and the $^{16}O/^{18}O$ ratios of peptides found in this mixed digest will be determined by using mass spectrometry (17, 22). If a modification takes place in a given region of HSA due to glycation, an increase in the mass of peptides from this region should occur resulting in a reduced concentration of non-modified peptide in the glycated sample. This reduction in concentration would cause the $^{16}O/^{18}O$ ratio to increase above the expected $^{16}O/^{18}O$ ratio when no modification is present. In addition, the size of the $^{16}O/^{18}O$ ratio should make it possible to compare the relative extent of glycation-related modifications on different regions of HSA.

These experiments will be conducted by using these tools from quantitative proteomics to examine minimally glycated HSA that has been prepared *in vitro* by Sigma Aldrich (St Louis, MO). A similar type of HSA has been used previously to identify possible modifications in glycated HSA (5). This previous research will be expanded upon in this current study by using $^{16}O/^{18}O$-labeling and MALDI-TOF MS to rank the degree of modification that is occurring at such sites in HSA.

The theory behind the determination of $^{16}O/^{18}O$ ratios when using $^{16}O$- and $^{18}O$-labeling has been described previously (15, 23). Two different methods for calculating $^{16}O/^{18}O$ ratios were used in this report. The first method employed mass spectra for a $^{18}O$ and $^{16}O$ mixed digest along with the expected relative intensities for unmodified peptides, as predicted by using MSIsotope (24). The $^{16}O/^{18}O$ ratio for a peptide in this case was

determined by using the following equation, which was derived from a set of equations that are described in the published literature (23).

$$\frac{^{16}O}{^{18}O} = \frac{I_0}{\left(I_2 - I_0\frac{M_2}{M_0}\right) + \left(I_4 - I_0\frac{M_4}{M_0} - \left(I_2 - I_0\frac{M_2}{M_0}\right)\frac{M_2}{M_0}\right)} \quad (1)$$

In this equation, the terms $I_0$ through $I_5$ represent the relative intensities for the M+0 to M+5 peaks in an isotope cluster, where $M_0$ through $M_5$ represent the expected relative intensities for a digest in which no isotopic label is present. The terms in the first parentheses on the left side of the denominator represent the $^{18}O$ contribution for a peptide with one $^{18}O$ label. The terms in the second parentheses represent the $^{18}O$ contribution for a peptide with two $^{18}O$ labels. This approach for determining $^{16}O/^{18}O$ ratios will be referred to as "Method 1" throughout this chapter.

The second method employed an internal standard by using the mass spectra obtained by digesting HSA in $^{16}O$-enriched water instead of using the theoretical digest. For this method, the relative intensities of both the $^{16}O$ and $^{18}O$ digest were used to estimate the contribution of $^{16}O$ or $^{18}O$-labeled peptides to the corresponding mixed peptides. This is method is previously described in the literature (15, 19). A summary of the equation that was used to determine the $^{16}O/^{18}O$ ratio is shown below.

$$\frac{^{16}O}{^{18}O} = \frac{I'_0}{\left(I''_0 + I''_{2(c)} + I''_{4(c)}\right)} \quad (2)$$

In this equation, $I'_0$ represents the contribution of the $^{16}O$-labeled digest to the $I_0$ peak in the mixed digest. Similarly, $I''_0$, $I''_{2(c)}$, and $I''_{4(c)}$ represents the corrected $^{18}O$-

labeled digest contribution to the $I_0$, $I_2$, and $I_4$ peaks respectively, in the mixed digest. The advantage of using this method is that it allows corrections to be made for variations in the actual extent of $^{18}O$ incorporation and the type of instrumentation being employed. The disadvantage of this method is that peptides have to be detected in all three digests for the $^{18}O/^{16}O$ ratio to be determined. The use of this approach to determine $^{16}O/^{18}O$ ratios will be referred to as "Method 2" in the remainder of this chapter.

## 2.B. EXPERIMENTAL

**2.B.1. Materials.** The following chemicals were purchased for Sigma-Aldrich (St. Louis, MO): Des-Arg-bradykinin (97% pure), glu-fibrinopeptide (97%), angiotensin I (97%; acetate salt), HSA (99%, essentially fatty acid and globulin free), glycated HSA (95%; lot number 013K9166, containing 8 mol hexose/mol HSA), sequencing grade trypsin, sequencing grade Glu-C, sequencing grade Lys-C, guanidine HCl (99%), D/L dithiothreitol (99%), iodoacetamide (99%), formic acid (96%), 2,5-dihydroxybenzoic acid (98%), α-cyano-4-hydroxycinnamic acid (99%), $^{18}$O-enriched water (97%), $^{16}$O-enriched water (99.99%), tris-HCl (99%) and ammonium bicarbonate (99%). All of the chemicals that were used were reagent grade or better. The water used for these experiments (other than that utilized for $^{18}$O-labeling) was obtained from a Nanopure water system (Barnstead, Dubuque, IA, USA)

**2.B.2. Apparatus.** The following items were purchased from Thermo Fisher Scientific (Rockford, IL): Slide-A-Lyzer dialysis cassettes (7000 Da MW cutoff, 0.1 – 0.5 mL capacity) and a 0.5 – 10 µL digital pipette. Micro-C18 ZipTip pipette tips with 5.0 µg of bed material were obtained from Millipore (Billerica, MA). Plastic sheets were used to mix the sample and matrix mixing prior to MALDI-TOF MS. These sheets were purchased from C-Line Products (Des Plaines, IL). Mass spectra were acquired on a Voyager 6148 MALDI-TOF-MS system (Applied Biosystems, CA). The instrument settings were as follows: positive-ion delayed extraction reflection mode; delay time, 100 ns; accelerating voltage, 20 kV; guide wire voltage, 0.008% of accelerating voltage; grid voltage, 76% of accelerating voltage. The MSIsotope software was obtained from the UCSF Protein Prospector webpage (24).

**2.B.3. Sample Pretreatment, Digestions & Peptide Fractionation.** The HSA samples were pretreated as described previously in work with immobilized HSA supports (15), but with the following modifications being made for this study. A 5 mg/mL HSA solution prepared in denaturing buffer was used as described in the previous study, however, the solution volume was now reduced from 1 mL to 300 μL to reduce the amount of HSA that was used. The ratio of HSA to dithiothreitol and iodoacetamide was kept the same as in Ref. (15) while the solution volumes were adjusted accordingly to compensate for the change in the volume of the HSA solution. Digestion was carried out with normal HSA being placed into $^{16}$O-labeled water and glycated HSA being digested in $^{18}$O-labeled water. Zip-tip fractionation was performed as described previously by using aqueous mixtures of 5, 10, 20, 30 and 50% acetonitrile for elution in a series of step gradients (15).

**2.B.4. Mass Spectrometry & Data Acquisition.** The details regarding the preparation of mass calibrants and matrix solutions are described elsewhere (15). The mass spectrometer was externally calibrated using a mixture containing des-Arg-bradykinin (25 pmol/μL), glu-fibrinopeptide (32.5 pmol/μL), and angiotensin I (32.5 pmol/μL). A 4 μL portion of this solution was mixed with 96 μL of α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid mixed matrix solution (i.e., giving a final concentration of 0.1–5 pmol/μL for the final calibration mixture), and the resulting mixture was spotted adjacent to each sample well on a MALDI plate. A mass spectrum was obtained for the calibration mixture using Voyager Control Panel software (Applied Biosystems, Foster City, CA, USA) and the file was exported to the Data Explorer software (also from Applied Biosystems). Using Data Explorer, the monoisotopic M+0 peaks for des-Arg-

bradykinin, glu-fibrinopeptide, and angiotensin I were calibrated to masses of 904.4681 Da, 1570.6774 Da, and 1296.6853 Da, respectively. The calibration constants obtained were then saved and imported to the Voyager software. The mass spectrum for the sample was then obtained. The calibration process was repeated for every sample spot on the MALDI plate. The mass accuracy obtained when this method was used was typically <50 ppm.

**2.B.5. Peak Selection Criteria and Data Sorting.** The masses of the peptides detected in the HSA or glycated HSA digests were compared to the masses predicted for a theoretical digest by using PEPTIDEMASS software (25, 26). When generating the theoretical digest, the following considerations were made: the maximum amount of missed cleavages was 2; all cysteine residues were assumed to be treated with iodoacetamide; variable oxidation of methionine residues was allowed; and only monoisotopic masses were selected. All peptides that could be matched within 50 ppm of the theoretical digest were selected for further analysis (22). For all digests, only peptides having a signal-to-noise ratio >5 were used for further quantitative studies. In the $^{18}O$ digests, the M+4 peak was used to determine the signal-to-noise ratio for the resultant peptides, while the M+0 peak was used to determine the signal-to-noise ratio for the $^{16}O$ digests. Information about the p$K_a$ and the fractional accessible surface area (FAS) was used to predict the reactivity of a given amino acid residue in forming glycation products; the methods used for calculating these p$K_a$ and FAS values have been described previously in the literature (5, 27, 28).

## 2.C. RESULTS

**2.C.1. Sequence Coverage.** The usable sequence coverage (i.e., the sequence coverage representing the residues on HSA corresponding to calculated $^{16}O/^{18}O$ ratios) for both methods are summarized in Figure 2.1. Prior to looking at the relative extent of glycation at various regions on HSA, the sequence coverage for this protein was examined under the analysis conditions used in this study. When using only qualitative data and looking for m/z values that could be linked to specific peptides, the coverage obtained for the trypsin, Lys-C and Glu-C digests of HSA were found to be 83.6%, 61.0% and 54.0% respectively. The total coverage obtained when all three digests were considered was 92.3%, which is close to a previously reported value of 97.4% under similar pretreatment and analysis conditions for samples of soluble or immobilized HSA that were not glycated (15).

When Method 1 was used to obtain $^{16}O/^{18}O$ ratios, the usable sequence coverage in these quantitative studies was lowered to 51.5%, 53.9%, and 51.5% for the trypsin, Lys-C, and Glu-C digests respectively, which gave a total usable sequence coverage for HSA of 83.6% (see Figure 2.1). Similarly, when Method 2 was employed, the usable sequence coverage was 33.5%, 21.9%, and 42.1% for the trypsin, Lys-C, and Glu-C digests, respectively, which gave an overall usable sequence coverage of 66.3% for HSA. This reduction in usable sequence coverage in going from the qualitative to quantitative experiments was expected because several of the peptides that were detected in the qualitative studies did not have a high enough intensity to be used in the quantitative studies. This same type of reduction in sequence coverage has been observed in previous

Figure 2.1. Primary sequence of HSA and sequence coverage for modified residues obtained for Method 1 (bold) and Method 2 (underlined).

1– DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV
41– KLVNEVTEFA KTCVADEsae NCDKSLHTLF GDKLCTVATL
81– RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV
121– DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR
161– ykAAFTECCQ AADKAACLLP KLDELRDEGK assakqrlkc
201– aslqkFGERA FKAWAVARLS QRFPKAEfae vskLVTDLTK
241– VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE
281– KPLLEKSHCI AEVEndempa dlpslaadfv eskdvckNYA
321– EAKDVFLGMF LYEYARRHPD YSVVLLLRLA KTYEttlekC
361– CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE
401– YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH
441– PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES
481– LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKe
521– rqikkQTALV ELVKhkpkat keQLKAVMDD FAAFVEKcck
561– addketcfae egkklvaasq aalgl

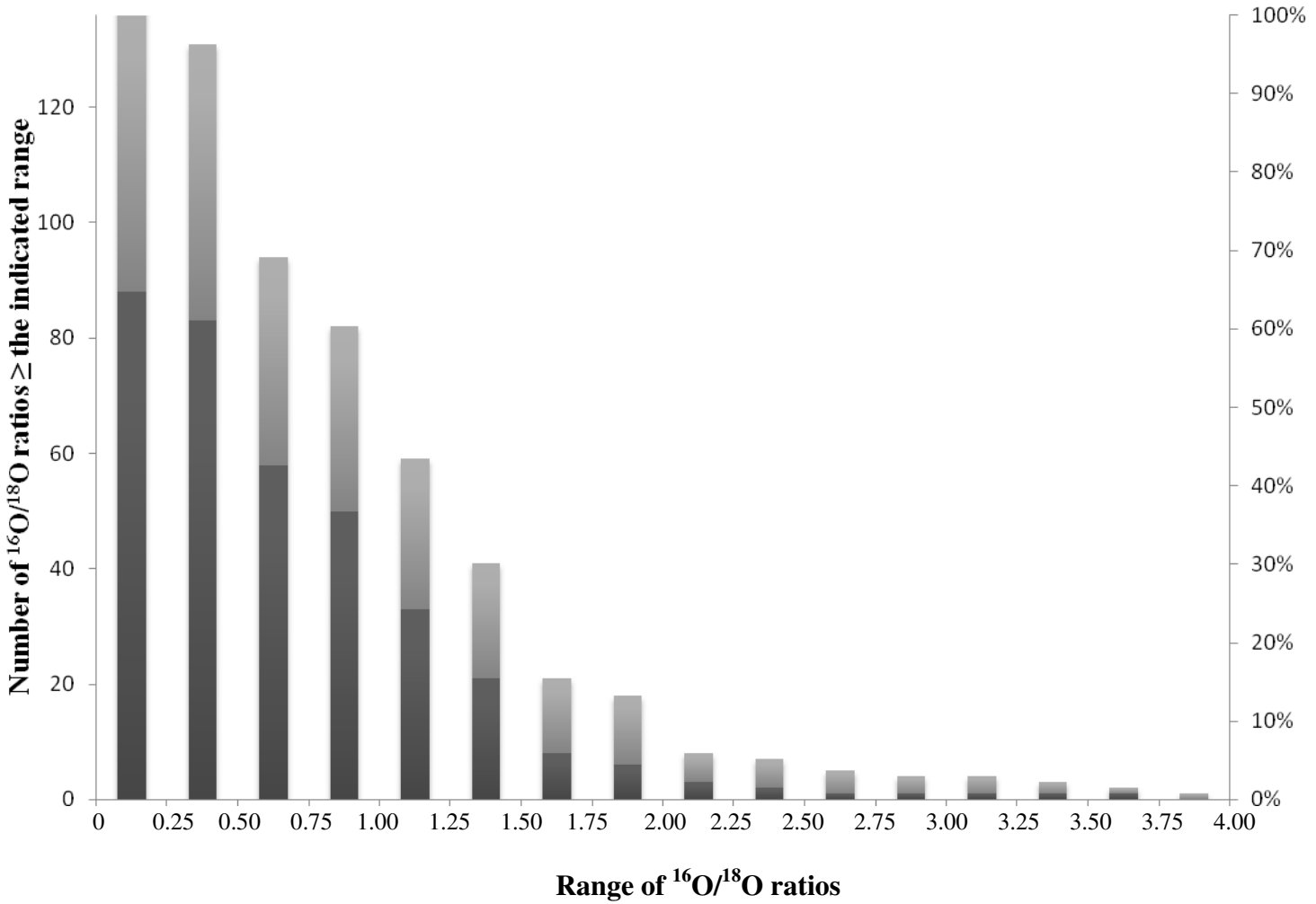
studies using quantitative proteomics for immobilized versus soluble HSA, where a total usable sequence coverage of 78% was obtained when quantitative data was considered (15).

When using both Methods 1 and 2 for the quantitative studies, $^{16}O/^{18}O$ ratios could be determined for peptides that encompassed 42 lysine and 21 arginine residues in HSA. By using a qualitative comparison between the $^{16}O$- and $^{18}O$-labeled samples, K525 (found in residues 525-534 in the tryptic digest obtained in this study), was also found to have significant levels of modification. The remaining 16 lysine and 3 arginine residues that were not included in these peptides are located within residues 58–61, 160–163, 191–206, 295–317, 355–359, 520–524, 535–543 and 558–585. None of these residues, with the exception of K199 (a known site of major glycation), are thought to undergo any significant amounts of modification during the glycation of HSA (11).

**2.C.2. Measurements of $^{16}O/^{18}O$ Ratios for Peptide Digests from Glycated HSA.**

Peptides containing a $^{16}O/^{18}O$ ratio higher than a cutoff value of 2.00 were selected and treated as potential modification sites in the digests of glycated HSA. This cutoff value was determined using a cumulative sums plot (Figure 2.2), where it was determined that 6% of the detected peptide peaks had a $^{16}O/^{18}O$ ratio greater 2.00. This cutoff value is similar to an upper reference value for $^{16}O/^{18}O$ ratios of 2.0–3.0 (depending on the type of digest being used) that has been previously reported for this digestion and labeling approach with normal HSA (15). The $^{16}O/^{18}O$ ratio range of 2.0–3.0 was previously obtained by comparing two control samples of the same preparation of normal HSA (15). When both Methods 1 and 2 were used for quantitative analysis, the $^{16}O/^{18}O$ ratios for 83 peptides were determined. Of these peptides, eight peptides containing four lysine

Figure 2.2. Bar chart showing the cumulative sum of all the detected peaks as a function of $^{16}O/^{18}O$ ratio range for Methods 1 (■) and 2(■). The number of $^{16}O/^{18}O$ ratios that were greater than or equal to the indicated$^{16}O/^{18}O$ ratio range was used for this plot. The relative amount of the number of $^{16}O/^{18}O$ ratios is also given as a percentage.


Number of $^{16}O/^{18}O$ ratios ≥ the indicated range
120
100
80
60
40
20
0
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
3.25
3.50
3.75
4.00
Range of $^{16}O/^{18}O$ ratios

residues and five arginine residues were identified as major sites of modification because their $^{16}O/^{18}O$ ratios were above the selected cutoff ratio of 2.0. Two peptides that encompass residues 525–534 and 415-439 on HSA were found to have significant levels of modification; however, the extent of modification was too large to be quantified by this current approach.

There were two cases where the $^{16}O/^{18}O$ ratio was calculated for peptides with identical residues where one peptide had a cysteine that had been modified by iodoacetamide to form carbamidomethyl-cysteine (CAM-cysteine) and the other peptide remained unmodified. These two peptides were detected in the Glu-C digest and corresponded to residues 466–479 and residues 101–119. In this situation, a pair of $^{16}O/^{18}O$ ratios was calculated for each residue, where one ratio corresponds to the CAM-modified peptide and the other to the unmodified peptide. In both cases, the unmodified peptides had a much higher $^{16}O/^{18}O$ ratio (1.27–1.86) than their CAM-modified counterparts (0.31–0.37). This anomaly may indicate that CAM-incorporation was not consistent between the glycated HSA and the reference HSA samples. The $^{16}O/^{18}O$ ratio for the CAM-modified peptide in this case was assumed to be the $^{16}O/^{18}O$ ratio that most closely represented glycation because the formation of CAM cysteine should be highly favored under the reactions conditions used for alkylation in this study.

The $^{16}O/^{18}O$ ratios obtained for the tryptic digest using Method 1 and 2 are given in Tables 2.1 and 2.2, respectively. In the tryptic digest, five peptides were found to have $^{16}O/^{18}O$ ratios greater than 2.00. A total of three peptides were also noted in the Lys-C digest that had $^{16}O/^{18}O$ ratios greater than 2.00, but the Glu-C digest gave no peptides with $^{16}O/^{18}O$ ratios above this cutoff value (see Supplemental Information).

Table 2.1. Calculated $^{16}O/^{18}O$ ratios for the tryptic digest using Method 1

| | | Observed peak ratios for each peptide at given %ACN in water | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio** [a] |
| 74-81 | 933.52 | 3.68 | | | | | 3.68 |
| 146-159 | 1742.89 | | | | 2.40 | | 2.40 |
| 318-323 | 695.34 | 2.15 | | | | | 2.15 |
| 324-336 | 1623.79 | | | | | 1.94 | 1.94 |
| 1-10/42-51 | 1149.62 | 1.43 | | | | | 1.43 |
| 145-159 | 1899.00 | | | | 1.30 | 1.17 | 1.23 (± 0.09) |
| 337-348 | 1467.84 | 1.08 | 1.39 | 1.11 | 1.18 | 1.12 | 1.17 (± 0.13) |
| 65-81 | 1932.04 | | | | 1.07 | 1.10 | 1.09 (± 0.02) |
| 11-20 | 1226.61 | 1.07 | | | | | 1.07 |
| 213-218 | 673.38 | 1.02 | | | | | 1.02 |
| 65-73 | 1017.54 | 0.81 | 1.15 | | | | 0.98 (± 0.24) |
| 138-144 | 927.49 | | 0.97 | 0.94 | | | 0.95 (± 0.02) |
| 403-410 | 960.56 | 0.93 | 0.93 | 0.96 | | | 0.94 (± 0.02) |
| 414-428 | 1639.94 | | 1.12 | 0.75 | | | 0.94 (± 0.26) |
| 546-557 | 1342.63 | | | 0.93 | | | 0.93 |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This method for determining the final standard deviation was determined using the method of error propagation).

Table 2.1 (continued). Calculated $^{16}O/^{18}O$ ratios for the tryptic digest using Method 1

| | | Observed peak ratios for each peptide at given %ACN in water | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio** [a] |
| 373-389 | 2045.10 | | | 0.85 | 0.95 | | 0.90 (± 0.07) |
| 107-114 | 940.45 | 0.89 | | | | | 0.89 |
| 338-348 | 1311.74 | | | 0.87 | 0.8 | | 0.84 (± 0.05) |
| 13-20 | 951.44 | 0.83 | | | | | 0.83 |
| 137-144 | 1055.59 | 0.68 | 0.78 | | | | 0.73 (± 0.07) |
| 263-274 | 1443.64 | 0.7 | | | | | 0.7 |
| 234-240 | 789.47 | 0.51 | | | | | 0.51 |
| 82-93 | 1434.53 | 0.36 | | | | | 0.36 |
| 182-186 | 645.36 | 0.34 | | | | | 0.34 |
| 433-444 | 1400.68 | 0.34 | | | | | 0.34 |
| 390-402 | 1657.75 | | | 0.32 | | | 0.32 |
| 275-286 | 1546.80 | 0.26 | | | | | 0.26 |
| 115-136 | 2778.36 | | | 0.2 | 0.12 | 0.39 | 0.24 (± 0.13) |
| 94-106 | 1714.80 | 0.21 | 0.12 | | | | 0.16 (± 0.06) |
| 485-500 | 1910.93 | | | 0.13 | 0.17 | | 0.15 (± 0.03) |
| 163-174 | 1371.57 | 0.1 | | | | | 0.1 |
| 115-136 | 2650.26 | | | | | 0.1 | 0.1 |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation(Note: This method for determining the final standard deviation was determined using the method of error propagation).

Table 2.2. Calculated $^{16}O/^{18}O$ ratios for the tryptic digest using Method 1

| | | **Observed peak ratios for each peptide at given %ACN in water** | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio** [a] |
| 107-114 | 940.45 | 3.77 | | | | | 3.77 |
| 213-218 | 673.38 | 3.48 | | | | | 3.48 |
| 145-159 | 1899.00 | | | | 1.8 | | 1.8 |
| 414-428 | 1639.94 | | 1.41 | 1.8 | | | 1.60 (± 0.28) |
| 137-144 | 1055.59 | 1.35 | 1.28 | | | | 1.32 (± 0.05) |
| 138-144 | 927.49 | 1.25 | 1.27 | | | | 1.26 (± 0.14) |
| 403-410 | 960.56 | 1.25 | 1.27 | | | | 1.26 (± 0.14) |
| 21-41 | 2490.29 | | | | | 0.94 | 0.94 |
| 373-389 | 2045.10 | | | 0.9 | 0.87 | | 0.88 (± 0.02) |
| 337-348 | 1467.84 | | | 0.81 | | | 0.81 |
| 94-106 | 1714.80 | 0.43 | 0.43 | | | | 0.43 (± 0.00) |
| 338-348 | 1311.74 | | | 0.41 | | | 0.41 |
| 11-20 | 1226.61 | 0.39 | | | | | 0.39 |
| 390-402 | 1657.75 | | | 0.36 | | | 0.36 |
| 433-444 | 1400.68 | 0.35 | | | | | 0.35 |
| 485-500 | 1910.93 | | | 0.23 | 0.19 | | 0.21 (± 0.03) |
| 115-136 | 2778.36 | | | 0.14 | 0.12 | | 0.13 (± 0.01) |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This method for determining the final standard deviation was determined using the method of error propagation).

Table 2.3. Peptides corresponding to the top 8 peak ratios (Using a cutoff ratio of 2.0) and their potential modifications are reported[a]

| Amino Acids | p$K_a$/FAS[b] | Residue (digest, method) | Peak ratio | Detected[c] | Residue [13][c] | Modification |
|---|---|---|---|---|---|---|
| R114 | 12.29/0.80 | 107-114 (Trypsin, 2) | 3.77 | This Study | n/a | none detected |
| R81 | 12.50/0.76 | 71-84 (Trypsin, 1) | 3.68 | This Study | n/a | none detected |
| R218 | 9.64/0.20 | 213-218 (Trypsin, 2) | 3.48 | [13] | 200-218 (Try.) | AFGP |
| **K413**, R410 | 10.01/0.23, 12.46/0.33 | 403-413 (Lys-C, 2) | 3.29 (± 0.89) | This Study | 403-413 | FL-2$H_2O$ |
| **K432**, R428 | 9.92/0.36, 11.29/0.18 | 415-432 (Lys-C, 2) | 2.57 (± 1.48) | [13] | 426-442 (Glu-C) | FL |
| ″ | ″ | ″ | ″ | This Study | 415-432 | FL-2$H_2O$ |
| K159 | 9.94/0.44 | 146-159 (Trypsin, 1) | 2.40 | [13] | 145-160 (Try.) | PYR |
| **K212**, R209 | 10.43/0.40, 12.08/0.61 | 206-212 (Lys-C, 2) | 2.32 (± 0.76) | [13] | 200-218 (Try.) | AFGP |
| ″ | ″ | ″ | ″ | ″ | 206-225 (Lys-C) | FL |
| K323 | 10.50/0.54 | 318-323 (Trypsin, 1) | 2.15 | This Study | n/a | none detected |
| | | **Significantly modified peptides as indicated by a qualitative comparison** | | | | |
| K233 | 10.29/0.43 | 226-233 (Lys-C, 1) | 4.32[d] | This Study | n/a | none detected |
| K436, **K439** | 9.67/0.51, 10.50/0.94 | 415-439 (Lys-C) | n/a | [13] | 426-442 (Glu-C) | FL |
| **K525**, K534 | 10.06/0.07, 11.11/0.12 | 525-534 (Trypsin) | n/a | [13] | 520-525 (Lys-C) | CEL |

[a] – The values in bold represent the most likely modification sites.

[b] – This column shows the p$K_a$ and fractional accessible surface area for the protein calculated using PROPKA [33] and VADAR [34].

[c] – This column shows whether a similar amino acid or peptide was detected previously [13]. The corresponding residue and modification are also shown.

[d] – The corresponding peptide in the mixed digest had a mass accuracy that was outside the 50 ppm threshold.

Table 2.3 shows the ranking of the peptides that were found in the quantitative studies to have major degrees of modification, as listed according to the order of their $^{16}O/^{18}O$ ratios. These peptides contained amino acids R114, R81, R218, K413, K432, K159, K212 and R209, and K323. Of these residues R218, K159 and R209 were found to be modified to produce the AGEs 1-alkyl-2-formyl-3,4-glycoyl-pyrole (AFGP), pyrraline (Pyr) and AFGP, respectively. Similarly, residues 212, 413 and 432 were modified to form fructosyl-lysine (FL), FL-$2H_2O$ and FL-$2H_2O$, respectively, all of which are early stage glycation products.

There were two instances where the m/z values for a given peptide were clearly observed in the $^{16}O$ digest, but the same m/z values were not found in the $^{18}O$ digests or mixed labeled digest (see Table 2.3). The absence of these characteristic peptides in the $^{18}O$ digests is indicative of significant glycation, because the concentration of the unmodified forms of these peptides were too low to be detected at the selected signal-to-noise threshold of 5 (i.e., only peaks with intensities below this threshold were noted in the $^{18}O$ and mixed oxygen labeled digests). Low intensity m/z values corresponding to residues 525–534 (1128.69 Da) in the tryptic digest and residues 415–439 (2629.42 Da) in the Lys-C digest were found only in the $^{16}O$-labeled digest. Interestingly, these residues also correspond to lysines 439 and 525, which are amongst the most commonly cited glycated lysine residues in HSA (11). This finding is consistent with our previous work, where K525 was shown to be modified to form $N_\varepsilon$–carboxyethyl-lysine (CEL) and K439 was shown to be modified to form FL (5). In addition to these peptides, K233 which is found on residues 226–233 of HSA is also believed to have significant levels of modification. A set of m/z values corresponding to this peptide (880.44 Da), were clearly

identified in the $^{16}O$- and $^{18}O$-labeled digests; however, the corresponding m/z value in the mixed digest seemed to shift to 881.29 Da. This shift would result in a value that falls outside the assigned mass accuracy. Nevertheless, a $^{16}O/^{18}O$ ratio was calculated for this peptide, giving a $^{16}O/^{18}O$ ratio of 4.32, indicating that this peptide could potentially contain high levels of glycation. Previous studies have indicated that K233, which is found within residues 226-233 of HSA, may be glycated (11).

The $^{16}O/^{18}O$ ratios for other modifications that occurred on glycated HSA which did not exceed the 2.00 cutoff value were also considered, based on possible modifications that have been noted in the literature in qualitative studies with HSA (5). If multiple $^{16}O/^{18}O$ ratios could be linked to a given lysine or arginine residue, the resulting $^{16}O/^{18}O$ ratios that are modified to form CAM-cysteine and peptides that have a minimum combination of lysine and/or arginine residues were used to assign the most likely glycation sites. This filtering was done to provide the most realistic representation of the $^{16}O/^{18}O$ ratio for a given amino acid, without the influence of CAM modified cysteine or assigning $^{16}O/^{18}O$ ratios that result from multiple minor modifications to a single amino acid. When CAM modification and lysine/arginine heterogeneity are factored in, the range of $^{16}O/^{18}O$ ratios for lysines 12 and 51 (or the *N*-terminus) is 1.18–1.32 and 1.28–1.78 respectively. These results indicate that K51 (or the *N*-terminus) contains the most early glycation products, followed by K12. These amino acid residues have previously been linked to the formation of early stage glycation products (i.e., FL-related modifications). Similarly, the range of $^{16}O/^{18}O$ ratios for AGE-linked peptides (5) containing K159, K286, K378, R472 and R222 is 1.80–2.40, 0.26–0.28, 0.45–1.29, 0.31–0.50 and 1.62–1.86, respectively. When these residues are ranked in order of decreasing

$^{16}O/^{18}O$ ratios, K159 appears to have the largest extent of modification followed by R222, K378, R472 and K286. Lysines 159, 378, and 286 were linked to Pyr, CEL, and CML respectively. Similarly, arginines 222 and 472 were linked to G-H1 and ArgP respectively.

The peptide with the highest detected $^{16}O/^{18}O$ ratio of 3.77 was detected in the trypsin digest and had a mass of 940.45 Da. This mass corresponds to residues 107–114 on HSA, which indicates that R114 is modified to form AGEs. The particular type of modification could not be identified by m/z shifts in this study or in our previous study (5); however, early glycation does not occur on arginine residues so it likely being modified to form AGEs. Similarly, a peptide (933.52 Da) with a high $^{16}O/^{18}O$ ratio of 3.68 was detected in the trypsin digest. This mass corresponds to residues 74–81 on HSA, indicating that R81 may be involved in AGE formation. Several peptides were identified in this analysis that had high $^{16}O/^{18}O$ ratios with corresponding glycation related m/z shifts as identified previously (5). For example, a high $^{16}O/^{18}O$ ratio peptide corresponding to residues 213–218 (673.38 Da) on HSA has previously been shown to be modified by AFGP. Similar cases where high $^{16}O/^{18}O$ ratios and glycation related modifications can be linked are shown in Table 2.3.

**2.C.3. Peptides with High $^{16}O/^{18}O$ Ratios Representing Early or Advanced Glycation Products.** Some of the peptides containing $^{16}O/^{18}O$ ratios above the cutoff ratio had multiple sites for possible modification. As a result it was difficult to link an increased $^{16}O/^{18}O$ ratio to a particular amino acid in the respective peptide. For peptides with several possible modification sites, the detected mass shifts were used to help assign the sites that were being modified. For instance, peptide 403–413 (1352.77 Da) detected

in the Lys-C digest (See Appendix) had a $^{16}O/^{18}O$ ratio of 3.29 and contained both K413 and R410. A separate peptide was found from the same region (residues 403-413) that had a mass shift of 126 Da. This mass shift corresponds to modification by FL-$2H_2O$, suggesting that K413 is being modified during glycation to form this type of adduct. Similarly, lysines 432 and 212 were clearly identified as likely modification sites because the m/z shift data that was obtained previously (5) show that these amino acids were being modified by early glycation products.

There were some glycation-related mass shifts that were identified in this study (Table 2.4) and linked to peptides with intermediate $^{16}O/^{18}O$ ratios. The resulting mass shifts were similar to the mass shifts that were obtained previously (5). For instance, amino acids K106, K136, K240, K174, K466, and R336 are suspected to be modified to produce CML, FL-$2H_2O$, FL-$1H_2O$, FL and CEL, CEL, and AFGP respectively. Of these residues, lysines 136, 174, and 240 were previously shown to be modified by FL-$1H_2O$, CEL, and FL-$1H_2O$, respectively (5). All of the other modifications that were linked to intermediate $^{16}O/^{18}O$ ratios are new to this current study. Of these residues, K432 and K413 were linked to $^{16}O/^{18}O$ ratios that were above the cutoff ratio of 2.00. The number of m/z shifts that were linked to intermediate $^{16}O/^{18}O$ ratios in this study and in our previous study (5) illustrates the heterogeneous nature of glycation-related modifications and the need for improved characterization techniques.

**2.C.4. Correlation of Modifications with Local p$K_a$ or FAS Values.** Additional sources of information can be used to determine which lysine or arginine residues on HSA might be prone to modifications by glycation. This information includes the

Table 2.4. Glycation adducts (found in this study) that are linked to peptides with calculated $^{16}O/^{18}O$ ratios.

| Amino Acids | Residue | Detected Digest | Mass Shift | Modification | $^{16}O/^{18}O$ (Method 1) | $^{16}O/^{18}O$ (Method 2) |
|---|---|---|---|---|---|---|
| **K413**, R410 | 403-413 | Lys-C | 126.03 | FL-$2H_2O$ | 1.29 (± 0.19) | 3.29 (± 0.89) |
| **K106**, R98 | 94-106 | Lys-C | 58.01 | CML | 0.77 (± 0.08) | 1.42 (± 0.13) |
| K174 | 163-174 | Lys-C | 162.05 | FL | 0.72 (± 0.06) | na |
| K174 | 163-174 | Lys-C | 72.02 | CEL | 0.72 (± 0.06) | na |
| **K432**, R428 | 415-432 | Lys-C | 126.03 | FL-$2H_2O$ | 1.37 (± 0.35) | 2.57 (± 1.48) |
| **K106**, R114, R117 | 101-119 | Glu-C | 58.01 | CML | 1.27 (± 0.24), 0.37 (± 0.29) | 0.47 (± 0.20) |
| **K466**, R472, K475 | 466-479 | Glu-C | 72.02 | CEL | 1.86 (± 1.39), 0.31 (± 0.14) | 1.19 (± 0.14) |
| K240 | 234-240 | Trypsin | 144.04 | FL-$1H_2O$ | 0.51 | na |
| R336 | 324-336 | Trypsin | 270.07 | AFGP | 1.94 | na |
| **K136**, R117 | 115-136 | Trypsin | 126.03 | FL-$2H_2O$ | 0.39, 0.10 | 0.13 |

estimated p$K_a$ values for the individual amino acid side chains, the fractional accessible surface are (FAS) for each of these residues, and the location (surface or buried) of the amino acid residue. Lysine or arginine residues on HSA having low p$K_a$ values would be expected to be more susceptible to glycation reactions because a larger portion of these residues would exist in a non-protonated state, favoring nucleophilic addition or substitution (5). Amino acid side chains with high FAS values and residues that occur on the surface of HSA would be more reactive because they have greater access to reactants compared to sites buried within the protein's structure. Another factor to consider is that HSA is a flexible protein (3) in which changes in confirmation may facilitate or hinder glycation reactions. As a result, the aforementioned variables are not static, and residues with low FAS values, high p$K_a$ values or that are normally buried within the protein may become more reactive as glycation-induced conformational changes occur.

Of the various residues with high $^{16}O/^{18}O$ ratios that were noted in this study (Table 2.3), residues R218, R410, and R428 are all buried within the structure of HSA, while the majority of the remaining residues are located on the surface of HSA. In addition, lysines 233, 439, and 525, which are believed to have undergone significant glycation based on a qualitative comparison of the digest, are all found on the surface of HSA. This result makes sense because the majority of modification that occurs on HSA is expected to occur with residues that are easily accessible to sugars in the surrounding medium. When looking at the modifications which were identified by mass shifts (Table 2.4), six of these modifications occur on residues that are found on the surface of the protein and five occur on residues that are more buried in the structure of HSA. The location of AGEs or glycation products in the buried locations could be explained by

changes in the structure of HSA when glycation occurs (29). This change would increase the accessibility of the glycation-related reactants to lysine or arginine residues. Further evidence of this was indicated previously, when K199 was identified as a prominent glycation site. Even though the location of this residue is buried within the structure of HSA, it has consistently been identified as a glycation site that occurs on HSA (5, 30). Additional information, such as the calculated p$K_a$ and FAS values for these residues, were next used to provide further insight into these modification processes.

The range of calculated p$K_a$ values for the lysine and arginine residues on HSA are 6.23–11.11 and 7.56–16.02, respectively, where the lysine residues have an average p$K_a$ value of 10.08 and the arginine residues have an average p$K_a$ of 12.34 (Note: details on how these calculated results were obtained are provided in the literature (15)). The peptides listed in Table 2.3 appear to follow a trend in which the most frequently occurring modifications took place on residues that either have p$K_a$ values below the average or have high FAS values. For instance, R114, R81, and R209 have FAS values of 0.80, 0.76, and 0.61, respectively, which are on the high end of the range of FAS values for arginine. These residues are also suspected modification sites based on their $^{16}O/^{18}O$ ratio values. R218 has the highest FAS value of all arginines on HSA, a p$K_a$ value which is lower than average, and the highest calculated $^{16}O/^{18}O$ ratio in this study. All of these features provide strong evidence that this residue is a likely site for modification due to glycation-related reactions. Similarly, lysines 413, 432, and 212 were assigned as glycation sites because of the corresponding low p$K_a$ values, lysine specific modifications were identified on these peptides, and peptides corresponding to these residues all had high $^{16}O/^{18}O$ ratios. Similar assignments were made for m/z shifts

that were identified in this study that corresponded to early or late stage glycation adduct formation (Table 2.4). Of the glycation sites that were identified through a qualitative comparison of the digests, lysine 439 had a high FAS value of 0.94, which partially explains the reactivity of this residue. The reactivity of K233 and K525, however, cannot be explained based on the p$K_a$ or FAS values alone because these values are more typical of those seen for other lysines on HSA. It is therefore likely that some other process such as a conformational change upon glycation (29) or localized acid-base catalysis (11) may increase the amount of modification that occurs on these residues.

**2.C.5. Location of Modified Sites versus Major Drug Binding Sites on HSA.** It has been suggested that the extent of glycation of HSA may affect how tightly HSA binds drugs and small solutes in the body (31, 32). The two major drug binding sites on HSA, commonly referred to as the Sudlow sites 1 and 2, contain several amino acids which facilitate the binding of these compounds to HSA (2). When the results from Tables 2.3 and 2.4 are compared to the location of these binding sites (3), many of the potential glycation sites are at or near the same regions on HSA that facilitate drug and solute binding. For instance, K212, K233, R209, and R218 (noted in Table 3 to be potentially important sites for modifications due to glycation) occur in the same vicinity as the key residues W214, Y211, and Y211 that are found in Sudlow site 1. Similarly, K413, K432, K439, are located in the same region of HSA as R410 and Y411, which are important residues taking part in drug interactions at Sudlow site 2. Of the major glycation adducts that were noted in this study (Table 2.4), K240 and K432 occur in the same region of HSA as L430 (Sudlow site 2) and L234/L238 (Sudlow site 1), respectively. These results suggest that the glycation of HSA, even at the minimal levels examined in this report,

may have a potential effect on the binding of drugs and other small solutes at these sites on HSA.

**2.C.6. Comparisons of Methods 1 and 2 for $^{16}O/^{18}O$ ratio Determination.** The consistency of Methods 1 and 2, as used in this report for the quantitative studies, was also examined. As was mentioned earlier, Method 1 consistently gave a larger sequence coverage than Method 2. This result is expected, because the $^{16}O/^{18}O$ ratio calculations using Method 1 only involved the use of a combined $^{16}O$ and $^{18}O$ digest, while method 2 required that a $^{16}O$, $^{18}O$, and combined digest all be employed. Ideally, in this scenario all of the digests will behave similarly, but in practice not all peptides detected in the combined digest are detected in the $^{16}O$ and/or $^{18}O$ digest. A consequence of using Method 2, therefore, is that the sequence coverage of HSA will be reduced compared to Method 1. The advantage of using Method 2 is that the changes in $^{16}O/^{18}O$ ratios are more sensitive to changes in the extent of glycation, because corrections for back exchange are being made.

