CHEMICAL OXIDATION OF TRYPTIC DIGESTS TO IMPROVE SEQUENCE COVERAGE IN PEPTIDE MASS FINGERPRINT PROTEIN IDENTIFICATION

A Thesis

by

JESSICA ELAINE LUCAS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2003

Major Subject: Chemistry

CHEMICAL OXIDATION OF TRYPTIC DIGESTS TO IMPROVE SEQUENCE COVERAGE IN PEPTIDE MASS FINGERPRINT PROTEIN IDENTIFICATION

A Thesis

by

JESSICA ELAINE LUCAS

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

David H. Russell (Chair of Committee)	Emile A. Schweikert (Member)
Max D. Summers (Member)	Gyula Vigh (Member)

Emile A. Schweikert
(Head of Department)

December 2003

Major Subject: Chemistry

ABSTRACT

Chemical Oxidation of Tryptic Digests to Improve

Sequence Coverage in Peptide Mass Fingerprint

Protein Identification. (December 2003)

Jessica Elaine Lucas, B.S., University of Texas at Dallas

Chair of Advisory Committee: Dr. David H. Russell

Peptide mass fingerprinting (PMF) of protein digests is a widely-accepted method for protein identification in MS-based proteomic studies. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) is the technique of choice in PMF experiments. The success of protein identification in a PMF experiment is directly related to the amount of amino acid sequence coverage. In an effort to increase the amount of sequence information obtained in a MALDI PMF experiment, performic acid oxidation is performed on tryptic digests of known proteins. Performic acid was chosen as the chemical oxidant due to the ease of use and to the selective oxidation of cysteine, methionine, and tryptophan residues. In experiments performed in our laboratory, performic acid oxidation either increased or did not affect protein sequence coverage in PMF experiments when oxidized tryptic digests were analyzed by MALDI. Negative mode MALDI data were acquired, as well as positive mode MALDI data, due to the enhanced ionization of cysteic acid-containing peptides in negative mode. Furthermore, the confidence in a protein match is increased by observation of mass shifts indicative of cysteine, methionine, and/or tryptophan in oxidized peptide ion signals when comparing

MALDI spectra prior to performic acid oxidation and after oxidation due to the low abundance of these residues in the majority of all known and hypothetical proteins.

DEDICATION

I would like to dedicate this thesis to several people who have influenced me throughout the years. First, there are my parents who encouraged me to keep going to school, and without their love and support for my decisions, I would surely be lost. I would also like to thank Brandon Ruotolo for constantly telling me that I am smart even though I felt dim-witted and for his emotional support when "I lost my marbles". Thanks to my fellow Russell group members, Holly Sawyer, Dr. William (Billy) Russell, and Joseph Morgan who always helped me in their own ways. I must also thank Dr. Peter Barnes and Dr. Brian Williams for their advice during the last, yet challenging, several months here. Of course, I thank Dr. David Russell for giving me "a way out" when I was, unfortunately, too blind and chicken to do it myself. I would like to especially thank Dr. Gyula Vigh for telling me, in a way only he can, that it would be "bloody dumb" if I did not change my current "experimental design" when I was unhappy, and the current "parameters" did not work for me.

ACKNOWLEDGMENTS

I must thank my advisor, Dr. David Russell, for his advice and guidance during my time here. My committee members deserve thanks for their time and suggestions to improve my research. I must thank Dr. Sharon Braunagel for substituting for my defense at short notice. The Russell group members are thoroughly appreciated for answering my research-related questions. I also thank Dr. William (Billy) Russell for acquiring TOF-TOF data for me, and most of all, allowing me to work with him on various research projects.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Proteomics is the comprehensive study of the identity, function, and location of all proteins that comprise a unicellular or multicellular organism.[1] Unlike genomics, where function is inferred through homology, proteomics directly studies the biochemical pathways necessary for life in order to determine biochemical details, such as: protein localization within a cell or tissue, interactions of proteins with substrates or other proteins, and state-specific post-translational modifications.[2] Identification of a protein or a mixture of proteins in a biological sample can be undertaken by several different analytical techniques (e.g., immunoassays, electrophoresis, mass spectrometry). Mass spectrometry (MS) provides the highest degree of selectivity, resolution, and sensitivity for the analysis of complex biological mixtures when compared with other standard methods.[1,2] As such, MS has been applied to the analysis of a wide variety of proteins including lipoproteins[3], glycoproteins[4], and DNA-binding domains[5].

Peptide mass fingerprinting (PMF) is the most common MS-based experiment for protein identification in current research. PMF identification[6-8] involves chemical or enzymatic digestion of a protein or protein mixture to yield peptides characteristic of a specific protein or proteins. The resulting peptides are analyzed by MS as mass over charge (m/z), and the peptide masses are then submitted to an online protein search engine, such as Mascot[9], ProFound[10,11], and MSFit[12], for protein identification.

This thesis follows the style and format of *Analytical Chemistry.*

These search engines "mine" annotated protein databases (e.g., Swiss-Prot[13], NCBI[14]) that contain the amino acid sequences of known and hypothetical proteins from numerous organisms to match the peptide m/z data to theoretical digests. Owing to the large amount of information contained in a protein database, additional experimental conditions can be specified to reduce the number of potential protein matches or "hits". Search parameters that can be limited include: specific organisms of sample origin, m/z error tolerance, post-translational or chemical modifications, and the proteolytic enzyme and/or chemical used in the digestion process. Protein "scores" are included in search results to provide a measure of confidence for a possible protein match. MOWSE, the scoring algorithm used in Mascot PMF searches, is based on the number of matching peptides per 10 kDa of a protein.[7] The extent of percent sequence coverage, which is reported for all possible protein matches, adds to the confidence of a PMF identification in conjunction with the score. In general, 15 to 30 percent sequence coverage has been determined to be the minimal amount necessary for a confident protein identification.[8] High confidence levels in PMF identification relate directly to the percent sequence coverage of a possible match.

The mass spectrometric technique of choice for PMF experiments is matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. MALDI is less sensitive to impurities within a sample that can often interfere with analyte detection in other MS techniques[15]; therefore, MALDI requires less sample preparation and clean up making this technique generally more robust. Ionization and desorption of analytes in MALDI requires the use of matrix, *i.e.*, a molecule that absorbs at the wavelength of the

ionization laser (typically 337 nm produced by a N_2 laser). Theories pertaining to the mechanism of analyte protonation (or deprotonation for $[M-H]^-$) have been proposed by several researchers.[16] For example, excited-state proton transfer (ESPT) has been used to describe the ionization/desorption process. That is, absorption of 337 nm radiation yields an electronically excited matrix molecule having a greater acidity which undergoes H^+ transfer to an analyte.[17] Matrix is typically in excess of 100 to 10,000 fold. The matrices most commonly used are organic acids, such as α-cyano-4-hydroxycinnamic acid (CHCA) and sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid), which are dissolved in organic or aqueous solutions. In a typical "dried droplet" MALDI sample spot, small aliquots of matrix solution and an aqueous sample are mixed in an Eppendorf tube by aspirating and expelling the mixture with a microliter pipet. A microliter of the matrix-sample solution is deposited onto a MALDI sample plate and allowed to dry prior to analysis.

Proteolytic digestion by trypsin produces peptides with C-terminal lysine (K) or arginine (R). A serum protein, for example, on average can generate ~ 1 tryptic peptide for every ~ 1500 Da, and tryptic peptides typically fall within a mass range of 500-4000 Da.[6,18-20] The molecular weight (MW) of a protein, in general, dictates the number of possible tryptic peptides observed (*i.e*, high MW proteins produce more tryptic peptides). Although there are many peptides generated from digestion, some peptides that are deposited in a MALDI spot may not ionize and thus not detected. There have been several reports that have attempted to identify the most important causal factors for these "analyte suppression effects." For example, hydrophobicity[21], acid-base

character[22] and sample morphology[23] have all been identified as potential causal factors for peptide ion suppression. In regards to acid-base character of an analyte in MALDI, an analyte with high basicity will, most often, be observed in a MALDI spectrum.[22]

Although an active area of research, analyte suppression is still a significant limitation to effective complex mixture analysis by mass spectrometry. Suppression effects are most apparent in the highly complex mixtures commonly encountered in proteomics experiments, *e.g.*, mixtures with greater than 10 protein components. Therefore, methods developed for complex peptide mixtures are optimized in terms of sensitivity/limit of detection to minimize the influence of ion suppression effects. Thus, even though the minimal requirement for protein identification, as determined by Jensen and co-workers, is 15 to 30% coverage[8], complex mixture analysis requires that the maximum number of peptides be detected.[24]

In an effort to enhance ionization of peptides in MALDI and thereby attain maximum percent coverage of a protein, covalent modifications of peptides and altering the solvent system of the matrix/sample solution have been studied by several researchers. Arginine-terminated peptides ionize efficiently in MALDI and appear to suppress ionization of lysine-terminated peptides.[21,25] Guanidation of lysine-terminated peptides, for example, was shown to increase the efficiency of ionization by increasing the basicity of lysine to be comparable to that of arginine-terminated peptides, thus allowing a higher probability of detection.[26] In addition, since a lysine-terminated peptide is modified to cause a shift in mass, the presence of lysine in the peptide can be assumed and used in refinement of a protein database search.[26] However, guanidation

experiments require sample preparation steps which degrade the throughput of the method and results in loss of peptide ion signal. MALDI spectra have also been shown to be sensitive to the solvent and additives used in matrix/sample preparation.[27,28] Previous experiments performed in our laboratory indicate that MALDI ion signals are influenced by the polarity of the solvent system used in matrix/sample preparations in addition to what, if any, ion-pairing agent is added to the matrix/sample solution.[28] Thereby, the extent of interaction between matrix and analyte molecules is directly influenced by matrix/sample solvent composition.[28]