As shown in Figure 2.2, the distribution of $^{16}O/^{18}O$ ratios for both methods is similar. 94% of the detected peaks in both methods 1 and 2 had $^{16}O/^{18}O$ ratio less than 2.00, which was the rationale for choosing this value as a cutoff $^{16}O/^{18}O$ ratio. When looking at $^{16}O/^{18}O$ ratios calculated using methods 1 and 2 one observation is that the $^{16}O/^{18}O$ results were not particularly correlated. One explanation for the low correlation is that the theoretical isotopic distribution is different than the actual distribution that was obtained and used in Method 2. Such cases are more evident when looking at larger mass peptides or peptides that overlap and have a similar mass at 50 ppm. It is expected that using Method 2 would be advantageous in both of these cases because the actual isotopic

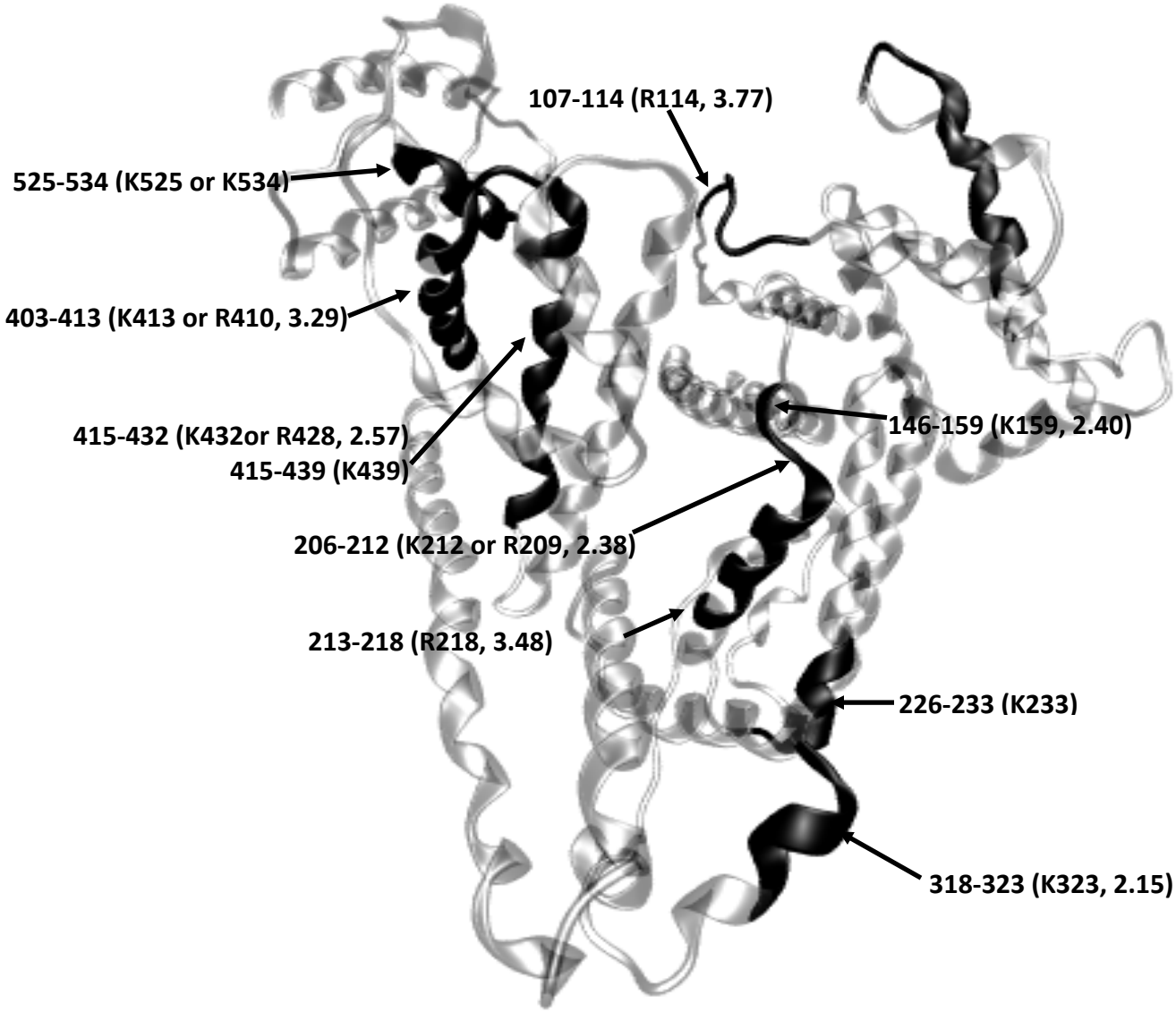
distributions in the $^{16}O$ and $^{18}O$ digests can be factored in, hence the ratio obtained would be a true value, and not an inferred result. Overall a larger number of high ratio peptides were detected using Method 2, and a larger amount of the peptides found in Method 2 were linked to modifications which were previously identified [13]. Both of these reasons indicate that Method 2 would be a better suited for quantitation, especially under conditions where small amounts of modification are taking place.

## 2.D. SUMMARY

Figure 2.3 shows an image of HSA (as created using the Visual Molecular Dynamics software (33)) that summarizes the results found in this study. This figure includes the major glycation/modification sites identified in this study and lists their $^{16}O/^{18}O$ ratios. When ranked in order of decreasing $^{16}O/^{18}O$ ratios, R114, R81, R218, K413, K432, K159, K212, and K323 were found to give significant changes in their $^{16}O/^{18}O$ ratios due to glycation or a reaction that produced AGEs. In addition to these peptides, lysines 439, 525, and 233 are also believed to contain significant amounts of modification based on this work. These possible modification sites were obtained by combining information on the observed mass shifts for the affected peptides in the digests, the $^{16}O/^{18}O$ ratios for these peptides, the location of the peptides in HSA, and the corresponding p$K_a$ and FAS values that were calculated for the given residues.

The approach discussed in this report is novel because it allows for the simultaneous quantification of glycation-related products that occur on lysine and arginine residues in HSA. It was clearly established in this work that significant glycation and AGE-modification occur at or near the two major drug binding sites of HSA, which explains the changes in binding affinity that have been noted for some drugs with glycated HSA (4, 29, 32). The results in this chapter also complement and build on previous work in the analysis of glycation related modifications (5, 34, 35), by providing a relative ranking of the amount of glycation that is found in various portions of minimally glycated HSA.

Figure 2.3. Structure of HSA, showing the peptides with significant glycation-related modifications that were found in minimally glycated HSA and the corresponding arginine or lysine that was modified in this region. The values of the measured indicated $^{16}O/^{18}O$ ratios are also provided, where applicable. This image was generated using VMD and a PDB file with accession number 1AO6 (36).


107-114 (R114, 3.77)
525-534 (K525 or K534)
403-413 (K413 or R410, 3.29)
415-432 (K432or R428, 2.57)
415-439 (K439)
146-159 (K159, 2.40)
206-212 (K212 or R209, 2.38)
213-218 (R218, 3.48)
226-233 (K233)
318-323 (K323, 2.15)

While the results from this current analysis are promising, the lack of usable sequence coverage for sites with the highest amount of glycation and the heterogeneity of the glycation products are both areas where further examination is needed. In addition, the range of $^{16}O/^{18}O$ ratios for a reference HSA digest is approximately 0.5 – 3.0 (15). This wide range, combined with the fact that the "high" $^{16}O/^{18}O$ ratios were marginally greater than the assigned cutoff value of 2.0 indicates that true quantitation is not being obtained with this current approach. Issues related to glycation product heterogeneity, lack of usable sequence coverage for sites with the highest amount of glycation, and quantitation errors are examined more thoroughly in the remaining chapters using HSA that was glycated *in vitro*.

## 2.E. REFERENCES

(1) Koyama, H.; Sugioka, N.; Uno, A.; Mori, S.; Nakajima, K. Effects of glycosylation of hypoglycemic drug binding to serum albumin. *Biopharm. Drug. Dispos.* **1997***, 18*, 791-801.

(2) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattachara, A. A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **2005***, 353*, 38-52.

(3) Peters, T. In *All About Albumin: Biochemistry, Genetics, and Medical Applications.* Academic Press: San Diego, 1996; , pp 102-126.

(4) Nakajou, K.; Watanabe, H.; Kragh-Hansen, U.; Maruyama, T.; Otagiri, M. The effect of glycation on the structure, function and biological fate of human serum albumin as revealed by recombinant mutants. *Biochim Biophys Acta.* **2003***, 1623*, 88-97.

(5) Wa, C.; Cerny, R. L.; Clarke, W. A.; Hage, D. S. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta.* **2007***, 385*, 48-60.

(6) Lapolla, A.; Fedele, D.; Seraglia, R.; Traldi, P. The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes: An update. *Mass. Spectrom. Rev.* **2006***, 25*, 775.

(7) Lapolla, A.; Fedele, D.; Traldi, P. The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes. *Mass. Spectrom. Rev.* **2000***, 19*, 279-304.

(8) Thornalley, P. J.; Langborg, A.; Minhas, H. S. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **1999***, 344*, 109-116.

(9) Kislinger, T.; Humeny, A.; Pischetsrieder, M. Analysis of protein glycation products by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Curr. Med. Chem.* **2004***, 11*, 2185-2193.

(10) Kislinger, T.; Humeny, A.; Peich, C. C.; Zhang, X.; Niwa, T.; Pischetsrieder, M.; Becker, C. M. Relative quantification of Nε-(carboxymethyl)lysine, imadazolone A, and the amadori product in glycated lysozyme by MALDI-TOF mass spectrometry. *J. Agr. Food Chem.* **2003***, 51*, 51-57.

(11) Iberg, N.; Fluckiger, R. Nonenzymatic glycosylation of albumin in vivo. *J. Biol. Chem.* **1986***, 261*, 13542-13545.

(12) Bidasee, K. R.; Nallani, K.; Yu, Y.; Cocklin, R. R.; Zhang, Y.; Wang, M.; Dincer, U. D.; Besch, H. R. Chronic diabetes increases advanced glycation end products on cardiac ryanodine receptors/calcium-release channels. *Diabetes.* **2003***, 52*, 1825-1836.

(13) Monnier, V. M.; Kohn, R. R.; Cerami, A. Accelerated age related browning of human collagen in diabetes mellitus. *Proc. Natl. Acad. Sci.* **1984***, 81*, 583-587.

(14) Zhang, Q.; Tang, N.; Schepmoes, A. A.; Phillips, L. S.; Smith, R. S.; Metz, T. O. Proteomic profiling of nonenzymatically glycated proteins in human plasma and erythrocyte membranes. *J. Proteome Res.* **2008***, 7*, 2025-2032.

(15) Wa, C.; Cerny, R. L.; Hage, D. S. Identification and quantitative studies of protein immobilization sites by stable isotope labeling and mass spectrometry. *Anal. Chem.* **2006***, 78*, 7967-7977.

(16) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic $^{18}O$ labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* **2004***, 73*, 2836-2842.

(17) Stewart, I. I.; Thomson, T.; Figeys, D. $^{18}O$ labeling: A tool for proteomics. *Rapid. Comm. Mass. Spec.* **2001***, 15*, 2456-2465.

(18) Sun, G.; Anderson, V. E. A strategy for distinguishing modified peptides based on post-digestion $^{18}O$ labeling and mass spectrometry. *Rapid. Comm. Mass. Spec.* **2005***, 19*, 2849-2856.

(19) Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using $^{18}O$-labeled internal standards. *Rapid. Comm. Mass. Spec.* **2000***, 14*, 1226-1232.

(20) Johnson, K. L.; Muddiman, D. C. A method for calculating $^{16}O/^{18}O$ peptide ion ratios for the relative quantification of proteomes. *J. Am. Soc. Mass. Spectrom.* **2004***, 15*, 437-445.

(21) Eckel-Passow, J. E.; Oberg, A. L.; Therneau, T. M.; Mason, C. J.; Mahoney, D. W.; Johnson, K. L.; Olson, J. E.; Bergen, H. R., III. Regression analysis for comparing protein samples with $^{16}O/^{18}O$ stable-isotope labeled mass spectrometry. *Bioinformatics.* **2006***, 22*, 2739.

(22) Wa, C.; Cerny, R.; Hage, D. S. Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: Analysis of human serum albumin as a model. *Anal. Biochem.* **2006***, 349*, 229-241.

(23) Stewart, I. I.; Thomson, T.; Figeys, D.; Duewel, H. S. In *The use of $^{18}O$ labeling as a tool for proteomic applications.* Conn, P. M., Ed.; Handbook of Proteomic Methods; Humana Press: Totowa, N.J., 2003; pp 145-175.

(24) Baker, P. R.; Clauser, K. R. http://prospector.ucsf.edu. Accessed: Feb 2, 2010.

(25) Wilkins, M. R.; Lindskog, I.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Hochstrasser, D. F.; Appel, R. D. Detailed peptide characterization using PEPTIDEMASS - a world-wide web accessible tool. *Electrophoresis.* **199***7, 1*8, 403-408.

(26) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *Protein identification and analysis tools on the ExPASy server.* Walker, J. M., Ed.; The Proteomics Protocols Handbook; Humana Press: Totowa, N. J., 2005; pp 571-607.

(27) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalizaion of protein pKa values. *Proteins: Struct. Funct. Bioinf.* **200***5, 6*1, 704-721.

(28) Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **200***3, 3*1, 3316-3319.

(29) Shaklai, N.; Garlick, R. L.; Bunn, H. F. Nonenzymatic glycosylation of human serum albumin alters its confirmation and function. *J. Biol. Chem.* **198***4, 25*9, 3812-3817.

(30) Takátsy, A.; Böddi, K.; Nagy, L.; Nagy, G.; Szabó, S.; Markó, L.; Wittmann, I.; Ohmacht, R.; Ringer, T.; Bonn, G. K.; Gjerde, D.; Szabó, Z. Enrichment of Amadori products derived from the nonenzymatic glycation of proteins using microscale boronate affinity chromatography. *Anal. Biochem.* **200***9, 39*3, 8-22.

(31) Rodkey, F. L. Direct spectrophotometric determination of albumin in human serum. *Clin. Chem.* **196***5, 1*1, 478-487.

(32) Vorum, H.; Fisker, K.; Otagiri, M.; Pedersen, A. O.; Hansen, U. K. Calcium ion binding to clinically relevant chemical modifications of human serum albumin. *Clin. Chem.* **199***5, 4*1, 1654-1661.

(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **199***6, 1*4, 33-38.

(34) Lapolla, A.; Fedele, D.; Reitano, R.; Bonfante, L.; Guizzo, M.; Seraglia, R.; Tubaro, M.; Traldi, P. Mass spectrometric study of in vivo production of advanced glycation end-products/peptides. *J. Mass. Spectrom.* **200***5, 4*0, 969-972.

(35) Robb, D. A.; Olufemi, S. O.; Williams, D. A.; Midgley, J. M. Identification of glycation at the N-terminus of albumin by gas chromatography - mass spectrometry. *Biochem J.* **198***9, 26*1, 871-878.

(36) Sugio, S.; Kashima, A.; Mochizuki, S.; Kobayashi, K. Crystal structure of human serum albumin at 2.5 A resolution. *Protein Eng.* **199***9, 1*2, 439-446.

## 2.F. APPENDIX

One figure and five tables are given in this section. Figure 2.1a shows the distribution of $^{16}O/^{18}O$ ratios that were obtained for Methods 1 and 2. The first four tables summarize the $^{16}O/^{18}O$ ratios that were obtained for the Glu-C digests (Tables 2.1a and 2.2a) and the Lys-C digests (Tables 2.3a and 2.4a) when using Methods 1 and 2.

**FIGURE LEGEND**

Figure 2.1a. Distribution of $^{16}O/^{18}O$ ratios found for peptides in minimally glycated HSA when using Methods 1 (■) and 2 (■).

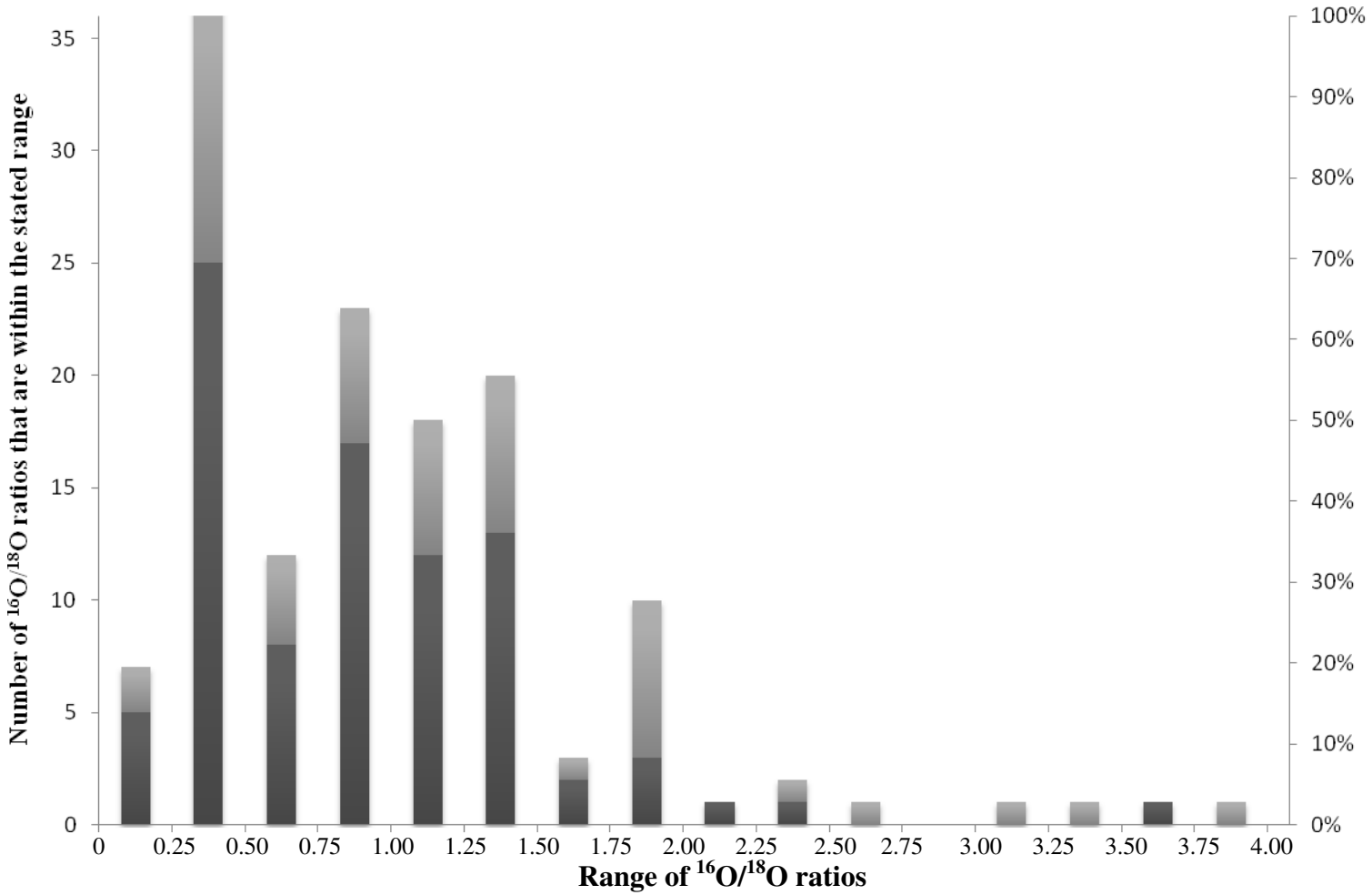

Number of $^{16}O/^{18}O$ ratios that are within the stated range
0
5
10
15
20
25
30
35
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
3.25
3.50
3.75
4.00
Range of $^{16}O/^{18}O$ ratios

Figure 2.1a

Table 2.1a. Calculated $^{16}O/^{18}O$ ratios for the Glu-C digest using Method 1

| | | **Observed peak ratios for each peptide at given %ACN in water** | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio** [a] |
| 466-479[b] | 1566.76 | | | | | 1.86 | 1.86 |
| 334-354 | 2563.41 | | | 1.41 | | | 1.41 |
| 209-227 | 2232.25 | | | | 1.31 | | 1.31 |
| 101-119[b] | 2291.17 | | | | 1.27 | | 1.27 |
| 133-141 | 1204.66 | 1.26 | | | | | 1.26 |
| 543-556 | 1583.78 | | | | | 1.19 | 1.19 |
| 7-17 | 1300.66 | 1.22 | 1.15 | | | | 1.18 (± 0.05) |
| 426-442 | 1955.99 | | 1.11 | | | | 1.11 |
| 142-153 | 1519.78 | 0.93 | 0.96 | 0.93 | 0.94 | | 0.94 (± 0.02) |
| 443-450 | 962.45 | 0.85 | | | | | 0.75 |
| 496-505 | 1197.58 | 0.8 | 0.76 | | | | 0.78 (± 0.03) |
| 369-376 | 1031.45 | 0.71 | 0.58 | | | | 0.64 (± 0.09) |
| 120-132 | 1566.62 | 0.77 | 0.47 | | | | 0.62 (± 0.21) |
| 502-518 | 1945.86 | | | | 0.54 | | 0.54 |
| 377-393 | 2086.06 | | | 0.45 | 0.45 | | 0.45 (± 0.01) |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This was determined using the method of error propagation).

[b] – CAM-cysteine was also detected on this peptide. The higher mass peptides would correspond to this modification.

Table 2.1a (continued). Calculated $^{16}O/^{18}O$ ratios for the Glu-C digest using Method 1

| | | Observed peak ratios for each peptide at given %ACN in water | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio** |
| 253-266 | 1699.74 | | 0.45 | | | | 0.45 |
| 451-465 | 1801.92 | | | | | 0.44 | 0.44 |
| 480-492 | 1548.8 | 0.45 | 0.46 | 0.48 | 0.44 | 0.36 | 0.44 (± 0.05) |
| 496-501 | 736.39 | 0.38 | | | | | 0.38 |
| 101-119[b] | 2348.19 | | 0.4 | 0.34 | 0.33 | 0.42 | 0.37 (± 0.04) |
| 61-82 | 2578.28 | | | | 0.38 | 0.31 | 0.35 (± 0.05) |
| 466-479[b] | 1680.8 | 0.31 | | | | | 0.31 |
| 87-100 | 1737.7 | 0.29 | | | | | 0.29 |
| 286-294 | 1015.49 | 0.28 | | | | | 0.28 |
| 49-57 | 983.45 | 0.26 | | | | | 0.26 |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This was determined using the method of error propagation).

[b] – CAM-cysteine was also detected on this peptide. The higher mass peptides would correspond to this modification.

Table 2.2a. Calculated $^{16}O/^{18}O$ ratios for the Glu-C digest using Method 2

| | | **Observed peak ratios for each peptide at given %ACN in water** | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio[a]** |
| | 1583.78 | | | | | 1.83 | 1.83 |
| 443-450 | 962.45 | 1.73 | | | | | 1.73 |
| 7-17 | 1300.66 | 1.32 | | | | | 1.32 |
| 466-479 | 1680.8 | 1.19 | | | | | 1.19 |
| 209-227 | 2232.25 | | | | 1.09 | | 1.09 |
| 142-153 | 1519.78 | 1.00 | 1.05 | 0.9 | | | 0.98 (± 0.08) |
| 502-518 | 1945.86 | | | | 0.98 | | 0.98 |
| 334-354 | 2563.41 | | | | | 0.89 | 0.89 |
| 49-57 | 983.45 | 0.89 | | | | | 0.89 |
| 369-376 | 1031.45 | 0.87 | 0.82 | | | | 0.85 (± 0.03) |
| 480-492 | 1548.8 | 0.83 | 0.91 | 0.85 | 0.77 | | 0.84 (± 0.06) |
| 377-393 | 2086.06 | | | 0.75 | 0.77 | | 0.76 (± 0.01) |
| 451-465 | 1801.92 | 0.57 | | | | 0.64 | 0.61 (± 0.05) |
| 120-132 | 1566.62 | 0.58 | 0.49 | | | | 0.54 (± 0.06) |
| 61-82 | 2578.28 | | | | 0.49 | 0.53 | 0.51 (± 0.03) |
| 87-100 | 1737.7 | 0.51 | | | | | 0.51 |
| 466-479 | 1566.76 | 0.53 | 0.46 | | | | 0.50 (± 0.05) |
| 101-119 | 2348.19 | | 0.39 | 0.45 | 0.46 | 0.57 | 0.47 (± 0.08) |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This method for determining the final standard deviation was determined using the method of error propagation).

Table 2.3a. Calculated $^{16}O/^{18}O$ ratios for the Lys-C digest using Method 1

| | | Observed peak ratios for each peptide at given %ACN in water | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio[a]** |
| 213-225 | 1529.87 | 1.75 | 1.51 | 1.59 | 2.64 | | 1.87 (± 0.52) |
| 546-557 | 1342.63 | | 1.7 | | | | 1.7 |
| 13-20 | 951.44 | 1.63 | | | | | 1.63 |
| 21-41 | 2490.16 | | | | | 1.48 | 1.48 |
| 138-159 | 2807.39 | | | | 1.53 | 1.22 | 1.38 (± 0.22) |
| 415-432 | 1924.09 | 1.22 | 1.49 | 1.42 | | | 1.37 (± 0.14) |
| 42-51 | 1149.62 | 1.21 | 1.52 | | | | 1.36 (± 0.21) |
| 65-73 | 1017.54 | 1.29 | 1.43 | | | | 1.36 (± 0.09) |
| 7-12 | 973.52 | 1.34 | | | | | 1.34 |
| 414-432 | 2052.18 | 1.16 | 1.34 | | | | 1.25 (± 0.13) |
| 403-413 | 1352.77 | 1.23 | 1.27 | 1.38 | | 1.06 | 1.24 (± 0.13) |
| 373-389 | 2045.1 | | | 1.29 | 1.07 | | 1.18 (± 0.16) |
| 526-534 | 1000.6 | 1.17 | | | | | 1.17 |
| 373-378 | 784.39 | 1.16 | | | | | 1.16 |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This method for determining the final standard deviation was determined using the method of error propagation).

Table 2.3a (continued). Calculated $^{16}O/^{18}O$ ratios for the Lys-C digest using Method 1

| | | Observed peak ratios for each peptide at given %ACN in water | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN** | **50% ACN** | **Average $^{16}O/^{18}O$ ratio[a]** |
| 206-212 | 854.45 | 1.4 | 1.45 | | | | 1.42 (± 0.04) |
| 403-414 | 1480.86 | 0.88 | | | | | 0.88 |
| 263-274 | 1443.64 | 0.82 | | | | | 0.82 |
| 94-106 | 1714.8 | 0.77 | | | | | 0.77 |
| 175-190 | 1827.96 | 0.75 | 0.74 | 0.78 | | | 0.76 (± 0.02) |
| 163-174 | 1371.57 | 0.72 | | | | | 0.72 |
| 501-519 | 2260.02 | 0.57 | | | | 0.51 | 0.54 (± 0.04) |
| 445-466 | 2674.31 | 0.35 | | | 0.54 | 0.35 | 0.35 (± 0.10) |
| 241-262 | 2585.12 | 0.36 | 0.34 | | | | 0.35 (± 0.01) |
| 74-93 | 2349.03 | | | 0.3 | | | 0.3 |
| 360-372 | 1552.6 | 0.27 | | | | | 0.27 |

[a] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This method for determining the final standard deviation was determined using the method of error propagation).

Table 2.4a. Calculated $^{16}O/^{18}O$ ratios for the Lys-C digest using Method 2

| | | **Observed peak ratios for each peptide at given %ACN in water** | | | | | |
|---|---|---|---|---|---|---|---|
| **Residue** | **m/z** | **5% ACN** | **10% ACN** | **20% ACN** | **30% ACN[a]** | **50% ACN[a]** | **Average $^{16}O/^{18}O$ ratio[b]** |
| 403-413 | 1352.77 | 3.41 | 2.96 | 3.50 | | | 3.29 (± 0.29) |
| 415-432 | 1924.09 | 2.46 | 2.60 | 2.64 | | | 2.57 (± 0.09) |
| 206-212 | 854.45 | 2.32 | | | | | 2.32 |
| 213-225 | 1529.87 | 2.11 | 1.66 | 1.80 | | | 1.86 (± 0.23) |
| 7-12 | 973.52 | 1.84 | | | | | 1.84 |
| 65-73 | 1017.54 | 1.78 | | | | | 1.78 |
| 42-51 | 1149.62 | 1.77 | 1.79 | | | | 1.78 (± 0.01) |
| 414-432 | 2052.18 | 1.47 | 1.42 | | | | 1.45 (± 0.03) |
| 94-106 | 1657.78 | | 1.42 | | | | 1.42 |
| 94-106 | 1714.8 | 1.29 | 1.05 | | | | 1.17 (± 0.17) |
| 175-190 | 1827.96 | 1.07 | 1.19 | | | | 1.13 (± 0.08) |
| 241-262 | 2585.12 | 0.47 | 0.46 | | | | 0.46 (± 0.01) |

[a] – Peptides in zip-tip fractions 30 % and 50 % are mislabeled. These values were not included in subsequent calculations.

[b] – For consistency, the standard deviations shown were calculated using the $^{16}O/^{18}O$ ratios obtained in the different ACN fractions. During data processing there are some instances where standard deviations were calculated for each Zip-tip fraction. In these instances, the square root of the sum of the standard deviations squared was used to determine the final standard deviation (Note: This method for determining the final standard deviation was determined using the method of error propagation).

# CHAPTER 3

# IMPROVED TECHNIQUES FOR IDENTIFYING EARLY AND ADVANCED GLYCATION ADDUCTS USING MASS SPECTROMETRY

## 3.A. INTRODUCTION

During diabetes, the formation of early and advanced glycation products can be enhanced due to the increased concentration of reducing sugars in the body. The process by which proteins are modified by glycation products is known as the Maillard reaction (1, 2). The extent of modification that occurs on a given protein depends on the amount of time that the protein is incubated with a reducing sugar (3), the type of reducing sugar that is present (4), and the environment (i.e. location in the body) where the Maillard reaction is occurring (5). The identification of the types of modification that are occurring on a given protein is often difficult because the sugars that modify serum proteins often fragment, self-oxidize, and rearrange to form reactive dicarbonyl compounds that are also capable of modifying proteins to form advanced glycation end products (AGEs) (6-9). There have been many reports indicating that many of the advanced glycation products that are found *in vivo* are linked to the production of reactive dicarbonyl compounds, namely glyoxal, methylglyoxal, or 3-deoxyglucosone (8). The end result of these of reactions is that a given protein may be modified by all of the aforementioned compounds

to form a heterogeneous set of advanced glycation end products (10), which may alter the normal function of a given protein (6, 11-13). For instance, blindness and heart failure have both linked to AGE formation that occurs on the lens crystallin protein (14) and vascular collagen protein respectively (15).

Under physiological conditions, human serum albumin (HSA) is modified to form these products; however, the extent of modification that occurs on HSA has been shown to be accelerated during diabetes (16). The heterogeneous nature of these compounds (17) makes it difficult to identify advanced glycation products that occur on HSA because many peptide and glycation product combinations overlap at mass values that are close to each other. An example of this overlapping modification pattern is shown in a previous study (1) where some of the m/z shifts that were identified in commercially glycated HSA were found to be linked to multiple glycation products. This chapter therefore investigates techniques that can be used to increase the certainty by which these glycation products are assigned.

The heterogeneous nature of many post-translational modifications in proteins has often led researchers to develop their own database search tools or programs based on a custom set of criteria (1). As an example, tools such as GlycoX (18), GlycoMod (19, 20), and VEMS (21) are commonly used for determining *N*- and *O*-linked glycosylation sites. However, there are few such tools that have been utilized in previous work to address the degree of heterogeneity that is present in adducts that result from glycation. The main objective of this project was to test a method using Matlab for determining the location of glycation sites on proteins, using human serum albumin (HSA) as a model. This study was done in collaboration with Dr. Ala Qadi (Department of Computer Science, UNL),

who generously provided the programs that were used for examining glycation in this report. Advantages of using Matlab for such work include the fact that it is commonly used for numerical computations, it is easy to learn, and it has pre-written libraries (i.e., a bioinformatics package) that can be integrated with other programs for such research. This report examined the use and several key aspects of a Matlab program that was designed for use with MALDI-TOF-MS to determine glycation adducts in a protein such as HSA.

In this report, HSA was first glycated *in vitro* and then digested using Glu-C, followed by the use of MALDI-TOF-MS to examine the resulting digest. A control sample of non-glycated HSA was treated in a similar manner. The mass values obtained from the glycated sample were compared to the control sample to identify modified peptides which only occurred in the glycated sample. The mass spectra were then compared to a theoretical digest and a sample list of modifications (i.e., see Table 3.1) using an in-house program developed in Matlab. These results were compared to the locations on HSA that are known from previous studies to be modified to form glycation products (e.g., lysines 525, 199, 281, 439, and the *N*-terminus) (22-24). The effects of mass accuracy and glycation product heterogeneity in this process were considered, as well as the possible extension of this approach to other glycated proteins.

## 3.B. EXPERIMENTAL

**3.B.1. Materials.** The following items were purchased from Sigma-Aldrich (St. Louis, MO): HSA (99% pure, essentially fatty acid and globulin free), sequence grade trypsin, guanidine HCl (99%), dithiothreitol (99%), iodoacetamide (99%), formic acid (96%), trifluoroacetic acid (98%), 2,3-dihyroxybenzoic acid (98%), α-cyano-4-hydroxycinnamic acid (99%), D-glucose (99%), potassium phosphate dibasic (99%), potassium phosphate monobasic (99%), ammonium bicarbonate (99%), sodium azide (99%) and molecular biology grade water. The glycated serum protein enzymatic assay was purchased from Diazyme Laboratories (Poway, CA).

**3.B.2. Apparatus.** The following items of equipment were obtained from Fisher Scientific (Pittsburgh, PA): Pierce Slide-A-Lyzer dialysis cassettes (7 kDa MW cutoff, 0.1–0.5 mL and 0.5–3 mL capacity), Bio-Rad 10 DG size exclusion columns (30 x 10 mL) and Millipore μ–C18 ZipTip pipette tips (5.0 μg). The mass spectra were acquired on a Voyager 6148 MALDI-TOF-MS system (Applied Biosystems, CA). The instrument settings were as follows: positive-ion delayed extraction reflection mode; delay time, 100 ns; accelerating voltage, 25 kV; guide wire voltage, 0.008% of accelerating voltage; grid voltage, 77% of accelerating voltage. Mascot Wizard was obtained from the Matrix Science website (25). Matlab 2009a, including the bioinformatics and statistics toolboxes, was purchased from Mathworks (Natick, MA). All aqueous solutions were prepared using water from a Nanopure water system (Barnstead, Dubuque, IA).

**3.B.3. Preparation of glycated HSA.** Glycated HSA was prepared by dissolving HSA to a final concentration of 42 g/L in a 3 mL solution that contained 200 mM potassium

phosphate (pH 7.4), 15 mM D-glucose, and 1 mM sodium azide in water (Note: sodium azide was added to prohibit bacterial growth; this agent is toxic, and care must be used to avoid contact of this compound with skin) (9, 26, 27). The concentration of HSA and D-glucose selected for this process represented the physiological concentration of HSA (23) and a typical blood glucose level found in diabetic patients (i.e., a level of 7 mM when fasting or a level of 11.1 mM used in the diagnosis of diabetes) (28). This mixture was incubated at 37 °C for 5 weeks. A control HSA sample was prepared in a similar manner without the addition of glucose. Following the incubation, the low mass reagents were removed by using size-exclusion chromatography. The size-exclusion column was washed with 20 mL of water and loaded with 3 mL of the HSA/glucose mixture. The protein was eluted by applying 5 mL of water to the column, with 5 mL fractions being collected. Additional purification of the collected protein fraction was obtained by performing equilibrium dialysis against two 1 L portions of water for 2 h at room temperature. An additional dialysis was performed against 1 L of water at 4 °C overnight. Lyophilization was used to remove water from the HSA for long-term storage. The resulting glycated HSA samples were stored at 4 °C until later use. The level of glycation was determined using a commercial fructosamine assay kit from Diazyme.

**3.B.4. Sample pretreatment and digestion.** Each HSA or glycated HSA sample was denatured by suspending 3 mg of the protein in a 1 mL solution containing 6 M guanidine HCl and 100 mM ammonium bicarbonate (pH 8.5). A 1 M dithiothreitol reagent solution was prepared in pH 8.5, 1 M ammonium bicarbonate buffer. A 15 μL portion of the dithiothreitol reagent was added to the protein suspension. The protein mixture was incubated at 37 °C for one hour. A 36 μL aliquot of 1 M iodoacetamide prepared in 1 M

sodium hydroxide was added to this protein mixture. This mixture was incubated in the dark for 30 min at room temperature, followed by the addition of 150 µL dithiothreitol reagent to terminate alkylation. A 0.5 mL portion of this solution was transferred to a 0.5–1 mL dialysis cassette and dialyzed against two 3 L portions of water for 4 h per cycle. The resulting solution was divided into 50 µL aliquots and brought down to dryness by using a Speedvac and then stored at -80 ºC.

Glu-C was selected for the enzymatic digestion because it cleaves specifically at glutamyl residues, and its digestion pattern is less affected by glycation related modifications than enzymes which cleave at lysine or arginine residues (29, 30). A working solution of this enzyme was prepared by dissolving Glu-C to a final concentration of 1 mg/mL in water. A 200 µL volume of 100 mM ammonium bicarbonate (pH 7.8) was added to a vial of dried HSA. The digestion with this enzyme was performed by adding 10 µL of this Glu-C solution to the reconstituted HSA or glycated HSA, followed by incubation of this mixture at 37 °C for 8 h. An additional 5 µL of the 1 mg/mL Glu-C reagent was later added to the protein/Glu-C mixture, and the solution was incubated for another 18 h. Following these incubation steps, the digestion was terminated by adding 10 µL of concentrated formic acid. The resulting digest was separated into 10 µL aliquots and stored at -80 °C until use.

**3.B.5. Mass spectrometry.** A mixture of α-cyano-4-hydroxycinnamic acid (CHCA) and 2,3-dihyroxybenzoic acid (DHB) was used as the MALDI matrix. To prepare this mixture, 20 mg/mL CHCA was prepared by dissolving CHCA in a solution containing 5% formic acid and a 70:30 (v/v) mixture of acetonitrile and water; 20 mg/mL DHB was similarly prepared by dissolving DHB in a solution containing 0.1% TFA and a 70:30

(v/v) mixture of acetonitrile and water. These DHB and CHCA solutions were then mixed in a 50:50 (v/v) ratio. A calibrant solution was also prepared that contained 25 pmol/μL des-Arg-bradykinin (904.4681 Da), 32.5 pmol/μL glu-fibrinopeptide (1570.6774 Da), and 32.5 pmol/μL angiotensin I (1296.6853 Da) in acetonitrile. A 4 μL portion of this calibrant mixture was combined with 96 μL of the CHCA and DHB matrix mixture. This calibrant-spiked matrix solution served two purposes. The first purpose was to perform external calibration by spotting this calibrant-spiked matrix in MALDI plate positions that were adjacent to the protein digests. The other purpose was to use the spiked compounds for internal calibration.

Zip-tip pipettes were used to fractionate the peptides in each digest prior to obtaining mass spectra. Peptide fractionation was accomplished by wetting a Zip-tip pipette with two 10 μL portions of 50% acetonitrile and then equilibrated with two 10 μL portions of 0.1% TFA. A protein digest sample was loaded into the Zip-tip by performing ten aspiration and dispensing cycles using a 10 μL portion of the desired digest. The sample was then desalted by washing the resin with two 10 μL portions of 0.1% TFA solution containing 5% methanol. The peptide fractions were eluted from the Zip-tip by using 1 μL solutions containing 5, 10, 20, 30, or 50% acetonitrile and 0.1% TFA in water. Each collected fraction was mixed in a 50:50 (v/v) ratio with the-calibrant-spiked matrix solution and spotted onto a 96-well plate for analysis by MALDI-TOF-MS. The MALDI plate was first air-dried, and mass spectra were then acquired.

The HSA and glycated HSA samples were first externally calibrated using a spiked-matrix solution that was adjacent to the respective protein digest on the MALDI plate. Following external calibration, the mass spectrometry data files were copied and

placed into a separate computer folder. These new copies were internally calibrated using the spiked calibrant mass values. A mass accuracy of 50 ppm was typically obtained with this approach when external calibration was used with the delayed extraction and reflector mode for mass analysis (1, 31, 32). A mass accuracy of 25 ppm was typically obtained under the same conditions when internal calibration was used (33). Based on these results, 25 ppm and 50 ppm error thresholds were used for peptide mass matching in this work.

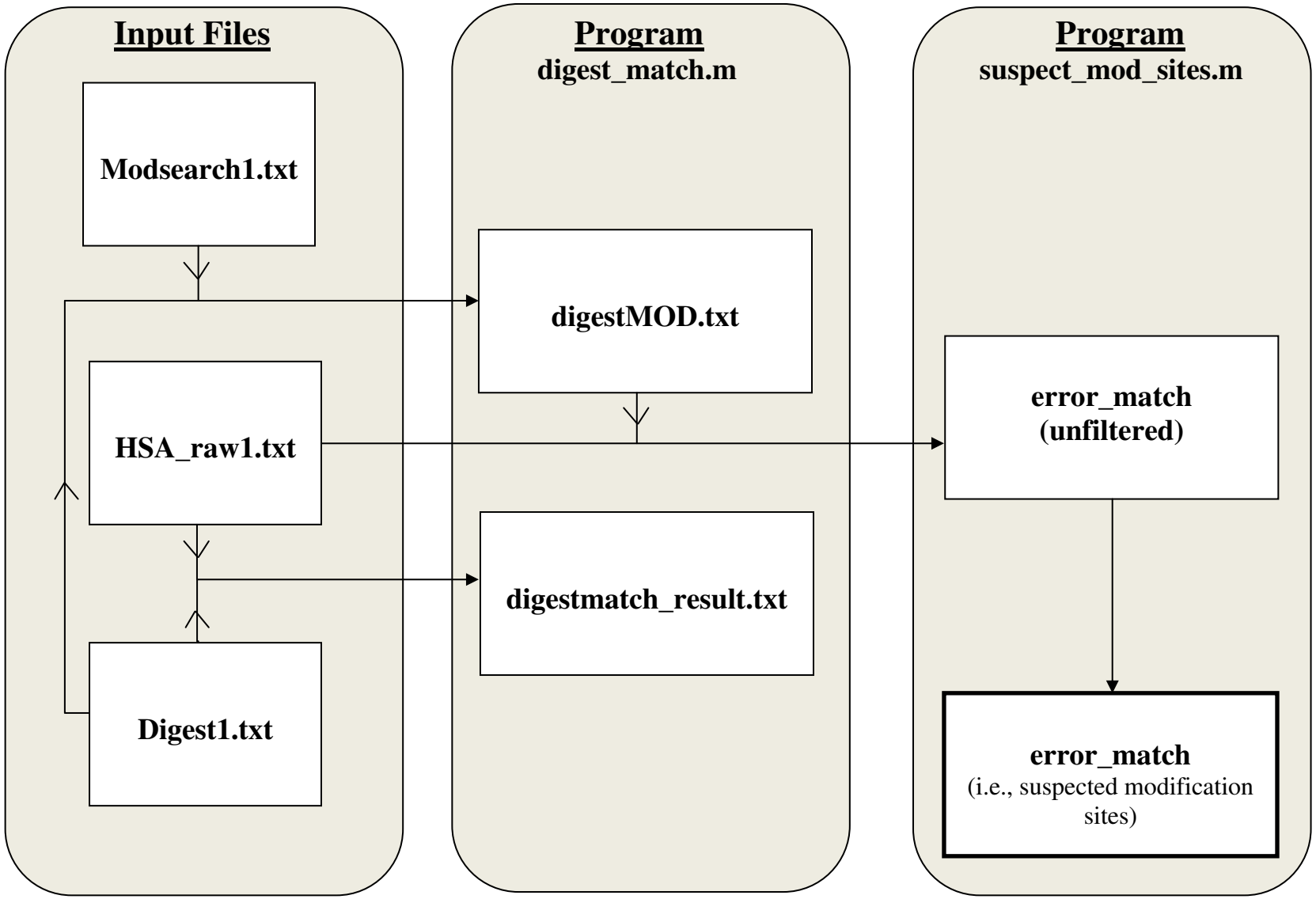
**3.B.6. Peak extraction and matlab program development.** A diagram that summarizes the assignment of glycation products based on m/z shifts is summarized in Figure 1.3. Mascot Wizard (25) is a program that is designed to automatically perform peptide mass fingerprinting directly from an acquired mass spectra. This peptide mass assignment is done by comparing a list of peptide masses (that is automatically generated) to a given database with known proteins. For this current approach, Mascot Wizard was used to generate a deisotoped mass list from each mass spectrum by first performing a database search using the raw mass spectrometry data files. Upon completion of the database search, one of the probability scores was expanded and all of the observed mass values (including the mass values that were not mapped to any given protein) were transferred to a tab delimited text file (HSA_raw1.txt). In this file, column 1 contained the mass values and column 2 contained the identity of the respective Zip-tip fraction (Note: see example in Appendix, Table 3.1a). A theoretical digest table was generated using PeptideMass, where a maximum of two missed cleavages was allowed. Carboxyamidomethyl-cysteine (cys-CAM) formation and methionine oxidation were both assigned as variable modifications during this process (34). This digest table was

copied directly to a spreadsheet, the formatting and titles were removed, and the resulting table was rearranged (Note: see example in Appendix, Table 3.2a) and saved as a tab-delimited text file (Digest1.txt).