Selective oxidation of amino acids cysteine, methionine, and tryptophan has previously been utilized for protein denaturation and characterization.[29-35] Cystine oxidation disrupts the disulfide bonds within a protein by cleaving the disulfide and results in two cysteic acid residues. Free cysteine residues are oxidized as well. The pK_a of cysteic acid is 1.89[36] versus that of cysteine, 8.33[36]; thereby dramatically altering the acidic character of the modified peptide. Oxidation of methionine residues in peptides to methionine sulfoxide (addition of one oxygen) has been used by Gevaert, et al. to enhance the selectivity in RP-HPLC separations of protein digests, by increasing the hydrophilicity of the modified peptide.[37] The complete oxidation of methionine results in methionine sulfone which further increases the hydrophilicity of the peptide. Products arising from hydrogen peroxide oxidation of free tryptophan amino acid and tryptophan-containing peptides have also been characterized in RP-HPLC separations.[38] The hydrophilicity of singly- and doubly-oxidized species is higher than that of native tryptophan. Methionine, tryptophan, and cysteine oxidation have been applied recently

in biological mass spectrometry as well.[32-35] Fast atom bombardment mass spectrometry (FAB MS) spectra of peptides containing methionine and tryptophan residues before and after chemical oxidation were compared by Wagner and Fraser (1987).[33] Methionine and tryptophan both incorporated one oxygen atom to cause a shift in m/z; therefore, a 16 Da mass shift indicates the presence of either methionine or tryptophan after oxidation.

Amino acid residues can be oxidized by a variety of chemical reagents. DMSO (Dimethylsulfoxide)/HCl/acetic acid[33,35,38], Fe(III)/ascorbic acid/O_2[38], hydrogen peroxide[37,38] and peracids[29-32,38] are a few reagents commonly used for oxidation. Unfortunately, performing oxidations with DMSO, for example, only leads to partial oxidation of methionine and/or tryptophan[33] and DMSO has been shown to degrade MALDI spectral quality.[15] During performic acid oxidation[29-32], cysteine (or cystine) incorporates three (or six) atoms of oxygen (Figure 1A), methionine acquires two atoms of oxygen to become methionine sulfone (Figure 1B), and tryptophan, shown in Figure 2, oxidizes by incorporation of one, two, or three atoms of oxygen, β-(3-oxyindolyl)alanine, N-formylkynurenine, and hydroxy-N-formylkynurenine, respectively. In addition, cysteine, methionine, and tryptophan have a low frequency of occurrence in proteins: cysteine abundance ~ 1.75%; methionine abundance ~ 2.25%; tryptophan abundance ~ 1.25%.[39] Thus, the selectivity of performic acid oxidation can be used to refine PMF identifications, particularly in instances where a PMF query returns several possible protein matches. To verify the sites of oxidation accessible to performic acid oxidation, a set of preliminary studies were performed in our laboratory.

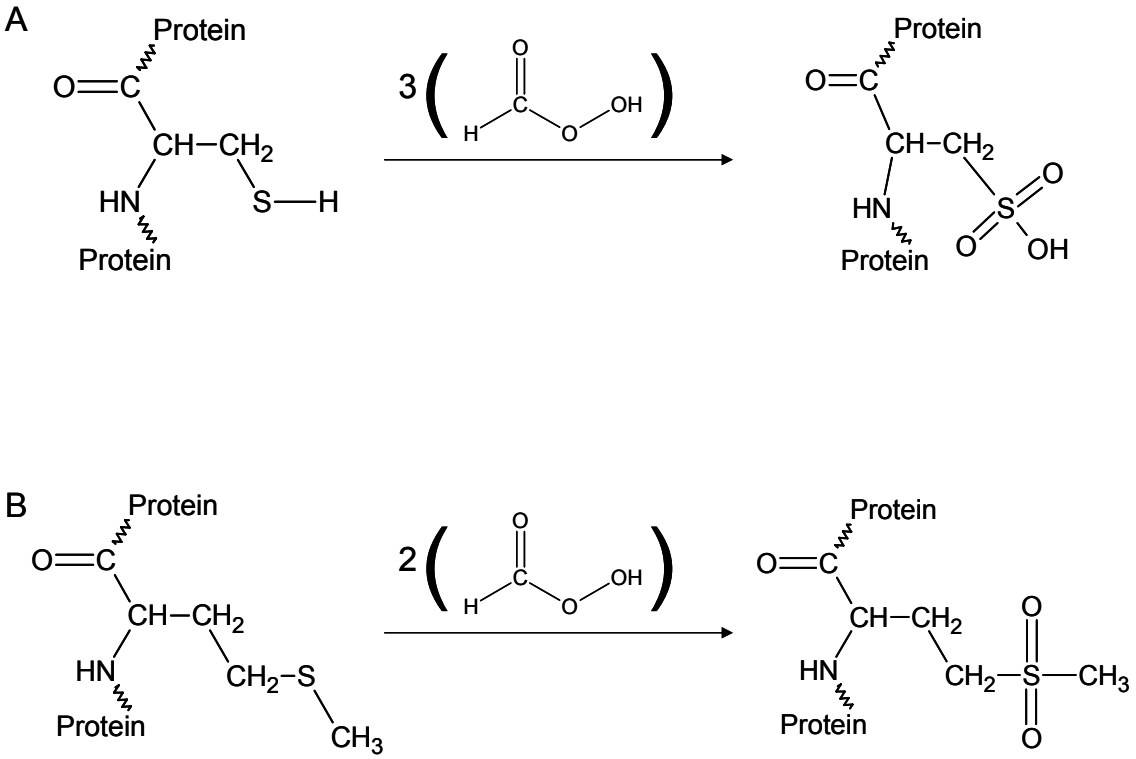

Figure 1. Effect of performic acid on cysteine and methionine. A) Oxidation of cysteine to cysteic acid using performic acid. B) Oxidation of methionine to methionine sulfone using performic acid.

Three peptides, KPVTDAENCHLAR, AMTNLR, and TSDANINWNNLK, were isolated using tandem mass spectrometry from a MALDI tryptic digest sample of transferrin. TOF-TOF data for three peaks corresponding to oxidized peptides from an oxidized transferrin tryptic digest are shown in Figures 3-5. Figure 3 shows the fragmentation of KPVTDAENC(O_3)HLAR; Figure 4 shows AM(O_2)TNLR; Figure 5 shows TSDANINW(O)NNLK. In Figure 3, the mass difference between y_5 ([C(O_3)HLAR+H]$^+$) and y_4 ([HLAR+H]$^+$) corresponds to the mass of a cysteic acid residue (151 Da). The fragment ions corresponding to the loss of methionine sulfone are not observed in Figure 4, although with the partial sequence information generated from y-ion and b-ion series, the site of oxidation can be limited to methionine or alanine. Under performic acid oxidation, alanine is not oxidized; this is illustrated by Figure 3 where the mass difference between y_8 ([AENC(O_3)HLAR+H]$^+$) and y_7 ([ENC(O_3)HLAR+H]$^+$) corresponds to the residual mass of alanine (71 Da). Therefore, methionine is the only amino acid residue oxidized by performic acid oxidation in AMTNLR. In Figure 5, the mass difference between y_5 ([W(O)NNLK+H]$^+$) and y_4 ([NNLK+H]$^+$) equals the mass of a β-(3-oxyindolyl)alanine residue. MS/MS data indicates that the specific sites of oxidation is cysteine, methionine, and tryptophan; hence, changes in mass of a peptide containing one or more of an oxidized residue can provide amino acid composition information for a tryptic peptide. That is, the additional knowledge of the presence of cysteine, methionine, and/or tryptophan, each of which have a low frequency of occurrence in known and hypothetical proteins, greatly enhances the confidence of proteins identified by PMF experiments.

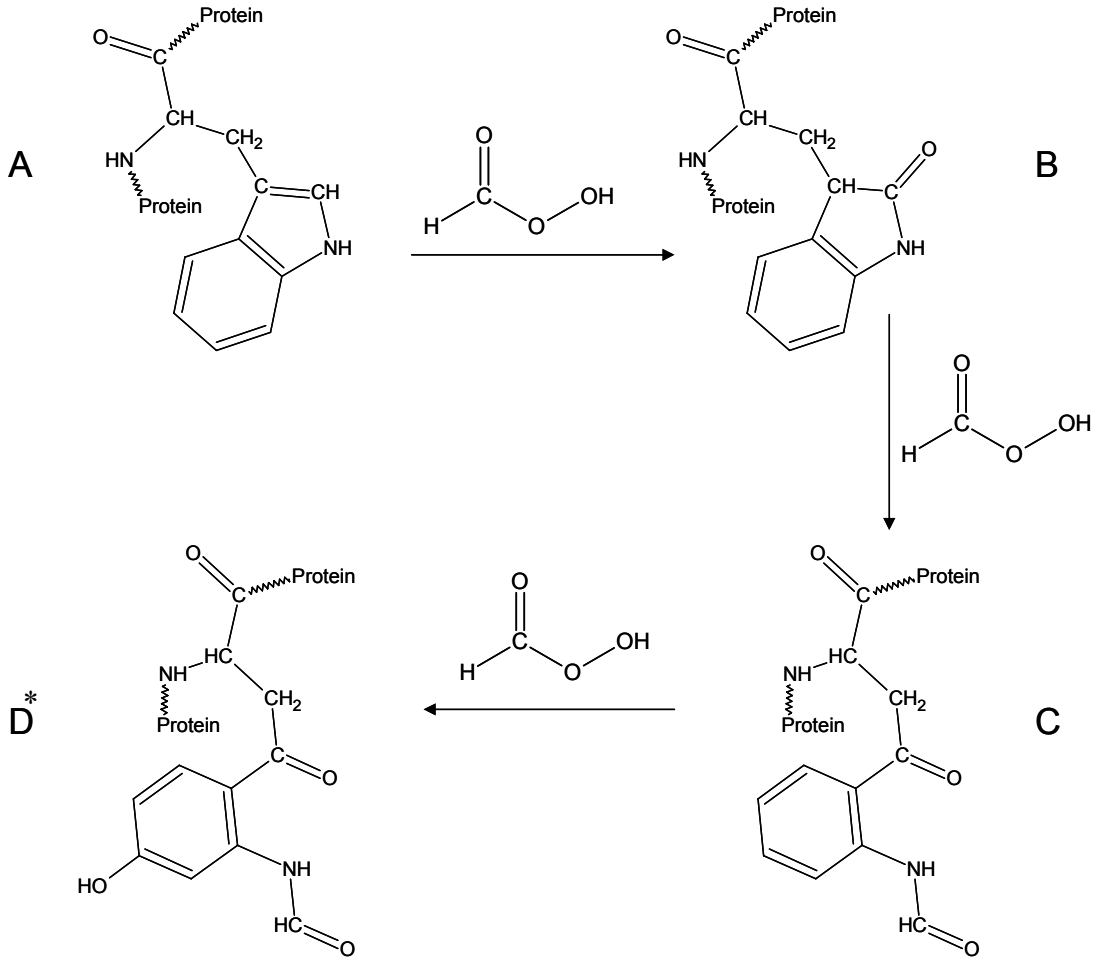

Figure 2. Oxidation of tryptophan with performic acid. A) Tryptophan residue in unoxidized state; B) Singly-oxidized tryptophan, β-(3-oxyindolyl)alanine; C) Double-oxidized tryptophan, N-formylkynurenine; D) Triply-oxidized tryptophan, hydroxy-N-formylkynurenine. *Position of phenyl hydroxyl group variable.

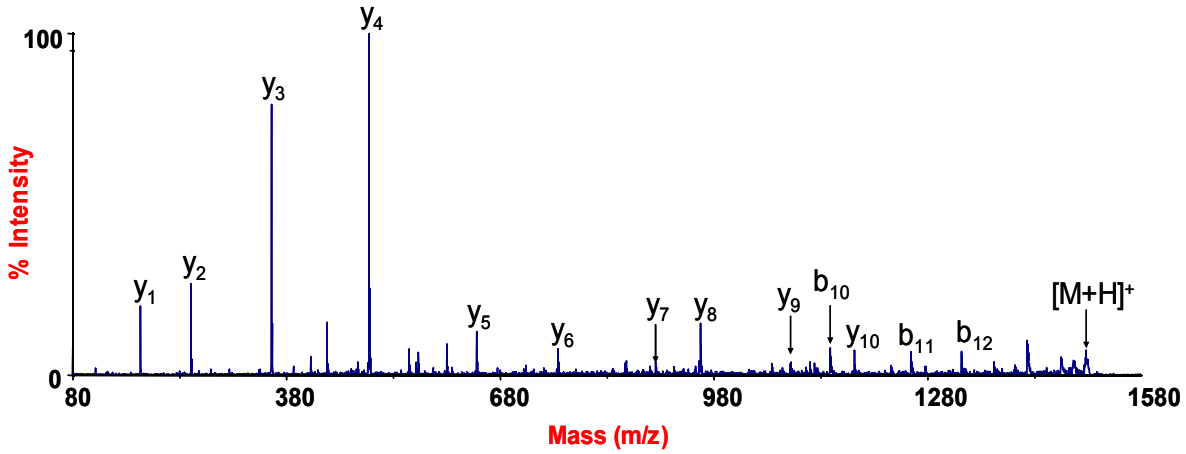

Figure 3. TOF-TOF spectrum of the peptide sequence KPVTDAENC(O_3)HLAR.

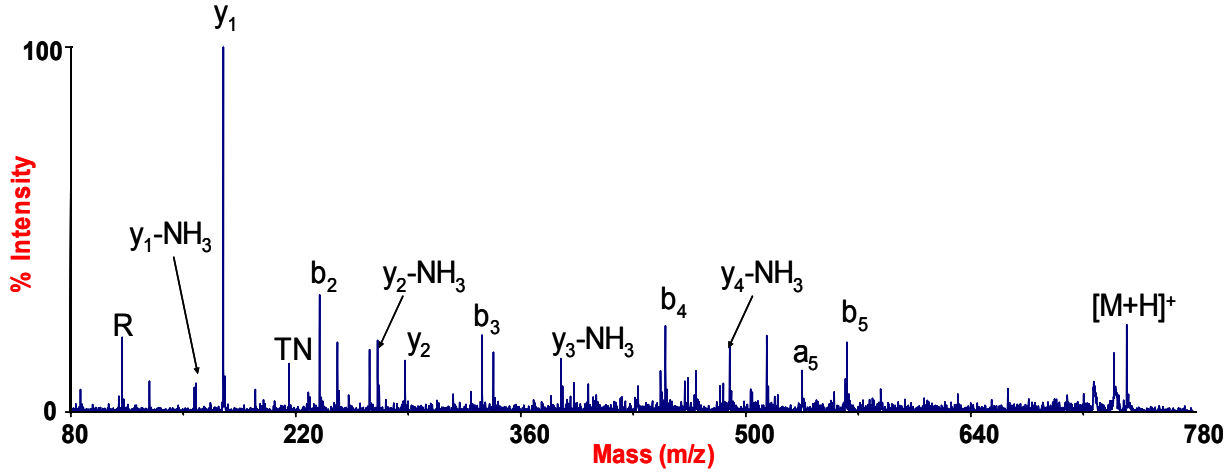

Figure 4. TOF-TOF spectrum of the peptide sequence AM(O_2)TNLR.

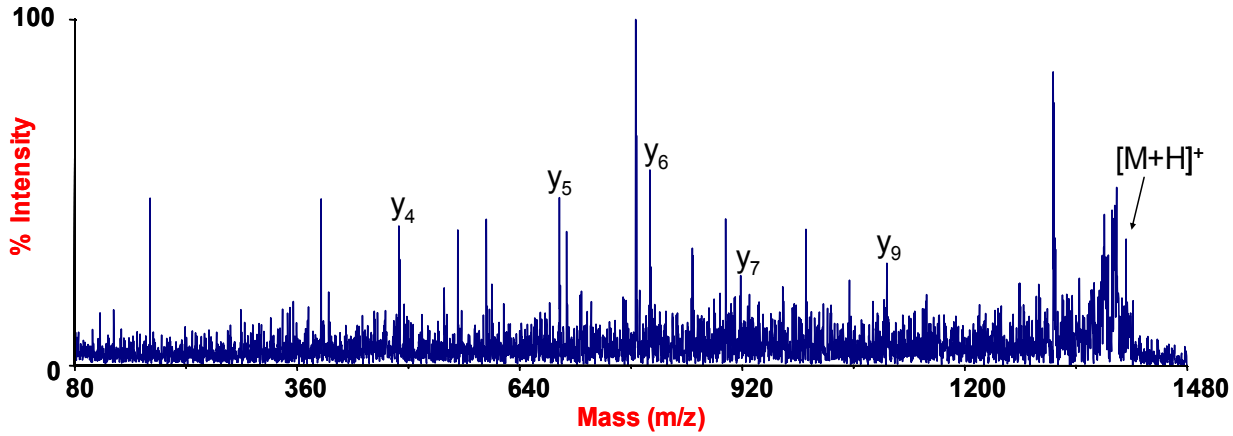

Figure 5. TOF-TOF spectrum of the peptide sequence TSDANINW(O)NNLK.

Here, oxidation, using performic acid, is assessed to determine if percent coverage in PMF identification yields a higher confidence level. Performic acid was chosen as the oxidant due to its high vapor pressure, ease of use, lack of additional sample preparation prior to MALDI analysis to remove any contaminants, and the relative amino acid specificity when compared with other methods of chemical oxidation. Although data obtained in negative mode MALDI are typically not used in PMF identification, negative ions are included in the PMF identification experiments to determine the effect of oxidation on peptide ions observed in negative mode MALDI. With the ability to observe peptides before and after oxidation, additional information about the primary structure of a peptide is inferred, thus adding to the confidence level of protein identification and the ability to refine PMF protein identifications. Further development of a high-throughput approach using "on-target" oxidation[32] allows MALDI analysis of the same dried-droplet sample spot with similar results obtained using the "in-solution" oxidation.

CHAPTER II

EXPERIMENTAL

SAMPLE PREPARATION

Protein Denaturation and Tryptic Digestion

Commercially available proteins (Table 1) were obtained from Sigma-Aldrich and stock solutions were made for each with concentrations varying from 1 to 6 mg/mL distilled and deionized (ddI) H_2O (Millipore, in-house). Beta casein utilized in these studies contains phosphorylated sites; alpha casein was purchased in the dephosphorylated form, and all other proteins are assumed to contain no post-translational modifications. Molarity of protein stock solutions was calculated, and the volume corresponding to 200 pmol of each bulk protein was transferred to 500 mL Eppendorf vials. 50 mM NH_4HCO_3 was added to each 200 pmol protein solution to reach a 50 μL total volume. For thermal denaturation according to Park et al, protein vials were heated in a 90°C water bath for 20 minutes and then quickly transferred to the freezer for 5 minutes.[40]

Sequencing-grade trypsin (Promega) was diluted in 50 mM acetic acid to a concentration of 0.05 μg/mL. The optimal amount of trypsin for digestion of each protein was calculated based on the 1:50 (g trypsin: g protein) ratio. The respective amount of trypsin was added to each protein solution, and the vials were placed in a 37°C water bath and allowed to digest overnight (~14 h). Trypsin activity was quenched

Table 1. Proteins used and their properties.