A third table containing 119 possible early and late stage glycation adducts was generated in a similar fashion. For this table, a maximum of two modifications was allowed per peptide, as based on the possible modifications listed in Table 1.1. Each modification was assigned four integer values based on whether this modification occurred at one lysine (K1), two lysines (K2), one arginine (R1), two arginines (R2), or both a lysine and arginine (K and R residues) (see example in Appendix, Table 3.3a). This file was then saved as a tab-delimited text file (ModSearch1.txt).

Two programs, which were developed by Dr. Ala Qadi (Department of Computer Science), were used to identify the glycation-modified peptides (see Figure 3.1). The first program (digestmatch.m) was used to match the detected mass values to the theoretical digest and generate a theoretical list of modified peptides (see Appendix, Table 3.4a). The second program (suspect_mod_sites.m) was used to match the extracted mass values to the list of modified peptides (see Appendix, Table 3.5a). Further details regarding the development and execution of the script are provided in the Appendix.

Figure 3.1. Flow chart showing the programs that were used for glycation adduct identification and the respective input tables that were used with these programs.


Input Files
Program
digest_match.m
Program
suspect_mod_sites.m
Modsearch1.txt
digestMOD.txt
HSA_raw1.txt
error_match
(unfiltered)
digestmatch_result.txt
Digest1.txt
error_match
(i.e., suspected modification sites)

## 3.C. RESULTS

**3.C.1. Sequence coverage and identified modification sites.** Using the fructosamine assay, the glycated HSA sample used in this work was found to contain an average of 1.2 (± 0.1) mol hexose per mol HSA. A total of 77 m/z values were obtained from the internally-calibrated mass spectra for this sample when using Mascot Wizard. In some instances, a given peptide could be identified in more than one Zip-tip fraction. When this occurred, the detected m/z values for a given peptide were averaged prior to running the Matlab program. This approach simplified the analysis because the detected glycation products could be identified in a single run step of the program instead of running the program multiple times. In addition, averaging these m/z values served a secondary purpose in that using several individual mass measurements to calculate an average mass gave a value that was less susceptible to random variations and outliers in the results.

When the masses of these peptides were matched to a theoretical Glu-C digest at 25 ppm mass threshold, nineteen of these peptides were found to have matches (see Table 3.1), representing peptides from 44% of the primary sequence of HSA (Figure 3.2). While it is possible to obtain a higher sequence coverage by using multienzyme digestion (32), the level of coverage that was obtained in this current study was sufficient for illustrating the factors that may affect the identification of glycation sites. Seventeen of the 77 peptides that were obtained using Mascot Wizard that were found in the glycated HSA digest were not present in the control HSA digest. The program "suspect_mod_sites.m" was then executed using the masses of these unique peptides and

Table 3.1. Peptides in the internally-calibrated glycated HSA sample that matched the theoretical GluC digest at 25 ppm

| Residues | Amino Acid Sequence | Modification | MC | Mass (Da) | Error (ppm) |
|---|---|---|---|---|---|
| 369-376 | CYAKVFDE | CAM | 0 | 1031.4477 | -2 |
| 496-505 | TYVPKEFNAE | | 1 | 1197.586 | 6 |
| 133-141 | TFLKKYLYE | | 0 | 1204.647 | -12 |
| 7-17 | VAHRFKDLGEE | | 0 | 1300.658 | -5 |
| 521-531 | RQIKKQTALVE | | 0 | 1313.772 | -14 |
| 142-153 | IARRHPYFYAPE | | 0 | 1519.7815 | 1 |
| 480-492 | SLVNRRPCFSALE | CAM | 0 | 1548.79 | -3 |
| 120-132 | VDVMCTAFHDNEE | CAM | 0 | 1566.636 | 10 |
| 466-479 | KTPVSDRVTKCCTE | CAM | 0 | 1680.803 | -1 |
| 154-167 | LLFFAKRYKAAFTE | | 0 | 1704.937 | -6 |
| 87-100 | MADCCAKQEPERNE | CAM | 1 | 1737.7013 | 1 |

[a]CAM and MC are abbreviations for “carboxyamidomethyl-cysteine” and “missed cleavages”, respectively.

Table 3.1 (continued). Peptides in the internally-calibrated glycated HSA sample that matched the theoretical Glu-C digest at 25 ppm

| Residues | Amino Acid Sequence | Modification | MC | Mass (Da) | Error (ppm) |
|---|---|---|---|---|---|
| 502-518 | FNAETFTFHADICTLSE | | 1 | 1945.885 | 11 |
| 426-442 | VSRNLGKVGSKCCKHPE | CAM | 0 | 1955.9737 | -8 |
| 377-393 | FKPLVEEPQNLIKQNCE | CAM | 0 | 2086.071 | 3 |
| 189-208 | GKASSAKQRLKCASLQKFGE | CAM | 0 | 2194.146 | -14 |
| 209-227 | RAFKAWAVARLSQRFPKAE | | 0 | 2232.23 | -9 |
| 101-119 | CFLQHKDDNPNLPRLVRPE | | 0 | 2291.193 | 9 |
| 101-119 | CFLQHKDDNPNLPRLVRPE | CAM | 0 | 2348.1945 | 1 |
| 61-82 | NCDKSLHTLFGDKLCTVATLRE | CAM | 0 | 2578.282 | 3 |

[a]CAM and MC are abbreviations for “carboxyamidomethyl-cysteine” and “missed cleavages”, respectively.

Figure 3.2. Sequence coverage obtained at a mass accuracy of 25 ppm for an internally calibrated glycated HSA sample after digestion with Glu-C. The bold letters represent the sequence coverage that was obtained for the Glu-C digest. The lower case letters represent the sequence that was not covered.

dahkse**VAHR FKDLGEE**nfk alvliafaqy lqqcpfedhv
klvnevtefa ktcvadesae **NCDKSLHTLF GDKLCTVATL**
**RE**tyge**MADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV**
**DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR**
**YKAAFTE**ccq aadkaacllp kldelrde**GK ASSAKQRLKC**
**ASLQKFGERA FKAWAVARLS QRFPKAE**fae vsklvtdltk
vhtecchgdl lecaddradl akyicenqds issklkecce
kpllekshci aevendempa dlpslaadfv eskdvcknya
eakdvflgmf lyeyarrhpd ysvvlllrla ktyettlekc
caaadphe**CY AKVFDEFKPL VEEPQNLIKQ NCE**lfeqlge
ykfqnallvr ytkkvpqvst ptlve**VSRNL GKVGSKCCKH**
**PE**akrmpcae dylsvvlnql cvlhe**KTPVS DRVTKCCTES**
**LVNRRPCFSA LE**vde**TYVPK EFNAETFTFH ADICTLSE**ke
**RQIKKQTALV E**lvkhkpkat keqlkavmdd faafvekcck
addketcfae egkklvaasq aal

the modification search list containing all possible combinations of the glycation products shown in Table 1.1. When these modification sites were considered, the sequence coverage increased to 46% and four peptides were found to be modified to form glycation products (Note: of the glycated peptides that are identified, a single peptide corresponding to residues 281-294 on HSA contributed to the noted increase in sequence coverage). The modification sites were confirmed by a manual examination of the mass spectra. An example of a typical result that was obtained through this manual observation is shown in Figure 3.3(i), where the mass of an unmodified peptide, encompassing residues 426-442, is shown along with its corresponding modification that occurs at 2107.99 Da (Figure 3.3(ii)). When the control spectra were examined, the same modification was barely detectable (Figure 3.3(iii)) and there was no evidence of an isotopic cluster.

The four suspected modification sites encompass residues 281-294, 426-442, 426-442 and 189-208 on HSA, which were being modified in this case to give $N_{\varepsilon}$-carboxyethyl-lysine (CEL) plus fructosyl-lysine (FL-2$H_2O$), pyralline (Pyr) plus $N_{\varepsilon}$-(5-hydro-4-imidazolon-2-yl)ornithine (G-H1), CEL plus argpyrimidine (ArgP), and Pyr plus imadazolone B (IB), respectively (see Table 3.2(i)). Lysines 281 and 286 are the only lysines within residues 281-294. Peptide modifications that form CEL and FL-2$H_2O$ can only occur on lysine residues, so it appears that K281 and K286 are being modified in the region 281-294 to form these adducts. The region containing residues 426-442 was identified as a potential modification site in two instances, leading to the formation of Pyr + G-H1 and CEL + ArgP. Lysines 432, 436, and 439 and arginine 428 are found within residues 426-442. Because there is only a single arginine on this peptide, the arginine-

Figure 3.3. Several representative mass spectra obtained during the analysis of glycation adducts in HSA.

(i) The mass spectrum represents a peptide (from residues 426 – 442) that was found during analysis of the internally-calibrated control HSA sample; a peptide with the given mass was identified in the same Zip-tip fraction (20% acetonitrile) using both control and glycated HSA).

(ii) The mass spectrum represents a peptide from the internally-calibrated glycated HSA sample that corresponded to residues 426 – 442, with a mass shift of 152 Da versus non-glycated HSA. These results indicate that the detected peptide contains CEL and ArgP. This peptide was only identified in the 30% acetonitrile Zip-tip fraction.

(iii) The mass spectrum represents the internally-calibrated control HSA sample and shows that a mass shift of 152 Da on residues 426 – 442 has a very low intensity. In this case, there is a peak that occurs at 2109 Da, however, this peak is not associated with an isotope cluster and it is 1 Da above the expected mass of 2108 Da. This indicates that either there is an overlapping peptide, or the signal of the modified peptide is too low to be observed.

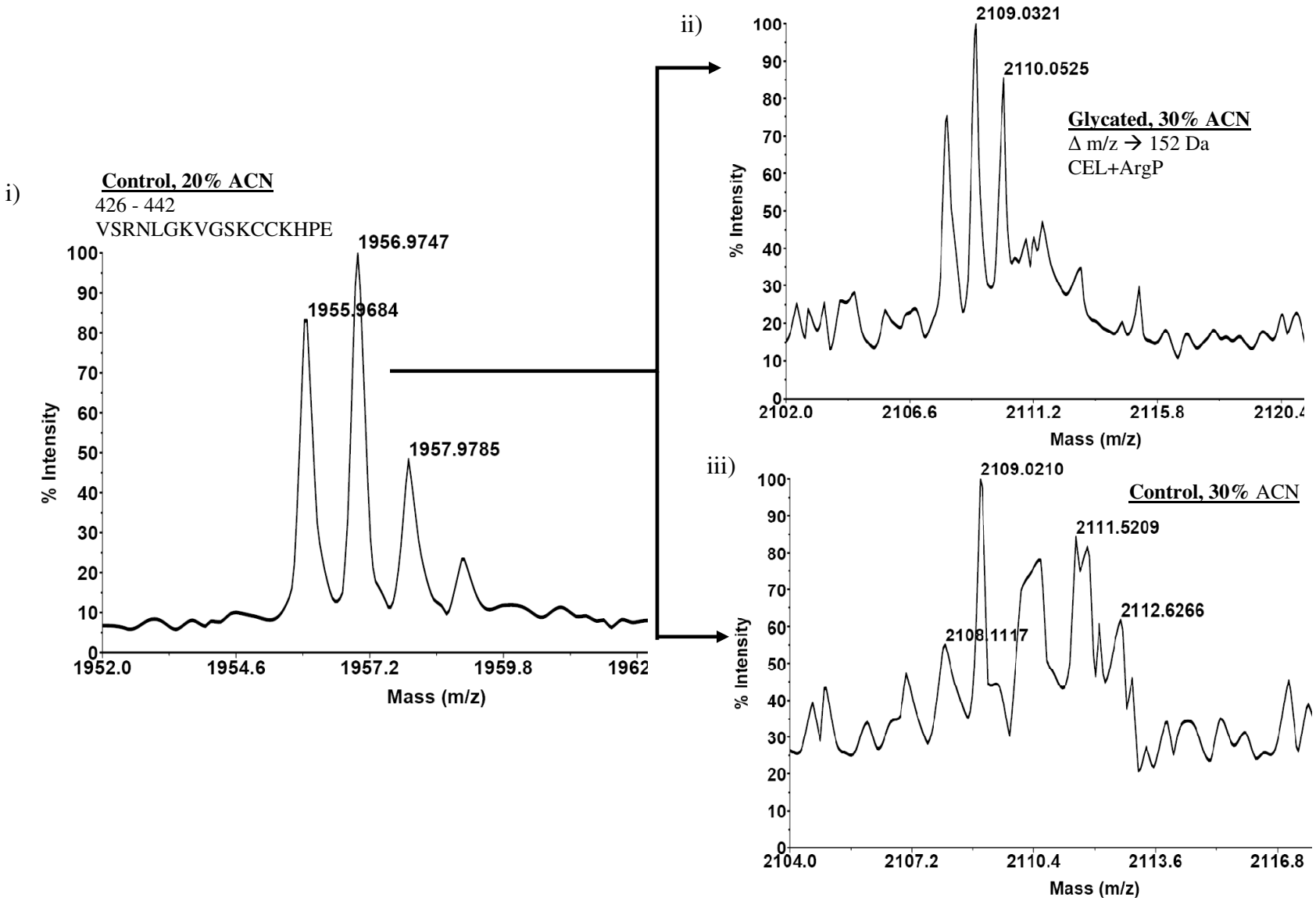

i)
Control, 20% ACN
426 - 442
VSRNLGKVGSKCCKHPE
1956.9747
1955.9684
1957.9785
% Intensity
Mass (m/z)
ii)
2109.0321
2110.0525
Glycated, 30% ACN
Δ m/z → 152 Da
CEL+ArgP
% Intensity
Mass (m/z)
iii)
2109.0210
Control, 30% ACN
2111.5209
2112.6266
2108.1117
% Intensity
Mass (m/z)

Table 3.2. Suspected glycation sites that were identified using a 25 ppm and 50 ppm mass accuracy[a]

i) HSA (internal calibration) mass values matched to DigestMOD at 25 ppm

| **Residues** | **Amino Acid Sequence** | **Raw Mass (Da)** | **ΔMass (Da)** | **Modifications** | **Error (ppm)** |
|---|---|---|---|---|---|
| 281-294 | KPLLEKSHCIAEVE | 1850.9650 | 198.0528 | CEL+FL-2$H_2O$ | -24 |
| 426-442 | VSRNLGKVGSKCCKHPE | 1989.9650 | 148.0160 | Pyr+G-H1 | -1 |
| 426-442 | VSRNLGKVGSKCCKHPE | 2108.0300 | 152.0473 | CEL+ArgP | 4 |
| 189-208 | GKASSAKQRLKCASLQKFGE | 2444.2070 | 250.0477 | Pyr+IB | 7 |

ii) HSA (external calibration) mass values that were modified by glycation at 50 ppm

| **Residues** | **Amino Acid Sequence** | **Raw Mass (Da)** | **Δ Mass (Da)** | **Modifications** | **Error (ppm)** |
|---|---|---|---|---|---|
| 267-277 | NQDSISSKLKE | 1374.7370 | 126.0317 | FL-2$H_2O$ | -45 |
| *426-442* | *VSRNLGKVGSKCCKHPE* | *1990.0300* | *148.0160* | *Pyr+G-H1* | *-34* |
| 278-292 | CCEKPLLEKSHCIAE | 2107.9890 | 234.0528 | FL-2$H_2O$+Pyr | -36 |
| 278-292 | CCEKPLLEKSHCIAE | 2107.9890 | 234.0739 | FL+CEL | -26 |
| *426-442* | *VSRNLGKVGSKCCKHPE* | *2107.9890* | *152.0473* | *CEL+ArgP* | *23* |
| 426-442 | VSRNLGKVGSKCCKHPE | 2176.1140 | 220.0583 | FL+CML | -30 |
| 267-285 | NQDSISSKLKECCEKPLLE | 2444.2180 | 166.0266 | CML+Pyr | -35 |
| *189-208* | *GKASSAKQRLKCASLQKFGE* | *2444.2180* | *250.0477* | *Pyr+IB* | *2* |

[a]Each raw mass value listed is the mass that was obtained using Mascot Wizard. The suspected modifications and their corresponding mass shifts are also shown. The values in the boxes highlight where a mass value could be assigned to multiple glycation-related modifications. The italicized values represent glycation-related modifications that were identified in both the internally and externally-calibrated samples.

specific adducts G-H1 and ArgP can safely be assigned to R428 in this example. It has been shown previously that the proximity of positively-charged residues and disulfide bonds to lysine residues increases the reactivity of lysine towards glucose (22). By looking for lysine residues that are close to positively-charged residues, it was possible to identify possible reactive sites. Lysine 439 is expected to be more reactive than other lysines found on this peptide because of its proximity to H440. It is therefore likely that K439 was being modified by both Pyr and CEL.

Residues 189-208 on HSA have lysines 190, 195, 199, and 205, as well as arginine 197. R197 is the only arginine present on in this region, so this residue is likely being modified to give IB based on the results obtained in this study. Lysine 199 is most likely being modified by Pyr because of its proximity to C200, which places H247 in a position that is adjacent to K199. The estimated p$K_a$ values for the lysine residues can also be used to indicate the reactivity of a lysine residue to give glycating products, where lysines with lower p$K_a$ values would be expected to be more reactive with glucose or other sugars at a physiological pH. The p$K_a$ values for the lysine residues in HSA were calculated using PROPKA (35). K199 was estimated to have a relatively low p$K_a$ of 7.9 (compared to an average lysine p$K_a$ of 10.1 in HSA), which further supports the belief that this reside should have good reactivity to glucose during glycation.

**3.C.2. Effects of mass accuracy and glycation product heterogeneity.** The effects of mass accuracy were examined by comparing the modifications identified in the internally-calibrated sample (i.e., with a 25 ppm error limit) to the modifications identified when external calibration was used (i.e., with a 50 ppm error limit) using only peptides that were found in the glycated HSA samples. The sequence coverage for both

samples where quite close, where the externally-calibrated sample had 43% coverage and the internally-calibrated sample had 44% coverage. This small difference occurred because three unmodified peptides occurring on residues 521-531, 466-479, and 87-100 that were identified when internal calibration was used were not identified when external calibration was used. There were also two unmodified peptides occurring on residues 334-354 and 451-465 that were only identified in the externally-calibrated sample. With the exception of these five peptides, similar peptides were identified when both internal and external calibration were used.

The heterogeneous nature of glycation products can have a significant effect on adducts that are assigned at any given level of mass accuracy. An example of this effect occurred after the program was executed and five glycated peptides were found in the externally-calibrated sample (see Table 3.2(ii)) compared to three glycated peptides that were found in the internally-calibrated sample. Of these peptides, only two peptides (426-442 and 189-208) were identified at both levels of mass accuracy. The CEL and FL-$2H_2O$ modifications that were found on residues 281-294 in the internally-calibrated sample were not identified when external calibration was used. In addition, there were five modifications located on three peptides that were identified only when external calibration was used. These results indicated that using a technique with higher mass accuracy resulted in increased sequence coverage and reduced the number of assigned glycation products, thus enhancing the confidence of glycation product assignment.

Mass accuracy considerations can be especially important when a m/z value can be linked to multiple modifications that have close mass values or when different peptide masses and modification mass combinations have similar m/z values. For instance, a

peptide with a m/z value of 2107.99 Da was identified in an externally-calibrated HSA sample. This mass could be linked to residues 278-292 with either a combination of FL-$2H_2O$ plus Pyr modifications or FL plus CEL. The resulting difference in mass between these two combinations is 0.0281 Da, or 13 ppm. The same mass could also be linked to residues 426-442 on HSA after modification to form CEL plus ArgP adducts. This latter set of modifications would result in a peptide that has a calculated mass of 2108.038 Da, representing a 49 - 58 ppm mass shift compared to the modifications occurring on residues 278-292. When the mass accuracy was increased to 25 ppm, 426-442 was the only residue which was could be matched to a glycation product, showing that these potential glycation sites could be better differentiated when a higher mass accuracy is used.

**3.C.3. Effects of peak selection criteria and adduct heterogeneity.** Another effect that was examined was the effect of peak selection on the amount and types of modification sites that were identified. The examination of this effect was done by comparing modifications found in the entire data versus those peptides that were unique to the glycated HSA. Protein samples that were calibrated internally were used for both of these comparisons. When the entire data set was considered, a total of fourteen m/z values (see Table 3.3) were identified as being modified. Of these fourteen identified m/z values, there were four instances where a given m/z value could be mapped to either a modified or unmodified peptide. When only the unique peptides were considered, four peptides were identified as being modified and none of these four peptides overlapped with the predicted m/z values for unmodified peptides. The peptide overlap pattern

clearly illustrates that it is advantageous to use unique peptides for the assignment of modification sites because these sites can be assigned with very little ambiguity.

It is expected that if a greater number of masses are being used to search for glycation adducts, a larger number of false-hits would occur because of the sheer number of glycation-modified peptides that may be present. For instance, the theoretical digest for Glu-C has 204 values and the glycation adduct list has 119 possible values, resulting in a modified peptide table containing 24,276 values. This issue would be further compounded when techniques such as multi-enzyme digestion are used to increase the sequence coverage. Using the unique peptides in this instance would be useful because the changes in glycation which occur on HSA can be directly linked to these modifications, increasing the confidence at which these assignments can be made.

Table 3.3. Modifications detected in internally-calibrated glycated HSA sample at 25 ppm when all peptides were used[a]

| Residue | Amino Acid Sequence | Raw Mass (Da) | ΔMass (Da) | Modifications | Error (ppm) |
|---|---|---|---|---|---|
| *443-450* | *AKRMPCAE* | *1031.4477* | *126.0317* | *FL-2H2O* | *17* |
| *443-450* | *AKRMPCAE* | *1031.4477* | *126.0317* | *CEL+MG-H1* | *17* |
| 312-321 | SKDVCKNYAE | 1552.6660 | 396.1057 | FL-2H2O+AFGP | -19 |
| *267-277* | *NQDSISSKLKE* | *1680.8030* | *432.1268* | *FL+AFGP* | *-20* |
| *572-585* | *GKKLVAASQAALGL* | *1704.9370* | *378.0951* | *Pyr+AFGP* | *-19* |
| 480-492 | SLVNRRPCFSALE | 1801.8590 | 310.0689 | G-H1+AFGP | -9 |
| 281-294 | KPLLEKSHCIAEVE | 1850.9650 | 198.0528 | CEL+FL-2H2O | -24 |
| *278-292* | *CCEKPLLEKSHCIAE* | *1945.8850* | *72.0211* | *CEL* | *-2* |

[a]The values in boxes indicate where a mass value could be assigned to multiple glycation products. The italicized values represent mass values that overlapped with those for some unmodified peptides in the theoretical digest.

Table 3.3 (continued). Modifications detected in internally-calibrated glycated HSA sample at 25 ppm when all peptides were used[a]

| Residue | Amino Acid Sequence | Raw Mass (Da) | ΔMass (Da) | Modifications | Error (ppm) |
|---|---|---|---|---|---|
| 426-442 | VSRNLGKVGSKCCKHPE | 1989.9650 | 148.0160 | Pyr+G-H1 | -1 |
| 426-442 | VSRNLGKVGSKCCKHPE | 2108.0300 | 152.0473 | CEL+ArgP | 4 |
| 566-585 | TCFAEEGKKLVAASQAALGL | 2194.1460 | 130.0266 | CEL+CML | -18 |
| 480-495 | SLVNRRPCFSALEVDE | 2216.0210 | 324.0846 | MG-H1+AFGP | -2 |
| 359-376 | KCCAAADPHECYAKVFDE | 2314.9440 | 144.0423 | FL-1H2O | -1 |
| 359-376 | KCCAAADPHECYAKVFDE | 2314.9440 | 144.0422 | CEL*2 | -1 |
| 228-244 | FAEVSKLVTDLTKVHTE | 2331.1490 | 414.1163 | FL-1H2O+AFGP | 0 |
| 566-585 | TCFAEEGKKLVAASQAALGL | 2331.1490 | 324.1056 | FL*2 | 6 |
| 101-119 | CFLQHKDDNPNLPRLVRPE | 2331.1490 | 39.9949 | G-H1 | 7 |
| 189-208 | GKASSAKQRLKCASLQKFGE | 2444.2070 | 250.0477 | Pyr+IB | 7 |

[a]The values in boxes indicate where a mass value could be assigned to multiple glycation products. The italicized values represent mass values that overlapped with those for some unmodified peptides in the theoretical digest.

## 3.D. SUMMARY

This chapter illustrates the use of MALDI-TOF-MS and Matlab for rapidly identifying potential glycation-related modifications that occur on HSA. By selecting the unique peptides and using a 25 ppm mass accuracy, residues 281-294, 426-442 (modified twice) and 189-208 of HSA were clearly shown as being modified to form CEL + FL-$2H_2O$, Pyr + G-H1, CEL + ArgP, and Pyr + IB respectively. The most likely modification sites based on the structure of HSA, and the proximity of these residues to positively charged amino acids would be lysines 281, 286, 439, and 199 as well as arginines 197 and 428.

Three key considerations were described in this chapter which affect the accuracy and certainty by which glycation adducts can be assigned. The first consideration was the effect of mass accuracy, where a reduction in mass accuracy was found to result in peptide modifications being assigned with greater certainty because glycation products with similar masses could be differentiated. The second consideration examined was the effect of adduct heterogeneity, where it was found that using an expanded glycation product search list is necessary for an assignment to be made because many peptide and glycation product combinations may have similar overlapping masses at a given level of mass accuracy. The third consideration explored was that some form of peak selection criteria is necessary for increasing the confidence of assigning glycation products. By looking at the unique peptides that were present in a glycated HSA sample, it was possible to link modifications directly to the experimental conditions.

While the overall level of selectivity for this method is good, in some instances it is difficult to provide information about the particular amino acid residues that are being modified in the given protein. In addition, other information, such as the location of positively-charged residues with respect to glycation sites, can be used to help infer which amino acids are the most likely sites for the production of glycation adducts (22). Other information based on p$K_a$ values and fractional-accessible surface area for a given lysine or arginine residue can also be used to aid in these assignments (1, 31). It is also possible to provide additional selectivity by coupling these techniques with tandem mass spectrometry (9, 36), fluorescence (37), and quantitative proteomics (29, 31). The mass accuracy considerations, peak selection criteria, and amino-acid reactivity guidelines that are outlined in this chapter are not limited to glycation occurring on HSA. As such, other proteins can be analyzed using the same guidelines.

## 3.E. REFERENCES

(1) Wa, C.; Cerny, R. L.; Clarke, W. A.; Hage, D. S. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta.* **2007***, 385*, 48-60.

(2) Nursten, H. In *The Maillard Reaction;* Royal Society of Chemistry: Cambridge, UK, 2005, pp 214.

(3) Bank, R. A.; Baylis, M. T.; Lafeber, F. P.; Maroudas, A.; Tekoppele, J. M. Ageing and zonal variation in post-translational modification of collagen in normal human articuar cartilage. The age-related increase in non-enzymatic glycation affects iomechanical properties of cartilage. *Biochem J.* **1998***, 330*, 345-351.

(4) Syrovy, I. Glycation of albumin: Reaction with glucose, fructose, galactose, ribose or glyceraldehyde measured using four methods. *J. Biochem. Biophys. Meth.* **1994***, 28*, 115-121.

(5) Ulrich, P.; Cerami, A. Protein glycation, diabetes, and aging. *Recent Progr. Horm. Res.* **2001***, 56*, 1-21.

(6) Zhang, Q.; Ames, J. M.; Smith, R. D.; Baynes, J. W.; Metz, T. O. A perspective on the maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. *J. Proteome Res.* **2009***, 8*, 754-769.

(7) Thornalley, P. J.; Langborg, A.; Minhas, H. S. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **1999***, 344*, 109-116.

(8) Ahmed, N.; Argirov, O. K.; Minhas, H. S.; Cordeiro, C. A. A.; Thornalley, P. J. Assay of advanced glycation endproducts (AGEs): surveying AGEs by chromatographic assay with derivitization by 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and application to $N_{\varepsilon}$-carboxymethyl-lysine- and $N_{\varepsilon}$-(1-corboxyethyl)lysine-modified albumin. *Biochem. J.* **2002***, 364*, 1-14.

(9) Lapolla, A.; Fedele, D.; Reitano, R.; Arico, N. C. Enzymatic digestion and mass spectrometry in the studies of advanced glycation end products/peptides. *J. Am. Soc. Mass. Spectrom.* **2004***, 15*, 496-509.

(10) Henle, T.; Deppisch, R.; Beck, W.; Hergesell, O.; Hansch, G. M.; Ritz, E. Advanced glycation end-products (AGE) during haemodialysis treatment: discrepant results with different methadologies reflecting the heterogeneity of AGE compounds. *Nephrol. Dial. Transplant.* **1999***, 14*, 1968-1975.

(11) Hedrick, C. C.; Thorpe, S. R.; Fu, M. X.; Harper, C. M.; Yoo, J.; Kim, S. M.; Wong, H.; Peters, A. L. Glycation impairs high-density lipoprotein function. *Diabetologia.* **2000***, 43*, 312-320.

(12) Lui, J.; Masure, M. R.; Vatner, D. E.; Jyothirmayi, G. N.; Regan, T. J.; Vatner, S. F.; Meggs, L. G.; Malhora, A. Glycation end-products cross-link breaker reduces collagen and improves cardiac function in aging diabetic heart. *Am. J. Physiol. Heart Circ. Phsol.* **2003***, 285*, H2587-H2591.

(13) Magalhaes, P. M.; Appell, H. J.; Duarte, J. A. Involvement of advanced glycation end products in the pathogenesis of diabetic complications: the protective role of regular physical activity. *Eur. Rev. Aging. Phys. Act.* **2008***, 5*, 17-29.

(14) Jansirani; Anathanaryanan, P. H. A comparitive study of lens protein glycation in various forms of cataracht. *Indian J. Clin. Biochem.* **2004***, 19*, 110-112.

(15) Smit., A. J.; Hartog, J. W. L.; Voors, A. A.; Veldhuisen, D. J. V. Advanced glycation endproducts in chronic heart failure. *Ann. New York Acad. Sci.* **2008***, 1126*, 225-230.

(16) Cohen, M. P.; Chen, S.; Ziyadeh, F. N.; Shea, E.; Hud, E. A.; Lautenslager, G. T.; Shearman, C. W. Evidence linking glycated albumin to altered glomerular nephrin and VEGF expression, proteinuria, and diabetic nephropathy. *Kidney Intl.* **2005***, 68*, 1554-1561.

(17) Kislinger, T.; Humeny, A.; Pischetsrieder, M. Analysis of protein glycation products by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Curr. Med. Chem.* **2004***, 11*, 2185-2193.

(18) An, H. J.; Tillinghast, J. S.; Woodruff, D. L.; Rocke, D. M.; Lebrilla, C. B. A new computer program (GlycoX) to determine simultaneously the glycosylation sites and oligosaccharide heterogeneity of glycoproteins. *J. Proteome Res.* **2006***, 5*, 2800-2808.

(19) Cooper, C. A.; Gasteiger, E.; Packer, N. In *Predicting glycan composition from experimental mass using GlycoMod.* Conn, M. P., Ed.; Handbook of Proteomic Methods; Humana Press: Trenton, NJ, 2003; pp 225-232.

(20) Cooper, C. A.; Gasteiger, E.; Packer, N. GlycoMod - a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics.* **2001***, 1*, 340-349.

(21) Matthiesen, R.; Trelle, M. B.; Hojrup, P.; Bunkenborg, J.; Jensen, O. N. VEMS 3.0: Algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J. Proteome Res.* **2005***, 4*, 2338-2347.

(22) Iberg, N.; Fluckiger, R. Nonenzymatic glycosylation of albumin in vivo. *J. Biol. Chem.* **1986***, 261*, 13542-13545.

(23) Peters, T. In *All About Albumin: Biochemistry, Genetics, and Medical Applications.* Academic Press: San Diego, 1996, pp 102-126.

(24) Robb, D. A.; Olufemi, S. O.; Williams, D. A.; Midgley, J. M. Identification of glycation at the N-terminus of albumin by gas chromatography - mass spectrometry. *Biochem J.* **1989***, 261*, 871-878.

(25) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **1999***, 20*, 3551-3567.

(26) Ney, K. A.; Karen J. Colley, K. J.; Pizzob, S. V. The standardization of the thiobarbituric acid assay for nonenzymatic glucosylation of human serum albumin. *Anal. Biochem.* **1981***, 118*, 294-300.

(27) Voziyan, P. A.; Khalifah, R. G.; Thibaudeau, C.; Yildiz, A.; Jacob, J.; Serianni, A. S.; Hudson, B. G. Modification of proteins in vitro by physiological levels of glucose. *J. Biol. Chem.* **2003***, 278*, 46616-46624.

(28) Gavin, J. R.; et al. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care.* **2004***, 25*, S5-S20.

(29) Capote, F. P.; Scherl, A.; Mu''ller, M.; Waridel, P.; Lisacek, F.; Sanchez, J. C. Glycation isotopic labelling (GIL) with 13C-reducing sugars for quantitative analysis of glycated proteins in human plasma. *Mol Cell Proteomics.* **2009***, in press*.

(30) Capote, F. P.; Sanches, J. C. Strategies for proteomic analysis of non-enzymatically glycated proteins. *Mass Spectrom. Rev.* **2008***, 28*, 135-146.

(31) Wa, C.; Cerny, R. L.; Hage, D. S. Identification and quantitative studies of protein immobilization sites by stable isotope labeling and mass spectrometry. *Anal. Chem.* **2006***, 78*, 7967-7977.

(32) Wa, C.; Cerny, R.; Hage, D. S. Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: analysis of human serum albumin as a model. *Anal. Biochem.* **2006***, 349*, 229-241.

(33) Edmondson, R. D.; Russel, D. H. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass measurement accuracy by using delayed extraction. *J. Am. Soc. Mass. Spectrom.* **1996***, 7*, 995-1001.

(34) Wilkins, M. R.; Lindskog, I.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Hochstrasser, J. F.; Appel, R. D. Detailed peptide characterization using PEPTIDEMASS - a World-Wide-Web-accessible tool. *Electrophoresis.* **1997***, 18*, 403-408.

(35) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein $pK_a$ values. *Proteins.* **2005***, 61*, 704-721.

(36) Gadgil, H. S.; Bondarenko, P. V.; Treuheit, M. J.; Ren, D. Screening and sequencing of glycated proteins by neutral loss scan LC/MS/MS method. *Anal. Chem.* **2007***, 79*, 5991-5999.

(37) Miyata, T.; Taneda, S.; Kawai, R.; Yasuhiko, U.; Horiuchi, S.; Hara, M.; Maeda, K.; Monnier, V. M. Identification of pentosidine as a native structure for advanced glycation end products in microglobulin-containing amyloid fibrils in patients with dialysis-related amyloidosis. *Proc. Natl. Acad. Sci.* **1996***, 93*, 2353-2358.

## 3.F. APPENDIX

A detailed description of the programs that were written in Matlab for identifying glycation sites is given here. Examples of the raw data file, the theoretical digest table, and the modification search list that were read by the programs are given in Tables 3.1a, 3.2a, and 3.3a respectively. The code for the program “digestmatch.m” is given in Table 3.4a and the code for the program “suspectmodsites.m” is given in Table 3.5a.

**3.F.1. Development and execution of the matlab programs.** A flow chart for the programs "digest_match.m "and "suspect_mod_sites.m" is shown in Figure 3.1. The code for both of these programs are given in more detail in Tables 4s and 5s. The first of these programs (digestmatch.m) compared a raw list of masses with mass values that were predicted for a theoretical HSA digest in order to determine which unmodified peptides were present in a sample. The first step of this program involved assigning a matrix containing the number of raw mass values (i.e., *i* rows) and the number of values used for the theoretical digest (i.e., *j* columns) (lines 1-5). The difference between the raw mass value and the digest mass value were calculated (lines 9-13). Following this, a new matrix was assigned in a similar fashion to determine the error, in ppm. A table of results titled "digestmatch_result" was generated with the columns shown in Table 5s (lines 47-51). A second calculation was performed which added all of the modification masses to all of the theoretical digest masses. The results from this calculation (Mass2) were output to a table named "DigestMOD". Both of these tables were exported to text files named "digestmatch_result.txt" and "DigestMOD.txt", respectively.

A second program (suspect_mod_sites.m) was used to read the files DigestMOD.txt and HSA_Raw1.txt and set up a new matrix with Mass2 by MassHSA number of cells (lines 1-13). An error value (in ppm) is calculated for all of these mass values. If this error was less than the desired threshold, the amino acid sequence was searched with the aacount function (Note: This function is included in the Matlab bioinformatics toolbox) to determine whether the indicated modification was possible. When aacount was used, the sequence had to first be converted to characters and the aacount function had to be assigned (lines 17 and 18). In this case, the function was assigned to "AA", where AA.K and AA.R give a count of the total number of lysine and

arginine residues, respectively, for a given peptide in the theoretical digest. This lysine and arginine count was compared to the total number of lysine and arginine residues that were needed for a given modification. To determine the number of lysine or arginine residues that are present for a given modification, a table was developed where the number "1" was placed in K1, K2, R1, or R2 if the modification occurs on one lysine, two lysines, one arginine, or two arginines respectively. "If" statements can then be used to determine where the modification occurs, and to see if the assigned modification is possible given the lysine/arginine count. For example, an "if" statement could be used to determine that a modification was in K1, then AA.K (i.e. the total number of lysines) could be calculated, where any AA.K value greater than one will be assigned as a glycation site (lines 20-24). The final result was reported in five tables with the general name "error_matchn", in which n can be K1, R1, R2, K2, or KR. These values were then copied into a spreadsheet and filtered as desired. A separate program was created based on similar techniques that compared the control HSA digest to the glycated digest to determine which the m/z values were unique to the glycated HSA digest. Both the unique peptides and the entire dataset were analyzed.

Table 3.1a. Example of input for the file "HSA_raw1.txt"[a]

| **Mass (Da)** | **Zip-Tip Fraction**[a] |
|---|---|
| 650.0880 | 5 |
| 656.0970 | 5 |
| 760.3660 | 10 |
| 760.8700 | 10 |
| 777.0890 | 5 |
| 855.1110 | 5 |
| 860.4120 | 5 |
| 861.1160 | 5 |
| 912.5440 | 5 |
| 1014.3710 | 10 |
| 1019.5370 | 10 |
| 1031.4030 | 10 |
| 2086.0540 | 20 |
| 2087.0910 | 30 |
| 2176.2360 | 30 |
| 2192.3310 | 30 |
| 2447.3780 | 50 |
| 2523.3570 | 50 |
| 2547.1760 | 50 |
| 2563.4550 | 50 |

[a]The information in the top row of this table should not be included in the HSA_raw1.txt input file. The "Zip-tip fraction" column indicates the percent of acetonitrile that was used to elute the indicated peptides from the Zip-tip pipette tips.

Table 3.2a. Example of input for the file "Digest1.txt"[a]

| Residues | MC | Mass (Da) | Amino Acid Sequence | CAM | OX[a] |
|---|---|---|---|---|---|
| 496-518 | 2 | 2720.2548 | TYVPKEFNAETFTFHADICTLSE | CAM | |
| 543-565 | 1 | 2718.2571 | QLKAVMDDFAAFVEKCCKADDKE | CAM | |
| 120-141 | 1 | 2711.2367 | VDVMCTAFHDNEETFLKKYLYE | | OX |
| 185-208 | 1 | 2707.4307 | LRDEGKASSAKQRLKCASLQKFGE | CAM | |
| 133-153 | 1 | 2705.4238 | TFLKKYLYEIARRHPYFYAPE | | |
| 120-141 | 1 | 2695.2418 | VDVMCTAFHDNEETFLKKYLYE | | |
| 98-119 | 1 | 2690.3579 | RNECFLQHKDDNPNLPRLVRPE | | |
| 245-266 | 1 | 2684.1207 | CCHGDLLECADDRADLAKYICE | CAM | |
| 298-321 | 1 | 2670.2425 | MPADLPSLAADFVESKDVCKNYAE | CAM | |
| 496-518 | 2 | 2663.2333 | TYVPKEFNAETFTFHADICTLSE | | |
| 185-208 | 1 | 2650.4093 | LRDEGKASSAKQRLKCASLQKFGE | | |
| 443-465 | 1 | 2647.3040 | AKRMPCAEDYLSVVLNQLCVLHE | | |
| 443-465 | 1 | 2631.3091 | AKRMPCAEDYLSVVLNQLCVLHE | | |
| 298-321 | 1 | 2629.2160 | MPADLPSLAADFVESKDVCKNYAE | | OX |
| 312-333 | 1 | 2627.2520 | SKDVCKNYAEAKDVFLGMFLYE | CAM | |
| 543-565 | 1 | 2620.2091 | QLKAVMDDFAAFVEKCCKADDKE | | OX |

[a]The information in the top row of this table should not be included in the HSA_raw1.txt input file. "MC", "CAM", and "OX" are used here as abbreviations that stand for "missed cleavages", "carboxyamidomethyl-cysteine", and "methionine oxidation", respectively.

Table 3.3a. Example of the input for file "ModSearch1.txt"

| | Residues Involved | | | | |
|---|---|---|---|---|---|
| **Modification** | **K1** | **K2** | **R1** | **R2** | **Mass (Da)**[a] |
| FL | 1 | 0 | 0 | 0 | 162.0528 |
| CEL | 1 | 0 | 0 | 0 | 72.0211 |
| CML | 1 | 0 | 0 | 0 | 58.0055 |
| FL-1H2O | 1 | 0 | 0 | 0 | 144.0423 |
| FL-2H2O | 1 | 0 | 0 | 0 | 126.0317 |
| Pyr | 1 | 0 | 0 | 0 | 108.0211 |
| AFGP | 1 | 0 | 0 | 0 | 270.0740 |
| 3-DG-H1 | 0 | 0 | 1 | 0 | 144.0423 |
| THP | 0 | 0 | 1 | 0 | 144.0423 |
| IB | 0 | 0 | 1 | 0 | 142.0266 |
| ArgP | 0 | 0 | 1 | 0 | 80.0262 |
| MG-H1 | 0 | 0 | 1 | 0 | 54.0106 |
| G-H1 | 0 | 0 | 1 | 0 | 39.9949 |
| AFGP | 0 | 0 | 1 | 0 | 270.0740 |
| FL*2 | 0 | 1 | 0 | 0 | 324.1056 |
| CEL*2 | 0 | 1 | 0 | 0 | 144.0422 |
| CML*2 | 0 | 1 | 0 | 0 | 116.0110 |
| FL-1H2O*2 | 0 | 1 | 0 | 0 | 288.0846 |
| FL-2H2O*2 | 0 | 1 | 0 | 0 | 252.0634 |
| Pyr*2 | 0 | 1 | 0 | 0 | 216.0422 |
| AFGP*2 | 0 | 1 | 0 | 0 | 540.1480 |
| FL+CEL | 0 | 1 | 0 | 0 | 234.0739 |
| FL+CML | 0 | 1 | 0 | 0 | 220.0583 |
| FL+FL-1H2O | 0 | 1 | 0 | 0 | 306.0951 |
| FL+FL-2H2O | 0 | 1 | 0 | 0 | 288.0845 |
| FL+Pyr | 0 | 1 | 0 | 0 | 270.0739 |
| FL+AFGP | 0 | 1 | 0 | 0 | 432.1268 |
| CEL+CML | 0 | 1 | 0 | 0 | 130.0266 |
| CEL+FL-1H2O | 0 | 1 | 0 | 0 | 216.0634 |
| CEL+FL-2H2O | 0 | 1 | 0 | 0 | 198.0528 |

[a]The information in the top row of this table should not be included in the HSA_raw1.txt input file. The "residues involved" show where the modifications are expected to occur. A number "1" in positions K1 or K2 rows means that the modification shown occurs on one or two lysines respectively. The same is true for R1 and R2, which represents one arginine and two arginines respectively.

Table 3.4a. Code for program “digestmatch.m”

```
%MData = dlmread('data1.txt',' ')
[Residue,      MC,     MassDig,       Sequence,      Mod,    Mod2] =textread('digest1.txt','%s %f %f %s %s
%s');
[MassHSA       , ZT] =textread('HSA_Raw1.txt','%f %s');
[Gmod,  K1,     K2,     R1,     R2,     MassMod] = textread('ModSearch1.txt', '%s %f %f %f %f %f');
sizeHSAMass =size(MassHSA);
sizeDigMass = size(MassDig);

%calcualting the difference
for i=1:1:sizeHSAMass(1);
    for j=1:1:sizeDigMass(1);
        Difference(i,j) = MassHSA(i) - MassDig(j);
    end
end
%calculating the error
sizeDifference=size(Difference);
k=0;
for i=1:1:sizeHSAMass(1);
    for j=1:1:sizeDifference(2);
      error(i,j) = (Difference(i,j)/MassHSA(i)) * 1e6;
       if ((error(i,j) > -50) && (error(i,j) < 50));
           k=k+1;
          % digestmatch_result(k,:) =
[Residue(j),Sequence(j),Mod(j),Mod2(j),MC(j),MassDig(j),MassHSA(i),ZT(i),error(i,j)];
           digestmatch_result(k,:) = [Residue(j) Sequence(j) Mod(j) Mod2(j) MC(j) MassDig(j) MassHSA(i)
ZT(i) error(i,j)];
        end
    end
end

%This is for generating a list of modified peptides by adding the mass if
%the AGE to the mass of the theoretical digest.

sizeModMass = size(MassMod);
```

Table 3.4a, continued. Code for program “digestmatch.m”

```
for i=1:1:sizeDigMass(1);
    for j=1:1:sizeModMass (1);
        Mass2(i,j)=MassDig(i)+MassMod(j);
    end
end

sizeMass2=size(Mass2);

% Computes a new digest table based by adding the mass of all modifications
% to the mass of the digest. Outputs values as DigestMOD. Use DigestMOD to
% generate a new table (text file) for calculting and filtering error values

h=0;

for i=1:1:sizeDigMass(1);
    for j=1:1:sizeMass2(2);
        h=h+1;

       DigestMOD(h,:)=[Residue(i), MC(i),MassDig(i), Mass2(i,j),Sequence(i), Gmod(j), MassMod(j), K1(j),
K2(j), R1(j), R2(j)];
            end
end

outfile = fopen('DigestMod.txt','wt');
sizeDigesMod= size(DigestMOD);
for i=1:1:sizeDigesMod(1)
    fprintf(outfile, '%s \t %f\t %f\t %f\t %s\t %s\t %f\t %f\t %f\t %f\t %f\t \n',  DigestMOD{i,
1},DigestMOD{i, 2},DigestMOD{i, 3},DigestMOD{i, 4},DigestMOD{i, 5},DigestMOD{i, 6},DigestMOD{i,
7},DigestMOD{i, 8},DigestMOD{i, 9},DigestMOD{i, 10},DigestMOD{i, 11});
    % fprintf(outfile, '%s %s %s %s %s %s %s %s %s %s %s \n',  DigestMOD(i, 1),DigestMOD(i, 2),DigestMOD(i,
3),DigestMOD(i, 4),DigestMOD(i, 5),DigestMOD(i, 6),DigestMOD(i, 7),DigestMOD(i, 8),DigestMOD(i,
9),DigestMOD(i, 10),DigestMOD(i, 11));
  %  fprintf(outfile,'%f\n',DigestMOD{i, 2});
end
fclose(outfile);
```

Table 3.4a, continued. Code for program “digestmatch.m”

```
outfile = fopen('digestmatch_result.txt','wt');
sizedigestmatch_result=size(digestmatch_result);
for i=1:1:sizedigestmatch_result(1);
    fprintf(outfile, '%s\t %s\t %s\t %s\t %f\t %f\t %f\t %s\t %f\t \n',
digestmatch_result{i,1},digestmatch_result{i,2},digestmatch_result{i,3},digestmatch_result{i,4},digestmatch
_result{i,5},digestmatch_result{i,6},
digestmatch_result{i,7},digestmatch_result{i,8},digestmatch_result{i,9});
end
     fclose(outfile);
```

Table 3.5a. Code for program “suspect_mod_sites.m”

```
%Mdata=dlmread('data1.txt', ' ')
[Residue, MC, MassDig, Mass2, Sequence, Gmod, MassMod, K1, K2, R1, R2]=textread('digestMOD.txt', '%s %f %f
%f %s %s %f %f %f %f %f');
[MassHSA       , ZT] =textread('HSA_Raw1.txt','%f %s');

sizeHSAMass=size(MassHSA);
sizeMass2=size(Mass2);

%Caluculating error in ppm

h=0;

for i=1:1:sizeHSAMass;
    for j=1:1:sizeMass2;
        error(i,j)=((Mass2(j)-MassHSA(i))/MassHSA(i))*10^6;
        if ((error(i,j)>-50) && (error(i,j)<50));
            h=h+1;
            char_seq= char(Sequence(j));
           AA=aacount(char_seq);
            error_match(h,:)=[error(i,j), Residue(j), MC(j), Sequence(j), MassDig(j), MassHSA(i), ZT(i),
Gmod(j), MassMod(j), K1(j), K2(j), R1(j), R2(j)];

           if K1(j)==1 && K2(j)==0 && R1(j)==0 && R2(j)==0;
             if AA.K >= 1;
                 error_matchK(h,:)=[error(i,j), Residue(j), MC(j), Sequence(j), MassDig(j), MassHSA(i),
ZT(i), Gmod(j), MassMod(j), K1(j), K2(j), R1(j), R2(j)];
             end
           end

           if R1(j)==1 && K1(j)==0 && K2(j)==0 && R2(j)==0;
             if AA.R >= 1;

                 error_matchR(h,:)=[error(i,j), Residue(j), MC(j), Sequence(j), MassDig(j), MassHSA(i),
ZT(i), Gmod(j), MassMod(j), K1(j), K2(j), R1(j), R2(j)];
             end
```

Table 3.5a, continued. Code for program "suspect_mod_sites.m"

```
  end

              if K2(j)==1 && R1(j)==0 && K1(j)==0 && R2(j)==0;
              if AA.K >= 2;
                 error_matchK2(h,:)=[error(i,j), Residue(j), MC(j), Sequence(j), MassDig(j), MassHSA(i),
ZT(i), Gmod(j), MassMod(j), K1(j), K2(j), R1(j), R2(j)];
              end
         end

         if R2(j)==1 && K2(j)==0 && R1(j)==0 && K1(j)==0;
              if AA.R >= 2;
                 error_matchR2(h,:)=[error(i,j), Residue(j), MC(j), Sequence(j), MassDig(j), MassHSA(i),
ZT(i), Gmod(j), MassMod(j), K1(j), K2(j), R1(j), R2(j)];
              end
         end

         if K1(j) >= 1 && R1(j) >= 1 && K2(j)==0 && R2(j)==0;
              if AA.K >= 1 && AA.R >= 1;
                 error_matchKR(h,:)=[error(i,j), Residue(j), MC(j), Sequence(j), MassDig(j), MassHSA(i),
ZT(i), Gmod(j), MassMod(j), K1(j), K2(j), R1(j), R2(j)];
              end
         end

        end
      end
end
```

# CHAPTER 4

# CHARACTERIZATION OF GLYCATION SITES ON *IN VITRO* GLYCATED HUMAN SERUM ALBUMIN USING MASS SPECTROMETRY

## 4.A. INTRODUCTION

This chapter used the optimized methods for identifying glycation sites that were developed in Chapter 3, to identify early glycation adducts that occurred on human serum albumin with varying degrees of glycation. Typically, when external calibration is used, a maximum allowed error of 50 ppm is used for matching a theoretical digest table to m/z values within a given mass spectrum. By using internal calibration, a maximum allowed error of 25 ppm could be used. The increased mass accuracy that was obtained with internal calibration allowed for increased confidence in the assignment of many modified residues. Glycation product heterogeneity was also found to be a critical factor for glycation site assignment, because there were many peptide and glycation product combinations that overlapped at a very similar mass. In these cases, searching a broad list of modifications was necessary because it allowed these different combinations to be taken into account. The higher mass accuracy that was obtained when internal calibration was used allowed for better differentiation between these glycated peptides. Additionally, by searching for peptides that were unique to the glycated HSA samples, it was possible

to assign modification sites that arise from the experimental conditions. By applying the aforementioned techniques to this current analysis, glycation sites can be assigned to HSA with greater confidence.

This chapter focuses on early glycation, which occurs on the primary amine groups on proteins (i.e., lysines or the *N*-terminus) (1, 2). Early-stage glycation involves the formation of a Schiff base which forms when free sugars react with these primary amine groups. The Schiff-base is unstable and subsequently rearranges to form the more stable Amadori product (3-5). When glucose is used as a reagent, this Amadori product will take the form of fructosyl-lysine (FL) or dehydrated fructosyl-lysine (i.e., FL-1$H_2O$ or FL-2$H_2O$). While this process is has been well studied in numerous proteins (6-15), little has been done to characterize these reactions on HSA (2, 16-18). Additionally, comparisons to see whether glycation of HSA that occurs *in vitro* is similar to glycation of HSA that that occurs *in vivo* has not yet been done. This study therefore provides preliminary information that can be used for this comparison.

For this study, glycated HSA was prepared *in vitro* under physiological conditions. The HSA that was used for this study was collected after two, four, or five weeks of incubation. These time periods were selected because it should yield HSA that has similar glycation patterns to HSA that is identified *in vivo*, given its half life (~ 2 - 4 weeks) (1). Techniques that were previously used to obtain high sequence coverage for HSA (19) and accurate glycation site assignment (20) were both used in this approach.

## 4.B. EXPERIMENTAL

**4.B.1. Materials.** HSA (99%, essentially fatty acid and globulin free), guanidine-HCl (99%), ammonium bicarbonate (99%), tris-HCl (99%), potassium phosphate monobasic (99%), potassium phosphate dibasic (99%), dithiothreitol (99%), iodoacetamide (99%), D-glucose (99%), sodium azide (99%), trifluoroacetic acid (98%), formic acid (96%), molecular biology grade water (DNase, RNase, and protease free), methanol (HPLC grade, 99.9%), and acetonitrile (HPLC grade, 99.9%) were all purchased from Sigma-Aldrich (St. Louis, MO). The enzymes which were used for this study; sequence grade trypsin, Glu-C and Lys-C, were also purchased from Sigma-Aldrich. The glycated serum protein enzymatic assay was purchased from Diazyme Laboratories (Poway, CA).

**4.B.2. Apparatus.** Siliconized low retention microcentrifuge tubes (0.6 mL and 1.5 mL), Pierce Slide-A-Lyzer dialysis cassettes (7 kDa MWCO, 0.1 – 0.5 mL and 0.5 – 3 mL), Millipore μ-$C_{18}$ Zip-tip pipette tips (0.2 μL bed volume), 0.20 μm nylon filters, and Bio-Rad 10 DG size exclusion columns (10 mL) were all obtained from Fisher Scientific (Pittsburgh, PA). Mass spectrometry data was collected on a Voyager 6148 MALDI-TOF-MS system (Applied/Perspective Biosystems, CA). Positive ion delayed extraction mode was used with an accelerating voltage of 25 kV, a grid wire voltage which was 0.008% of the accelerating voltage, a grid voltage which was 77% of the accelerating voltage, and a delay time of 100 nanoseconds. Matlab 2009a which included the bioinformatics toolbox was purchased from Mathworks (Natick, MA). Mascot wizard was obtained from the matrix science website (21).

**4.B.3. Preparation of glycated HSA.** Glycation was initiated by incubating a physiological concentration of HSA (42 g/L) with a 15 mM glucose solution. The glucose solution was prepared in pH 7.4, 200 mM phosphate buffer which also contained 1 mM sodium azide (Note: sodium azide was added to prohibit bacterial growth. This HSA solution was divided into several portions and incubated at 37 °C. HSA samples containing three different levels of total glycation was examined in this report. To make these samples, the starting HSA/glucose solution was divided into several portions and incubated at 37 °C for two or five weeks, which gave the "HSA-2" and "HSA-5" samples used in this report. A control sample was prepared under identical conditions by using a solution of 42 g/L HSA dissolved in the same pH 7.4 buffer with sodium azide (but without the addition of glucose) and that was incubated for five weeks at 37 °C. A third glycated HSA sample was prepared under these conditions but now using a total added glucose concentration of 30.7 mM and an incubation time of 4 weeks; these conditions provided the "HSA-4" sample that was used in this report. A second control sample of HSA was also prepared under these latter conditions but in the absence of any glucose. The conditions used here were selected because HSA has a half-life in the body of approximately 3 weeks (1). The HSA incubated over a 2-week, 4-week, and 5-week periods were therefore selected because these samples would mimic the maximum amount of time that HSA spends in the body *in vivo*.

The HSA was isolated from the reagent solution using a combination of size-exclusion chromatography (SEC) and equilibrium dialysis. First, a 10 DG size exclusion column was washed with 20 mL of water followed by loading the column with 3 mL of the incubated HSA solution. A 5 mL portion of water was then added to the SEC column

to elute the protein, where this fraction was collected. A 2.5 mL aliquot of this eluent was then transferred to 0.5 – 3 mL dialysis cassettes, and these samples were dialyzed extensively against two 1000 mL portions of nanopure water for 2 h at room temperature. An additional dialysis was then performed overnight at 4 °C. The solution was then removed from the dialysis cassette, lyophilized, and stored at 4 ºC for later use.

**4.B.4 Assay of the extent of glycation.** The total amount of glycation in each HSA sample was determined using a glycated serum protein assay which was obtained from Diazyme Laboratories (Poway, CA). In this assay, glycated proteins were first cleaved into glycated peptides using a proteinase K. Fructosaminase was then added to oxidize these glycated peptides to form glucosone while catalyzing the production of hydrogen peroxide from oxygen. The amount of hydrogen peroxide produced was measured by observing the absorbance when hydrogen peroxide reacts with Trinder's reagent (i.e., a mixture of *N*-Ethyl-*N*-(2-hydroxy-3-sulfopropyl)-3-methylaniline sodium and 4-amino-1,5-dimethyl-2-phenyl-pyrazol-3-one) in the presence of horse radish peroxidase to form a colored compound (22, 23). The fructosamine assay kit included two proprietary reagents (i.e., reagent 1 and reagent 2) and standards containing known amounts of fructosamine.

To perform this assay, 20 µL of either a saline blank, standards with known fructosamine concentrations, or HSA dissolved in pH 7.4, 25 mM phosphate buffer (Note: The concentration of HSA was either 42 or 60 g/L depending on whether the absorbance fell within the range of the calibration curve) were mixed with 200 µL of Reagent 1. This mixture was incubated for 5 minutes using a temperature controlled UV that was set to 37 ºC. At this time, the absorbance at 600 nm ($A_{600}$) and 700 nm ($A_{700}$)

were obtained. A 50 µL aliquot of Reagent 2 was added to the mixture, and the resulting solution was incubated for another 5 minutes at 37 ºC. The $A_{600}$ and $A_{700}$ values were then obtained. Corrections were made for both light scattering and fructosamine dilution in this analysis. The correction for light scattering was made by subtracting the $A_{700}$ from the $A_{600}$, and the fructosamine dilution that occurred when Reagent 2 was added to the mixture was corrected by multiplying the 5 minute absorbance value (Note: this value was corrected for light scattering) by 0.815. A calibration curve was then developed using these corrected absorbance values and standard beers law calculations. The extent of glycation, as given in units of moles of hexose per mol of HSA, was then determined by dividing the fructosamine concentration by the known HSA concentration.

**4.B.5. HSA alkylation and digestion.** The glycated HSA samples and corresponding control samples were all pretreated and digested using the same procedure, equipment and reagents to minimize any differences in digestion patterns due to changes in the experimental conditions. Molecular biology grade water was used to prepare ph 8.5, 100 mM ammonium bicarbonate buffer. Solutions containing 6 M guanidine-HCl or 1 M dithiothreitol were prepared using this ammonium bicarbonate buffer. A solution containing 1 M iodoacetamide (IAM) and 1 M NaOH was also prepared using nanopure water. HSA was first reconstituted to a concentration of 3 mg/mL with guanidine-HCl solution. The resulting solution was gently mixed using a vortex genie mixer. A 15 µL aliquot of DTT reagent was then added to the HSA solution. This solution was incubated at 37 °C for 1 hr, after which 36 µL of the IAM/NaOH reagent was added to the HSA solution. The resulting solution was then incubated in the dark at room temperature for 30 minutes. Following this incubation, an additional 150 µL of DTT reagent was added to

the mixture. The HSA solution was then divided into 0.5 mL portions, where each portion was transferred to 0.1 – 0.5 mL dialysis cassettes. Dialysis was performed on these portions against three 1000 mL portions of nanopure water, where each dialysis was performed for 4 h. The dialyzed HSA samples were then divided into 50 μL aliquots, which contained 100 μg of HSA per aliquot, and placed into low-retention microcentrifuge tubes. These aliquots were then reduced to dryness by using Speedvac. The resulting HSA samples (i.e., pretreated HSA) were then stored at -80 °C until later use. The pretreatment procedure was done on both the control and glycated HSA samples.

Stock enzyme solutions containing 1 mg/mL of trypsin, Lys-C, or Glu-C were prepared in nanopure water. A pH 7.8, 100 mM ammonium bicarbonate buffer and a pH 9.0, 75 mM tris-HCl buffer was also prepared. For the trypsin digest, 200 μL of ammonium bicarbonate buffer was added to vials pretreated HSA or glycated HSA. Each vial of reconstituted HSA solution was vortexed for 3 – 5 minutes. A 3.3 μL portion of trypsin stock solution was then added to each reconstituted HSA sample, giving substrate to enzyme ratios of ~ 30 to 1. These solutions were incubated at 37 °C for 18 h. A 10 μL aliquot of concentrated formic acid was then added to each HSA solution. The resulting HSA digests were stored at -80 °C until later use.

For the Glu-C digest, 200 μL of a pH 7.8, 100 mM ammonium bicarbonate buffer was added to vials of pretreated HSA or glycated HSA. Each vial of reconstituted HSA solution was vortexed for 3 – 5 minutes. A 10 μL portion of Glu-C was then added to each reconstituted HSA sample and the solutions were incubated at 37 °C for 8 h. Another 5 μL portion of Glu-C was added to each HSA sample, giving a final substrate to

enzyme ratios of 7 to 1, and the resulting solutions were incubated for an additional 18 h. A 10 μL aliquot of concentrated formic acid was then added to these digests and the resulting solutions were stored at -80 °C.

For the Lys-C digest, 200 μL of 75 mM tris-HCl buffer (pH 9.0) was added to vials of pretreated HSA or glycated HSA. Each vial of reconstituted HSA solution was vortexed for 3–5 minutes. A 5 μL portion of Lys-C was added to each vial of reconstituted HSA and the resulting mixtures were incubated at 37 °C for 18 h. The digestions were terminated by acidifying and freezing the mixture as previously described.

**4.B.6. Collection and analysis of mass spectrometry data.** The matrix that was used for MALDI consisted of a mixture of α-cyano-4-hydroxycinnamic acid (CHCA) and 2,3-dihyroxybenzoic acid (DHB). A solution containing 20 mg/mL CHCA and 5% formic acid was prepared in a 70:30 (v/v) mixture of acetonitrile and nanopure water. Similarly, a solution containing 20 mg/mL DHB and 0.1% TFA was prepared in a 70:30 (v/v) mixture of acetonitrile and nanopure water. Both matrix solutions were mixed vigorously 2 – 3 minutes using a vortex-genie mixer and combined in a 50:50 (v/v) ratio. A stock calibration mixture (calmix), which contained 32.5 pM glu-fibrinopeptide (1570.6774 Da), 25 pM des-Arg-bradykinin (904.4681 Da), and 32.5 pM angiotensin I (1296.6853 Da) was prepared in 100% acetonitrile. A calibrant spiked matrix solution was prepared by adding 4 μL of the calmix solution to 96 μL of the matrix. This spiked matrix solution was placed on every adjacent row of a 96-well MALDI plate (i.e., these were placed in positions that were adjacent to the HSA samples).

Solutions of 5, 10, 20, 30 and 50 % ACN were prepared in nanopure water. Concentrated TFA was added to each of these solutions, yielding a final TFA concentration of 0.1%. A separate wetting solution of 50% ACN was prepared in filtered nanopure water. An equilibration solution of 0.1% TFA in nanopure water and a wash solution of 5% methanol and 0.1% TFA in nanopure water were also prepared. A Zip-tip was wet with two 10 μL portions of the wetting solution, and then equilibrated with two 10 μL portions of the equilibration solution. A 10 μL aliquot of the HSA digest was then loaded onto the Zip-tip by using 10 aspirate-dispense cycles. The Zip-tip was then immediately washed with two 10 μL portions of the wash solutions. A 1 μL portion of the 5% ACN solution was then placed into a low retention microcentrifuge tube, and peptide elution was performed by passing this solution through the Zip-tips with 3 aspirate-dispense cycles. The eluted peptides were mixed with 1 μL of the spiked matrix solution and 0.80 μL of the resulting mixture was spotted onto a 96-well MALDI plate. The Zip-tip was washed 3 times with 4 μL of the same ACN solution that was used to elute the peptides. The elution process was then repeated by using consecutively higher ACN gradients. All of the mass spectra acquired were first externally calibrated using the calibration spots on the adjacent rows, then internally calibrated by using the spiked calibrant masses that were found in the mass spectra for the HSA samples. An evaluation of unspiked HSA samples shows that there is no overlap in the m/z values for the calibrants and the HSA-derived peptides. An error of 25 ppm was found to be sufficient for peptide mass matching when this technique is used (20).

Raw mass values from each mass spectrum were extracted using Mascot Wizard and these values were compared to an expanded modification list by using a program

written in Matlab. Details of this program are described in Chapter 3 (20). The glycation adducts and AGE adducts that were examined are shown in Chapter 3 (Table 3.1). Only the unique peptides which were found in HSA-2, HSA-4, or HSA-5 samples were matched to this search list. To determine which peptides were unique, a separate Matlab program was used to compare the extracted mass list obtained for the glycated HSA samples to the control HSA samples and match these masses at 50 ppm. The peaks in the glycated sample that did not match were selected for further analysis, where any modified residues which matched at 25 ppm were considered to be potential glycation sites (See Figure 3.4 for an example of this assignment).

## 4.C. RESULTS

**4.C.1. Fructosamine assay and sequence coverage.** The moles of hexose condensed per moles of HSA increased in a linear fashion over a 3-week period, after which the rate of glucose incorporation began to slow down (See Figure 4.1). Using this assay, it was found that the HSA-2 sample had 0.7 (± 0.1) mol hexose/mol HSA, the HSA-4 sample had 3.4 (± 0.1) mol hexose/mol HSA, and the HSA-5 sample had 1.2 (± 0.1) mol hexose/mol HSA.

The sequence coverage obtained for the HSA and glycated HSA samples are summarized in Figure 4.2. For the HSA-C sample, the tryptic, Lys-C and Glu-C digests covered 78%, 47% and 56% of the primary sequence of HSA respectively. Similarly, multienzyme digestion of the HSA-2 gave a sequence coverage of 63%, 49% and 44% and digestion of the HSA-5 sample gave a sequence coverage of 54%, 44% and 49% for the trypsin, Lys-C, and Glu-C digests respectively. The sequence coverage obtained for the HSA-4 sample was 69%, 45%, and 55% for the trypsin, Lys-C, and Glu-C digest respectively. When multienzyme digestion was considered, the total sequence coverage for the HSA-C, HSA-2, HSA-4, and HSA-5 samples were 90%, 81%, 84%, and 82% respectively. A decreased coverage for the glycated HSA samples was expected because glycation induced mass shifts would result in some peptides being undetectable. The coverage for the HSA-2, HSA-4 and HSA-5 samples all increased to approximately 86% when the assigned glycated adducts were factored into the sequence coverage calculation.

**4.C.2. Mass shifts identified on the HSA samples.** Table 4.1 summarizes the peptides that were identified at all three stages of glycation. There some instances where early-

Figure 4.1. The extent of glycation that was determined using the enzymatic fructosamine assay. The results are expressed as the moles of hexose per moles of HSA as a function of time.

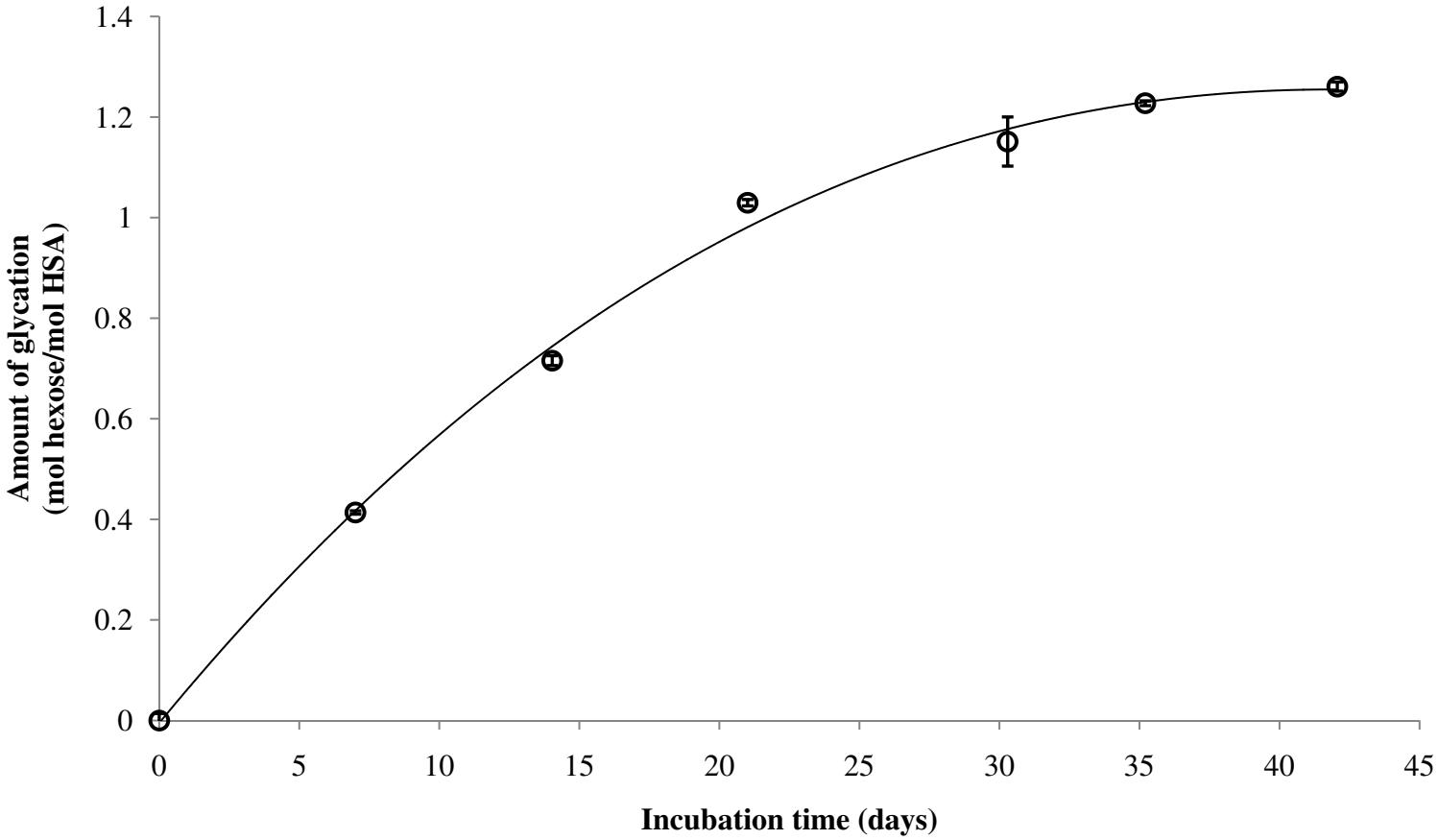

Amount of glycation
(mol hexose/mol HSA)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
40
45
Incubation time (days)

Figure 4.2. Sequence coverage for the control HSA and the glycated HSA samples. The sequence coverage obtained when all four HSA samples are considered is shown in **bold**. The sequence coverage obtained for the HSA control sample is shown in **CAPS**, the coverage obtained for the HSA-2 sample is **underlined**, the coverage obtained for the HSA-4 sample is **highlighted**, and the coverage obtained for the HSA-5 sample is designated with a **line** over the amino acid sequence. Peptides that were linked to modification sites were factored into the sequence coverage that is shown for the HSA-2, HSA-4, and HSA-5 samples.


1 – dahkSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV
41 – KLVNEVTEFA Ktcvadesae NCDKSLHTLF GDKLCTVATL
81 – RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV
121 – DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR
161 – YKAAFTECCQ AADKAACLLP KLDELRDEGK ASSAKQRLKC
201 – ASLQKFGERA FKAWAVARLS QRFPKAEfae vsklvtdltk
241 – VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE
281 – KPLLEKSHCI AEVENDEMPA DLPSLAADFV ESKdvckNYA
321 – EAKDVFLGMF LYEYARRHPD YSVVLLLRLA KTYETTLEKC
361 – CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE
401 – YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH
441 – PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES
481 – LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKE
521 – RQIKKQTALV Elvkhkpkat keqlkAVMDD FAAFVEKcck
561 – ADDKETCFAE EGKKlvaasq aalgl

Table 4.1. Suspect modification sites that were found on glycated HSA. The most likely modification sites are given in bold.

| | Digest | Residues | Lysine | Arginine | Suspect Modification Sites[a] |
|---|---|---|---|---|---|
| HSA-2 | Trypsin | 1-10 | N+, K4 | R10 | **N+** (FL-1$H_2O$) & R10 (AFGP) or N+ (AFGP) & R10 (3-DG-H1 or THP) |
| | Trypsin | 182-195 | K190, K195 | K186 | K190 & K195 (FL & CML) |
| | | 275-286 | K276, K281, K286 | n/a | K276 & K286 (FL & FL-1$H_2O$) |
| | Trypsin | 539-557 or 542-560 | K541, K545, K557, K560 | n/a | K545 & K557 (FL-2$H_2O$ & FL-2$H_2O$) or (FL-1$H_2O$ or Pyr) |
| | | 565-585 | K573, K574 | n/a | K573 & K574 (Pyr*2 or CEL & FL-1$H_2O$) |
| | Glu-C | 278-292 | K281, K286 | n/a | K281 & K286 (FL & CEL) or (FL-2$H_2O$ & Pyr) |
| | Glu-C | 83-100 | K93, R98 | R98 | **K93** (FL-2$H_2O$) or **K93** (CEL) & R98 (MG-H1) |
| | Lys-C | 196-205 | K199, K205 | R197 | **K199** & **K205** (FL & FL-1$H_2O$) or K199 (FL) & R197 (3-DG-H1 or THP) |
| | Lys-C | 414-432 | K414, K432 | R428 | **K414** & **K432** (FL-1$H_2O$ & AFGP) or K414 (AFGP) & R428 (3-DG-H1 or THP) |
| HSA-5 | Trypsin | 182-195 | K190, K195 | R186 | **K190** & **K195** (FL & CML) |
| | | 275-286 | K276, K281, K286 | n/a | K276 & K286 (FL & FL-1$H_2O$) |
| | Glu-C | 281-294 | K281, K286 | n/a | K281 & K286 (FL-2$H_2O$ & CEL) |
| | Lys-C | 5-12 | K12 | R10 | K12 (FL-1$H_2O$) & R10 (MG-H1) |
| | Lys-C | 525-534 | K525, K534 | n/a | **K534** (FL-1$H_2O$) or K525 (CEL) & **K534** (CEL) |
| | Lys-C | 5-20 | K12, K20 | R10 | **K12** (FL) |
| | Lys-C | 277-286 | K281, K286 | n/a | K281 & K286 (FL & FL-1$H_2O$) |
| | Lys-C | 414-432 | K414, K432 | R428 | **K414** & **K432** (FL-1$H_2O$ & AFGP) or K414 (AFGP) & R428 (3-DG-H1 or THP) |

Table 4.1 (continued). Suspect modification sites that were found on glycated HSA. The most likely modification sites are given in bold.

**HSA-4**

| Digest | Residues | Lysine | Arginine | Suspect Modification Sites |
|---|---|---|---|---|
| Trypsin | 206-212 | K212 | R209 | **K212** (FL-2$H_2O$) or R209 (MG-H1) & **K212** (CEL) |
| Trypsin | 1-12 | N+, K4, K12 | R10 | **N+** & **K12** (CEL & FL-2$H_2O$) or N+ (FL-1$H_2O$) & R10 (MG-H1) |
| | 200-212 | K205, K212 | R209 | K205 (CML) & R209 (ArgP) |
| | 403-413 | K413 | R410 | K413 or R410 (AFGP), or K413 (FL-2$H_2O$) & R410 (3-DG-H1 or THP) |
| | 175-186 | K181 | R186 | K181 (FL-1$H_2O$) & R186 (ArgP) |
| Trypsin | 546-560 | K557, K560 | n/a | K557 & K560 (CML & FL-1H2O) |
| Trypsin | 258-274 | K262, K274 | n/a | **K274** (FL-1$H_2O$) or K262 (CEL) & **K274** (CEL) |
| | 160-174 | K162, K174 | R160 | K162 or K174 (FL) |
| Lys-C | 526-536 | K534, K536 | n/a | K534 & K536 (CML & FL-1$H_2O$) |
| Lys-C | 501-519 | K519 | n/a | **K519** (FL) |
| Glu-C | 557-571 | K557, K560, K564 | n/a | K557 & K564 (FL & CML) |
| Glu-C | 426-442 | K432, K436, K439 | R428 | **K436** (FL) |
| Glu-C | 83-100 | K93, R98 | R98 | **K93** (FL-2$H_2O$) or **K93** (CEL) & R98 (MG-H1) |

[a] – In some instances, a given m/z value could be linked to multiple modifications. The most likely modification sites, which were identified based on p$K_a$ and FAS data, are given in **bold.**

stage glycation products could be assigned to peptides with a high level of certainty. For example, a m/z value of 1850.97 Da was identified in the HSA-5 sample (Glu-C digest). This mass was assigned to residues 281-294 on HSA which was modified to form *N*ε-Carboxyethyl-lysine (CEL) and dehydrated fructosyl-lysine (FL-2$H_2O$). The only lysines on this residue that are available for modification are K281 and K286, hence the formation of FL-2$H_2O$ or CEL could occur on either residue. Another example of this scenario was found on residues 5-12 in the HSA-5 sample. A mass of 1171.591 Da was detected, which corresponds to residues 5-12 being modified to form FL-1$H_2O$ and a *N*ε-(5-Hydro-5-methyl-4-imidazolon-2-yl)ornithine (MG-H1). In this case K12 is being modified to form FL-1$H_2O$ and R10 is being modified to form MG-H1 because these residues are the only lysine and arginine residues that are found in this peptide. Similar observations were made in the HSA-4 sample, where early glycation adducts could be assigned to K93, K557 or K560, and K519.

There were many cases where m/z values could be mapped to different glycation adducts that resulted in the same m/z shift. For example a raw mass of 2107.95 Da was identified in the Glu-C digest for the HSA-2 sample, which corresponds to early-glycation occurring on regions 278-292 on HSA. The first modification involved the addition of FL and CEL moieties, which would result in a mass shift of 234.07 Da, and the second modification involved the addition of FL-2$H_2O$ and Pyr moieties, which would result in a mass shift of 234.05 Da. At the assigned mass accuracy of 25 ppm, these two mass values couldn't be differentiated, so it is possible that this m/z value could represent either of the two possibilities. Similarly a peptide encompassing residues 414-432 (2466.27 Da) in the Lys-C digest was found to be modified to form FL-1$H_2O$ and 1-

Alkyl-2-formyl-3,4-glycosyl-pyrrole (AFGP). This peptide contains two lysines (K414 and K432) and one arginine (R428) which are able to undergo modification. Since AFGP is reactive with both lysine and arginine residues, there is one possible scenario where these modifications are occurring on K414 and K432 and there is another scenario where FL-1$H_2O$ is occurring on either K414 or K432, and AFGP is occurring on R428.

Situations were also present, where different peptides having varying modifications could also overlap at the same mass, at the given mass accuracy of 25 ppm. For example, a peptide having a m/z value of 1622.81 Da was detected in the trypsin digest of the HSA-4 sample. When only early glycation is considered, this mass could result from residues 1-12, 403-413, and 175-186 being modified to form FL-2$H_2O$ and CEL, FL-2$H_2O$ and $N_\varepsilon$-[5-(2,3,4-Trihydroxybutyl)-5-hydro-4-imidazolon-2-yl]ornithine (3-DG-H1) or tetrahydropyrimidine (THP), and FL-1$H_2O$ and argpyrimidine (ArgP) respectively. The HSA-2 and HSA-5 samples also had peptides which showed the same characteristics. For example, two residue clusters which encompass regions 182-195 and 275-286, or regions 539-560 and 565-585 were identified at the same m/z value on the HSA-2 sample. Similarly, residues 182-195 and 275-286 were both linked to the same m/z value in the HSA-5 sample.

When looking at the total number of early-stage glycation products that were identified in the glycated HSA samples, the HSA-2, HSA-5, and HSA-4 samples were modified to form 7, 7, or 9 glycated peptides respectively. The assignment of many of these peptides, however, was not certain because some early-stage and late-stage glycation adduct combinations overlap at the same mass. If these overlapping peptides

are not considered, the number of early glycation adducts that were assigned to the HSA-2, HSA-5, and HSA-4 sample reduced to 3, 5, and 5 glycated peptides respectively.

**4.C.3. Location of the most likely glycation sites based on p$K_a$ and FAS data.** It is clear from the above results that assigning glycation adducts from only the mass shift data is not adequate for identifying the most likely glycation sites. Information regarding the glycation-induced mass shifts can be combined with known information about the p$K_a$ and fractional accessible surface area (FAS) of each residue to determine the most probable source of glycation. The p$K_a$ and FAS values for residues that were linked to m/z shifts are summarized in Table 4.2, where these values were determined using methods that have been described in the literature (2, 24, 25). It is expected the lysines having high FAS and low p$K_a$ values would facilitate glycation because these lysines are more exposed to the solvent and because a greater number of lysines would be deprotonated. The p$K_a$ range of lysine residues on HSA is between 6.2 and 11 with an average p$K_a$ of 10.08. The FAS range for lysine residues is between 0.01 – 0.96 with an average FAS of 0.51.

An example of p$K_a$ assignment is shown on residues 1-10 on HSA 2 sample. This residue contains the *N*-terminus and K4, both of which are able to interact with the early-glycation products. This peptide also contains R10 which is able to interact with late-stage Maillard products. The p$K_a$ of the *N*-terminus is 7.79 compared to a p$K_a$ of 10.53 for K4, indicating that the *N*-terminus is the most likely modification site. In situations where multiple peptides and glycation product combinations overlapped at the same m/z value, p$K_a$ assignment was also used to determine which peptide had the highest chance of

Table 4.2. Calculated p$K_a$ and FAS values for amino-acids which are believed to be glycated.

| | Residue | p$K_a$ | FAS |
|---|---|---|---|
| Lysine | N+ | 7.79 | |
| | 4 | 10.53 | n/a |
| | 12 | 10.43 | 0.59 |
| | 20 | 10.43 | 0.32 |
| | 93 | 10.36 | 0.5 |
| | 137 | 10.43 | 0.52 |
| | 162 | 10.43 | 0.43 |
| | 174 | 10.22 | 0.34 |
| | 181 | 11.92 | 0.37 |
| | 190 | 6.23 | 0.42 |
| | 195 | 10.76 | 0.5 |
| | 199 | 7.47 | 0.16 |
| | 205 | 10.43 | 0.58 |
| | 212 | 10.43 | 0.4 |
| | 218 | 9.64 | 0.2 |
| | 262 | 10.43 | 0.66 |
| | 274 | 9.87 | 0.27 |
| | 276 | 10.29 | 0.66 |
| | 281 | 10.5 | 0.52 |
| | 286 | 10.29 | 0.2 |
| | 313 | 10.5 | 0.8 |
| | 317 | 10.36 | 0.8 |
| | 323 | 10.5 | 0.54 |
| | 402 | 10.43 | 0.61 |
| | 413 | 10.01 | 0.23 |
| | 414 | 7.66 | 0.08 |
| | 432 | 9.92 | 0.36 |
| | 436 | 9.67 | 0.51 |
| | 439 | 10.43 | 0.94 |

| | Residue | p$K_a$ | FAS |
|---|---|---|---|
| Lysine | 444 | 10.36 | 0.48 |
| | 519 | 10.43 | 0.72 |
| | 524 | 10.43 | 0.54 |
| | 525 | 15 | 0.07 |
| | 534 | 11.11 | 0.12 |
| | 536 | 10.08 | 0.24 |
| | 541 | 10.5 | 0.9 |
| | 545 | 10.36 | 0.56 |
| | 557 | 10.43 | 0.68 |
| | 560 | 11.63 | 0.82 |
| | 564 | 10.5 | 0.93 |
| | 573 | 10.5 | 0.72 |
| | 574 | 10.5 | 0.82 |
| Arginine | 10 | 16.02 | 0.13 |
| | 98 | 11.38 | 0.08 |
| | 144 | 15.69 | 0.11 |
| | 145 | 11.21 | 0.21 |
| | 160 | 13.84 | 0.32 |
| | 186 | 11.75 | 0.6 |
| | 197 | 14.37 | 0.13 |
| | 209 | 12.08 | 0.61 |
| | 218 | 9.64 | 0.2 |
| | 336 | 15.7 | 0.07 |
| | 410 | 12.46 | 0.33 |
| | 428 | 11.29 | 0.18 |
| | 445 | 11.94 | 0.21 |
| | 484 | 10.09 | 0.06 |
| | 521 | 14.01 | 0.15 |

being glycated. For example, in the tryptic digest of the HSA-5 sample, a m/z value of 1738.82 Da was observed which corresponds to residues 182-195 as well as residues 275-286 on HSA. Residues 182-195 were chosen as the most likely glycated peptide because K190 had favorable p$K_a$ and FAS combinations of 6.23 and 0.42 respectively, compared to K281 and K286 which had p$K$a values greater than 10.29. In other cases the combined p$K_a$ and FAS data can be used to differentiate the likely glycation sites. For instance a mass of 2199.80 Da was observed in the Glu-C digest of both the HSA-2 and HSA-4 samples. This mass could have resulted from modifications occurring on residues 83-100 on HSA. This region of HSA contains K93 and R98, where the detected mass shift corresponded to the formation of either FL-2$H_2O$ or CEL and MG-H1. Arginine 98 has a FAS value of 0.08 which is low compared to the average arginine FAS values on HSA. It is therefore unlikely that R98 is being modified to form MG-H1. Additionally, K93 has p$K_a$ and FAS values of 10.36 and 0.50 respectively, indicating that this residue is most likely being modified to form FL-2$H_2O$. Similar assignments are made throughout Table 4.1, where the most likely modification sites are shown.