Abbreviation	Protein	Source	MW (kDa)	Oxidizable Residues			Correction
				C	M	W	
AC-S1	Alpha Casein-Subunit One	Bos taurus	22.9	0	5	2	+7
AC-S2	Alpha Casein-Subunit Two		24.3	2	4	2	+7
ADH	Alcohol Dehydrogenase I	Saccharomyces cerevisiae	36.7	8	6	5	+19
Ald	Aldolase A	Oryctolagus cuniculus	39.2	7	3	2	+0
BC	Beta Casein	Bos taurus	23.5	0	6	1	+54
BSA	Serum Albumin	Bos taurus	66.4	34	4	2	+4
CA	Carbonic Anhydrase II	Bos taurus	29.0	0	3	7	+0
CytC	Cytochrome C	Equus caballus	12.4	2	2	1	+0
Hemo-α	Hemoglobin-Alpha Chain	Bos taurus	15.1	0	2	2	+0
Hemo-β	Hemoglobin-Beta Chain		15.9	2	6	4	+0
Lys	Lysozyme C	Gallus gallus	14.3	8	2	6	+12
Myo	Myoglobin	Equus caballus	16.9	0	2	2	+0
Ova	Ovalbumin	Gallus gallus	44.4	6	16	3	+0
Phos	Glycogen Phosphorylase	Oryctolagus cuniculus	97.1	9	21	12	+0
Ribo	Ribonuclease A	Bos taurus	13.7	8	4	0	+17
TF	Transferrin	Bos taurus	75.8	35	9	8	+3

by placing the protein vial in the freezer for 20 minutes. 2 μL of each sample was used for MALDI MS analysis.

Oxidation

Performic acid was prepared based on the method described by Hirs.[31] Formic acid (88%, EM Science) and hydrogen peroxide (30%, EM Science) were mixed 32:1 (v/v) in a 1.5 mL Eppendorf tube and allowed to "age" for 120 minutes at room temperature. Half of the digested protein samples (25 μL) were evaporated in a speed-vac to dryness, then reconstituted with 50 μL 88% formic acid and 10 μL methanol (EM Science, HPLC grade). All samples were then placed in a freezer at -8°C for 30 minutes. 40 μL of performic acid, in excess, was added to each sample and then returned to -8°C for 180 minutes to oxidize. Immediately after the incubation period, ~ 0.5 mL of cold ddI water was added to each sample, and all samples were evaporated to dryness in a speed-vac to halt oxidation. Protein samples were then reconstituted with 25 μL of ddI water for MALDI analysis.

For "on-target" oxidation, performic acid was made in a 1:9 (v/v) of 30% hydrogen peroxide and 88% formic acid respectively. This ratio was chosen for "on-target" oxidation due to the higher concentration of performic acid formed (~1 μmol/μL) and the length of the incubation period (~10 min) stemming from evaporation of the oxidizing solution. "Aged" performic acid was mixed in methanol (EM Science, HPLC grade) at ratios of 1:1, 2:1, 3:1 (v/v) respectively and deposited on a dried-droplet tryptic digest of thermally-denatured BSA or transferrin. Other ratios were made of performic acid, acetonitrile (EM Science, HPLC grade), and methanol at 1:1:1, 2:2:3, 1:2:1, 1:2:2

(v/v/v) respectively prior to deposition onto a dried-droplet tryptic digest MALDI sample. Another procedure of first applying 1:1 (v/v) methanol and acetonitrile to the dried-droplet sample and then adding 1 μL of performic acid to the droplet was also utilized for the studies reported in Chapter VI.

Protein Digest Mixtures

Protein digest mixtures were made with five, ten, and fifteen protein components. Unused portions of thermally denatured protein digest samples used in Chapter IV were used to make multiple-component protein digest mixtures. Unoxidized and oxidized protein digests were added at 40 pmol protein content (exceptions being the subunits of alpha casein at 20 pmol) to have a total protein content of 200, 400, and 600 pmol for the 5, 10 and 15 component digest mixtures, respectively. The total protein molarity of 4 μM was kept constant for each mixture. One unoxidized and oxidized mixture was made for each of the five, ten, and fifteen protein digest mixtures.

MALDI Sample Preparation

For MALDI analysis of the samples, 2 μL of each tryptic digest was mixed in separate vials with 2 μL of ddI water and 4 μL of matrix solution (10-25 mg α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich, recrystallized) in 1 mL methanol (EM Science, HPLC grade)). 1 μL, equivalent to 1 pmol of digested protein, of each sample was then deposited on a stainless-steel MALDI sample stage (Applied Biosystems) and allowed to crystallize at room temperature. All samples were analyzed in both positive and negative modes (*i.e.*, both positive and negative ions produced by the MALDI event were detected in separate experiments).

INSTRUMENTATION

MALDI Mass Spectrometry

A commercially available matrix-assisted laser desorption/ionization (MALDI) delayed-extraction reflected time-of-flight (DE-RTOF) mass spectrometer, Voyager STR (Applied Biosystems), was used for PMF experiments in positive and negative ion modes. Extraction delay time was kept in the range of 150-165 ns for optimum resolution and sensitivity. Grid voltage was set at 68% of the accelerating voltage (± 20 kV). Laser power was adjusted within the range of 2100-2400 arbitrary units (a.u.) for positive mode and 2200-2500 (a.u.) for negative mode. Mass range was 500-5000 m/z for peptide detection. The matrix used was α-cyano-4-hydroxycinnamic acid (CHCA) in concentrations of 10-20 mg/mL methanol (HPLC grade).

All spectra were averages of 1000 to 1500 laser shots at a frequency of 20 Hz. Instrument calibration was performed externally in both polarities using the monoisotopic masses of des-Arg_1-bradykinin ($[M+H]^+$ = 904.4681; $[M-H]^-$ = 902.4525), angiotensin I ($[M+H]^+$ = 1296.6853; $[M-H]^-$ = 1294.6697 m/z), glu_1-fibrinopeptide B ($[M+H]^+$ = 1570.6774; $[M-H]^-$ = 1568.6618), adrenocorticotropin hormone 1-17 ($[M+H]^+$ = 2093.0867; $[M-H]^-$ = 2091.0711), and adrenocorticotropin hormone 18-39 ($[M+H]^+$ = 2465.1989; $[M-H]^-$ = 2463.1833).

In Chapter V, internal calibration was performed using femtomolar amounts of des-Arg_1-bradykinin, angiotensin I, glu_1-fibrinopeptide B, and adrenocorticotropin 1-17. In MALDI spectra where the internal calibrants were not observed, positive mode was calibrated with one low m/z monoisotopic peak: protonated CHCA matrix (trimer)

($[M+H]^+$ = 568.1356), sodiated and hydrated CHCA matrix (tetramer) with 3 Na and H_2O ($[M+H]^+$ = 841.1347), or trypsin auto-cleavage peptide (100-107) VATVSLPR ($[M+H]^+$ = 842.5100), and one high m/z monoisotopic peak: α-hemoglobin chain tryptic peptide (17-31) VGGHAAEYGAEALER ($[M+H]^+$ = 1529.7342), phosphorylase tryptic peptide (256-269) DFNVGGYIQAVLDR ($[M+H]^+$ = 1566.7910), or phosphorylase tryptic peptide (507-519) IGEEYISDLDQLR ($[M+H]^+$ = 1550.7696). Negative mode was calibrated with CHCA matrix (trimer) with K ($[M-H]^-$ = 604.0759) and transferrin tryptic oxidized peptide (595-607) KPVTDAENC(O_3)HLAR ($[M-H]^-$ = 1499.6907) or α-hemoglobin tryptic peptide (17-31) VGGHAAEYGAEALER ($[M-H]^-$ = 1527.7186). MALDI data was manipulated with Data Explorer software (Applied Biosystems).

Tandem MS data was generated with an Applied Biosystems 4700 Proteomics Analyzer (TOF-TOF). Positive mode data was acquired with a 200 Hz Nd-YAG laser at 355 nm; 500 laser pulses were averaged for each spectrum. CHCA at 10 mg/mL methanol (EM Science, HPLC grade) was mixed 1:1 with the sample and deposited onto a stainless-steel MALDI sample plate in dried-droplet fashion. Collision-induced dissociation (CID) was done using air as the collision gas.

PROTEIN DATABASE SEARCHING

MALDI data was submitted to Mascot and/or ProFound protein database search engines. ProFound was utilized primarily for oxidized samples due to the option available for user-defined chemical modifications. Monoisotopic masses of all data were queried with a mass error of ± 150 ppm and allowed up to one tryptic missed cleavage. Taxonomic categories specified in the PMF queries were either “Other

mammalia" or "Other chordata" depending on the origin of the protein digest data. Negative ion data were first converted to [M] in Excel by addition of 1.0078 Da for hydrogen prior to submission due to the absence of an $[M-H]^-$ option for queried data. Percent coverage is automatically calculated by the search engine, although percent coverage was "corrected" for instances where the amino acid sequence encoded for the "pro-protein" indicating the signal peptide was incorporated into the overall percent coverage by the search engine. Percent coverage was also corrected for proteins whose tryptic peptides, with zero missed cleavages, fell out of the m/z 500-5000 data range. The total correction factor for each protein is listed in Table 1. For example, beta casein (BC in Table 1) has a total correction factor of +54% due to tryptic peptides that fall out of the mass range (m/z 500-5000) and due to a signal peptide. Two tryptic peptides of BC have a mass above 5000 Da that together contain a total of 105 amino acids out of the entire sequence of 224 amino acids in BC which is calculated as 47%. In addition, BC contains a signal sequence of 15 amino acids that is cleaved off in expressed protein; therefore, the total correction factor of BC is 47% + 7% = 54%.