**4.C.4. Semi-quantitative information obtained for the most likely glycation sites and comparison with previous data.** When p$K_a$ and FAS sorting were employed, 7 peptides were identified which could be linked to early-stage glycation products in the HSA-2 digest. Similarly 7 peptides were also identified in the HSA-5 digest and 9 peptides were identified in the HSA-4 digest. The increase in the amount of detected peptides in the HSA-4 digest was most likely due to the increased glucose concentration, which would subsequently increase the concentration of all glycation related products. There were

several cases in this study in which a given lysine or arginine residue was found to be modified in multiple samples, such as in different digests of a particular glycated HSA sample or in HSA samples representing different stages of glycation.. The degree of overlap is indicative of the reactivity of a particular residue. The data that was obtained can therefore be sorted based on the degree of overlap, in order to semi-quantitatively rank these glycation sites. A structure of HSA showing this semi-quantitative assignment is given in Figure 4.3. When this sorting was done, several residues were identified with overlapping modification sites in two or more glycated HSA samples. These residues include the *N*-terminus and lysines 12, 93, 190, 195, 281, 286, 414, 432, 534 and 557. A number of lysine residues on region 182-212 on HSA were found to be modified by early glycation adducts. These residues include K190, K195, K199, K205, and K212, where K199 was previously identified as having significant amounts of glycation (1, 18). The low p$K_a$ or high FAS values found on these lysines and the fact that these sites were all assigned as glycation sites, indicate that this region of HSA may be particularly prone to glycation. The *N*-terminus and K281, which were identified as modification sites in this analysis, were also previously reported as being privileged glycation sites (1, 26).

The glycation sites that were identified in this report were compared to sites that were previously identified using commercially glycated HSA (2, 27). Two comparisons were made, where the first comparison was performed with mass shifts that were identified using commercially glycated HSA (2) and the second comparison was performed with data that was obtained from the quantitative labeling of glycated peptides (27). The results obtained from this analysis are summarized in Table 4.3, where eight of the peptides that were identified using this current approach were also found to be

Figure 4.3. Structure of HSA showing the most likely glycation sites based on semi-quantitative data. The most reactive residues are indicated in red followed by the second most reactive regions indicated in purple.

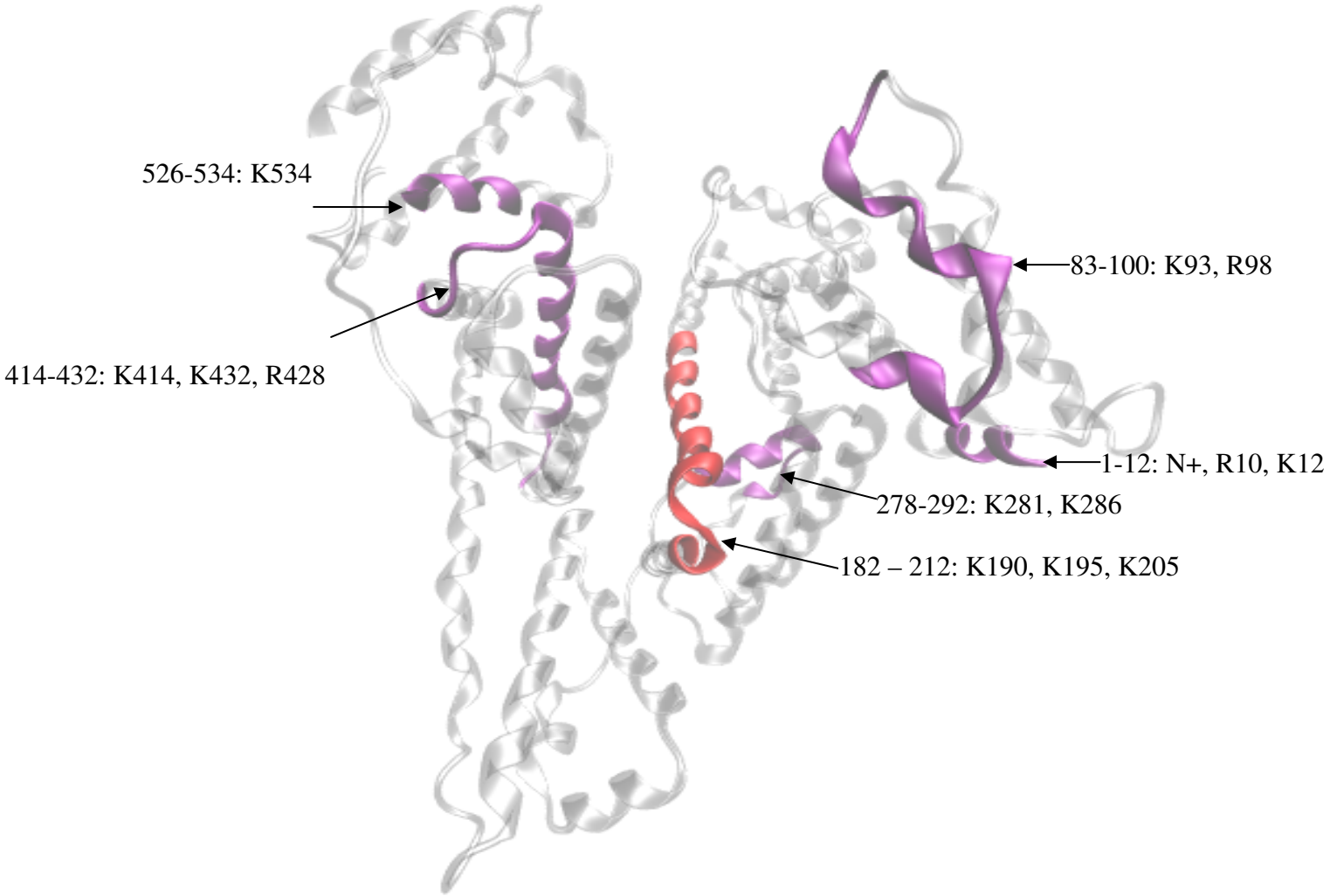

526-534: K534
414-432: K414, K432, R428
83-100: K93, R98
1-12: N+, R10, K12
278-292: K281, K286
182 – 212: K190, K195, K205

Table 4.3. Comparison of glycation sites that were found in this study, with glycation sites that were previously found on commercially glycated HSA.

| HSA Sample | Residue 1[a] | Residue 2[b] | Residue 3[c] | Overlapping Residues | Modification |
|---|---|---|---|---|---|
| HSA-5 | 525-534 | 520-525 | 525-534 | K525, K534 | FL, CEL |
| HSA-2, HSA-5 | 414-432 | 426-442 | 415-439 | K432 | FL or AFGP |
| HSA-4 | 206-212 | 200-218 | 206-212 | K212 | FL or AFGP |
| HSA-4 | 426-442 | 426-442 | 415-439 | K436 or K439 | FL |
| HSA-2 | 196-205 | 196-209 | n/a | K199, K205 | FL, FL |
| HSA-5 | 281-294 | 286-294 | n/a | K286 | FL or CEL |
| HSA-4, HSA-5 | 1-12, 5-12 | 7-17 | n/a | K12 | FL |
| HSA-2, HSA-5 | 182-195 | 191-205 | n/a | K190, K195 | FL, CML[d] |

[a] – Residue 1 represents the region of HSA that was found to be modified by early-stage glycation products in this study.

[b] – Residue 2 represents the region of HSA that was found to be modified by early-stage glycation products that was identified for commercially glycated HSA (2).

[c] – Residue 3 represents the region of HSA that was previously found to have significant levels of glycation which were quantified using $^{18}$O-labeling mass spectrometry (27).

[d] – Modifications by FL and CML that were identified using this current approach, however, CEL was found to be condensed on this region of HSA in a previous attempt where commercially glycated HSA was used (2).

modified in commercially glycated HSA. Of these modifications, five of these overlapping peptides contain K525, K439, and K199, which are residues that are expected to contain the largest amount of modification on HSA (1). While these peptides did overlap with these privileged glycation sites, the p*K*a and FAS data indicated that K534 and K436 were the most likely modification sites on their respective peptides. Additionally, peptides containing the *N*-terminus and K281 were also identified within glycated peptides using this current approach. The degree of overlap as well as the p$K_a$ and FAS data, however, indicated that K12 and K286 are the most likely regions of modification within these peptides. Similar cases are noted by other authors (28, 29), where the p*K*a and FAS data do not always agree with the actual glycation pattern. This disagreement can partially be explained by glycation enhancement that occurs when basic residues which are located in the vicinity of lysines (26), or conformational changes (1) that may alter the accessibility and p$K_a$ of a given lysine.

## 4.D. SUMMARY

When the semi-quantitative sorting and the p$K_a$/FAS trends were considered, regions 196-205 on HSA was found to be particularly prone to glycation. This region is also found within a major drug binding pocket on HSA; known as Sudlow site 1 (1, 30). Lysines 190 and 199, which are found within this highly glycated region of HSA, are known to be directly involved in the binding of certain of compounds to HSA. Similarly, F211 and W214 are also directly involved with drug interaction within Sudlow site 1, where modifications that were found on regions 206-212 overlap with these binding sites. The overlap of glycation sites with known drug interaction sites on HSA indicates that glycation may alter the interaction of many site 1 analytes with HSA. Additionally, residues 182-195 and 281-294, which were found to be glycated in multiple instances (See Table 4.3), are also found within Sudlow site 1 on HSA. There is also some overlap between glycation sites and Sudlow site 2 (1, 30) on HSA. Regions 414-432 and 426-442, which were identified as having high amounts of glycation in this analysis, overlap with key drug interaction sites within Sudlow sites 2 on HSA. These interaction sites include R410, W411, and V433, indicating that glycation would alter drug interaction that occurs within Sudlow site 2 on HSA. This finding is in agreement with what other authors have found, where glycation does alter the binding of HSA to various analytes (31-33).

Several glycated peptides were identified using this current approach. The semi-quantitative sorting of these residues and the p*K*a/FAS data indicated that *N*-terminus along with lysines 12, 93, 190, 195, 281, 286, 414, 432, 534 and 557 may all contain significant amounts of glycation. All of the privileged glycation sites (i.e., *N*-terminus,

K199, K281, K439, K525) were found on glycated peptides in this analysis, however, the p*K*a and FAS data does not indicate that these residues are the most likely modification sites. This discrepancy was also noted by other authors (26, 28, 29, 34), where the predicted pattern which is based on p$K_a$ and FAS is not always observed through experimentation.

Several modification sites were identified in all of the glycated-HSA samples that were examined in this chapter. Even though a high mass accuracy and unique peptides were used for this analysis, the data that was obtained for this analysis was quite complex. To simplify the analysis, information relating to early-stage glycation and late-stage glycation were split into two chapters, where information regarding late-stage glycation is presented given in Chapter 5. Further verification of the proposed modification sites is given in Chapter 6, where $^{18}$O-labeling mass spectrometry is used quantify the relative extent of modification occurring on different regions of HSA.

## 4.E. REFERENCES

(1) Peters, T. In *All About Albumin: Biochemistry, Genetics, and Medical Applications.* Academic Press: San Diego, 1996, pp 102-126.

(2) Wa, C.; Cerny, R. L.; Clarke, W. A.; Hage, D. S. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta.* **2007***, 385*, 48-60.

(3) Hayashi, T.; Mase, S.; Namiki, M. Formation of three-carbon sugar fragment at an early stage of the browning reaction of sugar with amines or amino acids. *Agric. Biol. Chem.* **1986***, 50*, 1959-1964.

(4) Monnier, V. M. Nonenzymatic glycosylation, the maillard reaction and the aging process. *J. Gerontol.* **1990***, 45*, B105-B111.

(5) Nursten, H. In *The Maillard Reaction;* Royal Society of Chemistry: Cambridge, UK, 2005, pp 214.

(6) Saito, M.; Marumo, K. Collagen cross-links as a determinant of bone quality: a possible explanation for bone fragity in aging , osteoporosis, and diabetes mellitus. *Osteoporos. Int.* **2010***, 21*, 195-214.

(7) Sell, D. R.; Nemet, I.; Monnier, V. M. Partial characterization of the molecular nature of collagen-linked flourescence: Role of diabetes and end-stage renal disease. *Arch. Biochem. Biophys.* **2010***, 493*, 192-206.

(8) Bank, R. A.; Baylis, M. T.; Lafeber, F. P.; Maroudas, A.; Tekoppele, J. M. Ageing and zonal variation in post-translational modification of collagen in normal human articuar cartilage. The age-related increase in non-enzymatic

glycation affects iomechanical properties of cartilage. *Biochem J.* **1998***, 330*, 345-351.

(9) Monnier, V. M.; Kohn, R. R.; Cerami, A. Accelerated age related browning of human collagen in diabetes mellitus. *Proc. Natl. Acad. Sci.* **1984***, 81*, 583-587.

(10) Jansirani; Anathanaryanan, P. H. A comparitive study of lens protein glycation in various forms of cataracht. *Indian J. Clin. Biochem.* **2004***, 19*, 110-112.

(11) Lewis, B. S.; Harding, J. J. The effects of aminoguanidine on the glycation (non-enzymatic glycosylation) of lens proteins. *Exp. Eye Res.* **1990***, 50*, 463-467.

(12) Turk, Z.; Misur, I.; Turk, N. Temporal association between lens protein glycation and cataract development in diabetic rats. *Acta Diabetol.* **1997***, 34*, 49-54.

(13) Biroccio, A.; Urbani, A.; Massoud, R.; di Ilio, C.; Sacchetta, P.; Bernardini, S.; Cortese, C.; Federici, G. A quantitative method for the analysis of glycated and glutathionylated hemoglobin by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Anal. Biochem.* **2005***, 336*, 279-288.

(14) Nakanishi, T.; Miyazaki, A.; Kishikawa, M.; Yasuda, M. Quantification of glycated hemoglobin by electrospray ionization mass spectrometry. *J. Mass. Spectrom.* **1997***, 32*, 773-778.

(15) Chandalia, H. B.; Krishnaswamy, P. R. Glycated hemoglobin. *Curr. Sci.* **2002***, 83*, 1522-1532.

(16) Zhang, Q.; Ames, J. M.; Smith, R. D.; Baynes, J. W.; Metz, T. O. A perspective on the maillard reaction and the analysis of protein glycation by mass spectrometry: Probing the pathogenesis of chronic disease. *J. Proteome Res.* **2009***, 8*, 754-769.

(17) Shaklai, N.; Garlick, R. L.; Bunn, H. F. Nonenzymatic glycosylation of human serum albumin alters its confirmation and function. *J. Biol. Chem.* **1984***, 259*, 3812-3817.

(18) Robb, D. A.; Olufemi, S. O.; Williams, D. A.; Midgley, J. M. Identification of glycation at the N-terminus of albumin by gas chromatography - mass spectrometry. *Biochem J.* **1989***, 261*, 871-878.

(19) Wa, C.; Cerny, R.; Hage, D. S. Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: analysis of human serum albumin as a model. *Anal. Biochem.* **2006***, 349*, 229-241.

(20) Barnaby, O. S.; Qadi, A.; Cerny, R. L.; Hage, D. S. Identification of glycation adducts using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and matlab. *Submitted to J. Proteome Res.* **2010**.

(21) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **1999***, 20*, 3551-3567.

(22) Armbuster, D. A. Fructosaminase: Structure, analysis and clinical usefulness. *Clin. Chem.* **1987***, 33*, 2153-2163.

(23) Kouzuma, T.; Usami, T.; Yamakoshi, M.; Takahashi, M.; Imamura, S. An enzymatic method for the measurement of glycated albumin in biological samples. *Clin. Chim. Acta.* **2002***, 324*, 61-71.

(24) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein $pK_a$ values. *Proteins.* **2005***, 61*, 704-721.

(25) Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **2003***, 31*, 3316-3319.

(26) Iberg, N.; Fluckiger, R. Nonenzymatic glycosylation of albumin in vivo. *J. Biol. Chem.* **1986***, 261*, 13542-13545.

(27) Barnaby, O. S.; Wa, C.; Cerny, R. L.; Clarke, W.; Hage, D. S. Quantitative analysis of glycation sites on human serum labeling using $^{16}O/^{18}O$ labeling and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Submitted to Clin. Chim. Acta.* **2010**.

(28) Bai, Y.; Ueno, H.; Manning, J. M. Some factors that influence the nonenzymatic glycation of peptides and polypeptides by glyceraldehyde. *J. Protein Chem.* **1989***, 8*, 299-315.

(29) Acharya, A. S.; Cho, Y. J.; Dorai, B. Nonreductive modification of proteins by glyceraldehyde. *Ann. New York Acad. Sci.* **2006***, 565*, 349-352.

(30) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattachara, A. A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **2005***, 353*, 38-52.

(31) Koyama, H.; Sugioka, N.; Uno, A.; Mori, S.; Nakajima, K. Effects of glycosylation of hypoglycemic drug binding to serum albumin. *Biopharm. Drug. Dispos.* **1997***, 18*, 791-801.

(32) Vorum, H.; Fisker, K.; Otagiri, M.; Pedersen, A. O.; Hansen, U. K. Calcium ion binding to clinically relevant chemical modifications of human serum albumin. *Clin. Chem.* **1995***, 41*, 1654-1661.

(33) Chiou, Y. J.; Tomer, K. B.; Smith, P. C. Effect of Nonenzymatic Glycation of Albumin and Superoxide Dismutase by Glucuronic Acid and Suprofen Acyl Glucuronide on their Functions in Vitro. *Chem. Biol. Interact.* **1999**, *121*, 141-159.

(34) Lapolla, A.; Fedele, D.; Reitano, R.; Aricò, N. C.; Seraglia, R.; Traldi, P.; Marotta, E.; Tonani, R. Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. *J. Am. Soc. Mass. Spectrom.* **2004**, *15*, 496-509.

## 4.F. APPENDIX

A table showing the m/z shifts that were identified in the glycated HSA samples is given. This table was generated using the raw output from the program without p$K_a$ or FAS sorting. This table also shows the m/z values of the detected glycated peptides, the mass of the related peptide when no modifications are present, as well as the shift in these mass values. Instances where peptides contained both early and advanced glycation products are also given this table.

Table 4.1a. Peptides that were modified to form early and advanced glycation products on glycated HSA

| | Digest | Raw Mass (Da)[a] | Residue | Unmodified Mass (Da)[b] | Suspect Adduct | Δ Mass (Da) |
|---|---|---|---|---|---|---|
| HSA-2 | Trypsin | 1563.71 | 1-10 | 1149.58 | FL-1$H_2O$+AFGP, AFGP+3-DG-H1, AFGP+THP | 414.12 |
| | Trypsin | **1738.83** | **182-195** | **1518.78** | **FL+CML** | **220.06** |
| | Trypsin | **1738.83** | **275-286** | **1432.76** | **FL+FL-1$H_2O$** | **306.10** |
| | Trypsin | **2409.13** | **542-560** | **2175.03** | **FL-2$H_2O$+Pyr, FL+CEL** | **234.05** |
| | Trypsin | **2409.13** | **539-557** | **2157.09** | **FL-2$H_2O$*2, FL-1$H_2O$+Pyr** | **252.06** |
| | Trypsin | **2409.13** | **565-585** | **2193.12** | **Pyr*2, CEL+FL-1H2O** | **216.04** |
| | Glu-C | 2107.96 | 278-292 | 1873.86 | FL+CEL, FL-2H2O+Pyr | 234.07 |
| | Glu-C | 2199.81 | 83-100 | 2073.83 | FL-2$H_2O$, CEL+MG-H1 | 126.03 |
| | Lys-C | 1480.76 | 196-205 | 1174.67 | FL+FL-1$H_2O$, FL+3-DG-H1, FL+THP | 306.10 |
| | Lys-C | 2466.27 | 414-432 | 2052.18 | FL-1$H_2O$+AFGP, AFGP+3-DG-H1, AFGP+THP | 414.12 |
| HSA-5 | Trypsin | **1738.83** | **182-195** | **1518.78** | **FL+CML** | **220.06** |
| | Trypsin | **1738.83** | **275-286** | **1432.75** | **FL+FL-1H2O** | **306.10** |
| | Glu-C | 1850.97 | 281-294 | 1652.87 | CEL+FL-2$H_2O$ | 198.05 |
| | Lys-C | 1171.60 | 5-12 | 973.52 | FL-1$H_2O$+MG-H1 | 198.05 |
| | Lys-C | 1272.74 | 525-534 | 1128.70 | FL-1$H_2O$, CEL*2 | 144.04 |
| | Lys-C | 2068.05 | 5-20 | 1905.95 | FL | 162.05 |
| | Lys-C | 1719.74 | 277-286 | 1305.62 | FL-1$H_2O$+AFGP | 414.12 |
| | Lys-C | 2466.27 | 414-432 | 2052.18 | FL-1$H_2O$+AFGP, AFGP+3-DG-H1, AFGP+THP | 414.12 |

Table 4.1a (continued). Peptides that were modified to form early and advanced glycation products on glycated HSA

| | Digest | Raw Mass (Da) [a] | Residue | Unmodified Mass (Da) [b] | Suspect Adduct [c] | Δ Mass (Da) |
|---|---|---|---|---|---|---|
| HSA-4 | Trypsin | 980.51 | 206-212 | 854.45 | FL-2$H_2O$, CEL+MG-H1 | 126.03 |
| | Trypsin | **1622.81** | **1-12** | **1424.74** | **CEL+FL-2$H_2O$, FL-1$H_2O$+MG-H1** | **198.05** |
| | Trypsin | **1622.81** | **200-212** | **1484.77** | **CEL+ArgP** | **138.03** |
| | Trypsin | **1622.81** | **403-413** | **1352.77** | **AFGP, FL-2$H_2O$+3-DG-H1, FL-2$H_2O$+THP** | **270.07** |
| | Trypsin | **1622.81** | **175-186** | **1398.78** | **FL-1$H_2O$+ArgP** | **224.07** |
| | Trypsin | 1894.76 | 546-560 | 1692.74 | CML+FL-1$H_2O$ | 202.05 |
| | Trypsin | **2028.93** | **160-174** | **1704.78** | **FL*2** | **324.11** |
| | Trypsin | **2028.93** | **258-274** | **1884.90** | **CEL*2, FL-1$H_2O$** | **144.04** |
| | Lys-C | 1467.81 | 526-536 | 1265.76 | CML+FL-1$H_2O$ | 202.05 |
| | Lys-C | 2365.07 | 501-519 | 2203.00 | FL | 162.05 |
| | Glu-C | 1939.80 | 557-571 | 1719.70 | FL+CML | 220.06 |
| | Glu-C | 2118.02 | 426-442 | 1955.99 | FL, Pyr+MG-H1 | 162.05 |
| | Glu-C | 2199.89 | 83-100 | 2073.83 | FL-2$H_2O$, CEL+MG-H1 | 126.03 |

[a] – This column shows the detected m/z values that correspond to glycated peptides.

[b] – This column shows the corresponding mass of the peptide when no modifications are present.

[c] – m/z values where multiple modifications could be assigned at the same mass are shown in bold.

# CHAPTER 5

# CHARACTERIZATION OF ADVANCED GLYCATION END PRODUCTS ON *IN VITRO* GLYCATED HUMAN SERUM ALBUMIN USING MASS SPECTROMETRY

## 5.A. INTRODUCTION

This chapter examines the advanced glycation products that are found on *in vitro* glycated human serum albumin (HSA). In Chapter 4, several peptides were identified on glycated HSA that contained early glycation products (i.e., the Amadori Product and its dehydrated forms). Based on a semi-quantitative comparison of the data, it appeared that residues 1-12, 83-100, 182-212, 278-292, 414-432, and 526-534 are being modified. The data indicated that lysines 93, 281, 286, 190, 195, 205, 534, 414, and 432 as well as the *N*-terminus are being modified to form variants of fructosyl-lysine. When the entire data set including early and late stage glycation were considered, the glycation pattern obtained was very complex, where a single m/z value could yield multiple peptide and glycation-product combinations. To simplify the analysis, the data was therefore sorted using the known p$K_a$ and fractional accessible surface area data. Additionally the data was separated based on whether early or late stage glycation was occurring. The data presented in this Chapter, therefore focuses on the identification of advanced glycation end products (AGEs) that occur, at different stages of glycation.

Following the initial stages of early glycation (1-5), any glucose in solution is slowly converted to α-oxaloaldehydes. These compounds, which have been shown to be significantly more reactive than their glucose counterparts (6-9) lead to the formation of AGE's. This study examines AGE's that result from the interaction of proteins with glyoxal, methylglyoxal, and 3-deoxyglucosone (10, 11). In some cases, reactions that occur on the Schiff-base or the Amadori Product can also lead to the formation of AGEs.

In a previous study that looked at the rate of α-oxaloaldehyde formation from a 3 week incubation of glucose (at 37 ºC), it was found that glyoxal had the highest concentration followed by 3-deoxyglucosone then methylglyoxal (11). From this study, they found that incubation of a 50 mM D-glucose solution yielded μM concentrations of α-oxaloaldehydes, where glyoxal had a 20 fold molar excess compared to methylglyoxal, and 3-deoxyglucosone had a 2.5 fold molar excess compared to methylglyoxal. Even though there is a large concentration difference between the initial sugar compound and α-oxaloaldehydes, previous reports indicate that the moles of AGEs found on glycated proteins is similar to the extent of modification found in early glycation (i.e., approximately 1 – 5 moles of modification per mol of HSA) (12, 13). It is therefore expected that the HSA samples examined in this chapter should have an extent of AGE modification that is similar to the extent of early glycation. In addition, AGEs that are derived from glyoxal should be quite abundant, followed by 3-deoxyglucosone and methylglyoxal derived AGEs (see Table 5.1 for an example of adducts derived from these α-oxaloaldehyde).

Table 5.1. List of early and advanced-stage glycation products and reactive dicarbonyl compounds that result in there formation

| **Early-stage glycation** | **Abbreviated** | **Residue** | **ΔM (Da)** |
|---|---|---|---|
| Fructosyl-lysine | FL | Lys | 162.05 |
| Fructosyl-lysine-1$H_2O$ | FL-1$H_2O$ | Lys | 144.04 |
| Fructosyl-lysine-2$H_2O$ | FL-2$H_2O$ | Lys | 126.03 |
| | | | |
| **Glyoxal derived adducts** | **Abbrev.** | **Residue** | **ΔM (Da)** |
| $N_\varepsilon$-Carboxymethyl-lysine | CML | Lys | 58.01 |
| $N_\varepsilon$-(5-Hydro-4-imidazolon-2-yl)ornithine | G-H1 | Arg | 39.99 |
| | | | |
| **Methylglyoxal derived adducts** | **Abbrev.** | **Residue** | **ΔM (Da)** |
| $N_\varepsilon$-Carboxyethyl-lysine | CEL | Lys | 72.02 |
| $N_\varepsilon$-(5-Hydro-5-methyl-4-imidazolon-2-yl)ornithine | MG-H1 | Arg | 54.01 |
| Tetrahydropyrimidine | THP | Arg | 144.04 |
| Argpyrimidine | ArgP | Arg | 80.03 |
| | | | |
| **3-deoxyglucosone derived adducts** | **Abbrev.** | **Residue** | **ΔM (Da)** |
| Pyrraline | Pyr | Lys | 108.02 |
| $N_\varepsilon$-[5-(2,3,4-Trihydroxybutyl)-5-hydro-4-imidazolon-2-yl]ornithine | 3-DG-H1 | Arg | 144.04 |
| Imadazolone B | IB | Arg | 142.03 |
| 1-Alkyl-2-formyl-3,4-glycosyl-pyrrole | AFGP | Arg/Lys | 270.07 |

The same sample preparation and data analysis steps were used for this Chapter and Chapter 4, therefore a detailed explanation of methods that were used are not given here. The optimizations that were examined in Chapter 3 are also applied here. In Chapter 4, emphasis is placed on the identification and location of early-glycation adducts that were present on glycated HSA, where both peptides that were linked specifically to early-stage glycation products or a mixture of early and late-stage glycation products were considered. In contrast, this chapter examines modifications that were identified that were due to AGEs or peptides where the majority of modifications were mostly AGEs.

## 5.B. RESULTS

**5.B.1. Frequency of AGE Formation on Glycated HSA.** A summary of the AGEs identified using this approach is given in Table 5.2. When only the peptides that were modified to form AGEs were considered, 2 AGE modified peptides were found in the HSA-2 sample, 8 AGE modified peptides were identified in the HSA-5 sample, and 5 AGE modified peptides were identified in the HSA-4 sample. When peptides that include a mixture of early and late stage glycation products were considered, 7, 6, and 9 peptides were found to be modified in the HSA-2, HSA-5, and HSA-4 samples respectively.

There does appear to be a trend where the number of peptides that are only modified by AGEs, increases with increasing glycation periods times. This trend is less obvious when both early and late-stage glycation is considered. This observation is consistent with the fact that late-stage Maillard products are formed at a fairly slow rate, and accumulate on proteins over time (14-16). Interestingly, the HSA-5 sample had more AGE modification compared to the HSA-4 sample, even though the HSA-4 sample was incubated in a 2-fold excess of D-glucose compared to the HSA-5 sample. The increased AGE content of the HSA-5 sample was an unexpected result, considering there was only a 1 week difference in the incubation period.

The dependence of AGE formation on the total extent of glycation is a pattern that was generally not observed when looking at the formation of early stage glycation products (see Chapter 4). The fact that a greatest amount of AGEs were identified in the

Table 5.2. List of AGEs that were identified in the glycated HSA samples. All three stages of glycation are shown here, which is represented by the HSA-2, HSA-4, and HSA-5 samples

| | Digest | Residues | Lysine | Arginine | Suspect Modification Sites[a] |
|---|---|---|---|---|---|
| HSA-2 | Trypsin | 390-410 | K402 | R410 | K402 (Pyr) & R410 (G-H1) |
| | Lys-C | 206-212 | K212 | R209 | **K212** (CEL) |
| HSA-5 | Trypsin | 520-525 | K524, K525 | R521 | K524 (Pyr) & R521 (G-H1) |
| | Trypsin | 210-218 | K212 | R218 | **K212** (CEL) & **R218** (ArgP) |
| | | 5-12 | K12 | R10 | K12 (FL-$1H_2O$) & R5 (MG-H1) |
| | Trypsin | 187-197 | K190, K195 | R197 | K190 (Pyr) & R197 (ArgP) |
| | Trypsin | 137-144 | K137 | R144 | K137 (AFGP) & R144 (ArgP) |
| | Trypsin | 324-336 | n/a | R336 | R336 (G-H1) |
| | Trypsin | 318-336 | K323 | R336 | K323 (CEL) & R336 (AFGP) |
| | Glu-C | 189-208 | K190, K195, K199, K205 | R197 | K190 or K199 (Pyr) & R197 (IB) |
| | Glu-C | 426-442 | K432, K436, K439 | R428 | **K436** (Pyr) & **R428** (G-H1) or K436 (CEL) & R428 (ArgP) |

[a] – The most likely region of modification within a given peptide, as indicated by the p$K_a$ and fractional accessible surface area (FAS) data, is also shown here in bold

Table 5.2 (continued). List of AGEs that were identified in the glycated HSA samples. All three stages of glycation are shown here, which is represented by the HSA-2, HSA-4, and HSA-5 samples

| | Digest | Residues | Lysine | Arginine | Suspect Modification Sites[a] |
|---|---|---|---|---|---|
| HSA-4 | Trypsin | 476-484 | n/a | R484 | R484(IB) |
| | Trypsin | 437-445 | K439, K444 | R445 | **K439**(AFGP), **R445**(G-H1) |
| | | 138-145 | n/a | R144, R145 | R144 & R145 (IB & AFGP) |
| | Trypsin | 206-218 | K212 | R209, R218 | K199 (FL-1$H_2O$) & R218 (3-MG-H1) or R209 & R218 (THP & 3-MG-H1) |
| | | 182-195 | K190, K195 | R186 | K190(Pyr), R186(ArgP) |
| | | 198-209 | K199, K205 | R209 | **K199**(AFGP) or K199&K205 (FL & Pyr) or K199&K205 (FL-1$H_2O$ & FL-1$H_2O$) or K199 & R209 (FL-1$H_2O$ & 3-DG-H1 or THP) |
| | Glu-C | 312-321 | K313, K317 | n/a | K313 & K317 (AFGP) |
| | Glu-C | 189-208 | K190, K195, K199, K205 | R197 | K190 (Pyr) & R197 (IB) |

[a] – The most likely region of modification within a given peptide, as indicated by the p$K_a$ and FAS data, is also shown here in bold

HSA-5 sample indicated that the amount of AGEs formed was more dependent on the amount of time used for HSA-glucose incubations than the concentration of D-glucose that was used. There were some cases where peptides could be assigned to more than one glycation product. This ambiguity in the assignment of modification sites occurred because certain m/z values could be linked to multiple peptide/AGE combinations at the maximum allowed error limit of 25 ppm. The assignment of multiple peptide/AGE combinations to the same mass value is related to a range of processes including the mass accuracy (17), level of AGE heterogeneity (18), and the selectivity of this current approach (see Chapter 3).

**5.B.2. AGEs Identified on the glycated HSA samples.** There were several cases where AGE formation could be identified solely based on the mass shift data. This type of assignment was common in this chapter because only AGE-related compounds were considered, and repeat assignments (i.e, for m/z values that were linked to both early and late stage glycation in Chapter 4) were not considered. For instance, a peptide having a mass of 2690.4 Da was detected in the trypsin digest of the HSA-2 sample. This mass value indicated that residues 390-410 was being modified to form pyrraline (Pyr) and *N*_-(5-Hydro-4-imidazolon-2-yl)ornithine (G-H1). Since this peptide contained one lysine and one arginine residue, G-H1 and Pyr were assigned to R410 and K402 respectively. Similarly a mass of 1280.49 Da was identified in the tryptic digest of the HSA-4 sample. This m/z value corresponded to a mass shift of 142 Da that occurred on residues 476-484 on HSA. This mass shift indicated that residues 476-484 were being modified to form imadizolone B (IB), which is specific for arginine residues. Since residues 476-484 contain a single arginine, R428 was being modified to form IB. AGEs could also be

assigned with little ambiguity to lysine or arginine residues in the HSA-5 sample. For example, a m/z value of 1679.8 Da was identified in the tryptic digest of the HSA-5 sample. This m/z value corresponded to residues 324-336 on HSA was being modified to form G-H1. Since there is only a single arginine on this peptide, R336 was found to be modified to form G-H1. Similar assignments were made on several peptides in the glycated HSA samples, where AGE assignment on residues 206-212, 137-144, and 318-336 were fairly straight forward.

There were several instances where assigning a particular modification site was difficult, because either multiple modifications overlapped to give the same mass, or because the modified peptide had multiple lysine or arginine residues that can be modified. For instance, a mass of 2444.2 Da was identified in the Glu-C digest of the HSA-5 sample. This mass value corresponded to a mass shift of 250.0 Da that occurred on residues 198-208 on HSA. The mass shift indicated that residues 198-208 were being modified to form Pyr and IB. Arginine 197 was found to be modified to form IB because it is the only arginine residue that's available for modification. The assignment of Pyr, which is a lysine specific modification, was not as straightforward because lysines 190, 195, 199, and 205 are all present within regions 189-208 of HSA. An identical modification was also found in the HSA-4 sample, where a peptide spanning residues 198-208 was found to be modified to form Pyr and IB. An example of a scenario where multiple AGEs could be assigned to the same m/z value was found on residues 206-218 in the tryptic digest of the HSA-4 sample. A mass shift of 198.1 Da was identified on these residues, which corresponded to the formation of $N_{\varepsilon}$-(5-Hydro-5-methyl-4-imidazolon-2-yl)ornithine (MG-H1) and dehydrated fructosyl-lysine (FL-1$H_2O$) or MG-

H1 and tetrahydropyrimidine (THP). This indicated that either a combination of K205 and K218 or R209 and R218 are being modified.

There were also cases where a detected m/z shift could be linked to several peptide/AGE combinations. For instance, a m/z value of 1171.6 Da was identified in the HSA-2 sample, which could be due to the modification of residues 210-218 or 5-12 on HSA. In this case, modification of residues 210-218 to form CEL and ArgP, or modification of residues 5-12 to form FL-1$H_2O$ and MG-H1, would have associated mass shifts of 152.0 Da and 198.1 Da respectively. Two other examples of this ambiguity in glycation site assignment were identified in the HSA-4 sample, where m/z values of 1495.6 Da and 1706.8 Da were identified as being potential modification sites. The m/z value of 1495.6 Da resulted from modifications that occurred on regions 437-445 or 138-145 on HSA. Similarly, the m/z value of 1706.8 Da corresponded to modifications that occurred on regions 182-105, 198-209, or 206-218 on HSA.

In cases where multiple modification sites, overlapping glycation products, or multiple combinations of modified peptides occurred, additional information was used to determine the most likely modification sites. In Chapter 4, the most likely modification sites were identified by comparing the suspected modification sites to the p$K_a$ and fractional accessible surface area (FAS) of a given amino-acid residue. Similar approaches are used in this current chapter, where p$K_a$ and FAS data is used to indicate AGEs on HSA, and the most likely location of these AGEs.

**5.B.3. Location of the most likely modification sites based on p$K_a$ and FAS data.** To overcome the complexity of AGE assignment, the p$K_a$ and the FAS values for AGE-

linked residues used to select the most likely modification sites. These values were calculated using tools that are available online, as described in the literature (19, 20). In general, lysine or arginine sidechains which have higher p$K_a$ values are expected to be less reactive with α-oxoaldehydes because a greater percentage of the amino-sidechains exists in the protonated state, precluding AGE formation. The FAS on the other hand is a measure of the accessibility of a given amino acid residue. Residues that have very low FAS values are therefore expected to be less modified than residues that are more accessible to the solvent. The detected peptides were sorted first by the respective sidechain p$K_a$ values and then by the respective FAS to determine which AGE and peptide combination was the most likely to occur at a given m/z value.

The assignment of AGEs on peptides resulting from the HSA-2 digest didn't require additional p$K_a$ or FAS based sorting because the modifications occurring could only be linked to specific residues on a given peptide. In the HSA-5 sample, there were 5 peptides where the p$K_a$ and FAS data were used to differentiate the most likely modification sites. For example, residues 520-525 on HSA-5 sample contained K524 and K525, where one of these residues is being modified to for Pyr. Lysine 524 has p$K_a$ and FAS values of 10.4 and 0.54 respectively, compared to p$K_a$ and FAS values of 15 and 0.07 that were obtained for K525. The p$K_a$/FAS data is strong indicator that K524 was being modified to form Pyr. Similarly, residues 189-208 was found to be modified on the HSA-5 sample, where this region of HSA is believed to be modified to form Pyr. This region of HSA includes K190, K195, K199, and K205, which made it difficult to pinpoint where this lysine specific modification was occurring. Of these residues, K190 and K199 both had low p$K_a$ values of 6.23 and 7.47, indicating that they are both fairly

reactive with α-oxaloaldehydes. Lysines 190 and 199 also have FAS values of 0.42 and 0.16 respectively. The p$K_a$ and FAS data in this case indicated that K190 was being modified to form Pyr. Similar assignments were made for residues 426-442 and residues 189-208 in the HSA-5 and HSA-4 digests respectively.

In some cases, the p$K_a$ and FAS were used to differentiate between multiple combinations of peptides and AGEs. There were a total of 3 peptides in the HSA-4 and HSA-5 sample where this type of sorting was used. For the HSA-5 sample, a m/z value of 1171.6 Da was identified which could result from the modification of residues 210-218 or 5-12 on HSA. Residues 210-218 on HSA contained K212 and R218 and residues 5-12 contain R10 and K12. When both lysines are compared, the FAS values for lysines 12 and 212 were 0.59 and 0.40 respectively and the p$K_a$ in both cases was 10.43. In comparison, there was a large difference in the p$K_a$ for arginines 10 and 218, where R10 had a p$K_a$ of 16.02 and R218 had a p$K_a$ of 9.64. The p$K_a$ of R218 was well below the average arginine p$K_a$ on HSA, which indicated that the peptide containing residues 5-12 is the most likely modification site. Based on this information, K212 was being modified to form CEL and R10 was being modified to form ArgP. Similar assignments were made, where residues 437-445 and 198-209 were both identified as being likely modification sites, based on the p$K_a$ and FAS data. Based on p$K_a$/FAS sorting, K439 and R445 were found to be modified to form AFGP and G-H1 respectively. Similarly, K199 and R209 were found to be modified to form FL-1$H_2O$ and 3-DG-H1 or THP respectively.

As was indicated in Chapter 4, there were situations where the most commonly identified modification pattern doesn't match the modification pattern that was determined based on p$K_a$ and FAS data. This was certainly the case for residues 189-208,

426-442, 437-445, and 520-525 where the most likely modifications are expected to occur on K199, K439, and K525 respectively. This discrepancy was also noted by other authors (see Chapter 4 for the relevant references), where the predicted pattern which is based on p$K_a$ and FAS data is not always observed through experimentation.

**5.B.4. Semi-quantitative analysis of AGEs.** In Chapter 4, the rate of occurrence of modification sites in the glycated HSA samples was used to estimate the relative extent of modification that occurred on various parts of HSA. A similar approach was used for ranking the residues which were modified by AGEs in this chapter. In some cases, a peptide may be modified by both AGEs and early stage glycation products. When both types of modification were present, residues that were modified by early stage glycation products were factored into the analysis. For this classification, the data that is presented in this chapter and the data that was presented in Chapter 4 (i.e., peptides with amino-acid residues that could be linked to both early and late stage glycation) are both examined.

When only the data that is presented in this chapter is considered, lysines 190, 199, 212, and 439 as well as arginine 336 were found to have the highest amount of AGE modifications. Each of the residues listed above were found to be modified in at least 2 peptides. When both early and late stage glycation are considered, K190 and K439 were found to be modified in at least 3 different peptides, which indicated that extensive modification was occurring on those residues. Lysines 432, 436, 525 as well as arginine 336 were found to overlap with at most 2 peptides. A diagram summarizing the semi-quantitative ranking of AGE-modification sites is summarized in Figure 5.1. The AGEs that were detected in this study were compared to previous qualitative (21) and quantitative data (22) that was obtained for commercially glycated HSA (see

Figure 5.1. Structure of HSA showing the most likely AGE modification sites based on semi-quantitative data. The most reactive residues are indicated in red (i.e., for amino acids that were present in 3 overlapping peptides) followed by the second most reactive regions shown in purple (i.e., for amino acids that were present in 2 overlapping peptides).