All tabulated data shown is calculated from the three trials (N =3) performed for each experimental sample set. The standard deviation from the calculated average of each sample set shows the variation in the three experimental trials. Due to the small number of trials performed (N = 3), the sample average and deviation may not be the optimal estimators of the true percent coverage for each protein sample.

CHAPTER III

PERCENT SEQUENCE COVERAGE WITHOUT THERMAL DENATURATION OF PROTEINS AND IN-SOLUTION OXIDATION

BACKGROUND

Oxidation has been utilized as a method of protein denaturation, primarily with respect to disulfide bond cleavage[29,31], although methionine oxidation also alters hydrophobicity.[37] Previous studies have utilized oxidation in fast-atom bombardment mass spectrometry (FAB-MS) to identify sites of disulfide bonds (or cystines) in proteins. Van de Weert, et al., observed oxidized tryptophan-containing peptides in a MALDI experiment and noted the increased signal intensity of the peptide after the tryptophan oxidation.[35] Extensive oxidation, performed by Sharp, et al., utilizes chemically-generated hydroxyl radicals for aggressive and non-specific oxidation of proteins expressed on the surface of a cell.[41] Specific amino acid residues become oxidized under performic acid oxidation; tryptophan, methionine, and cysteine residues become oxidized.

Performic acid oxidation of tryptic peptides may increase overall sequence coverage by altering the ionization efficiency of the oxidized peptide. Although thermal denaturation of proteins prior to digestion greatly facilitates trypsin activity, performic acid oxidation of a tryptic protein digest, alone, is considered in this study of PMF protein identification utility.

RESULTS AND DISCUSSION

In general, the positive ion MALDI spectra for both oxidized and un-oxidized tryptic protein digests have a similar appearance (Figure 6). The peptide ion with the highest intensity signal remains the same in positive ion mass spectra before and after performic acid treatment. The relative peptide ion signal abundance of other peptides also approximately remain unchanged before and after performic acid oxidation. The effect of oxidation on tryptophan-containing peptides is illustrated by data shown in Figure 7. Simply by visual inspection of an oxidized tryptic digest MALDI spectrum, peaks corresponding to peptides with tryptophan can be identified. The first peak in the triplet of peaks observed from oxidization of tryptophan corresponds to the singly-oxidized species (Δ m/z = +15.99); the second peak is doubly-oxidized tryptophan (Δ m/z = +31.99); the third is the triply-oxidized species (Δ m/z = +47.99). In every oxidized tryptic digest in this study, the singly-oxidized tryptophan peptide peak is the most dominant of the three oxidized tryptophan peaks. The ion signal intensity corresponding to the doubly-oxidized species ranges from 20 to 85% of the singly-oxidized species; the third, triply-oxidized tryptophan peak can be completely absent in the spectrum or comprise up to 70% intensity relative to the singly-oxidized peak. In the study reported by Van der Waart et al., the ion signal intensities of peptides with oxidized tryptophan were greater versus the ion signal intensities of the unoxidized tryptophan-containing peptides.[35] From observations reported in this laboratory, peptides containing unoxidized tryptophan typically had higher relative signal intensity

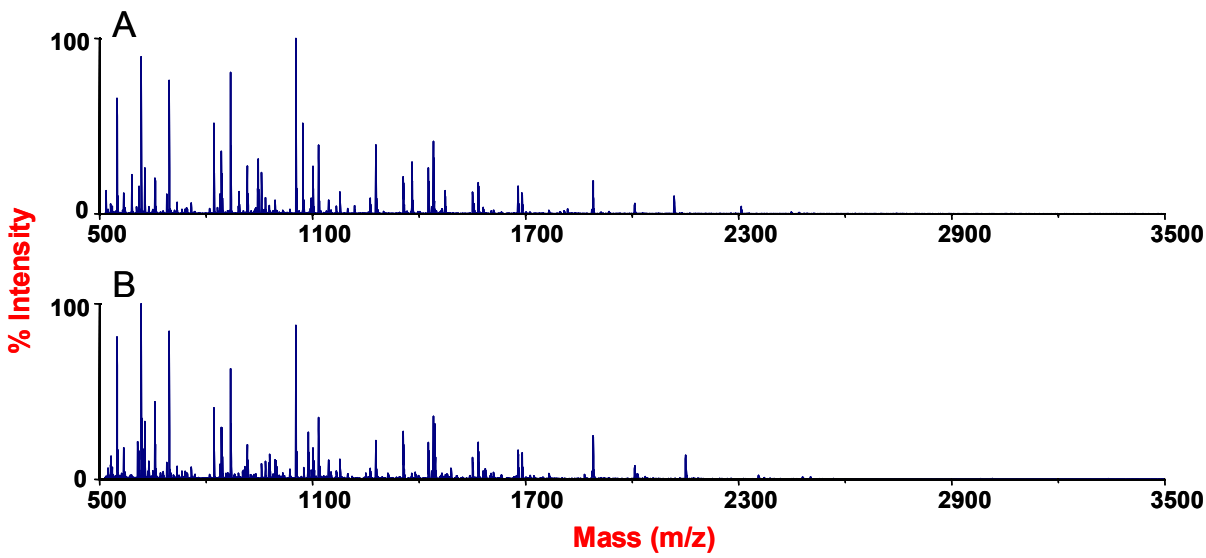

Figure 6. Positive mode MALDI spectra of Phos before (A) and after (B) in-solution performic acid oxidation.

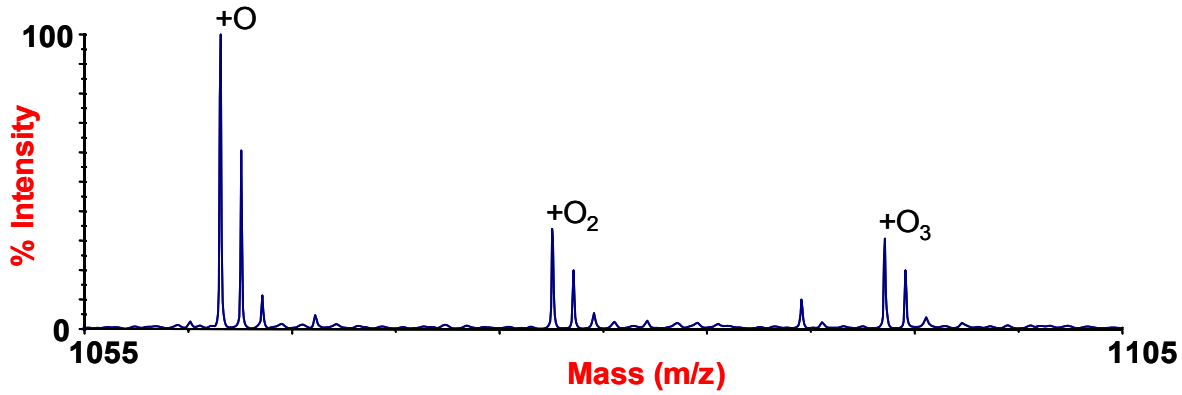

Figure 7. Positive mode MALDI spectra of Lys tryptic peptide GTDVQAWIR after in-solution performic acid oxidation.

than when oxidized. Although MALDI ion signal cannot be considered a quantitative measure of the peptides deposited on the sample plate[42], the probable reason for the decline in signal intensity is due to one peptide species being altered into three oxidized forms with separate ion signals, similar to the decrease in ESI signal observed due to multiple charge states. For the majority of protein digests studied here, more tryptophan-containing peptides were observed in unoxidized positive mode MALDI data.

The effect of performic acid oxidation of methionine residues in peptides appears to enhance the number of methionine sulfone-containing peptides detected in positive mode MALDI data. When comparing relative signal intensity of peaks corresponding to peptides with methionine residues, the respective oxidized species ion signal was similar before and after oxidation in both polarities. Adjacent peaks corresponding to unoxidizable peptides were used to gauge any change in relative signal intensity of the methionine-containing peptides. Despite no relative change in ion signal, more peptides containing methionine are observed in positive mode MALDI data after oxidation. An explanation for the observance of more methionine-containing peptides after oxidation could be due to the change in chemical properties of the sample mixture. The more dominant ion signal, in general, corresponded to peptides with no oxidizable residues before and after oxidation in positive mode MALDI data, although after oxidation, there was a slight decrease in signal intensity. Peptide-peptide suppression effects[22] that are caused by the peptides with high ion signal intensity (*i.e.*, peptides having high basicity) may have lessened after oxidation.

Cysteic acid-containing peptides, as well as cysteine-containing peptides, can be observed in positive ion MALDI spectra although not as often when compared to methionine sulfone-containing peptides. When PMF analysis was performed on unoxidized samples, a few small peptides (m/z 600-1000) contained cysteine. Very rarely was a larger peptide (m/z 1500-3500) with a cysteine residue detected. The lack of detection of peptides with cysteine residues could stem from the low efficiency of tryptic digestion due to non-denatured proteins. After oxidation, cysteic acid-peptides were observed in positive mode although some ions were sodiated and would not be identified in PMF identification unless the $[M+Na]^+$ ion was specified in the protein search parameters. Overall, the maximum detection of cysteic acid-containing peptide ions occurred when data acquisition was performed in negative mode MALDI.

PMF experiments for the majority of the proteins used yielded a percent coverage greater than 30% in unoxidized, positive mode as represented in Table 2. Ovalbumin, ribonuclease A, and carbonic anhydrase tryptic digests analyzed by MALDI had low percent coverage due to the need to denature the protein prior to digestion. Park et al. clearly observed this trend in prior experiments.[43] After undergoing performic acid oxidation, percent coverage in positive mode MALDI for each protein was within to slightly lower of the unoxidized percent coverage range with carbonic anhydrase being the exception.

Table 2. Average percent coverage and standard deviation for non-denatured protein samples digested with trypsin.

Protein	Percent Sequence Coverage*,♣			
	Positive	Negative	Oxidized Positive	Oxidized Negative
AC-S1	59 ± 4	62 ± 6	54 ± 13	55 ± 12
AC-S2	61 ± 6	30 ± 9	37 ± 13	19 ± 9
ADH	50 ± 1	48 ± 1	46 ± 6	50 ± 0
Ald	44 ± 2	27 ± 2	34 ± 11	36 ± 11
BC	86 ± 2	73 ± 3	81 ± 4	74 ± 3
BSA	38 ± 2	24 ± 3	43 ± 2	55 ± 7
CA	20 ± 7	10 ± 4	44 ± 9	26 ± 2
CytC	75 ± 2	52 ± 3	71 ± 5	54 ± 5
Hemo-α	95 ± 6	93 ± 4	97 ± 2	55 ± 5
Hemo-β	84 ± 1	83 ± 2	96 ± 2	81 ± 10
Lys	49 ± 6	37 ± 5	43 ± 2	40 ± 3
Myo	89 ± 2	70 ± 7	86 ± 3	62 ± 13
Ova	N/A	N/A	N/A	N/A
Phos	63 ± 2	54 ± 4	57 ± 8	59 ± 2
Ribo	9 ± 5	N/A	N/A	63 ± 4
TF	47 ± 11	28 ± 14	56 ± 9	51 ± 5

* Standard Deviation figured for N=3; ±1σ.

♣ Corrected percent coverage using Correction factors listed in Table 1.

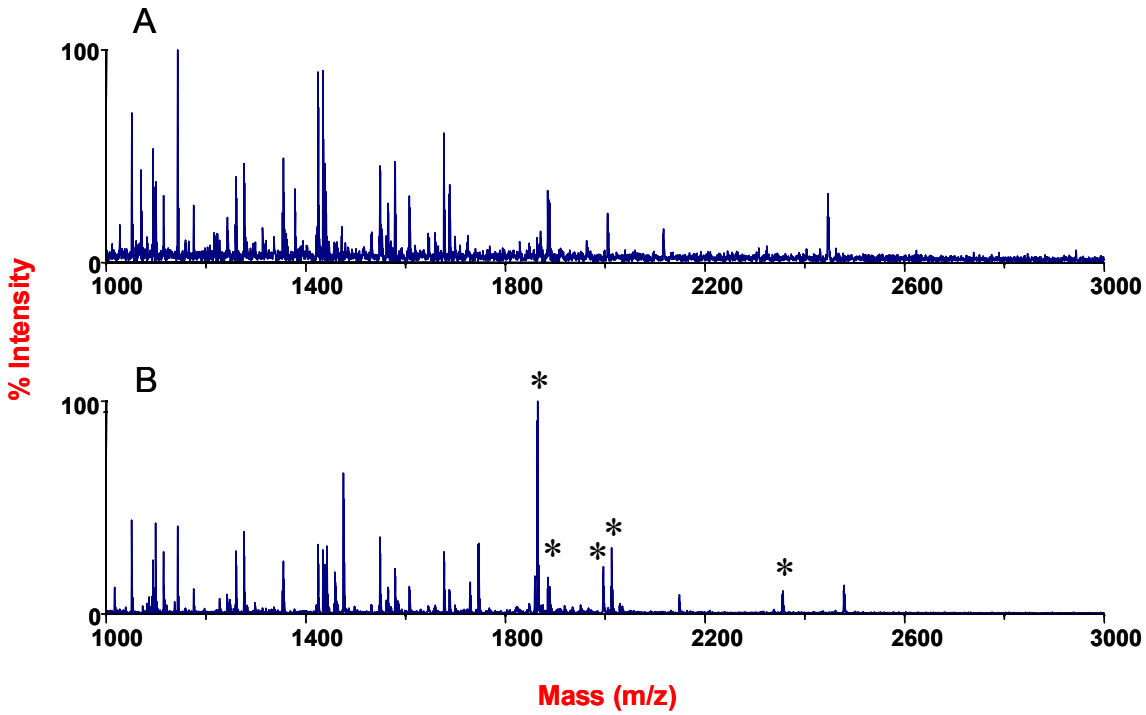

Figure 8. Negative mode MALDI spectra of Phos before (A) and after (B) in-solution performic acid oxidation. *Peaks corresponding to cysteine-containing peptides. No cysteine-containing peptides were identified in spectrum A.

Negative ion spectra for unoxidized samples differed significantly from oxidized samples. After PMF identification, more sequence information was typically obtained from negative mode MALDI of oxidized tryptic digests due to detection of peptides with cysteic acid. Cysteic acid-containing peptides dominate the negative ion spectra due to the high acidic character of the oxidized residue. Representative negative ion spectra of before and after oxidation are shown in Figure 8. Unoxidized, negative ion MALDI data for each protein queried in Mascot had a percent coverage similar to or significantly less than positive ion data. A possible explanation for this observation may be due to the basic C-terminal side-chain of the peptides as the primary charge carrier.[44] Typically, the peptides observed in negative mode are also present in positive mode of unoxidized digests; thus making data acquisition in positive mode MALDI more practical in detection of peptide ions for PMF identification. As indicated in Table 3, combining the data of positive and negative mode MALDI does not increase overall sequence coverage. Combined positive and negative data for oxidized digests were similar for: AC-S1, ADH, BC, CytC, Hemo-α, Myo, and TF, or higher for: Ald, BSA, CA, Hemo-β, Lys, Phos, and Ribo, than unoxidized data combined with the exception of AC-S2 where unoxidized data was slightly higher. For the higher percent coverage gained from oxidized digest data, the higher possible number of cysteine-containing tryptic peptides follow except TF. In cases where percent coverage is similar, the number of possible cysteine-containing peptides is low, with the exception of TF. The total-combined percent coverage (Table 3) indicates that when unoxidized and oxidized data are

Table 3. Combined percent coverage of non-denatured protein digests. Positive and negative MALDI data are combined for unoxidized, oxidized separately, and all data is combined to give a total-combined coverage.

Protein	Combined Percent Coverage*,♣		
	Unoxidized +/-	Oxidized +/-	Total-Combined
AC-S1	70 ± 5	60 ± 9	73 ± 1
AC-S2	61 ± 3	50 ± 2	65 ± 6
ADH	71 ± 1	73 ± 1	75 ± 1
Ald	43 ± 1	59 ± 4	62 ± 6
BC	86 ± 2	84 ± 1	87 ± 0
BSA	38 ± 1	62 ± 6	73 ± 9
CA	22 ± 8	44 ± 9	46 ± 13
CytC	69 ± 2	75 ± 5	76 ± 5
Hemo-α	98 ± 2	97 ± 2	99 ± 0
Hemo-β	85 ± 2	94 ± 5	94 ± 5
Lys	49 ± 6	60 ± 1	71 ± 6
Myo	89 ± 2	86 ± 3	99 ± 2
Ova	N/A	N/A	N/A
Phos	65 ± 2	73 ± 2	76 ± 2
Ribo	N/A	63 ± 4	63 ± 4
TF	52 ± 13	69 ± 4	78 ± 2

* Standard Deviation figured for N=3; ±1σ.

♣ Corrected percent coverage using Correction factors listed in Table 1.

combined both contribute somewhat complimentary sequence information to the sequence coverage of the protein.

CHAPTER IV

PERCENT SEQUENCE COVERAGE WITH THERMAL DENATURATION AND IN-SOLUTION OXIDATION

BACKGROUND

Protein denaturation by chemical or extrinsic means has proved to be helpful in enhancing proteolytic digestion.[45] Inter- and/or intramolecular interactions are disrupted allowing proteases access to internal cleavage sites usually buried within the tertiary or quaternary structure of a native protein. The primary step in increasing overall sequence coverage is to start at the enzymatic digest process. Chemical denaturants (e.g., urea, sodium dodecylsulfate, guanidine HCl) effectively denature proteins in solution although use of these additives can cause problems in down-stream analysis. MALDI has been shown to be tolerable of some chemical additives, yet relatively low concentrations of specific additives (*e.g.*, ionic detergents) can degrade spectral quality.[15] Thermal protein denaturation was proposed as a "mass spectrometry-friendly" approach to protein denaturation as no additives are required.[40,43] Comparing thermally-denatured protein digests to non-denatured digests, the overall percent coverage of a protein increases when thermal denaturation is performed prior to proteolytic digestion, although percent coverage of some proteins decline after thermal denaturation.[43]

Thermal denaturation has been shown to enhance the efficiency of protein enzymatic digestion relative to no increased temperature pretreatment. Performic acid oxidation of non-denatured protein digests has also shown utility in obtaining higher

sequence coverage by altering the chemical nature of the peptides. To compare both methods, tryptic digestion of thermally-denatured proteins before and after oxidation are compared to the results of Chapter III. The combination of thermal denaturation and performic acid oxidation further enhances overall sequence coverage and ultimately leads to higher confidence level protein identifications in PMF experiments.