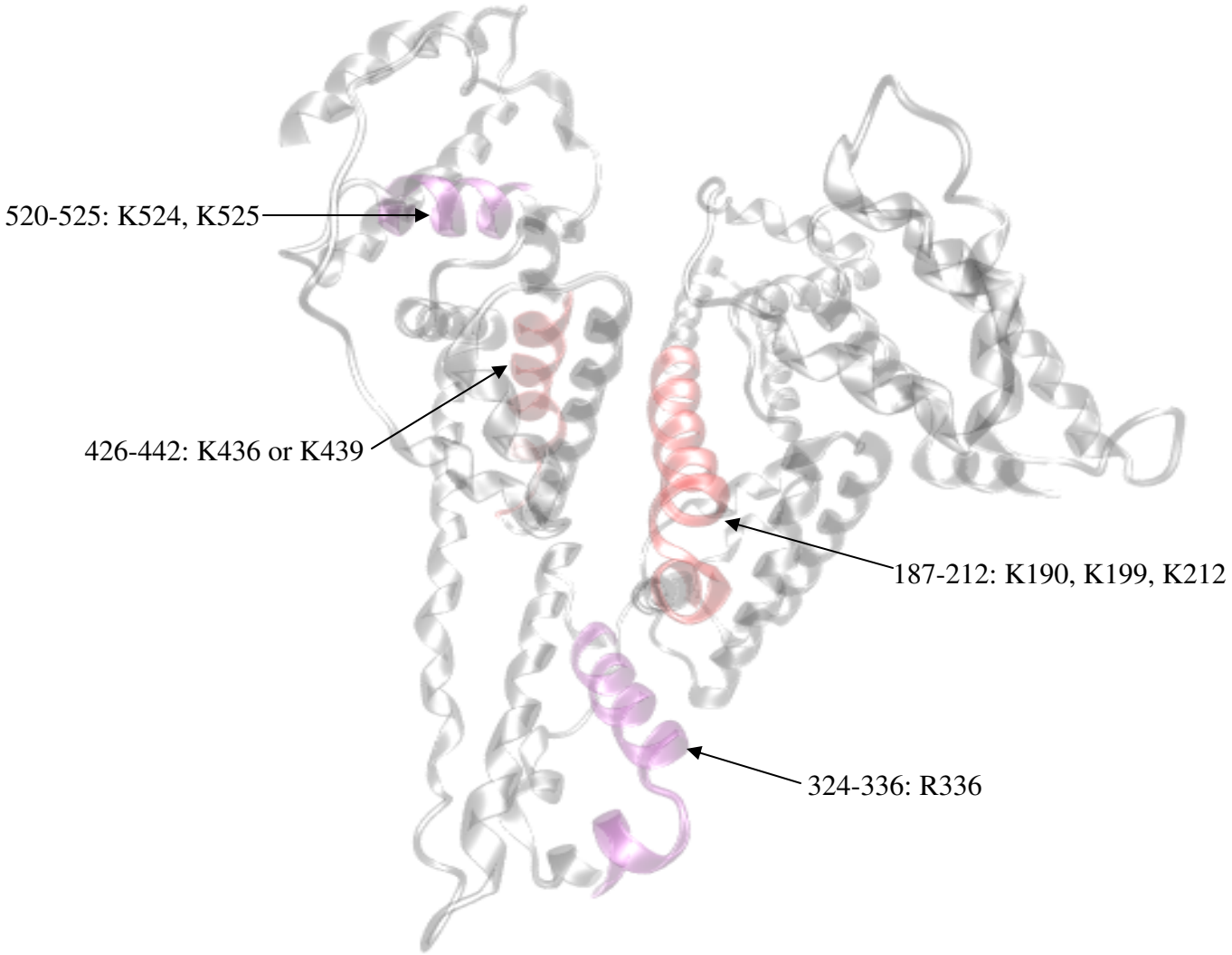

520-525: K524, K525
426-442: K436 or K439
187-212: K190, K199, K212
324-336: R336

Table 5.3. Comparison of AGE-modified peptides that were found in this study, with glycation sites that were previously found on commercially glycated HSA

| HSA Sample | Residue 1[a] | Residue 2[b] | Residue 3[c] | Overlapping Residues | Modification |
|---|---|---|---|---|---|
| HSA-5 | 520-525 | 520-525 | 525-534 | K525 | Pyr |
| HSA-2, HSA-5 | 206-212, 210-218 | 200-218 | 206-212, 213-218 | R209 | MG-H1, THP |
| HSA-5, HSA-4 | 426-442, 437-445 | 426-442 | 415-439 | K436, K439, & R428 | AFGP, Pyr, CEL, G-H1, ArgP |
| HSA-5, HSA-5 | 198-208, 189-208 | 196-209 | n/a | K199, K205, R197 | Pyr, 3-DG-H1, THP |
| HSA-2, HSA-5 | 189-208 | 191-205 | n/a | K190, K195, K199 | Pyr, IB |
| HSA-4 | 476-484 | 467-484 | n/a | R484 | G-H1 |
| HSA-5 | 137-144 | 133-141 | n/a | K137 | AFGP |
| HSA-5, HSA-4 | 312-321, 318-336 | n/a | 318-323 | K313, K317, K323 | AFGP, CEL |

[a] – Residue 1 represents the region of HSA that was found to be modified by early-stage glycation products in this study.

[b] – Residue 2 represents the region of HSA that was found to be modified by early-stage glycation products that was identified for commercially glycated HSA (21).

[c] – Residue 3 represents the region of HSA that was previously found to have significant levels of glycation which were quantified using $^{18}O$-labeling mass spectrometry (22).

Table 5.3). Eight peptides that were found to be modified in this current study were also found to have significant levels of modification in commercially glycated HSA. Four of these modified peptides contain K199, K439, and K525, which are residues that are expected to have the largest amount of modification in glycated HSA (23, 24). Interestingly, these peptides were also found to be linked to early-stage glycation products in the Chapter 4. Of these residues, K525 and K439 were both previously identified in our previous qualitative (21) and quantitative studies (see Chapter 2), while K199 was found to be modified by glycation products in commercially glycated HSA (21). Several peptides spanning residues 206-218, 476-484, 137-144, and 312-336 were found to be modified in this current study and in a previous studies using commercially glycated HSA (21). Of these peptides, the semi-quantitative ranking indicated that arginines 209 and 484, as well as lysines 137, 313, 317, and 323, are all possible modification sites.

## 5.C. SUMMARY

When p$K_a$ and FAS sorting as well as semi-quantitative ranking were used, a region spanning residues 189-212 on HSA was found to have extensive AGE modification. The same observation was made in the analysis of glycation sites on *in vitro* glycated HSA (see Chapter 4), where several amino acids within residues 196-205 were found to be modified by early glycation products. This finding indicated that a region on HSA spanning residues 189-212 may be highly modified during glycation. In addition, K199, which is a commonly identified modification site on HSA, is found within this same region on HSA. The semi-quantitative comparison also showed that residues 324-336, 426-442, and 520-525 were also found to have high levels of modification.

The location of AGEs on HSA was compared to the major drug binding sites on this protein (i.e., Sudlow sites I and II) (25, 26). In this comparison, several modified peptides were found within both Sudlow sites. Of the residues that were semi-quantitatively ranked, residues 189-212 is found within Sudlow site I. The proximity of AGE-modified residues (i.e., K190 or K199 and R197), to Sudlow site 1 is a strong indicator that AGE formation on these residues may interfere with Site 1 ligand interaction. Similarly K212 and R218, which were being modified to form AGEs, are also located within Sudlow site 1 of HSA. The proximity of K212 to Y211, which is a known drug interaction site, indicates that drug binding may be altered in this region of HSA during glycation. The AGE modification that occurred on R218 may also affect HSA-drug interaction, because this residue is also a known drug binding site on HSA.

Several modified peptides were also found in the vicinity of Sudlow site II on HSA. For instance, residues 426-442, which has a high level of AGE modification, was found in Sudlow site II. K436 or K439 are believed to be involved in AGE formation, where this modification may interfere with drug interaction that occurs at V433 or Cys-438 on HSA. Similarly, peptides spanning residues 390-410, 437-445, and 476-484 are all found within this region of HSA, where lysines 402 and 484 as well as arginines 410, 428, 445, and 484 are believed to have been modified to form AGEs. Arginine 410 is a known drug interaction site on HSA; therefore AGE formation on this residue would directly interfere with drug interaction at this site. The remaining AGE modified peptides are located in close proximity to residues Y411, L430, A449, and R485, indicating that AGE formation may alter the drug interaction that occur at these key drug binding sites.

The results indicate that AGE modification is indeed occurring in both Sudlow sites on HSA. The modifications that occur on HSA appear to be heterogeneous in nature, where a large number of both early and advanced stage glycation products are found within both drug interaction sites. Many regions of HSA that were found to have high levels of modification in this study were also identified in previous studies that examined commercially glycated HSA. When this comparison was done, peptides containing K525, R209, K439, and K199 were consistently identified as being modification sites. This finding is consistent with the most commonly reported glycation sites in the literature (23, 24, 27) and the results that were obtained in Chapter 4.

## 5.D. REFERENCES

(1) Li, W.; Ota, K.; Nakamura, J.; Naruse, K.; Nakashima, E.; Oiso, Y.; Hamada, Y. Antiglycation effect of glyclazide on in vitro AGE formation from glucose and methylglyoxal. *Exp. Biol. Med.* **2008**, *233*, 176-179.

(2) Makita, Z.; Vlassara, H.; Cerami, A.; Bucala, R. Immunochemical detection of advanced glycosylation end products in vivo. *J. Biol. Chem.* **1992**, *267*, 5133-5138.

(3) Magalhaes, P. M.; Appell, H. J.; Duarte, J. A. Involvement of advanced glycation end products in the pathogenesis of diabetic complications: the protective role of regular physical activity. *Eur. Rev. Aging. Phys. Act.* **2008**, *5*, 17-29.

(4) Hayashi, T.; Mase, S.; Namiki, M. Formation of three-carbon sugar fragment at an early stage of the browning reaction of sugar with amines or amino acids. *Agric. Biol. Chem.* **1986**, *50*, 1959-1964.

(5) Zhang, Q.; Ames, J. M.; Smith, R. D.; Baynes, J. W.; Metz, T. O. A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: Probing the pathogenesis of chronic disease. *J. Proteome Res.* **2009**, *8*, 754-769.

(6) American Diabetes Association. All about diabetes. http://www.diabetes.org/about-diabetes.jsp (accessed 9/2006, 2009).

(7) Lapolla, A.; Fedele, D.; Reitano, R.; Arico, N. C. Enzymatic digestion and mass spectrometry in the studies of advanced glycation end products/peptides. *J. Am. Soc. Mass. Spectrom.* **2004**, *15*, 496-509.

(8) Lapolla, A.; Fedele, D.; Reitano, R.; Bonfante, L.; Guizzo, M.; Seraglia, R.; Tubaro, M.; Traldi, P. Mass spectrometric study of in vivo production of advanced glycation end-products/peptides. *J. Mass. Spectrom.* **2005**, *4*0, 969-972.

(9) Lapolla, A.; Fedele, D.; Seraglia, R.; Traldi, P. The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes: An update. *Mass. Spectrom. Rev.* **2006**, *25*, 775.

(10) Takeuchi, M.; Kikuchi, S.; Sasaki, N.; Suzuki, T.; Watai, T.; Iwaki, M.; Bucala, R. l. Y., S. Involvement of advanced glycation end-products (AGEs) in Alzheimer's disease. *Curr. Alzheimer Res.* **2004**, 1, 39-46.

(11) Thornalley, P. J.; Langborg, A.; Minhas, H. S. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **1999**, *344*, 109-116.

(12) Abordo, E. A.; Thornalley, P. J. In *Pro-inflammatory cytokine synthesis by human monocytes induced by proteins minimally-modified by methylglyoxal.* O'Brien, J., Nursten, H. E., Crabbe, M. J. C. and Ames, J. M., Eds.; The Maillard Reaction in Foods and Medicine; The Royal Society of Chemistry: Cambridge, UK, 1998; pp 357-362.

(13) Lapolla, A.; Fedele, D.; Seraglia, R.; Catinella, S.; Baldo, L.; Aronica, R.; Traldi, P. A new effective method for the evaluation of glycated intact plasma proteins in diabetic subjects. *Diabetologia.* **1995***, 3*8, 1076-1081.

(14) Stitt, A. W.; Curtis, T. M. Advanced glycation and retinal pathology during diabetes. *Pharmacol. Rep.* **2005***, 5*7, 156-168.

(15) Ahmed, N.; Argirov, O. K.; Minhas, H. S.; Cordeiro, C. A. A.; Thornalley, P. J. Assay of advanced glycation endproducts (AGEs): Surveying AGEs by chromatographic assay with derivitization by 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and application to $N_{\varepsilon}$-carboxymethyl-lysine- and $N_{\varepsilon}$-(1-corboxyethyl)lysine-modified albumin. *Biochem. J.* **2002***, 36*4, 1-14.

(16) Thornalley, P. J. Dicarbonyl intermediates in the Maillard reaction. *Ann. N. Y. Acad. Sci.* **2005***, 104*3, 111-117.

(17) Edmondson, R. D.; Russel, D. H. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass measurement accuracy by using delayed extraction. *J. Am. Soc. Mass. Spectrom.* **1996***,* 7, 995-1001.

(18) Henle, T.; Deppisch, R.; Beck, W.; Hergesell, O.; Hansch, G. M.; Ritz, E. Advanced glycation end-products (AGE) during haemodialysis treatment: discrepant results with different methadologies reflecting the heterogeneity of AGE compounds. *Nephrol. Dial. Transplant.* **1999***, 1*4, 1968-1975.

(19) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein $pK_a$ values. *Proteins.* **2005***, 6*1, 704-721.

(20) Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **2003***, 3*1, 3316-3319.

(21) Wa, C.; Cerny, R. L.; Clarke, W. A.; Hage, D. S. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta.* **2007***, 38*5, 48-60.

(22) Barnaby, O. S.; Wa, C.; Cerny, R. L.; Clarke, W.; Hage, D. S. Quantitative analysis of glycation sites on human serum labeling using $^{16}O/^{18}O$ labeling and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Submitted to Clin. Chim. Acta.* **2010**.

(23) Iberg, N.; Fluckiger, R. Nonenzymatic glycosylation of albumin in vivo. *J. Biol. Chem.* **1986***, 26*1, 13542-13545.

(24) Peters, T. In *All About Albumin: Biochemistry, Genetics, and Medical Applications.* Academic Press: San Diego, 1996; pp 102-126.

(25) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattachara, A. A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **2005***, 35*3, 38-52.

(26) He, X. M.; Carter, D. C. Atomic structure and chemistry of human serum albumin. *Nature.* **1992***, 35*8, 209-215.

(27) Robb, D. A.; Olufemi, S. O.; Williams, D. A.; Midgley, J. M. Identification of glycation at the N-terminus of albumin by gas chromatography - mass spectrometry. *Biochem J.* **1989***, 26*1, 871-878.

## 5.E. APPENDIX

A table showing the AGE-related m/z shifts that were identified in the glycated HSA samples is given. This table was generated using the raw output from the program without p$K_a$ or FAS sorting. This table provides data that is complements the data provided in Table 4.1a. Instances where peptides are modified to form mixtures of overlapping early and advanced glycation products are given in Table 4.1a. Table 5.1a provides information for peptides that only contained AGEs.

Table 5.1a Peptides that were modified to form AGEs in the glycated HSA samples

| | Digest | Raw Mass (Da)[a] | Residue | Unmodified Mass (Da)[b] | Suspect Adduct[c] | Δ Mass (Da) |
|---|---|---|---|---|---|---|
| **HSA-2** | Trypsin | 2690.35 | 390-410 | 2542.28 | Pyr+G-H1 | 148.02 |
| | Glu-C | **2547.21** | **566-585** | **2007.06** | **AFGP*2** | **540.15** |
| | Glu-C | **2547.21** | **168-188** | **2305.10** | **FL+ArgP** | **242.08** |
| | Lys-C | 926.46 | 206-212 | 854.45 | CEL | 72.021 |
| **HSA-5** | Trypsin | 949.49 | 520-525 | 801.49 | Pyr+G-H1 | 148.02 |
| | Trypsin | **1171.60** | **5-12** | **973.52** | **FL-1$H_2O$+MG-H1** | **198.05** |
| | Trypsin | **1171.60** | **210-218** | **1019.58** | **CEL+ArgP** | **152.05** |
| | Trypsin | 1364.65 | 187-197 | 1176.60 | Pyr+ArgP | 188.05 |
| | Trypsin | 1405.72 | 137-144 | 1055.59 | AFGP+ArgP | 350.10 |
| | Trypsin | 1679.75 | 324-336 | 1639.78 | G-H1 | 39.99 |
| | Trypsin | 2658.16 | 318-336 | 2316.10 | CEL+AFGP | 342.10 |
| | Glu-C | 1989.97 | 426-442 | 1841.95 | Pyr+G-H1 | 148.01 |
| | Glu-C | 2108.03 | 426-442 | 1955.99 | CEL+ArgP | 152.05 |
| | Glu-C | 2444.21 | 189-208 | 2194.18 | Pyr+IB | 250.05 |

Table 5.1a (continued) Peptides that were modified to form AGEs in the glycated HSA samples

| | Digest | Raw Mass (Da)[a] | Residue | Unmodified Mass (Da)[b] | Suspect Adduct[c] | Δ Mass (Da) |
|---|---|---|---|---|---|---|
| HSA-4 | Trypsin | 1280.49 | 476-484 | 1138.50 | IB | 142.03 |
| | Trypsin | **1495.67** | **437-445** | **1185.56** | **AFGP+G-H1** | **310.07** |
| | Trypsin | **1495.67** | **138-145** | **1083.59** | **IB+AFGP** | **412.10** |
| | Trypsin | **1706.83** | **182-195** | **1518.78** | **Pyr+ArgP** | **188.05** |
| | Trypsin | **1706.83** | **198-209** | **1436.77** | **FL+Pyr, AFGP, FL-1$H_2O$+FL-2$H_2O$, FL-1$H_2O$+FL-2$H_2O$** | **270.07** |
| | Trypsin | **1706.83** | **198-209** | **1436.77** | **FL-2$H_2O$+3-DG-H1, FL-2$H_2O$+THP, AFGP** | **270.07** |
| | Trypsin | **1706.83** | **206-218** | **1508.81** | **FL-1$H_2O$+MG-H1, 3-DG-H1+MG-H1, THP+MG-H1** | **198.05** |
| | Glu-C | 1753.71 | 312-321 | 1213.55 | AFGP*2 | 540.15 |
| | Glu-C | 2444.24 | 189-208 | 2194.18 | Pyr+IB | 250.05 |

[a] – This column shows the detected m/z values that correspond to glycated peptides.

[b] – This column shows the corresponding mass of the peptide when no modifications are present.

[c] – m/z values where multiple modifications could be assigned at the same mass are shown in bold.

# CHAPTER 6

# QUANTITATIVE ANALYSIS OF GLYCATION SITES ON IN VITRO GLYCATED HUMAN SERUM ALBUMIN

## 6.A. INTRODUCTION

The glycation of various proteins in the body has been linked to many of the chronic complications that are commonly diagnosed in diabetes, including nephropathy (1), macro-vascular complications (2), and cataract formation (3). Glycation that occurs on human serum albumin (HSA) has been shown to alter the structure, function, and binding capabilities of this protein (4, 5); however, there is little quantitative information about the regions on HSA that are most prone to modification (5, 6). The objective of this study is to obtain quantitative information about the levels of glycation that occur on glycated human serum albumin (HSA) and link this information to the glycation products that were previously identified on HSA (see Chapters 4 and 5). This information can be used to model the expected glycation pattern on HSA that is glycated *in vivo*.

The same HSA and glycated HSA samples that were previously used for qualitative studies (see Chapters 4 and 5) were used to obtain quantitative data in this analysis. This quantification was performed using $^{18}$O-labeling (7, 8) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) (9). The HSA samples were glycated under physiological conditions for either two weeks or five weeks to yield glycated HSA that contained 0.72 (± 0.1) and 1.23 (± 0.1) mol hexose/mol

HSA respectively, as determined by an enzyme-based fructosamine assay (10, 11). This level of modification represents the expected amount of HSA glycation that on occurs during pre-diabetes (~0.5 - 1 mol hexose per mol HSA) and diabetes (~ 1-3 mol hexose per mol HSA) respectively (12, 13). This level of glycation is consistent with the 2-3 fold increase in the concentration of glycated peptides that is typically observed in diabetes (1).

To quantify the modification sites, the glycated HSA samples were digested in $^{18}O$-labeled water and the control HSA samples were digested in $^{16}O$-labeled water. The resulting peptides were then mixed in a 50:50 (v/v) ratio. Mass spectra for the $^{16}O$ digest, $^{18}O$ digest, and mixed digest were then acquired and the $^{16}O/^{18}O$ ratios were determined. The control sample was also digested in $^{18}O$ water, and this sample was used to obtain reference $^{16}O/^{18}O$ values for control HSA versus itself.

The coupling of $^{16}O/^{18}O$-labeling and MALDI-TOF-MS is advantageous for this analysis because the peptides can be labeled in a single step, reducing the chance of encountering systematic labeling errors (7). The use of MALDI-TOF-MS is also advantageous because it limits the amount of interaction that $^{18}O$-labeled peptides would have with aqueous solvents containing regular water, such as solvents that are used for LC-MS systems. In a previous attempt at quantifying immobilization sites (9), there was significant peptide-to-peptide variability for the reference $^{16}O/^{18}O$ ratios (i.e., $^{16}O/^{18}O$ ratios that were obtained when a control HSA sample was compared to itself), which indicated that the absolute $^{16}O/^{18}O$ ratios should not be used for quantifying the small amount of glycation that would occur on peptides. The wide variation in reference

$^{16}O/^{18}O$ ratios was confirmed in our previous attempt at measuring $^{16}O/^{18}O$ ratios for minimally glycated HSA, where there was some ambiguity in quantifying the glycation sites because few peptides had $^{16}O/^{18}O$ ratios that were above the assigned cutoff value. This effect had also been observed by other authors, where the peptide-to-peptide labeling efficiency was found to be influenced by the size of the peptide, resulting in inefficient enzyme-peptide interaction (8). This difference in enzyme-peptide substrate formation has been shown to significantly impact the amount of $^{18}O$ label that gets incorporated during digestion (7). Quantitation by $^{18}O$ labeling is also sensitive to variations in volume size, ion intensity, unresolved peptides, and differential sample loss (7, 14).

To counter the effects of variations in peptide-to-peptide labeling, the calculated $^{16}O/^{18}O$ ratios for the glycated HSA samples were compared directly to the $^{16}O/^{18}O$ ratios obtained for a control HSA sample (containing no glycation). An *f*-test (15) was used to determine whether the variance for both $^{16}O/^{18}O$ ratios were the same or not followed by the appropriate Student's *t*-test (15) to see whether the $^{16}O/^{18}O$ ratios were different at the 90% confidence level. The $^{16}O/^{18}O$ ratios in the glycated HSA sample that were higher at the 90% confidence level were assumed to be glycated. These values were subsequently ranked by the relative difference in their $^{16}O/^{18}O$ ratios for the glycated HSA and control HSA digests. By using this approach, it was possible to quantify the relative extent of modification occurring on different regions of HSA and assign glycation products to HSA with a high level of certainty.

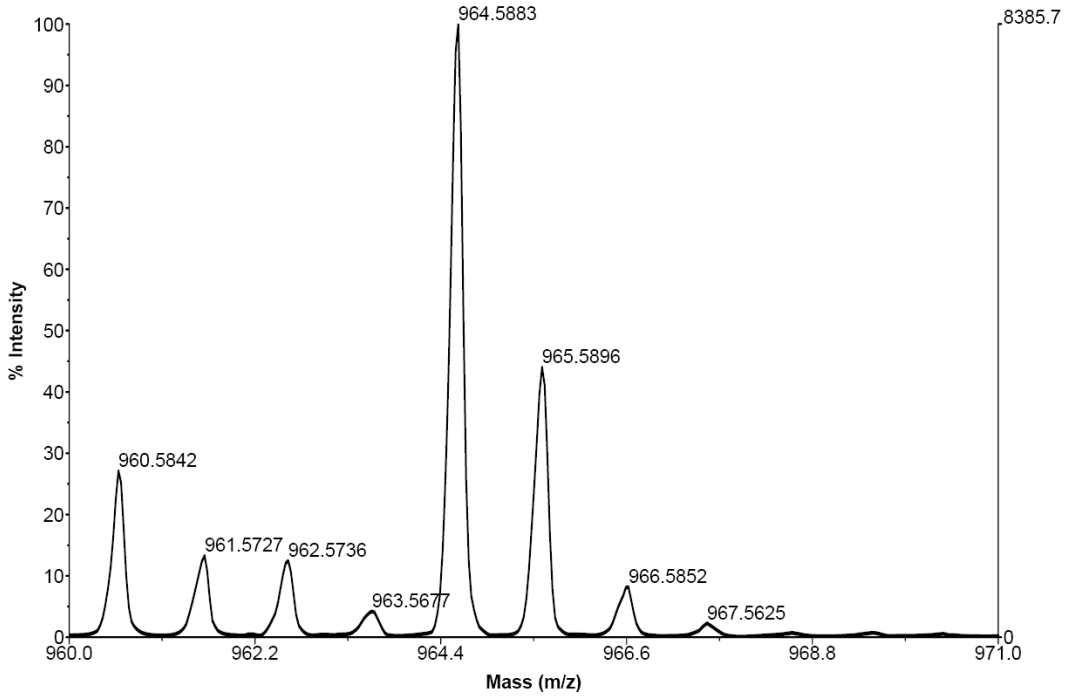
Variations in volume differences and sample loss were avoided by performing both the $^{16}O$ and $^{18}O$ digests side-by-side and by preparing samples with the same apparatus and reagents. The use of similar reagents was particularly important for the enzyme addition step where the use of separate vials of enzymes (even if they are from the same lot) has been shown to cause drastic changes in the $^{16}O/^{18}O$ ratios. This difference is clearly illustrated in Figures 6.1 (i) and 6.1 (ii).

Low retention micro-centrifuge tubes were used to reduce the degree of adsorption-based sample loss that can occur when the protein is being dried under vacuum. The MALDI laser intensity was manually adjusted to ensure that the majority of detected ions were of sufficient intensity for quantitation (i.e., a selected signal-to-noise ratio of 25) and that detector saturation was not occurring (i.e., the detector response was not allowed to go above 20,000 counts). By using the described adjustments, several peptides were identified as having significant levels of modification which could be linked to large increases in the $^{16}O/^{18}O$ ratios with respect to the reference $^{16}O/^{18}O$ ratios that were obtained for the control sample. All of the key glycation sites could be quantified and ranked in this analysis.

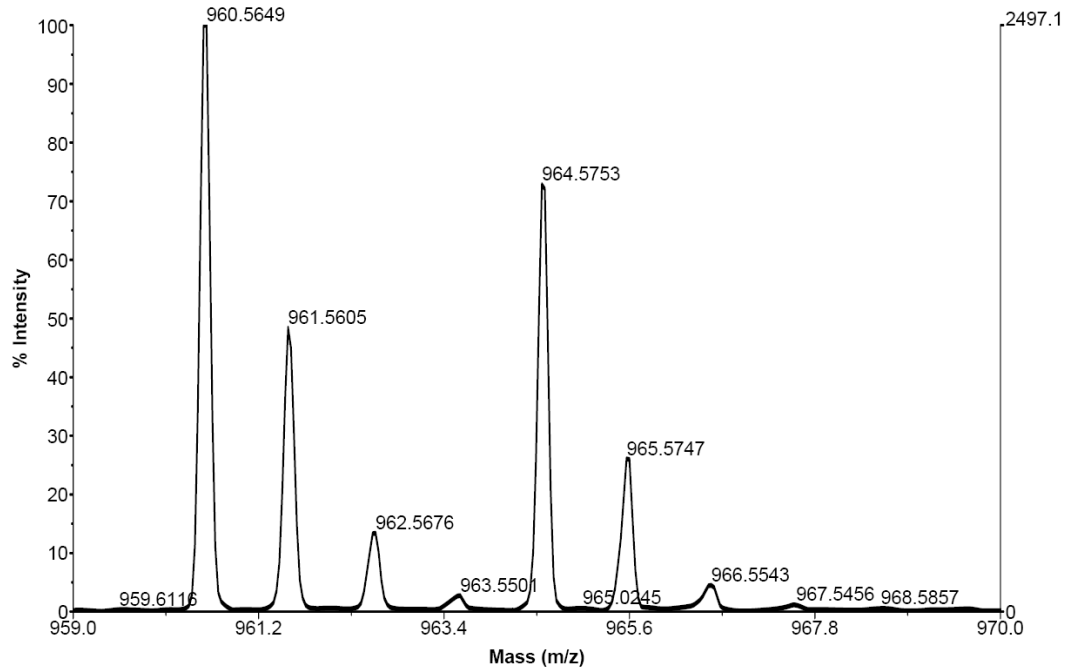
Figure 6.1. (i) Mass spectrum showing the isotopic distribution of a peptide when two different vials of trypsin are used to produce $^{16}O$- and $^{18}O$- labeled digests.

(ii) Mass spectrum obtained when an identical method to (i) is performed where the same vial of trypsin is being used for both the $^{16}O$ and $^{18}O$ labeled digest.

(i)


964.5883
8385.7
100
90
80
70
60
50
40
30
20
10
0
% Intensity
965.5896
960.5842
961.5727
962.5736
963.5677
966.5852
967.5625
960.0
962.2
964.4
966.6
968.8
971.0
Mass (m/z)

(ii)


960.5649
2497.1
100
90
80
70
60
50
40
30
20
10
0
% Intensity
964.5753
961.5605
965.5747
962.5676
966.5543
963.5501
967.5456
968.5857
959.6116
965.0245
959.0
961.2
963.4
965.6
967.8
970.0
Mass (m/z)

## 6.B. EXPERIMENTAL

**6.B.1. Materials.** The following chemicals were purchased for Sigma-Aldrich (St. Louis, MO): Des-Arg-bradykinin (97% pure), glu-fibrinopeptide (97%), angiotensin I (97%, acetate salt), 2,3-dihyroxybenzoic acid (98%), α-cyano-4-hydroxycinnamic acid (99%), HSA (99%, essentially fatty acid and globulin free), trypsin (sequence grade), Glu-C (sequence grade), Lys-C (sequence grade), guanidine-HCl (99%), dithiothreitol (99%), D-glucose (99%), iodoacetamide (99%), triflouroacetic acid (98%), sodium azide (99%), formic acid (96%), molecular biology grade water (DNase, RNase, and protease free), methanol (HPLC grade, 99.9%), $^{18}$O-enriched water (97%), $^{16}$O-enriched water (99.99%), molecular biology grade water, and acetonitrile (HPLC grade, 99.9%).

**6.B.2. Apparatus.** Siliconized low retention microcentrifuge tubes (0.6 mL and 1.5 mL) and 0.20 μm nylon filters were obtained from Fisher Scientific (Pittsburgh, PA). Slide-A-Lyzer dialysis cassettes (7000 Da MW cutoff, 0.1-0.5 mL and 0.5-3 mL) were obtained from Pierce (Rockford, IL), and Zip-tip pipette tips (μ-C18 bed, 0.2 μL bed) were obtained from Millipore (Billerica, MA). Mass spectra were collected on a Voyager 6148 MALDI-TOF-MS system (Applied/Perspective Biosystems, CA). The instrument settings were as follows: positive-ion delayed extraction reflection mode; delay time, 100 ns; accelerating voltage, 25 kV; guide wire voltage, 0.008% of accelerating voltage; grid voltage, 77% of accelerating voltage. Matlab 2009a, which included the bioinformatics toolbox, was purchased from Mathworks (Natick, MA). Mascot Wizard was obtained from the Matrix Science website (16).

**6.B.3 Digestion and labeling of HSA samples.** The same HSA and glycated HSA samples that were used previously for the identification of glycation-related modifications in Chapters 4 and 5 were also used for this analysis. The pretreatment of HSA (i.e., denaturation, reduction, and alkylation) was performed as described in Chapter 4. Following pretreatment, the HSA samples were placed into 0.1 – 0.5 mL dialysis cassettes, and dialysis was performed on the HSA samples versus 3, 1000 mL portions of water for 4 hours each. The resulting control and glycated HSA solutions were adjusted to the same volume. A 40 μL aliquot of each HSA solution was placed into several low retention microcentrifuge tubes and these samples were brought down to dryness using a Speedvac (unheated). These vials, which contained approximately 80 μg of either control HSA or glycated HSA, were stored at -80 °C until later use.

The tryptic digest was performed by first reconstituting a vial of control HSA in 200 μL of 0.1 M, pH 7.8 ammonium bicarbonate buffer that was prepared in $^{16}O$-enriched water. The glycated HSA samples were reconstituted in a similar manner, where $^{18}O$-enriched water was used in place of $^{16}O$-enriched water. A 3.33 μL aliquot of 1 μg/μL trypsin solution (prepared in Nanopure water) was then added to the reconstituted HSA samples. The HSA solutions were incubated at 37 °C for 18 h. Following this incubation, the digestions were terminated by adding 10 μL of formic acid to the HSA solutions. These digests were stored at -80 °C until later use.

The Glu-C digest was performed by reconstituting a vial of control HSA in 200 μL of 0.1 M, pH 7.8 ammonium bicarbonate buffer that was prepared in $^{16}O$-enriched water. The glycated HSA samples were reconstituted in a similar manner, where $^{18}O$-

enriched water was used in place of $^{16}$O-enriched water. A 10 μL aliquot of 1 μg/μL Glu-C solution (prepared in Nanopure water) was then added to the control HSA and glycated HSA samples and the samples were incubated for 8 h at 37 °C. An additional 5 μL of Glu-C was then added to the mixtures and the incubation was continued for an additional 18 h. Following this incubation, the digestions were terminated by adding 10 μL of formic acid and the samples were stored at -80 °C.

The Lys-C digest was performed by first reconstituting a vial of control HSA in 200 μL of pH 9.0 tris buffer (75 mM) that was prepared in $^{16}$O-enriched water. The glycated HSA samples were reconstituted in a similar manner, where $^{18}$O-enriched water was used in place of $^{16}$O-enriched water. A 5 μL aliquot of a 1 μg/μL Lys-C solution (prepared in Nanopure water) was then added to the reconstituted HSA samples. The mixtures were incubated for 18 h at 37 °C. Following this incubation, 10 μL of formic acid was added to terminate the digestion and the samples were stored at -80 °C (Note: the endoproteinase Lys-C that was obtained from Sigma-Aldrich was only available in 5 μg quantities, therefore, several of these vials were reconstituted individually and pooled prior to this analysis to ensure that the same solution was being used for each sample).

**6.B.4. MALDI-TOF-MS data collection.** The MALDI matrix consisted of a mixture of 2,3-dihyroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA). In previous studies, this mixture has been shown to enhance the sequence coverage and that is obtained for HSA (9). To make this matrix, 20 mg/mL solutions of CHCA and DHB were prepared in a 70:30 (v/v) mixture of acetonitrile and water. A 1 mL aliquot of the CHCA solution was spiked with 50 μL of formic acid and a 1 mL aliquot of the DHB

solution was spiked with 1 μL of trifluoroacetic acid. Both matrices were then mixed in a 50:50 (v/v) fashion, where the resulting mixture was used as the matrix solution for the remainder of this analysis. To calibrate the mass spectrometer, 96 μL of the mixed matrix was added to 4 μL of a standard calibrant solution that contained 32.5 pM glu-fibrinopeptide (1570.6774 Da), 25 pM des-arg-bradykinin (904.4681 Da), and 32.5 pM angiotensin I (1296.6853 Da) that was prepared in acetonitrile. The resulting solution was placed in positions on the MALDI plate that were adjacent to the HSA samples, where each calibrant spot was used to calibrate the mass spectrometer for each adjacent HSA sample. The HSA digests were fractionated by loading 10 μL of the HSA digest on Zip-tip pipette tips. These peptides were eluted with 1 μL of 5, 10, 20, 30, or 50% of acetonitrile in water, as previously described (9, 17, 18). The eluted fractions were then mixed with 1 μL of the mixed MALDI matrix and the resulting mixture was slowly vortexed for 15 seconds. A 0.75 μL portion of the resulting mixture was then immediately spotted on the MALDI plate.

**6.B.5. Peak selection using Matlab and data analysis techniques.** The mass spectra that were acquired represented the average of six individual spectra, where each spectrum represented the summation of 200 laser shots. Peptide mapping for the control HSA digest was performed by using Mascot Wizard with the default settings and a 50 ppm threshold. This analysis was limited to 2 missed cleavages, and carbamidomethyl-cysteine (fixed) and methionine oxidation (variable) were selected as possible modifications. The raw mass values that were obtained using Mascot Wizard were transferred to a tab delimited text file. After processing the data with Mascot Wizard, peptide masses that were identified in different Zip-tip fractions (i.e., for m/z values that

belonged to the same peptide) were averaged, and these averaged values were used to create a master list of mass values that were linked to peptides. Two programs were developed in Matlab that obtained areas for the M+0 to M+4 peaks from the raw data (see Figure 6.2 for a flow chart that represents both programs). The program that was developed by Dr. Ala Qadi (Department of Computer Science, UNL) was used as a template (see Chapter 2) for the development of program in this project.

The first program (see Figure 6.2) read two input files, where the first file contained the peptides that were obtained using Mascot wizard (list 1) and the second file contained the raw data (i.e., mass values and their corresponding area values) that was obtained from the mass spectra. In the first step, the mass values that were present in both lists were compared, to determine which mass values matched at a maximum allowed error of 50 ppm.

$$\text{error(ppm)} = \frac{\left(\text{mass(list1)} - \text{mass(list2)}\right)}{\text{mass(list2)}} * 10^6 \qquad (1)$$

For the peptides that matched, the mass values corresponding to the M+0 through M+4 peaks were assigned by adding 0, 1, 2, 3, and 4 to the masses that were present in list 1. The identified M+0 → M+4 values were exported to a separate text file (result.txt). This text file was then compared to the raw data to obtain area values for each mass within a given isotopic cluster (i.e., for the M+0 → M+4 peaks) (Note: A 100 ppm maximum allowed error was used for this comparison). Two programs are shown in the Appendix for obtaining the M+0 → M+4 values for the $^{16}$O-labeled digest (Tables 6.1a (i)) and the $^{18}$O-labeled digest (Table 6.1a (ii)).

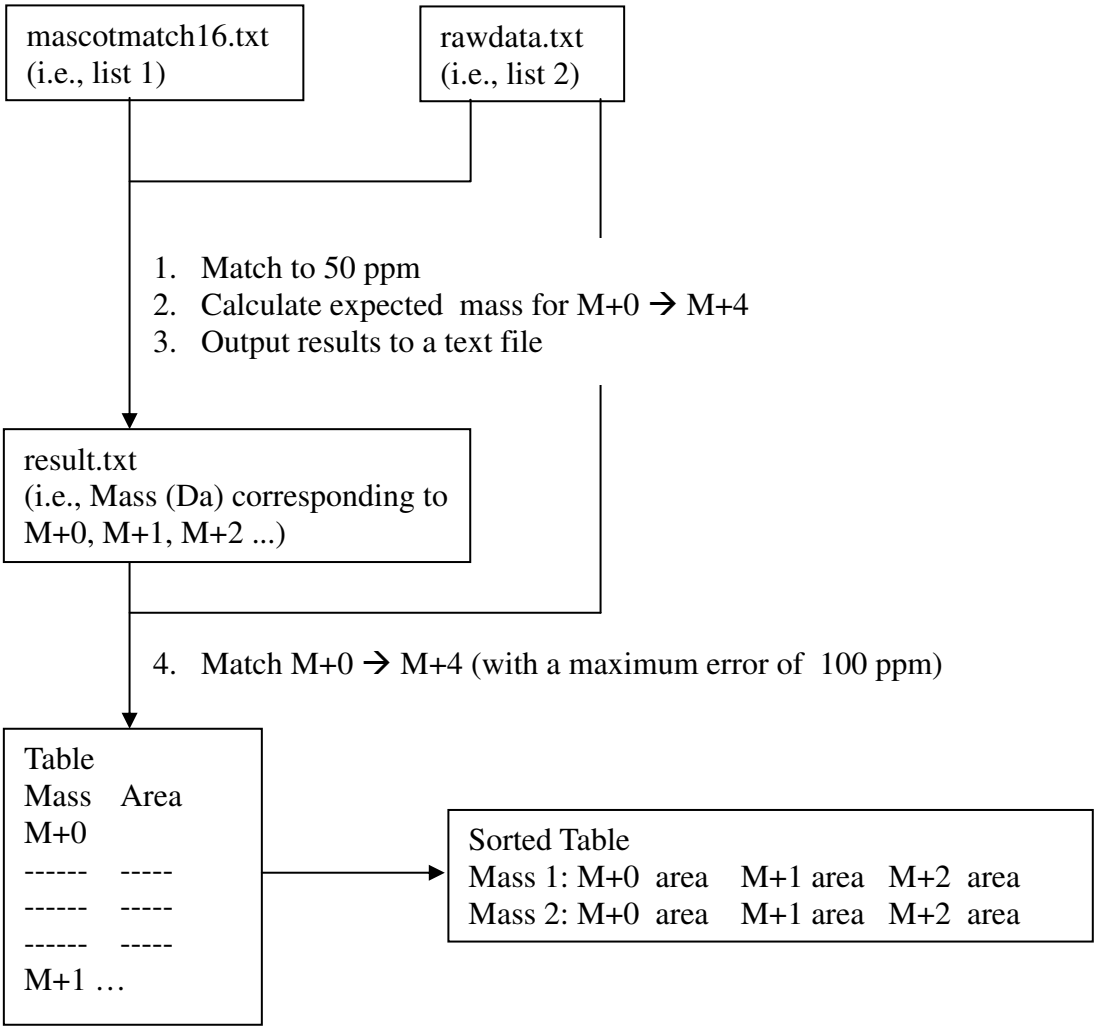
Ufuk Nalbantoğlu, a graduate student in Electrical Engineering at UNL, generously provided an addition to this program that was used to combine these five tables into a single table that was sorted by mass. The code that was used to develop this program is beyond the scope of this project; hence it is not shown here. This method was used to identify areas for the M+0 to M+4 peaks in the $^{16}O$- and mixed-O labeled digests, respectively. A similar approach was used for the $^{18}O$-labeled digests, where 4 Da was added to the mass values in list 1, and these adjusted mass values were used to search for peptides that were labeled with two $^{18}O$- atoms. Once these values were assigned, the M+0 to M+4 values were identified by subtracting 4, 3, 2, 1, and 0 from the matched mass values. All other steps for determining the peak area values were identical to the method that was used for the $^{16}O$-labeled digest.

The $^{16}O/^{18}O$ ratios were then determined as shown below (in Equation1). The

$$\frac{^{16}O}{^{18}O} = \frac{I'_0}{\left(I''_0 + I''_{2(c)} + I''_{4(c)}\right)} \quad (1)$$

terms I' and I" in the above equation represents the contribution from the $^{16}O$ and $^{18}O$ digests respectively. The subscripts 0, 2, and 4 show whether the contribution comes from peptides that contain zero (M+0 peak), one (M+2 peak), or two (M+4 peak) $^{18}O$ labels respectively. This equation is described in more detail elsewhere (9).

Figure 6.2. Flow chart showing the steps that were taken to extract the isotope peak areas for the mixed and $^{16}O$-labeled digests. The first four steps were performed using a single program (a_16_iso_select.m). A separate program (sort_mass.m) was used to sort the resulting table.


mascotmatch16.txt
(i.e., list 1)
rawdata.txt
(i.e., list 2)
1. Match to 50 ppm
2. Calculate expected mass for M+0 → M+4
3. Output results to a text file
result.txt
(i.e., Mass (Da) corresponding to
M+0, M+1, M+2 ...)
4. Match M+0 → M+4 (with a maximum error of 100 ppm)
Table
Mass Area
M+0
------ -----
------ -----
------ -----
M+1 …
Sorted Table
Mass 1: M+0 area M+1 area M+2 area
Mass 2: M+0 area M+1 area M+2 area

Each $^{16}O/^{18}O$ ratio that was obtained for glycated HSA was compared to its corresponding reference $^{16}O/^{18}O$ ratio to determine whether the glycated sample had significant amounts of modification (at the 90% confidence level). This comparison was done using a student's *t*-test (15). Prior to performing the student's *t*-test, an *F*-test was used to determine whether the variance of the $^{16}O/^{18}O$ ratio for a given peptide is the same at the 95% confidence level (15). If the $^{16}O/^{18}O$ ratio that was obtained for the glycated sample was significantly higher (at the 90% confidence level) than the $^{16}O/^{18}O$ ratio obtained for the control sample, the peptide corresponding to that ratio was considered to be a significant glycation site. A 90% confidence level was found to be more suitable than a 95% confidence level for assigning glycation sites with this approach because of the large variances that were associated with some of the $^{16}O/^{18}O$ ratios, especially in cases where the S/N ratio for the corresponding peptides was low.

## 6.C. RESULTS

**6.C.1. Detected peptides and sequence coverage.** The usable sequence coverage (i.e., the coverage representing the number of amino acid residues for which $^{16}O/^{18}O$ ratios could be determined) obtained for the control HSA sample digested with trypsin, Lys-C, or Glu-C was 76%, 62%, or 61%, respectively. When all of the digests were considered, a total of 106 peptides were detected which covered 91% of the amino acids in HSA (See Figures 6.3). There was however, a slight decrease in the usable sequence coverage once the calculated $^{16}O/^{18}O$ ratios was factored in, because some of the peptides that were identified in the control digest were not found in the glycated HSA digests. The resulting usable sequence coverage for the HSA-2, HSA-5, and control HSA samples (see Figure 6.1a) were reduced to 82%, 86%, and 86% respectively.

In the control and HSA-5 samples, the useable sequence coverage represented all regions of HSA except residues 52-60, 175-181 (K181), 228-240 (K233 and K240), 314-323 (K317 and K323), 349-359 (K351 and K359), 443-444 (K444), 532-545, 558-560 (K560), and 575-585. In addition to the above residues, residues 287-313 (K313) were not part of the useable sequence coverage for HSA-2 sample. The useable sequence coverage included 100% of the arginine residues and 84 – 85% of the lysine residues that were present on HSA. The lack of coverage in these regions was of little consequence to this study because the residues that are expected to have the highest amounts of modification (e.g., Lysines 525, 439, 281, 199 and the *N*-terminus) are not found in these regions of HSA (6, 19).

**6.C.2. Calculated $^{16}O/^{18}O$ ratios for the HSA samples.** The calculated $^{16}O/^{18}O$ ratios

Figure 6.3. The usable sequence coverage that was obtained for the control HSA digest is shown in bold and caps. The lower case letters represents the remainder of the sequence in HSA, where no $^{16}O/^{18}O$ ratio calculations were made.

1- **DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV**

41- **KLVNEVTEFA K**tcvadesae **NCDKSLHTLF GDKLCTVATL**

81- **RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV**

121- **DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR**

161- **YKAAFTECCQ AADK**aacllp k**LDELRDEGK ASSAKQRLKC**

201- **ASLQKFGERA FKAWAVARLS QRFPKAE**fae vsklvtdltk

241- **VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE**

281- **KPLLEKSHCI AEVENDEMPA DLPSLAADFV ESKDVCKNYA**

321- **EAKDVFLGMF LYEYARRHPD YSVVLLLRLA KTYETTLEKC**

361- **CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE**

401- **YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH**

441- **PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES**

481- **LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKE**

521- **RQIKKQTALV E**lvkhkpkat keqlk**AVMDD FAAFVEKCCK**

561- **ADDKETCFAE EGKK**lvaasq aalgl

for the tryptic HSA digest are given in Tables 6.1 (i) and (ii). From the tryptic digest table, it can be seen that the $^{16}O/^{18}O$ ratios were calculated for 31 peptides in both the control HSA and HSA-5 samples. In the HSA-2 sample, the $^{16}O/^{18}O$ ratio for the peptide corresponding to a mass of 2974.31 Da (representing residues 287-313) was not calculated because a sufficient signal was not detected in mixed digest that could be used for quantitation. The calculated $^{16}O/^{18}O$ ratios for the trypsin digest ranged from 0.59-1.79 with an average of 1.43 (Figure 6.4).

Similarly, the $^{16}O/^{18}O$ ratios for the Glu-C digest of HSA (see Appendix, Table 6.2a) were calculated at different stages of glycation. In this analysis, the $^{16}O/^{18}O$ ratios for 18 peptides were calculated for both the control HSA and HSA-5 digest, compared to 17 peptides that were found in the HSA-2 sample. The peptide that had a calculated $^{16}O/^{18}O$ ratio in the control HSA sample that did not have a calculated $^{16}O/^{18}O$ ratio in the HSA-2 sample had a mass of 1801.92 Da, which could be mapped to residues 451-465 on HSA. A ratio for this peptide could not be obtained due to the intensity of the same peptide in the mixed digest for the HSA-2 sample. The distribution of the calculated $^{16}O/^{18}O$ ratios are shown in Figure 6.2a (i). From this data set, it was determined that the $^{16}O/^{18}O$ ratio for the control HSA samples ranged anywhere from 0.12-2.57, with an average of 1.19.

Using similar approaches, the $^{16}O/^{18}O$ ratios for the peptides found in the Lys-C digest were arranged and compared (see Appendix, Table 6.3a). Using this format, the ratio for 20 peptides were calculated in the control HSA digest, compared to 19 peptides which were identified in the HSA-2 and HSA-5 samples, respectively. The reduction in

Table 6.1 (i). Calculated $^{16}O/^{18}O$ ratios for the tryptic digests.

| | | | | Control | | HSA-2 | | HSA-5 | |
|---|---|---|---|---|---|---|---|---|---|
| **Mass (Da)** | **Error (ppm)[a]** | **Residue** | **Amino Acid Sequence** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** |
| 927.50 | 2 | 138-144 | YLYEIAR | 1.66 | 0.24 | 1.94 | 0.41 | 2.02 | 0.41 |
| 960.57 | 6 | 403-410 | FQNALLVR | 1.51 | 0.31 | 1.30 | 0.18 | 1.53 | 0.25 |
| 1055.59 | -1 | 137-144 | KYLYEIAR | 1.34 | 0.39 | 1.59 | 0.55 | 1.55 | 0.20 |
| 1074.53 | -16 | 182-190 | LDELRDEGK | 1.67 | 0.93 | 2.55 | 0.44 | 2.02 | 1.58 |
| 1149.60 | -10 | 1-10 or 42-51 | DAHKSEVAHR | 1.99 | 0.81 | 1.57 | 0.81 | 3.20 | 1.69 |
| 1226.60 | -5 | 11-20 | FKDLGEENFK | 1.56 | 0.71 | 1.19 | 0.45 | 1.34 | 0.32 |
| 1342.62 | -11 | 546-557 | AVMDDFAAFVEK | 0.72 | 0.13 | 0.46 | 0.12 | 1.03 | 0.20 |
| 1371.56 | -5 | 163-174 | AAFTECCQAADK | 1.21 | 0.30 | 1.48 | 0.22 | 0.95 | 0.33 |
| 1434.53 | -6 | 82-93 | ETYGEMADCCAK | 1.10 | 0.17 | 1.36 | 0.77 | 1.89 | 0.93 |
| 1443.64 | -4 | 263-274 | YICENQDSISSK | 1.81 | 0.46 | 2.07 | 1.32 | 2.42 | 0.90 |
| 1467.82 | -17 | 337-348 | RHPDYSVVLLLR | 2.01 | 0.43 | 2.01 | 1.01 | 2.43 | 0.70 |
| 1546.80 | 2 | 275-286 | LKECCEKPLLEK | 1.37 | 0.51 | 2.02 | 0.58 | 1.49 | 0.80 |
| 1552.60 | -2 | 360-372 | CCAAADPHECYAK | 1.46 | 0.47 | 2.33 | 0.67 | 1.86 | 0.20 |
| 1623.80 | 5 | 324-336 | DVFLGMFLYEYAR | 1.85 | 0.06 | 2.43 | 0.36 | 3.05 | 0.38 |
| 1627.73 | 0 | 561-574 | ADDKETCFAEEGKK | 1.52 | 0.28 | 1.10 | 0.34 | 1.32 | 0.39 |
| 1639.94 | 0 | 414-428 | KVPQVSTPTLVEVSR | 1.31 | 0.20 | 1.69 | 0.46 | 1.38 | 0.19 |

[a] – "Error" represents the mass accuracy of the assigned peptide. For simplicity, the control HSA digest was used for peptide mass fingerprinting.

[b] – "SD" designates the standard deviation that was calculated for the given $^{16}O/^{18}O$ ratios. The value represents the square root of the sum of the standard deviations squared that were obtained across different Zip-tip fractions.

Table 6.1 (ii). Calculated $^{16}O/^{18}O$ ratios for the trypsin digest.

| | | | | Control | | HSA-2 | | HSA-5 | |
|---|---|---|---|---|---|---|---|---|---|
| **Mass (Da)** | **Error (ppm)[a]** | **Residue** | **Amino Acid Sequence** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** |
| 1657.72 | -21 | 390-402 | QNCELFEQLGEYK | 1.21 | 0.58 | 1.11 | 0.29 | 1.33 | 0.42 |
| 1714.80 | 0 | 94-106 | QEPERNECFLQHK | 1.46 | 0.21 | 1.59 | 0.16 | 1.86 | 0.33 |
| 1742.88 | -9 | 146-159 | HPYFYAPELLFFAK | 1.55 | 0.15 | 1.85 | 0.35 | 2.59 | 0.74 |
| 1899.01 | 7 | 146-160 | HPYFYAPELLFFAKR | 1.48 | 0.33 | 2.38 | 0.97 | 2.47 | 0.60 |
| 1910.92 | -5 | 485-500 | RPCFSALEVDETYVPK | 1.46 | 0.24 | 1.31 | 0.17 | 1.45 | 0.15 |
| 2045.10 | 0 | 373-389 | VFDEFKPLVEEPQNLIK | 1.43 | 0.08 | 1.68 | 0.08 | 1.95 | 0.41 |
| 2086.80 | -19 | 241-257 | VHTECCHGDLLECADDR | 1.27 | 0.20 | 1.66 | 0.29 | 1.65 | 0.25 |
| 2260.08 | 27 | 501-519 | EFNAETFTFHADICTLSEK | 1.73 | 0.47 | 1.37 | 0.47 | 0.75 | 0.24 |
| 2490.28 | -2 | 21-41 | ALVLIAFAQYLQQCPFEDHVK | 1.71 | 0.18 | 1.92 | 0.34 | 3.50 | 0.73 |
| 2518.20 | -5 | 352-372 | MPCAEDYLSVVLNQLCVLHEK | 1.82 | 0.30 | 1.23 | 0.53 | 1.43 | 0.68 |
| 2585.11 | -4 | 241-262 | VHTECCHGDLLECADDRADLAK | 0.92 | 0.06 | 0.81 | 0.04 | 0.84 | 0.15 |
| 2650.30 | 15 | 115-136 | LVRPEVDVMCTAFHDNEETFLK | 1.21 | 0.03 | 1.24 | 0.12 | 1.11 | 0.02 |
| 2674.31 | -1 | 445-466 | RMPCAEDYLSVVLNQLCVLHEK | 0.99 | 0.06 | 0.78 | 0.09 | 0.93 | 0.36 |
| 2778.43 | 25 | 115-137 | LVRPEVDVMCTAFHDNEETFLKK | 1.01 | 0.02 | 0.92 | 0.04 | 1.09 | 0.08 |
| 2974.31 | -10 | 287-313 | SHCIAEVENDEMPADLPSLAADFVESK | 1.05 | 0.05 | none detected | | 1.21 | 0.70 |

[a] – "Error" represents the mass accuracy of the assigned peptide. For simplicity, the control HSA digest was used for peptide mass fingerprinting.

[b] – "SD" designates the standard deviation that was calculated for the given $^{16}O/^{18}O$ ratios. The value represents the square root of the sum of the standard deviations squared that were obtained across different Zip-tip fractions.

Figure 6.4. Reference $^{16}O/^{18}O$ ratios that were obtained for the tryptic digest. The error bars represents the calculated standard deviation of a given $^{16}O/^{18}O$ ratio. The central dashed line represents the average $^{16}O/^{18}O$ ratio and outer dashed lines represents the range of $^{16}O/^{18}O$ ratios that were obtained.

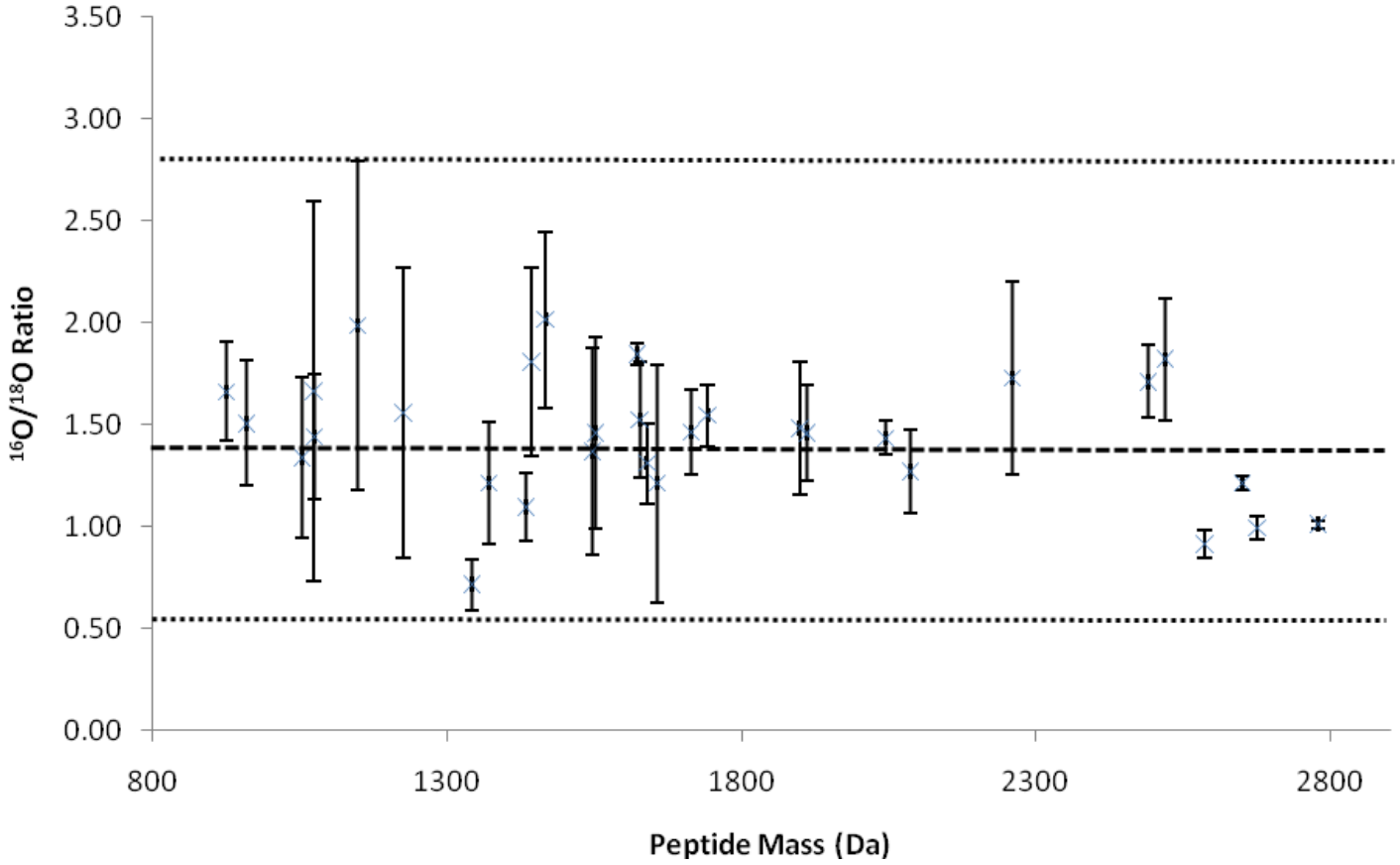

3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
$^{16}O/^{18}O$ Ratio
800
1300
1800
2300
2800
Peptide Mass (Da)

Figure 6.4

the number of detected peptides was due to a peptide with a m/z value of 2259.99 Da (residues 21-41) that was detected in the control digest but not in the HSA-2 or HSA-5 digests. The reference $^{16}O/^{18}O$ ratios in the Lys-C digest ranged anywhere from -0.04 to 2.11, with an average $^{16}O/^{18}O$ ratio of 1.13 (see Appendix, Figure 6.2a (ii)). The average $^{16}O/^{18}O$ ratios for the trypsin, Lys-C, and Glu-C control digests showed little deviation and varied from only 1.13-1.43; however, the individual $^{16}O/^{18}O$ ratios for the control digests did show significant variability.

**6.C.3. Peptides with high $^{16}O/^{18}O$ ratio peptides in the HSA-2 sample.** When the reference $^{16}O/^{18}O$ ratios were compared to the $^{16}O/^{18}O$ ratios that were obtained for glycated HSA, numerous peptides were identified in both glycated samples with high $^{16}O/^{18}O$ ratios. The $^{16}O/^{18}O$ ratios that were obtained for the glycated HSA samples are summarized in Table 6.2. In total, 17 peptides were identified in the HSA-2 sample that had significantly higher $^{16}O/^{18}O$ ratios (at the 90% confidence level) compared to their respective reference $^{16}O/^{18}O$ ratios. These peptides had a relative increase in $^{16}O/^{18}O$ ratios that ranged anywhere from 6% to 77%.

The highest change in $^{16}O/^{18}O$ ratios were found on residues 101-119, 1-10 or 42-51, 360-372, 521-537, and 275-286 on HSA, which had corresponding relative increases in$^{16}O/^{18}O$ ratios of 77%, 63%, 60%, 57%, and 48%, respectively versus normal HSA. The *N*-terminus, K525, and K281, which are believed to be the most likely glycation sites (5, 6, 19), were all located on residues that had the most significant levels of modification in this analysis. In addition to these sites, a peptide corresponding to residues 426-442 on HSA was also identified as having a relative change in $^{16}O/^{18}O$ ratio of 19%, which represented a significant increase at the 90% confidence level. This increase in $^{16}O/^{18}O$

Table 6.2. Peptides in the HSA-2 sample with significantly higher $^{16}O/^{18}O$ ratios at the 90% confidence level.

| | | | | Most Likely Location of Modified Residues[c] | |
|---|---|---|---|---|---|
| **Digest** | **Mass (Da)** | **Residue**[a] | **Relative Diff.**[b] | **Lysine** | **Arginine** |
| Glu-C | 2348.094 | 101-119 | 77% | 106 | 114, 117 |
| Lys-C | 1149.619 | **42-51/1-10** | 63% | N+ ,4 (FL) | 10 (AFGP, 3-DG-H1, or THP) |
| Tryptic | 1552.595 | 360-372 | 60% | 372 | n/a |
| Glu-C | 1313.796 | **521-537** | 57% | 524, 525, 536, 538 (Pyr, FL, or CEL) | 521 (G-H1) |
| Tryptic | 1546.800 | **275-286** | 48% | 276, 281, 286 (FL, CEL, Pyr) | n/a |
| Tryptic | 1623.795 | 324-336 | 31% | n/a | 336 |
| Tryptic | 2086.798 | 241-257 | 31% | n/a | 257 |
| Lys-C | 1552.592 | 360-372 | 30% | 372 | n/a |
| Tryptic | 1639.938 | 414-428 | 29% | 414 (FL, AFGP) | 428 (AFGP, 3-DG-H1, THP) |
| Glu-C | 1519.783 | 142-153 | 29% | n/a | 144 |
| Glu-C | 1945.862 | 502-518 | 28% | n/a | n/a |
| Glu-C | 2578.264 | 61-82 | 24% | 64, 73 | 81 |
| Glu-C | 1704.910 | 154-167 | 22% | 159 | 160 |
| Glu-C | 1955.975 | **426-442** | 19% | 432, 436, 439 (Pyr, CEL, FL, or AFGP) | 428 (G-H1, ArgP, AFGP, 3DG-H1, or THP) |
| Trypsin | 2045.096 | 373-389 | 17% | 378, 389 | n/a |
| Glu-C | 2232.208 | 209-227 | 16% | 212, 225 (CEL) | 209, 218, 222 (ArgP) |
| Lys-C | 1714.785 | 94-106 | 6% | 106 | 98 |

[a] – Residues that have been previously identified by other authors are shown in **bold**.

[b] – This value represents the percent relative difference of the $^{16}O/^{18}O$ ratio for the glycated HSA sample and the control HSA sample.

[c]- The most likely modification sites are underlined. The suspected modification; based on the data in previous chapters, is also shown.

ratio is on the lower end of increased $^{16}O/^{18}O$ ratio values that correspond to significant modification; however, this site was also previously identified as being one of the likely modification sites of *in vivo* glycated HSA (6). A peptide corresponding to residues 189-208 was also identified in the Glu-C digest, with a 24% increase in $^{16}O/^{18}O$ ratio compared to its reference $^{16}O/^{18}O$ ratio (i.e., 0.93 +/- 0.11). This peptide contains K199 which has previously been identified as one of the likely glycation sites on HSA (6). While an increase in the average $^{16}O/^{18}O$ ratio for this peptide was noted, the increase in the $^{16}O/^{18}O$ ratio for this peptide was not significant at the 90% confidence level because of the large variance in the $^{16}O/^{18}O$ ratio that was obtained for the HSA-2 sample (i.e., 1.16 +/- 0.99).

The regions of HSA that contained significant amounts of modification in this analysis were compared to the relative extent of modification that was found on commercially glycated HSA (see Chapter 2). When this comparison was done, 7 of the 11 peptides that were previously identified as having high amounts of modification in commercially glycated HSA were also found to have high amounts of modification in this current analysis. For instance, residues 107-114 was previously found to have the highest measurable $^{16}O/^{18}O$ ratio and thus contained the largest amount of modification in the commercially glycated HSA sample. This finding is consistent with the findings in this study, where a similar peptide (i.e., residues 101-119) was found to have the largest relative change in $^{16}O/^{18}O$ ratio. Other overlapping and likely modification sites included regions 521-537, 324-336, 414-428, 61-82, 426-442, and 209-227 on HSA. Some of the sites that were previously thought to have excessive amounts of modification (i.e., K233, K439, and K525) based on a qualitative analysis (see Chapter 2) were also found to have

large levels of modification in this current analysis. For instance, residues 521-537 (which contains K525) had the fourth largest increase in $^{16}O/^{18}O$ ratio and residues 426-442 (which contains K439) was found to have a moderate increase in $^{16}O/^{18}O$ ratio. A moderate change in $^{16}O/^{18}O$ ratio was also observed on residues 426-442 for the HSA-5 sample. Lysine 233, which was previously thought to have significant levels of modification based on a qualitative comparison (see Chapter 2), was not identified in either stage of glycation in this study as having significant levels of modification. These combined factors indicate that K525 and K439 may indeed contain high levels of modification. The difference in the ranked modification sites that were obtained in the previous study using commercially glycated HSA (see Chapter 2) and this current approach can be attributed to experimental design. By using low retention apparatus, the concentration of peptides in the HSA digests were increased, allowing for an improved estimation of $^{16}O/^{18}O$ ratios. The strength of the techniques used in this current chapter, is that changes in the $^{16}O/^{18}O$ ratio that are related to increased levels of modification can be distinguished from changes in $^{16}O/^{18}O$ ratios that are due to peptide labeling efficiency.

**6.C.4. Peptides with high $^{16}O/^{18}O$ ratios in the HSA-5 sample.** The peptides which were found to be significantly glycated in the HSA-5 sample are ranked and sorted in Table 6.3. In total, 21 peptides were found to have significant increases in $^{16}O/^{18}O$ ratio (compared to their respective reference $^{16}O/^{18}O$ ratio) at the 90% confidence level. These peptides had a relative increase in the $^{16}O/^{18}O$ ratios that ranged anywhere from 7% to 104%.

Table 6.3. Peptides in the HSA-5 sample with high $^{16}O/^{18}O$ ratios identified at the 90% confidence level.

| | | | | Most Likely Location of Modified Residues[c] | |
|---|---|---|---|---|---|
| **Digest** | **Mass (Da)** | **Residue[a]** | **Difference[b]** | **Lysine** | **Arginine** |
| Trypsin | 2490.280 | 21-41 | 104% | 41 | n/a |
| LysC | 1149.619 | **42-51/1-10** | 98% | 51/**N+** | n/a |
| GluC | 1313.796 | **521-531** | 75% | 524, **525** (Pyr, FL, or CEL) | 521 (G-H1) |
| Trypsin | 1434.525 | 82-93 | 72% | 93 | n/a |
| Trypsin | 1899.009 | 146-160 | 67% | 159 | 160 |
| Trypsin | 1149.604 | **1-10** | 61% | 4/**N+** | 10 (MG-H1) |
| Trypsin | 1342.620 | 546-557 | 44% | 557 | n/a |
| GluC | 1737.692 | 87-100 | 42% | 93 | 98 |
| GluC | 2194.152 | **189-208** | 40% | 190**,** 195, **199**, 205 (Pyr, FL, or CML) | 197 (ArgP, or IB) |
| Trypsin | 2045.096 | 373-389 | 36% | 378, 389 | n/a |
| Trypsin | 2086.798 | 241-257 | 30% | n/a | 257 |

[a] – Residues that have been previously identified by other authors are shown in **bold**.

[b] – This value represents the percent relative difference of the $^{16}O/^{18}O$ ratio for the glycated HSA sample and the control HSA sample.

[c] – The most likely modification sites are underlined. The suspected modifications (based on the data in previous chapters) are also shown in parenthesis.

Table 6.3 (continued). Peptides in the HSA-5 sample with high $^{16}O/^{18}O$ ratios identified at the 90% confidence level.

| | | | | Most Likely Location of Modified Residues[c] | |
|---|---|---|---|---|---|
| **Digest** | **Mass (Da)** | **Residue[a]** | **Difference[b]** | **Lysine** | **Arginine** |
| Trypsin | 1714.796 | 94-106 | 27% | 106 | 98 |
| Trypsin | 1552.595 | 360-372 | 27% | 372 | n/a |
| GluC | 1955.975 | **426-442** | 26% | 432, 436, **439** (Pyr, CEL, FL, or AFGP) | 428 (G-H1, ArgP, AFGP, 3DG-H1, or THP) |
| GluC | 1519.783 | 142-153 | 26% | n/a | 144 (ArgP), 145 |
| GluC | 2348.094 | 101-119 | 22% | 106 | 114, 117 |
| Trypsin | 927.495 | 138-144 | 22% | n/a | 144 (ArgP) |
| GluC | 2578.264 | 61-82 | 17% | 64, 73 | 81 |
| LysC | 2348.988 | 74-93 | 15% | 93 | 81 |
| GluC | 2232.103 | 209-227 | 10% | 212, 225 (CEL) | 209, 218, 222 (ArgP) |
| LysC | 1552.592 | 360-372 | 7% | 372 | n/a |

[a] – Residues that have been previously identified by other authors are shown in **bold**.

[b] – This value represents the percent relative difference of the $^{16}O/^{18}O$ ratio for the glycated HSA sample and the control HSA sample.

[c] – The most likely modification sites are underlined.

The largest relative changes in $^{16}O/^{18}O$ ratios occurred for residues 21-41, 1-10 or 42-51, 521-531, 82-93, and 146-160 on HSA, where the $^{16}O/^{18}O$ ratios for these residues increased by 104%, 98%, 75%, 72%, and 67%, respectively. Of these residues, the *N*-terminus and lysine 525 were both found on peptides that had the most significant levels of modification in this analysis. In addition, K199 and K439 were also found on peptides that had significantly increased $^{16}O/^{18}O$ ratios at the 90% confidence level; however, the extent of modification that was found on these residues was moderate. The ranking of the relative change in $^{16}O/^{18}O$ ratio for the HSA-5 digest was similar to the relative change in $^{16}O/^{18}O$ ratios that was obtained for the HSA-2 sample, where a peptide containing the *N*-terminus was found to have the highest amount of modification followed by lysines 525, 199 and 439 in the HSA-5 sample. Lysine 281, which was previously identified as one of the more likely glycation sites on HSA, was found on a peptide which had an associated relative increase in its $^{16}O/^{18}O$ ratio of 8% when compared to its reference $^{16}O/^{18}O$ ratio (i.e., 1.37 +/- 0.51). This amount of increase; however, did not represent a significant increase at the 90% confidence level because of the small difference in the average $^{16}O/^{18}O$ ratios and the large variance that was associated with the $^{16}O/^{18}O$ ratio measurement in the HSA-5 sample (1.49 +/- 0.80). This result was unexpected because this same peptide was identified with a 48% increase in $^{16}O/^{18}O$ ratio in the HSA-2 sample. A possible explanation for this result is that peptides found within residues 275-286 of HSA may be more susceptible to early glycation than later modification by advanced glycation end products. This finding also indicated that any early-stage glycation products that were present on this residue in the HSA-2 sample may have degraded after 3-weeks of incubation.

When this data was compared to the data that was previously obtained for commercially glycated HSA (see Chapter 2), six of the eleven peptides that were previously identified as having high amounts of modification in commercially glycated HSA were also found to have high amounts of modification in this current analysis. The previously identified regions of significant glycation overlap with residues 521-531, 146-160, 426-442, 101-119, 61-93, and 209-227, which were found to have relative increases in their $^{16}O/^{18}O$ ratios of 75%, 67%, 26%, 22%, 15-17%, and 10%, respectively. Lysine 525, which is found in residues 521-531 of HSA, gave a relative increase in the $^{16}O/^{18}O$ ratio of 75% using this current approach, where this increase was the third highest relative increase in $^{16}O/^{18}O$ ratio obtained of all the $^{16}O/^{18}O$ ratios of peptides found in the HSA-5 digest. This result is consistent with the findings for the HSA-2 sample, where a large relative increase in the $^{16}O/^{18}O$ ratio was also found for residue 521-531. The data presented here for the HSA-5 sample does not indicate that extensive modification is occurring on lysines 233 or 439.

**6.C.5. Determination of the most likely modification sites on glycated HSA.** To determine the most likely modification sites, peptides that were found to have high $^{16}O/^{18}O$ ratios in this current analysis were compared to modifications that were previously identified in the glycated HSA samples (see Chapters 4 and 5) as well as known information about localized acid-base catalysis. The p$K_a$ and fractional accessible surface (FAS) area for a given amino acid were previously used to identify the most likely early or late stage glycation products. There is however little consensus amongst other researchers as to whether the p$K_a$ values and the accessibility of these residues can be reliably used to assign glycation sites (6, 20, 21). An attempt was therefore made to

assign the most likely regions of modification based on the empirical $^{16}O/^{18}O$ data and compare this data with the previous assignments based on p$K_a$ and FAS data.

When the $^{18}O$-labeling data was factored into the identification of modified residues in the HSA-2 sample, several early and late stage glycation products could be linked to peptides with large changes in $^{16}O/^{18}O$ ratio values. A summary of these peptides is given in Table 6.4, where eight peptides were found to be modified by either early or late stage glycation products based on their mass shifts. The majority of modifications that were identified using $^{16}O/^{18}O$ ratio assignments were similar to peptides that were identified using p$K_a$ and FAS based assignment; however, modifications that occurred on residues 168-188 and 275-286 were both unique to this study. For example, a m/z value of 2547.21 Da indicated that residues 168-188 or 566-585 were being modified to form fructose-lysine (FL) plus argpyrimidine (ArgP) or 1-alkyl-2-formyl-3,4-glycosyl-pyrrole (AFGP) respectively. Since the mass value of the peptide/modification combination in both cases was the same, the most likely modification could not be differentiated based solely on mass accuracy. When p$K_a$ and FAS sorting was used, the later peptide (residues 566-585) was found to be the most likely modification region, however, when the $^{18}O$-labeling data was used, the peptide corresponding to residues 168-188 was found to be the most likely modification site. This finding is in agreement with a previous study where it was found that p$K_a$ and FAS trends do not always agree with the experimental data (6). A similar assignment was made for the HSA-5 sample, where peptides that were found to have high $^{16}O/^{18}O$ ratios in this analysis were compared to modifications that were previously identified in the glycated

Table 6.4. Identification of early and late stage glycation on the HSA-2 sample based on the change in $^{16}O/^{18}O$ ratios of a control sample versus a glycated sample.

| Digest | Raw Mass (Da) | Δ Mass (Da) | Residue | Suspected Modification[a] |
|---|---|---|---|---|
| Trypsin | 1563.709 | 414.116 | 1-10 | FL+AFGP, AFGP+(3-DG-H1or THP) |
| [b] | 1738.826 | 306.0951 | 275-286 | FL*2 |
| Glu-C | 2107.961 | 234.074 | 278-292 | FL+CEL, FL+Pyr |
| | 2199.812 | 126.032 | 83-100 | FL, CEL+MG-H1 |
| [b] | 2547.209 | 242.079 | 168-188 | FL+ArgP |
| Lys-C | 926.455 | 72.021 | 206-212 | CEL |
| | 1480.763 | 306.095 | 196-205 | FL+FL, FL+3-DG-H1, FL+THP |
| | 2466.271 | 414.116 | 414-428 | FL+AFGP, AFGP+3-DG-H1, AFGP+THP |

[a] – Due to the complexity and size of this dataset, all of the early stage glycation products (i.e., FL, FL-$1H_2O$, FL-$2H_2O$) are summarized as FL.

[b] – Multiple peptide assignment was possible due to overlapping m/z values resulting from peptides with different types of modification. The $^{16}O/^{18}O$ ratios were used to deduce the most likely combinations in these cases.

Table 6.5. Identification of early and late stage glycation in the HSA-5 sample based on the change in $^{16}O/^{18}O$ ratios of control HSA versus glycated HSA.

| Digest | Raw Mass (Da) | Δ Mass (Da) | Residue | Suspected Modification[a] |
|---|---|---|---|---|
| Trypsin | 949.492 | 148.016 | 520-525 | Pyr+G-H1 |
| [b] | 1171.6 | 198.0529 | 5-12 | FL-1H2O+MG-H1 |
| | 1405.715 | 350.1002 | 137-144 | AFGP+ArgP |
| [b] | 1738.829 | 220.0583 | 182-195 | FL+CML |
| | 2658.159 | 342.0951 | 318-336 | CEL+AFGP |
| Glu-C | 1989.965 | 148.016 | 426-442 | Pyr+G-H1 |
| | 2108.03 | 152.0473 | 426-442 | CEL+ArgP |
| | 2444.207 | 250.0477 | 189-208 | Pyr+IB |
| Lys-C | 1171.591 | 198.0529 | 5-12 | FL-1$H_2O$+MG-H1 |
| | 1272.743 | 144.0423 | 525-534 | FL-1$H_2O$, CEL+CEL |
| | 2466.27 | 414.1163 | 414-432 | AFGP+FL-1$H_2O$, AFGP+3-DG-H1, or AFGP+THP |

[a] – Due to the complexity and size of this dataset, all of the early stage glycation products (i.e., FL, FL-1$H_2O$, FL-2$H_2O$) are summarized as FL.

[b] – Multiple peptide assignment was possible due to overlapping m/z values resulting from peptides with different types of modification. The $^{16}O/^{18}O$ ratios were used to deduce the most likely combinations in these cases.

HSA samples (see Table 6.5). All of the modifications that were determined by $^{16}O/^{18}O$ ratio assignment were also found when the p$K_a$ and FAS values were factored in.

There were many situations in this current study where particular regions of HSA that had high $^{16}O/^{18}O$ ratios, contained many potential sites of modification. In these cases the location of a given lysine or arginine residue in relation to a positively charged residue (i.e., histidine and lysine residues) was used to help identify the most reactive regions on HSA. The distance between amino acids was calculated from a pdb file (22) using Visual Molecular Dynamics (23). If no basic residues were in the vicinity (i.e., within 20 Å) of the suspected modified residues, the p$K_a$ and FAS data were then used to further differentiate which amino acid residue was most likely to be modified. The results for this analysis are shown in Tables 6.2 and 6.3. For example, a peptide encompassing residues 189-208 on HSA was found to have a high $^{16}O/^{18}O$ ratio in the HSA-5 sample. This region of HSA contains four lysine and one arginine residue and was modified to form ArgP or imidazolone B (IB), which are arginine specific, and pyrraline (Pyr), FL, or $N_\varepsilon$-carboxymethyl-lysine (CML), which are lysine specific. Since there is only one arginine residue in this peptide, the assignment of ArgP or IB to R197 is straight-forward; however, the assignment of the lysine specific modifications was less obvious. The α-carbon of K199 is located 15 Å away from the α-carbon of H257, which is a basic residue at physiological pH. In addition, K199 has a low p$K_a$ of 7.47 versus an average p$K_a$ of 10.08 for all lysines on HSA. These factors strongly indicate that K199 was the most likely modification site in this region of HSA. The previously obtained m/z shift data (see Chapter 5); however, indicated that some modification was also occurring within regions 187-197 on HSA to form Pyr and ArgP adducts (see Table 6.5). This region contains

K190 and K195, where K190 has a p$K_a$ of 6.23 and K195 has a p$K_a$ of 10.76. The fact that K190 would be mostly deprotonated at a physiological pH indicates that this residue is more reactive than K195. The data, therefore, indicated that modifications that occurred on K190, K199, and R197 all contributed to the increased $^{16}O/^{18}O$ ratio that corresponded to residues 189-208 in the HSA-5 sample. Similar assignments were made in both the HSA-2 and HSA-5 samples, where the most likely modification sites are indicated in Tables 6.2 and 6.3

The glycated HSA samples were also compared to see whether there were any major differences in the total number of suspected early or late stage glycation products that were identified in these samples. For simplicity, a total count of the potential early or late stage glycation products was performed using the data presented in Tables 6.4 and 6.5. A total of twelve early stage glycation products were identified in the HSA-2 compared to five early stage glycation adducts in the HSA-5 sample. In addition, 16 late stage glycation adducts were found in the HSA-2 sample versus a total of 21 late stage glycation adducts identified in the HSA-5 sample. The total number of modifications was consistent at both stages of glycation, where a total of 28 and 26 modifications were identified in the HSA-2 and HSA-5 samples respectively. These trends are in line with the expectation that when proteins are glycated for extended periods of time, the concentration of late-stage glycation products should increase relative to the concentration of early stage glycation products. This trend also explains the results from the enzymatic glycation assay (see Chapter 4), where the concentration of fructosamine initially increased at a fairly steady rate (i.e., for a 0 to 3 week glycation period) followed by a decrease in the rate of fructosamine formation after a 3 week period of glycation.

## 6.D. SUMMARY

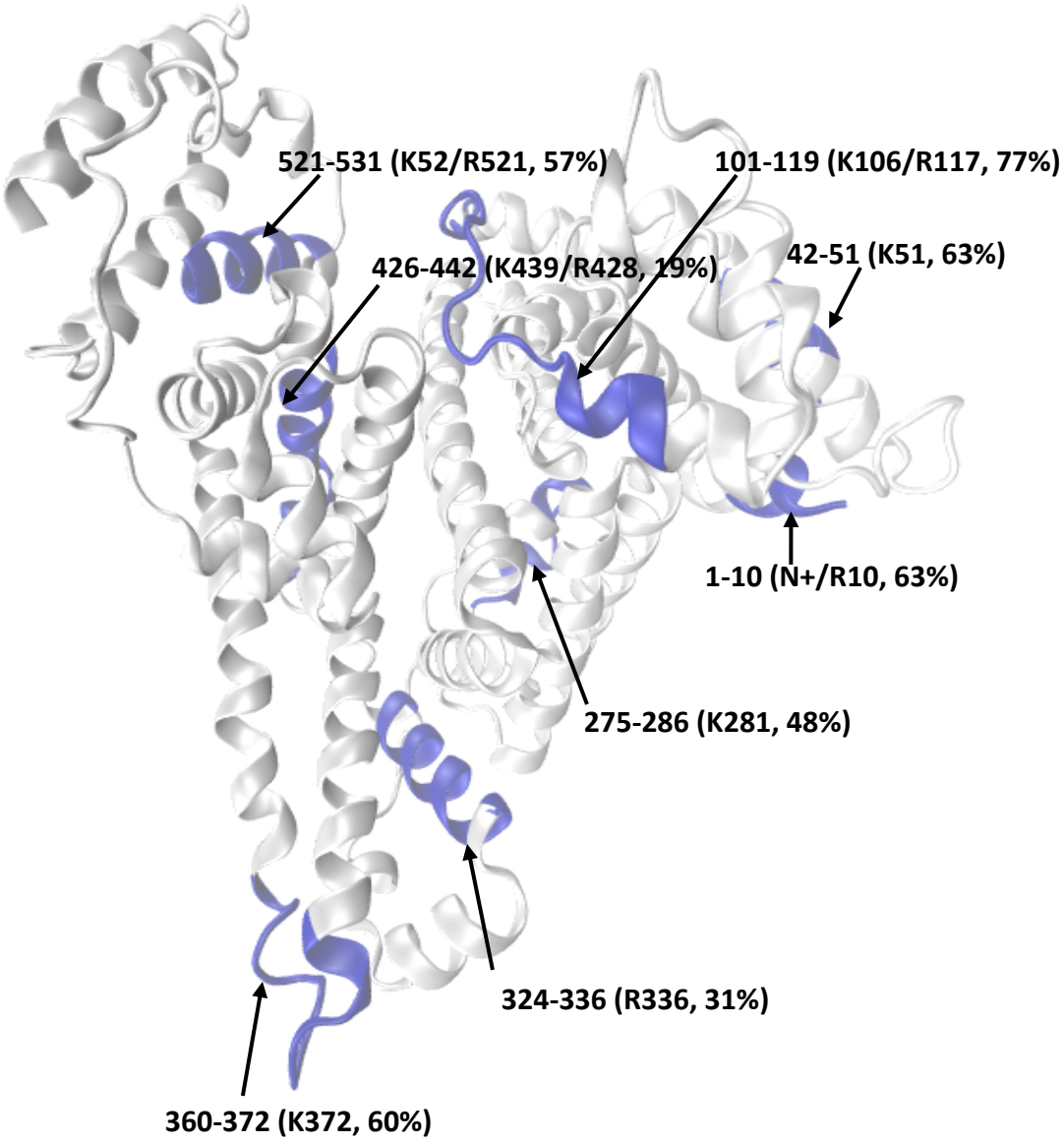
This chapter quantified the relative extent of modification that occurred on different regions of HSA. The major difference with this current approach, as compared to the previous attempt (Chapter 2), was that reference $^{16}O/^{18}O$ ratios were factored into the determination of modification sites. By obtaining a set of reference $^{16}O/^{18}O$ ratios, it was possible to do a direct comparison of these reference ratios to the $^{16}O/^{18}O$ ratios that were obtained for glycated HSA. These ratios were then ranked according to relative increase in $^{16}O/^{18}O$ ratio and used to determine the most likely regions of modification on HSA that contained varying amounts of total glycation. The data strongly indicated that the degree of modification that occurs of HSA during glycation is more widespread than was previously thought.