RESULTS AND DISCUSSION

Thermal denaturation prior to digestion by trypsin aided in increasing the percent sequence coverage (Table 4) of aldolase, carbonic anhydrase, albumin, and myoglobin, and allowed detection of ovalbumin and ribonuclease that were, for the most part, absent in the study done in Chapter III. The secondary/tertiary structure of the aforementioned proteins are compact and highly resistant to tryptic digestion.[45] The remainder of the proteins (AC-S1, AC-S2, ADH, BC, CytC, Hemo-α, Hemo-β, Lys, Phos, TF) yielded similar to lower average percent coverage. There are several potential explanations for the loss of percent coverage for some thermally-denatured proteins. The generation of more tryptic peptides may have elicited peptide-peptide suppression, where a peptide of higher ionization efficiency, not present in the non-denatured sample, causes the suppression of a previously detected peptide that has a lower ionization efficiency.

Combined sequence coverage of positive and negative ion MALDI data, unoxidized, oxidized, and both unoxidized/oxidized together (total-combined), was determined by PMF identification (Table 5). As stated in Chapter III, MALDI data acquired in negative mode contributed little to no additional peptide information to the combined percent sequence coverage when the tryptic digest is not oxidized. Thus, only

Table 4. Average percent coverage and standard deviation for thermally-denatured protein samples digested with trypsin.

Protein	Percent Sequence Coverage*,♣			
	Positive	Negative	Oxidized Positive	Oxidized Negative
AC-S1	68 ± 5	66 ± 2	62 ± 12	63 ± 8
AC-S2	54 ± 13	29 ± 11	45 ± 10	22 ± 7
ADH	54 ± 6	57 ± 1	48 ± 3	55 ± 5
Ald	51 ± 6	48 ± 8	58 ± 11	76 ± 8
BC	87 ± 0	72 ± 2	84 ± 2	73 ± 3
BSA	48 ± 8	38 ± 4	48 ± 9	68 ± 10
CA	71 ± 5	49 ± 18	67 ± 4	42 ± 2
CytC	79 ± 12	25 ± 11	62 ± 1	55 ± 6
Hemo-α	82 ± 14	50 ± 15	80 ± 13	62 ± 14
Hemo-β	69 ± 2	59 ± 4	86 ± 8	60 ± 12
Lys	57 ± 7	45 ± 9	55 ± 1	58 ± 5
Myo	96 ± 5	78 ± 5	85 ± 6	73 ± 6
Ova	49 ± 7	22 ± 9	47 ± 4	44 ± 7
Phos	59 ± 6	49 ± 4	59 ± 5	59 ± 4
Ribo	50 ± 4	39 ± 1	76 ± 4	77 ± 2
TF	66 ± 6	41 ± 3	50 ± 1	64 ± 8

* Standard Deviation figured for N=3; ±1σ.

♣ Corrected percent coverage using Correction factors listed in Table 1.

Table 5. Combined percent coverage of thermally-denatured protein digests. Positive and negative MALDI data are combined for unoxidized, oxidized separately, and all data is combined to give a total-combined coverage.

Protein	Combined Percent Coverage*,♣		
	Unoxidized +/-	Oxidized +/-	Total-Combined
AC-S1	73 ± 0	73 ± 3	74 ± 1
AC-S2	63 ± 6	47 ± 13	66 ± 4
ADH	75 ± 3	71 ± 5	77 ± 2
Ald	64 ± 3	90 ± 2	90 ± 2
BC	87 ± 0	83 ± 1	87 ± 0
BSA	50 ± 9	72 ± 3	78 ± 2
CA	73 ± 2	69 ± 5	75 ± 0
CytC	79 ± 12	70 ± 2	82 ± 6
Hemo-α	99 ± 0	83 ± 11	99 ± 0
Hemo-β	69 ± 2	86 ± 5	86 ± 5
Lys	63 ± 4	67 ± 7	68 ± 6
Myo	99 ± 1	86 ± 3	99 ± 1
Ova	44 ± 6	57 ± 6	69 ± 6
Phos	63 ± 4	70 ± 1	73 ± 1
Ribo	61 ± 13	81 ± 2	84 ± 4
TF	64 ± 2	57 ± 4	74 ± 3

* Standard Deviation figured for N=3; ±1σ.

♣ Corrected percent coverage using Correction factors listed in Table 1.

positive mode MALDI spectra need be acquired when a tryptic digest is unoxidized. On the other hand, the protein information retrieved when combining positive and negative ion data of an oxidized protein tryptic digest increased the overall percent sequence coverage relative to only positive ion data of oxidized samples. In an extreme case, that of transferrin, 80% of the peaks observed from negative mode MALDI of the oxidized digest corresponded to cysteic acid-containing peptides that were not observed in positive mode. Here, oxidized negative mode contributed more peptide information than oxidized positive mode data. The total-combined percent coverage of each protein when compared to that of Chapter III was within the average range for the majority of samples. Ald, CA, Ova, and Ribo, on the other hand, obtained a higher grand percent coverage due to thermal denaturation.

One reason for the enhanced percent coverage observed for oxidized relative to non-oxidized protein digest samples is related to the ability of chemical oxidation to reduce spectral overlap in the case of some isomass tryptic peptides. Figure 9 highlights the case of two near-isomass peptides from ovalbumin. Prior to chemical oxidation, the two sequences of DILNQITKPNDVYSFSLASR ($[M+H]^+$ = 2281.18) and VTEQESKPVQMMYQIGLFR ($[M+H]^+$ = 2284.15) exhibit a high degree of spectral overlap, however, upon performic acid oxidation the two signals are easily distinguishable as VTEQESKPVQMMYQIGLFR contains two methionine residues.

The high standard deviation between similar trials, in both Chapter III and IV, can be attributed to the "spot-to-spot" reproducibility[46] in MALDI spectra. Data generated for each sample in a trial was performed on different days, while instrumental

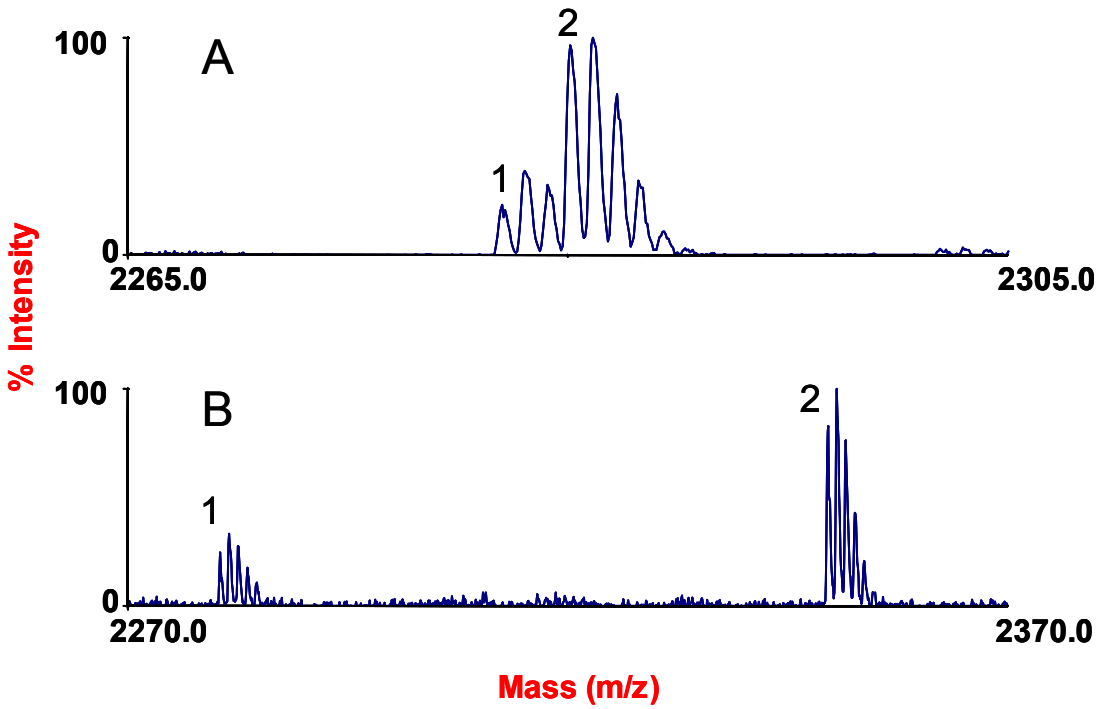

Figure 9. Ova tryptic peptides DILNQITKPNDVYSFSLASR (1) and VTEQESKPVQMMYQIGLFR (2) prior to performic acid oxidation (A) and after performic oxidation (B) in positive mode MALDI spectra.

parameters and sample preparation was kept constant in each sample. Variations in the surface of a MALDI sample plate can influence the uniformity of matrix/sample crystallization which, in turn, can influence the abundance/location of "sweet spots". Sweet spots are regions within a MALDI sample spot that yield spectra with high signal-to-noise of the analyte(s).[46]

Generally, chemical oxidation with performic acid increases the percent coverage for single component protein samples through decreased peptide ionization suppression effects and decreased spectral overlap. As previously noted by Park, et al., thermal denaturation generally increases the number of peptides observed by mass spectrometry, although it is apparently not optimal for all proteins.[43]

CHAPTER V

PERCENT SEQUENCE COVERAGE IN THERMALLY-DENATURED AND IN-SOLUTION OXIDIZED PROTEIN DIGEST MIXTURES

BACKGROUND

Complex biological matrices often contain large amounts (1 µg/mL to 50 mg/mL) of protein.[47] Blood, saliva, urine, etc. are all comprised of many proteins, along with lipids, salts, etc., at different concentrations. This analogy can be extended to a microscopic level with cells and the organelles within the cells. MS-based proteomics is continuously evolving to detect to identity and expression level of proteins in any given biological sample. The scope of proteomics extends further to determine protein-substrate (*e.g.*, protein-protein, protein-DNA, protein-lipid, etc.) interactions, protein post-translational modifications, and phenotype-dependant expression levels.