A summary of the largest changes in $^{16}O/^{18}O$ ratios at both sampled stages of glycation is given in Figure 6.5 (i) and (ii). It was determined that significant amounts of modification were taking place at or near both Sudlow sites 1 and 2 (i.e., the major binding sites on HSA) where Sudlow site 1 has a larger degree of modification in both tested HSA samples. For the HSA-2 sample, five significantly modified residues were found within Sudlow site 1, where the relative increase in $^{16}O/^{18}O$ ratio ranged anywhere from 16% - 48%. A lower degree of modification was found within Sudlow site 2 in the HSA-2 sample, where three significantly modified peptides were identified with relative increases in $^{16}O/^{18}O$ ratios that ranged between 17% and 29%. Similarly, a greater amount of modification was found in Sudlow site 1 for the HSA-5 sample, where three modification sites were identified with relative $^{16}O/^{18}O$ ratio increases of 30% and 67%,

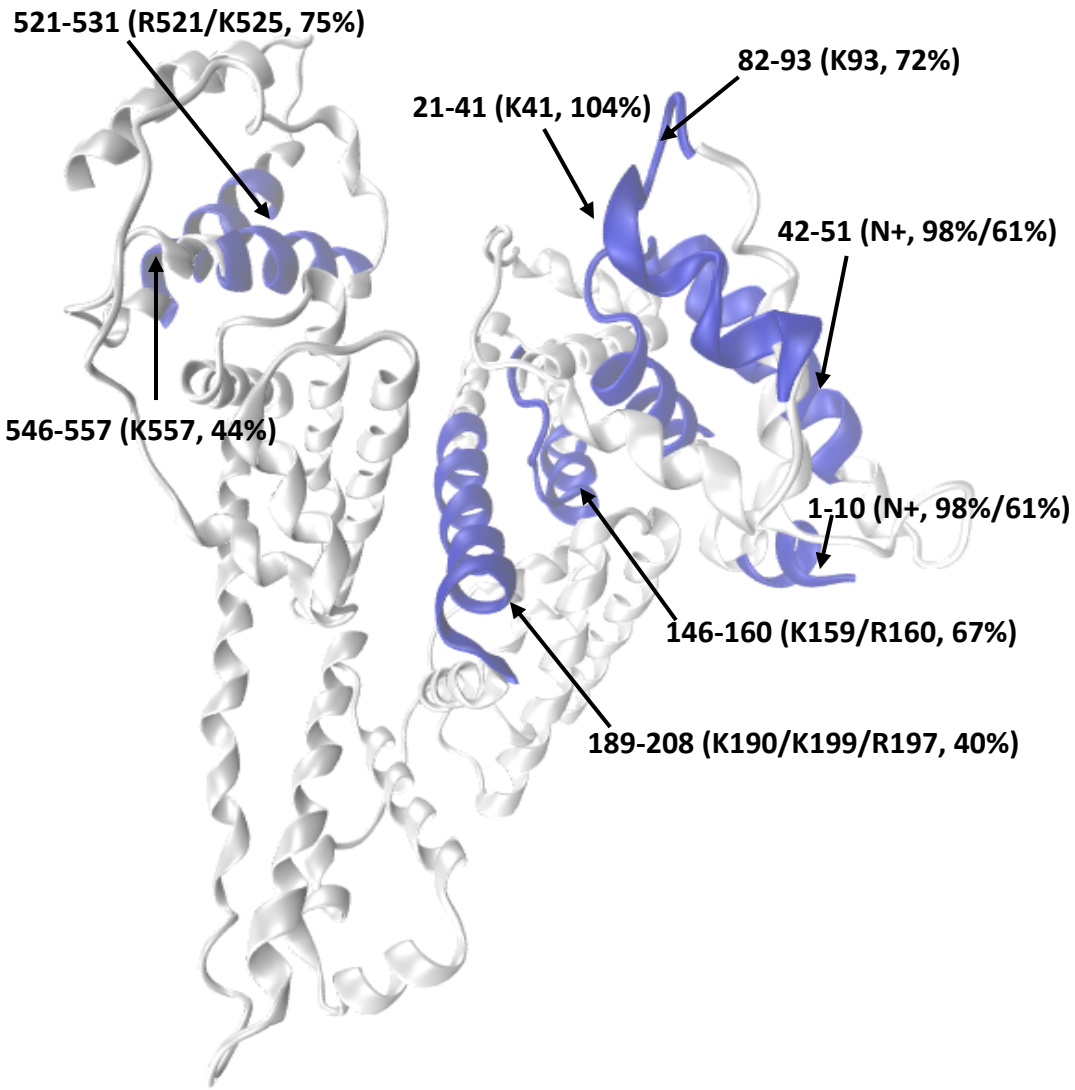
Figure 6.5 (i) A structure of HSA that shows the regions with the highest relative increases in $^{16}O/^{18}O$ ratios for the HSA-2 sample versus normal HSA. Several residues were identified as being highly glycated at both Sudlow sites of HSA. This structure was generating a pdb file and VMD (23).

(ii) A structure of HSA that shows the regions with the highest relative increases in $^{16}O/^{18}O$ ratios for the HSA-5 sample versus normal HSA. As was the case for the HSA-2 sample, several residues were identified as being highly glycated at both Sudlow sites of HSA (i.e., both major binding sites). There were also some differences in the ranking of these highly glycated regions, indicating that the length glycation alters the glycation pattern.

(i)


521-531 (K52/R521, 57%)
101-119 (K106/R117, 77%)
426-442 (K439/R428, 19%)
42-51 (K51, 63%)
1-10 (N+/R10, 63%)
275-286 (K281, 48%)
324-336 (R336, 31%)
360-372 (K372, 60%)

(ii)


521-531 (R521/K525, 75%)
82-93 (K93, 72%)
21-41 (K41, 104%)
42-51 (N+, 98%/61%)
546-557 (K557, 44%)
1-10 (N+, 98%/61%)
146-160 (K159/R160, 67%)
189-208 (K190/K199/R197, 40%)

respectively. Only one modification site was found in Sudlow site 2 for the HSA-5 sample, which had a relative $^{16}O/^{18}O$ ratio increase of 36%. The largest observed change in $^{16}O/^{18}O$ ratios in the HSA-2 sample were found for residues 101-119, 1-10 or 42-51, 360-372, 521-537, and 275-286. These findings are consistent with the glycation pattern seen previously, where K525, K281, and the *N*-terminus were found to have significant levels of glycation (5). These $^{16}O/^{18}O$ ratios indicated that the *N*-terminus or K51, K372, K525, and K281 and arginines 117, 10, 521 were being modified. Glycation products were found on the *N*-terminus, lysines 525 and 281, as well as arginines 10 and 521.

For the HSA-5 sample, the highest amount of modification occurred on residues 21-41, 1-10 or 42-51, 521-531, 82-93, and 146-160 on HSA. Of these residues, the high $^{16}O/^{18}O$ ratio obtained for residues 1-10 and 521-531 are consistent with the expectation that the *N*-terminus and K525 should have significant amounts of glycation. In addition, lysine 199, which is also a suspected glycation site (5), was also found to have significant levels of modification in this analysis. Lysine 281 was also identified as a glycation site with a relative $^{16}O/^{18}O$ ratio increase in the HSA-5 sample. The increase associated with this residue, however, was not significant at the 90% confidence level. Based on the m/z shift data that were sorted by $^{16}O/^{18}O$ ratios, the *N*-terminus/K51, and lysines 525, 199, and 378 as well as arginines 521, 10, and 197, could all be linked to residues that were modified to form glycation-related products.

## 6.E. REFERENCES

(1) Thornalley, P. J.; Langborg, A.; Minhas, H. S. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **1999***, 344*, 109-116.

(2) Horiuchi, S. Advanced glycation end products (AGE)-modified proteins and their potential relevance to atherosclerosis. *Trend Cardiovasc. Med.* **1996***, 6*, 163-168.

(3) Araki, N.; Ueno, N.; Chakrabarti, B.; Morino, Y.; Horiuchi, S. Immunochemical evidence for the presence of advanced glycation end products in human lens proteins and its positive correlation with aging. *J. Biol. Chem.* **1992***, 267*, 10211-10214.

(4) Vorum, H.; Fisker, K.; Otagiri, M.; Pedersen, A. O.; Hansen, U. K. Calcium ion binding to clinically relevant chemical modifications of human serum albumin. *Clin. Chem.* **1995***, 41*, 1654-1661.

(5) Peters, T. In *All About Albumin: Biochemistry, Genetics, and Medical Applications.* Academic Press: San Diego, 1996, pp 102-126.

(6) Iberg, N.; Fluckiger, R. Nonenzymatic glycosylation of albumin in vivo. *J. Biol. Chem.* **1986***, 261*, 13542-13545.

(7) Stewart, I. I.; Thomson, T.; Figeys, D. $^{18}O$ labeling: A tool for proteomics. *Rapid. Comm. Mass. Spec.* **2001***, 15*, 2456-2465.

(8) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. Protease-catalyzed incorporation of $^{18}O$ into peptide fragments and its application for protein

sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis.* **2005***, 17*, 945-953.

(9) Wa, C.; Cerny, R. L.; Hage, D. S. Identification and quantitative studies of protein immobilization sites by stable isotope labeling and mass spectrometry. *Anal. Chem.* **2006***, 78*, 7967-7977.

(10) Armbuster, D. A. Fructosaminase: Structure, analysis and clinical usefulness. *Clin. Chem.* **1987***, 33*, 2153-2163.

(11) Kouzuma, T.; Usami, T.; Yamakoshi, M.; Takahashi, M.; Imamura, S. An enzymatic method for the measurement of glycated albumin in biological samples. *Clin. Chim. Acta.* **2002***, 324*, 61-71.

(12) Abordo, E. A.; Thornalley, P. J. In *Pro-inflammatory cytokine synthesis by human monocytes induced by proteins minimally-modified by methylglyoxal.* O'Brien, J., Nursten, H. E., Crabbe, M. J. C. and Ames, J. M., Eds.; The Maillard Reaction in Foods and Medicine; The Royal Society of Chemistry: Cambridge, UK, 1998; pp 357-362.

(13) Lapolla, A.; Fedele, D.; Seraglia, R.; Catinella, S.; Baldo, L.; Aronica, R.; Traldi, P. A new effective method for the evaluation of glycated intact plasma proteins in diabetic subjects. *Diabetologia.* **1995***, 38*, 1076-1081.

(14) Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using $^{18}$O-labeled internal standards. *Rapid. Comm. Mass. Spec.* **2000***, 14*, 1226-1232.

(15) Harris, D. C., Ed.; In *Quantitative Chemical Analysis.* W.H. Freeman and Company: New York, 2003, pp 739.

(16) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **1999***, 20*, 3551-3567.

(17) Wa, C.; Cerny, R. L.; Clarke, W. A.; Hage, D. S. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta.* **2007***, 385*, 48-60.

(18) Wa, C.; Cerny, R.; Hage, D. S. Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: analysis of human serum albumin as a model. *Anal. Biochem.* **2006***, 349*, 229-241.

(19) Robb, D. A.; Olufemi, S. O.; Williams, D. A.; Midgley, J. M. Identification of glycation at the N-terminus of albumin by gas chromatography - mass spectrometry. *Biochem J.* **1989***, 261*, 871-878.

(20) Bai, Y.; Ueno, H.; Manning, J. M. Some factors that influence the nonenzymatic glycation of peptides and polypeptides by glyceraldehyde. *J. Protein Chem.* **1989***, 8*, 299-315.

(21) Lapolla, A.; Fedele, D.; Reitano, R.; Aricò, N. C.; Seraglia, R.; Traldi, P.; Marotta, E.; Tonani, R. Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. *J. Am. Soc. Mass. Spectrom.* **2004***, 15*, 496-509.

(22) Sugio, S.; Kashima, A.; Mochizuki, S.; Kobayashi, K. Crystal structure of human serum albumin at 2.5 A resolution. *Protein Eng.* **1999***, 12*, 439-446.

(23) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996***, 14*, 33-38.

## 6.F. APPENDIX

Table 6.1a shows the Matlab programs that were used to obtain area values for the M+0 to M+4 peaks within a given isotopic cluster. Two separate programs were developed for the $^{16}O$-labeled digest (i) and the $^{18}O$ labeled digest (ii) respectively. The $^{16}O/^{18}O$ ratios that were obtained for the Glu-C digest (Table 6.2a) and the Lys-C digest (Table 6.3a) are also given. The useable sequence coverage (i.e. the percentage of amino acids that correspond to calculated $^{16}O/^{18}O$ ratios) is also given in Figure 6.1a for the HSA-2 (i) and HSA-5 (ii) samples. Figure 6.2 summarizes the reference range of $^{16}O/^{18}O$ ratios that were obtained for the Glu-C (i) and Lys-C (ii) samples respectively.

Table 6.1a (i). Program used for obtaining the area values for the isotopes of a given peptide in the $^{16}O$ digest or mixed digest

```
[massa]=textread('mascotmatch16.txt', '%f');
[massb, area]=textread('rawdata.txt', '%f %f');

sizemassa=size(massa);
sizemassb=size(massb);

k=0;
h=0;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb(1);
       acc(i,j)=((massa(i)-massb(j))/massb(j))*10^6;
       if ((acc(i,j)>-70)&& (acc(i,j)<70));

       k=k+1;
       h=h+1;
       massb0(k,:)=(massb(j)+0);
       massb1(k,:)=(massb(j)+1);
       massb2(k,:)=(massb(j)+2);
       massb3(k,:)=(massb(j)+3);
       massb4(k,:)=(massb(j)+4);

       result(h,:)= [massb0(k,:), massb1(k,:), massb2(k,:),
massb3(k,:), massb4(k,:)];

       end;
    end;
end;


outfile=fopen('result.txt', 'wt');
sizeresult=size(result);
for i=1:1:sizeresult(1);
    fprintf(outfile, '%f\t %f\t %f\t %f\t %f\t \n', result(i,1),
result(i,2), result(i,3), result(i,4), result(i,5));
end
fclose(outfile);

clear all

[massb0, massb1, massb2, massb3, massb4]=textread('result.txt', '%f %f
%f %f %f');
[massa, area]=textread('rawdata.txt', '%f %f');

sizemassa=size(massa);
sizemassb0=size(massb0);
sizemassb1=size(massb1);
sizemassb2=size(massb2);
sizemassb3=size(massb3);
sizemassb4=size(massb4);
```

Table 6.1a (i), continued. Program used for obtaining the area values for the isotopes of a given peptide in the $^{16}O$ digest or mixed digest

```
k=0;
h=0;
m=0;
n=0;
o=0;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb0(1);
        acc0(i,j)=(((massa(i)-massb0(j))/massb0(j))*10^6);
        if ((acc0(i,j)>-100)&& (acc0(i,j)<100));
           k=k+1;
           a0(k,:)=[massa(i), area(i)];
        end;
    end;
end;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb1(1);
        acc1(i,j)=(((massa(i)-massb1(j))/massb1(j))*10^6);
        if ((acc1(i,j)>-100)&& (acc1(i,j)<100));
           h=h+1;
           a1(h,:)=[massa(i), area(i)];
        end;
    end;
end;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb2(1);
        acc2(i,j)=(((massa(i)-massb2(j))/massb2(j))*10^6);
        if ((acc2(i,j)>-100)&& (acc2(i,j)<100));
           m=m+1;
           a2(m,:)=[massa(i), area(i)];
        end;
    end;
end;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb3(1);
        acc3(i,j)=(((massa(i)-massb3(j))/massb3(j))*10^6);
        if ((acc3(i,j)>-100)&& (acc3(i,j)<100));
           n=n+1;
           a3(n,:)=[massa(i), area(i)];
        end;
    end;
end;
for i=1:1:sizemassa(1);
    for j=1:1:sizemassb4(1);
        acc4(i,j)=(((massa(i)-massb4(j))/massb4(j))*10^6);
        if ((acc4(i,j)>-100)&& (acc4(i,j)<100));
           o=o+1;
           a4(o,:)=[massa(i), area(i)];
```

Table 6.1a (i), continued. Program used for obtaining the area values for the isotopes of a given peptide in the $^{16}O$ digest or mixed digest

```
        end;
    end;
end;

%The a0 through a4 values should represent the distribution for the
oxygen
%16 digests. b0 through b4 are being used as random variables, not
linked to any distribution.

sort_mass
```

Table 6.1a (ii). Program used for obtaining the area values for the isotopes of a given peptide in the $^{18}O$ digest

```
[massc]=textread('mascotmatch16.txt', '%f');
[massb, area]=textread('rawdata.txt', '%f %f');

massa=massc+4;

sizemassa=size(massa);
sizemassb=size(massb);
sizemassc=size(massc);

k=0;
h=0;

%modified massa to represent the b4 peak
for i=1:1:sizemassa(1);
    for j=1:1:sizemassb(1);

       acc(i,j)=((massa(i)-massb(j))/massb(j))*10^6;
       if ((acc(i,j)>-60)&& (acc(i,j)<60));

       k=k+1;
       h=h+1;

       massb0(k,:)=(massb(j)-4);
       massb1(k,:)=(massb(j)-3);
       massb2(k,:)=(massb(j)-2);
       massb3(k,:)=(massb(j)-1);
       massb4(k,:)=(massb(j)-0);

       result(h,:)= [massb0(k,:), massb1(k,:), massb2(k,:),
massb3(k,:), massb4(k,:)];

       end;
    end;
end;

outfile=fopen('result.txt', 'wt');
sizeresult=size(result);
for i=1:1:sizeresult(1);
    fprintf(outfile, '%f\t %f\t %f\t %f\t %f\t \n', result(i,1),
result(i,2), result(i,3), result(i,4), result(i,5));
end
fclose(outfile);

clear all
[massb0, massb1, massb2, massb3, massb4]=textread('result.txt', '%f %f
%f %f %f');
[massa, area]=textread('rawdata.txt', '%f %f');

sizemassa=size(massa);
sizemassb0=size(massb0);
sizemassb1=size(massb1);
```

Table 6.1a (ii), continued. Program used for obtaining the area values for the isotopes of a given peptide in the $^{18}O$ digest

```
sizemassb2=size(massb2);
sizemassb3=size(massb3);
sizemassb4=size(massb4);

k=0;
h=0;
m=0;
n=0;
o=0;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb0(1);
        acc0(i,j)=(((massa(i)-massb0(j))/massb0(j))*10^6);
        if ((acc0(i,j)>-100)&& (acc0(i,j)<100));
           k=k+1;
           b0(k,:)=[massa(i), area(i)];
        end;
    end;
end;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb1(1);
        acc1(i,j)=(((massa(i)-massb1(j))/massb1(j))*10^6);
        if ((acc1(i,j)>-100)&& (acc1(i,j)<100));
           h=h+1;
           b1(h,:)=[massa(i), area(i)];
        end;
    end;
end;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb2(1);
        acc2(i,j)=(((massa(i)-massb2(j))/massb2(j))*10^6);
        if ((acc2(i,j)>-100)&& (acc2(i,j)<100));
           m=m+1;
           b2(m,:)=[massa(i), area(i)];
        end;
    end;
end;

for i=1:1:sizemassa(1);
    for j=1:1:sizemassb3(1);
        acc3(i,j)=(((massa(i)-massb3(j))/massb3(j))*10^6);
        if ((acc3(i,j)>-100)&& (acc3(i,j)<100));
           n=n+1;
           b3(n,:)=[massa(i), area(i)];
        end;
    end;
end;
for i=1:1:sizemassa(1);
    for j=1:1:sizemassb4(1);
```

Table 6.1a (ii), continued. Program used for obtaining the area values for the isotopes of a given peptide in the $^{18}O$ digest

```
        acc4(i,j)=(((massa(i)-massb4(j))/massb4(j))*10^6);
        if ((acc4(i,j)>-100)&& (acc4(i,j)<100));
           o=o+1;
           b4(o,:)=[massa(i), area(i)];
        end;
    end;
end;

%The b0 through b4 values should represent the distribution for the
oxygen
%16 digests. b0 through b4 are being used as random variables, not
linked
%to any distribution.
```

Table 6.2a. $^{16}O/^{18}O$ ratios that were obtained for the Glu-C digest

| | | | | Control | | HSA-2 | | HSA-5 | |
|---|---|---|---|---|---|---|---|---|---|
| **Mass (Da)** | **Error (ppm)[a]** | **Residue** | **Amino Acid Sequence** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** |
| 1031.453 | 3 | 369-376 | CYAKVFDE | 1.92 | 0.65 | 1.44 | 0.79 | 1.43 | 0.51 |
| 1197.571 | -6 | 496-505 | TYVPKEFNAE | 0.96 | 0.83 | 0.46 | 0.07 | 1.20 | 0.47 |
| 1300.674 | 7 | 7-17 | VAHRFKDLGEE | 1.57 | 0.43 | 2.00 | 1.00 | 1.87 | 0.59 |
| 1313.796 | 4 | 521-531 | RQIKKQTALVE | 1.15 | 0.30 | 1.80 | 1.13 | 2.00 | 0.79 |
| 1519.783 | 2 | 142-153 | IARRHPYFYAPE | 1.40 | 0.30 | 1.80 | 0.78 | 1.76 | 0.19 |
| 1541.749 | 36 | 506-518 | TFTFHADICTLSE | 1.42 | 0.52 | 1.41 | 0.68 | 1.66 | 0.16 |
| 1548.770 | -16 | 480-492 | SLVNRRPCFSALE | 1.29 | 0.22 | 1.40 | 0.16 | 1.36 | 0.28 |
| 1680.796 | -5 | 466-479 | KTPVSDRVTKCCTE | 1.15 | 0.18 | 1.33 | 0.55 | 1.40 | 0.65 |
| 1704.910 | -22 | 154-167 | LLFFAKRYKAAFTE | 1.63 | 0.26 | 1.98 | 0.08 | 1.84 | 0.35 |
| 1737.692 | -4 | 87-100 | MADCCAKQEPERNE | 1.19 | 0.31 | 1.76 | 0.70 | 1.70 | 0.49 |
| 1801.915 | 0 | 451-465 | DYLSVVLNQLCVLHE | 0.49 | 0.00 | none detected | | 2.23 | 0.00 |
| 1945.862 | -1 | 502-518 | FNAETFTFHADICTLSE | 1.16 | 0.03 | 1.48 | 0.07 | 1.19 | 0.34 |
| 1955.975 | -8 | 426-442 | VSRNLGKVGSKCCKHPE | 1.15 | 0.08 | 1.38 | 0.12 | 1.46 | 0.11 |
| 2085.985 | -38 | 377-393 | FKPLVEEPQNLIKQNCE | 1.22 | 0.25 | 1.46 | 0.28 | 1.43 | 0.34 |
| 2194.152 | -11 | 189-208 | GKASSAKQRLKCASLQKFGE | 0.93 | 0.11 | 1.16 | 0.99 | 1.29 | 0.54 |
| 2232.208 | -19 | 209-227 | RAFKAWAVARLSQRFPKAE | 1.12 | 0.09 | 1.30 | 0.23 | 1.24 | 0.07 |
| 2348.094 | -42 | 101-119 | CFLQHKDDNPNLPRLVRPE | 0.99 | 0.05 | 1.75 | 1.11 | 1.21 | 0.08 |
| 2578.264 | -4 | 61-82 | NCDKSLHTLFGDKLCTVATLRE | 0.68 | 0.01 | 0.84 | 0.05 | 0.80 | 0.14 |

[a] – "Error" represents the mass accuracy of the assigned peptide. For simplicity, the control HSA digest was used for peptide mass fingerprinting.

[b] – "SD" designates the standard deviation that was calculated for the given $^{16}O/^{18}O$ ratios. The value represents the square root of the sum of the standard deviations squared that were obtained across different Zip-tip fractions.

Table 6.3a. $^{16}O/^{18}O$ ratios that were obtained for the Lys-C digest

| | | | | Control | | HSA-2 | | HSA-5 | |
|---|---|---|---|---|---|---|---|---|---|
| **Mass (Da)** | **Error (ppm)[a]** | **Residue** | **Amino Acid Sequence** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** | **$^{16}O/^{18}O$** | **SD[b]** |
| 854.457 | 6 | 206-212 | FGERAFK | 0.51 | 0.54 | 1.40 | 0.58 | 1.76 | 1.11 |
| 973.520 | -1 | 5-12 | SEVAHRFK | 1.13 | 0.32 | 0.93 | 0.07 | 1.04 | 0.15 |
| 1149.619 | 4 | 42-51 | LVNEVTEFAK | 1.00 | 0.23 | 1.64 | 0.77 | 1.99 | 0.64 |
| 1342.628 | -5 | 546-557 | AVMDDFAAFVEK | 1.48 | 0.30 | 1.10 | 0.20 | 1.07 | 0.16 |
| 1352.764 | -3 | 403-413 | FQNALLVRYTK | 1.29 | 0.14 | 1.09 | 0.23 | 1.32 | 0.76 |
| 1371.553 | -10 | 163-174 | AAFTECCQAADK | 1.44 | 0.44 | 1.43 | 0.35 | 1.23 | 0.23 |
| 1443.630 | -8 | 263-274 | YICENQDSISSK | 1.37 | 0.48 | 1.36 | 0.12 | 1.28 | 0.28 |
| 1529.860 | -7 | 213-225 | AWAVARLSQRFPK | 1.27 | 0.07 | 1.26 | 0.07 | 1.21 | 0.08 |
| 1552.592 | -4 | 360-372 | CCAAADPHECYAK | 1.31 | 0.09 | 1.71 | 0.13 | 1.40 | 0.06 |
| 1657.738 | -9 | 390-402 | QNCELFEQLGEYK | 1.54 | 0.53 | 1.48 | 0.71 | 1.34 | 0.56 |
| 1714.785 | -7 | 94-106 | QEPERNECFLQHK | 1.17 | 0.06 | 1.25 | 0.04 | 1.17 | 0.08 |
| 1924.053 | -17 | 415-432 | VPQVSTPTLVEVSRNLGK | 0.73 | 0.22 | 0.52 | 0.01 | 0.39 | 0.12 |
| 2045.051 | -22 | 373-389 | VFDEFKPLVEEPQNLIK | 1.28 | 0.06 | 1.17 | 0.05 | 1.30 | 0.08 |
| 2052.147 | -17 | 414-432 | KVPQVSTPTLVEVSRNLGK | 1.26 | 0.15 | 1.01 | 0.15 | 1.15 | 0.08 |
| 2259.987 | -16 | 21-41 | EFNAETFTFHADICTLSEK | 1.53 | 0.57 | none detected | | | |
| 2348.988 | -20 | 74-93 | LCTVATLRETYGEMADCCAK | 0.83 | 0.01 | 0.80 | 0.25 | 0.95 | 0.25 |
| 2490.244 | -17 | 21-41 | ALVLIAFAQYLQQCPFEDHVK | 0.79 | 0.10 | 0.65 | 0.12 | 0.98 | 0.13 |
| 2585.088 | -11 | 241-262 | VHTECCHGDLLECADDRADLAK | 0.93 | 0.21 | 0.88 | 0.04 | 0.92 | 0.04 |
| 2674.285 | -11 | 445-466 | RMPCAEDYLSVVLNQLCVLHEK | 0.88 | 0.04 | 0.86 | 0.09 | 0.70 | 0.09 |
| 2807.453 | -6 | 138-159 | YLYEIARRHPYFYAPELLFFAK | 0.84 | 0.01 | 0.84 | 0.07 | 0.75 | 0.03 |

[a] – "Error" represents the mass accuracy of the assigned peptide. For simplicity, the control HSA digest was used for peptide mass fingerprinting.

[b] – "SD" designates the standard deviation that was calculated for the given $^{16}O/^{18}O$ ratios. The value represents the square root of the sum of the standard deviations squared that were obtained across different zip-tip fractions.

**FIGURE LEGEND**

Figure 6.1a (i) Usable sequence coverage obtained for the HSA-2 sample when the calculated $^{16}O/^{18}O$ ratios were considered. A 12% reduction in coverage was observed for this sample compared to the control HSA digest.

(ii) Usable sequence coverage obtained for the HSA-5 sample when the calculated $^{16}O/^{18}O$ ratios were considered. A 5% reduction in sequence coverage was observed for this sample compared to the control HSA digest.

Figure 6.2a (i) Reference $^{16}O/^{18}O$ ratios for the Glu-C digest. The central dashed line represents the average $^{16}O/^{18}O$ ratio and the outer dashed lines represent the range of $^{16}O/^{18}O$ ratios that were obtained.

(ii) Reference $^{16}O/^{18}O$ ratios for the Lys-C digest. The central dashed line represents the average $^{16}O/^{18}O$ ratio and the outer dashed lines represent the range of $^{16}O/^{18}O$ ratios that were obtained.

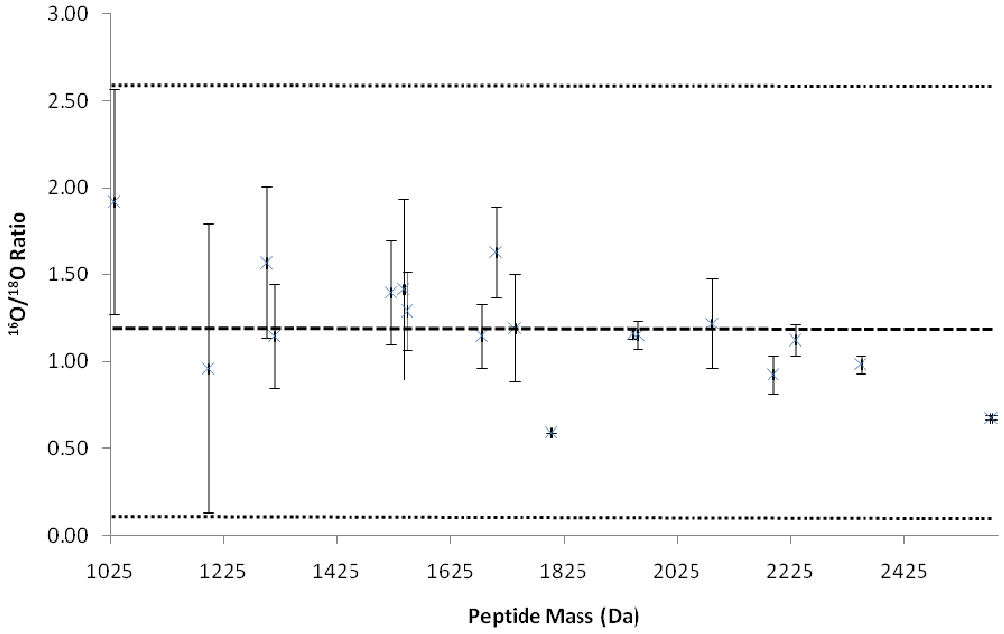
(i) Sequence coverage obtained for the HSA-2 digest

**1-** **DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV**
**41-** **KLVNEVTEFA K**tcvadesae **NCDKSLHTLF GDKLCTVATL**
**81-** **RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV**
**121-** **DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR**
**161-** **YKAAFTECCQ AADK**aacllp k**LDELRDEGK** assakqrlkc
**201-** aslqk**FGERA FKAWAVARLS QRFPKAE**fae vsklvtdltk
**241-** **VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE**
**281-** **KPLLEK**shci aevendempa dlpslaadfv eskdvcknya
**321-** eak**DVFLGMF LYEYARRHPD YSVVLLLR**la ktyettlek**C**
**361-** **CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE**
**400-** **YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH**
**441-** **PE**ak**RMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES**
**481-** **LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEK**e
**521-** **RQIKKQTALV E**lvkhkpkat keqlk**AVMDD FAAFVEK**cck
**561-** **ADDKETCFAE EGKK**lvaasq aalgl

(ii) Sequence coverage obtained for the HSA-5 digest

**1-** **DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV**
**41-** **KLVNEVTEFA K**tcvadesae **NCDKSLHTLF GDKLCTVATL**
**81-** **RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV**
**121-** **DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR**
**161-** **YKAAFTECCQ AADK**aacllp k**LDELRDEGK ASSAKQRLKC**
**201-** **ASLQKFGERA FKAWAVARLS QRFPKAE**fae vsklvtdltk
**241-** **VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE**
**281-** **KPLLEKSHCI AEVENDEMPA DLPSLAADFV ESK**dvcknya
**321-** eak**DVFLGMF LYEYARRHPD YSVVLLLR**la ktyettlek**C**
**361-** **CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE**
**401-** **YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH**
**441-** **PE**ak**RMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES**
**481-** **LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEK**e
**521-** **RQIKKQTALV E**lvkhkpkat keqlk**AVMDD FAAFVEK**cck
**561-** **ADDKETCFAE EGKK**lvaasq aalgl

(i)




(ii)

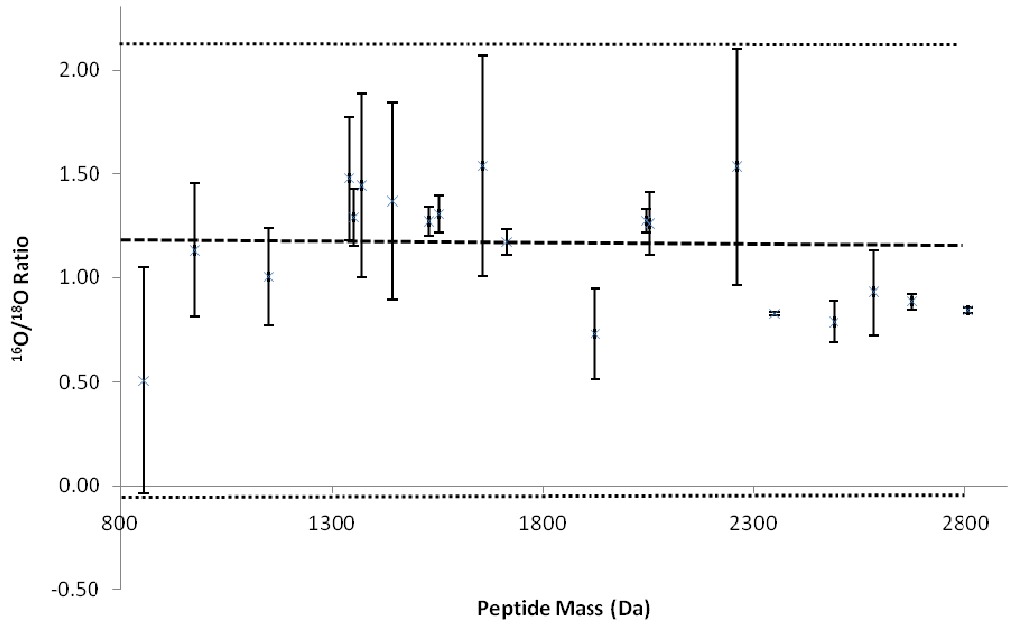



Figure 6.2a

# CHAPTER 7

# CONCLUSIONS AND FUTURE TRENDS

## 7.A. SUMMARY

The objective of the work presented in this dissertation was to provide qualitative and quantitative information about the modifications that occur on glycated human serum albumin (HSA). Techniques based on matrix-assisted laser desorption/ionization mass spectrometry (MALDI-TOF MS) and $^{18}O$-labeling were used for these studies.

The first chapter in this dissertation gave a brief introduction to protein glycation and techniques that are typically used for qualitatively and quantitatively characterizing glycation products. A discussion of the techniques that were examined in this dissertation was also provided here. The second chapter of this dissertation examined the use of $^{18}O$-labeling combined with MALDI-TOF MS as a means of determining the relative extent of modification that occurs on different regions of glycated HSA. This examination was done by digesting a HSA sample in $^{16}O$-labeled water and digesting a glycated HSA sample in $^{18}O$-labeled water, mixing these samples, and determining the $^{16}O/^{18}O$ ratio for each detected peptide. Under this regime, peptides that were glycated displayed an increase in their $^{16}O/^{18}O$ ratio. Several residues were identified in this study as having high levels of glycation. A qualitative comparison of the glycation sites indicated that K525, K439, and K233 may have significant levels of glycation. In addition, eight peptides were found to have high $^{16}O/^{18}O$ ratios, including R14, R81, R218, K413, K432, K159, K212, and K323. An observation was made in this study that the increases in

$^{16}O/^{18}O$ ratios were small compared to the typical observed range (~0.5 - 3.0) of $^{16}O/^{18}O$ ratios in control samples, therefore the use of reference $^{16}O/^{18}O$ ratios were needed.

The third chapter examined techniques that could be used for improving the assignment of glycation products on HSA. This chapter examined several variables, including the effects of mass accuracy, glycation product heterogeneity, and peptide selectivity, on the assignment of glycation sites. In this examination, it was found that increasing the mass accuracy by using internal calibration, using an expanded modification search list, and selecting the peptides that were only due to glycation could be used to reduce the complexity and improve the assignment of glycation sites. When these approaches were applied, lysines 281, 439, and 199 were found to be modified in glycated HSA.

Chapters 4 and 5 examined the locations and types of modifications that occur on glycated HSA. Chapter 4 focused on the identification of early-stage glycation products and Chapter 5 focused on the analysis of advanced glycation endproducts. HSA samples that contained three different levels of glycation were examined to see whether the modifications were consistent at different levels of total glycation. These samples were also used to semi-quantitatively rank the frequency of modification for a given peptide and compare this information to qualitative and quantitative data that was previously obtained for commercially glycated HSA. In the analysis of early-stage glycation that occurred on glycated HSA, regions 83-100, 1-12, 182-212, 278-292, 414-432, and 526-534 were found to have a high amount of modification. Similarly, regions 187-212, 324-336, 426-442, and 520-525, were found to have a high amount of modification when late-stage glycation was considered. This indicated that a region spanning residues 187-212

could have extensive modification during glycation. In addition, when all of the modification sites were compared to the quantitative and qualitative information obtained for commercially glycated HSA, lysines 199 or 205, 436 or 439, and 525 were consistently identified as modification sites.

In Chapter 6, $^{18}$O-labeling and MALDI-TOF MS were used to quantify the extent of modification that occurred on different regions of HSA at different stages of glycation. This attempt was different from prior attempts (Chapter 2) in that a relative comparison with a set of reference $^{16}$O/$^{18}$O ratios was performed, and a *t*-test was used to determine whether these modifications were significantly different or not. When this was done, significant increases in the $^{16}$O/$^{18}$O ratios were clearly evident at both stages of glycation with larger increases being seen for glycated HSA that was incubated with D-glucose for five weeks. All of the previously identified glycation sites were quantified using this approach. Of these previously identified sites, the *N*-terminus was consistently found to have found to have the largest amount of modification followed by lysines 525 and 439. The relative amount of modifications that occurred on lysine 281 and 199 seemed to fluctuate based on the total degree of glycation. This was an unexpected result, because K525 is commonly identified as having the highest amount of glycation on HSA. This strongly indicated that AGE modification plays an important role in the modification of HSA. Several arginine residues were also found to contain significant amounts of modification. The quantitative data obtained in Chapter 6 was used to select the most likely modified peptides, as identified in Chapters 4 and 5, where a decrease in the total number of early glycation products and an increase in the total number of advanced glycation endproducts were found as the total level of glycation for HSA was increased.

## 7.B. FUTURE DIRECTION

There is no clear consensus as to whether HSA glycation that is performed *in vitro* is a good representation of HSA glycation that occurs in the human body during diabetes. The next step for this project would be to apply the same techniques that were used for characterizing modifications on *in vitro* glycated HSA to the analysis of glycated HSA obtained from the serum of diabetic patients. The total level of glycation in these patient samples can be obtained using the fructosamine assay, as described earlier, and the location of these modification sites can be determined by looking at the m/z shifts that are induced by glycation. $^{18}$O-Labeling can also be used to obtain data that can be used for a direct comparison with the HSA samples examined in this dissertation. The analysis of *in vivo* glycated HSA may be complicated, however, by the fact that there may be variations in the sequence of HSA based on ethnic and regional differences. The selection of a suitable control for qualitative and quantitative comparison would therefore need to be examined.

Differences in the modification pattern of *in vitro* versus *in vivo* glycated HSA would most likely be due to the presence of multiple analytes *in vivo*. Future work should therefore examine the effects of these analytes on the glycation of HSA. For instance, HSA plays a major role in the binding and transport of fatty acids throughout the body. The glycation of HSA can therefore be examined in the presence of fatty acids to see how the glycation pattern is altered, and whether the degree of modification that occurs on *in vitro* glycated HSA approaches the degree of modification that is found on HSA that is obtained from diabetics. This study can be expanded to include the effects of physiological concentrations of α-oxaloaldehydes on the glycation of HSA and methods

that can be used to stabilize the concentrations of glucose/α-oxaloaldehydes during HSA glycation. Another interesting approach based on the above principles would be to perform HSA-glucose incubations in the presence of certain drugs to see how the relative extent of modification is affected. This type of study would provide direct information about how the glycation of HSA may affect its drug binding properties.

The efficient examination of mixed modes of HSA glycation, as described in the paragraph above, will require improved sample throughput and data analysis techniques. The sample throughput can be improved by using on-line digestion and performing on-line peptide fraction. For the studies examined in this dissertation, HSA was incubated with a given enzyme (in solution phase) for 18 – 24 hours. The major advantage of using on-line digestion/fractionation is that complete proteolysis and fraction of HSA can be achieved in less than 30 minutes. Isobaric tags for relative and absolute quantitation (iTRAQ) can also be used to improve the sample throughput. This tag can be used to modify the *N*-terminal region of peptides from a proteolytic digest. The major advantage of using this technique is that multiple mass tags are available, allowing for the quantification of peptides from multiple glycated HSA samples. The other advantage of using iTRAQ is that the fragmentation pattern of the tag can be used to perform SRM based quantitation.

Data analysis can be improved by modifying the glycation product identification program (as described in Chapter 3) to include data about the p$K_a$ of a given amino-acid sidechain and include data about local acid-base catalysis on a given lysine residue. The modification search list can be expanded to include other AGEs such as 5-methylimidazol-4-one, C2-imines, 3,4-dihydroxy-2-imidazoline, and various protein

cross-links. By including these AGEs, it would be possible to track the progress of the Maillard reaction at different stages of glycation. Tandem LC-MS/MS combined with 2D affinity enrichment could also be used to improve the quality of glycation product assignment. For instance, a whole serum digestion could be performed followed by the affinity enrichment of modified peptides. The enrichment can be achieved by using antibodies that specifically interact with certain classes of AGEs (i.e., glyoxal, methylglyoxal, or 3-deoxyglucosone based AGEs) or by using phenylboronic acid chromatography. The modified peptides could then be trapped on a RPLC column followed by LC-MS/MS detection.