Previous research by Park et al.,[43] showed that a separation step prior to enzymatic digestion and MALDI analysis is not necessary in some circumstances of protein mixtures. Model protein mixtures made up of nine commercially-available proteins were thermally denatured and then digested with trypsin. MALDI data was then queried in a protein database search engine to identify the protein mixture components by PMF. All nine proteins were identified by PMF with sequence coverage varying from 14 to 48%. The time and data storage space needed to accurately identify each protein component are less than that required for a two-dimensional liquid chromatography-mass spectrometry experiment. Thus, thermally-denatured,

enzymatically-digested protein mixtures followed by MALDI analysis is more amendable for high-throughput analysis and minimal sample volume.

Confidence levels in a PMF experiment can be increased by several parameters. Obtaining high mass accuracy with internal calibrants aids in reducing the number of false positives retrieved from a PMF query.[48] Also, knowledge of the origin of the sample limits protein homologues or related protein-families from other species in being falsely identified. Short of MS-MS analysis for primary structure information, any knowledge of possible residues within a peptide greatly enhances the confidence level of proteins identified by PMF experiments. Performic acid oxidation, specifically, causes mass shifts in peptide ion peaks containing cysteine, methionine, and/or tryptophan. Moreover, the extent of the mass shift indicates what residue(s) is/are present by comparing data from both before and after oxidation; thus, performic acid oxidation gives crucial information when comparing several protein matches returned from a PMF experiment due to the low frequency of abundance of cysteine, methionine, and tryptophan residues. Use of negative polarity in MALDI also increases the confidence of protein identification by yielding cysteic acid-containing peptides that previously might have not been observed in the unoxidized sample which increases the total sequence coverage.

RESULTS AND DISCUSSION

Five Protein Mixture-Myo, Ova, Phos, Ribo, TF

Positive mode MALDI data was utilized in the unoxidized protein digest mixture alone due to the lack of additional information gathered in negative mode for

protein digests in the previous studies (Chapter III and IV). Monoisotopic masses for the observed peaks were submitted to ProFound for PMF identification. Iterative searches can be performed by eliminating masses that correspond to a protein and resubmitting the remaining masses.[48] Even with a mass error of 25 ppm, searches done in ProFound could not be performed by specifying all taxa due to the number of false positives. Due to Ova being from a different taxonomic category than the other four proteins, "other chordata" and "other mammalia" had to be interchanged to identify Ova and the other four proteins respectively.

Using the data obtained in positive mode MALDI of the unoxidized digest mixture, Phos, TF, Ova, and Myo were identified by PMF identification; Ribo was not returned as a possible match. After manual inspection of the observed masses, no peaks corresponding to Ribo tryptic peptides were detected. The oxidized mixture data obtained from positive mode MALDI was queried for PMF identification. Phos, TF, Ova, and Myo were, again, identified. Ribo was not retrieved in positive mode MALDI data after oxidation. With TF as the exception, Phos, Ova, and Myo all gained in percent coverage by 8%, 5%, and 7% respectively. Overall, ~50% more peptides with methionine sulfone were detected, and almost all peptides with tryptophan were not detected in positive mode MALDI of the oxidized sample. This observation indicates that methionine sulfone may aid in the positive mode ionization efficiency of some peptides.

Negative ion data from the oxidized digest, after database searching, yielded different results. All five protein components were identified. Ribo had a percent

coverage of 40% due to the detection of three peptides, all of which, contain cysteic acid. Of all the 56 peaks identified as a peptide by PMF experiments, over 50% of the peptides contained at least one cysteic acid. Peptides identified as containing methionine sulfone or singly-oxidized tryptophan typically contained a cysteic acid residue as well; the presence of a cysteic acid residue possibly aided in the detection of these peptides.

Ten Protein Mixture-AC-S1, AC-S2, Ald, BC, BSA, Myo, Ova, Phos, Ribo, TF

PMF identification with a larger number of proteins became more complicated when using either Mascot or ProFound. High molecular weight proteins, e.g., TF, Phos, and BSA, were favored over the lower weight proteins, e.g., Myo, AC-S1, AC-S2. The large amount of queried masses possibly caused the favoring of high kDa proteins due to the scoring algorithm used in the search engine. To correct for scoring bias, protein mass ranges were specified (in kDa) and increased as iterative searches progressed.

With positive mode MALDI of the unoxidized mixture, eight of the ten proteins were identified using the molecular weight range modification. Ribo and BC were not identified, and peaks corresponding to the tryptic peptides were not observed when manually inspecting the spectrum. Positive mode data from the oxidized sample did not yield a match to Ribo and BC as well. As seen from interpretation of the five protein mixture data, oxidation increased the number of peptides containing methionine roughly 50%, although the number of tryptophan-containing peptides observed remained the same when compared to positive mode MALDI data of the unoxidized sample.

Negative mode MALDI data of the oxidized mixture matched 8 of the 10 proteins by PMF identification. Ribo was identified by detection of two cysteic acid-

containing peptides. Masses corresponding to peptides of AC-S2 and BC were not observed. As observed in the five protein digest mixture, over half of the matched peptides contained cysteic acid, and the majority of peptides observed that contain methionine sulfone also contain cysteic acid.

Fifteen Protein Mixture- AC-S1, AC-S2, Ald, BC, BSA,CA, CytC, Hemo-α, Hemo-β, Lys, Myo, Ova, Phos, Ribo, TF

The complexity of the fifteen protein mixture also warranted limiting the protein mass range while performing iterative PMF identification queries. Positive mode MALDI data of the unoxidized mixture matched 13 of the 15 proteins present; no peaks were observed that corresponded to Ribo and AC-S2. Of the 83 peaks identified, only 18 contained oxidizable residues half of which contained tryptophan. For the oxidized 15 protein component mixture, positive mode MALDI data matched 13 of the 15 proteins. After submission of the oxidized mixture data to the search engine, 17 peptides contained oxidized residues, 9 of which contained methionine sulfone. Again, as in positive mode MALDI data of the unoxidized sample, AC-S2 and Ribo were not identified.

PMF identification of the oxidized mixture using negative mode MALDI data resulted in identification of only 10 of the 15 proteins. The proteins not identified were AC-S1, AC-S2, BC, CA, and CytC. When referring to Table 1, these five proteins are among those with the lesser amount of cysteine residues; thus, the lack of cysteic acid may be responsible for the lower number of protein identifications. Even though Myo and Hemo-α contain no cysteine, one or more of their tryptic peptides, whether or not an

oxidizable residue is present, typically had a strong signal intensity that allowed for identification. Over half of the peaks identified as peptides did contain cysteic acid, and of the peptides containing methionine sulfone, the majority also contained cysteic acid. Only two peptides contained singly-oxidized tryptophan alone.

Combined Unoxidized and Oxidized Sequence Coverage for Protein Digest Mixtures

Although sufficient numbers of protein are identified by a single mode of analysis, enhanced results can be achieved by combining several approaches. In addition, percent coverage for proteins identified in the unoxidized, positive mode data by PMF increased using the combined unoxidized/oxidized data set. For example, the percent coverage for Ald (fifteen protein mixture) obtained using positive mode unoxidized data was 12%; whereas, the percent coverage for Ald (fifteen protein mixture) was determined to be 43% using a combined unoxidized/oxidized data set. On average, percent coverage for the positive mode, unoxidized fifteen component protein mixture was 23%, and increased to 33% using the combined data set. Generally, percent coverage for each protein component decreased as the sample complexity increased, although a combined approach of unoxidized/oxidized data always provided the same or better percent coverage for individual proteins in a complex mixture.

Furthermore, by comparison of unoxidized and oxidized MALDI spectra of complex mixtures of proteins after trypsin digestion, the possible protein components in a biological mixture identified in PMF experiments can be further refined to eliminate any false positives. After performic acid oxidation, the elimination of erroneous protein matches can simply be done by the observation of a specific mass shift, or lack thereof,

in the peptide ion purported to have a sequence with cysteine, methionine, and/or tryptophan residues.

CHAPTER VI

"ON-TARGET" PERFORMIC ACID OXIDATION OF THERMALLY-DENATURED TRYPTIC DIGESTS

BACKGROUND

In-solution performic acid oxidation, according to the protocol utilized in this study (Chapter III, IV, V), consumes close to five hours of sample preparation time. In order to make this oxidation adaptable to a high-throughput approach, a quicker procedure that retains the oxidative power of the in-solution protocol must be created. In addition, the ability to conserve the amount of the available sample would greatly aid in situations where sample is limited. A dried-droplet MALDI sample can be analyzed in both polarities; the sample plate ejected from the instrument, and an oxidation solution applied directly onto the previously irradiated sample spot and allowed to dry. The same sample spot can now be analyzed as an oxidized digest. Unfortunately, the majority of matrices used in MALDI are soluble in organics and performic acid is an aqueous solution. In addition, other pitfalls include unwanted side products (i.e., chlorinated tyrosine), can occur when performic acid oxidation is performed at ambient temperatures.[29,31] Bondanarko et al.,[32] accomplished an "on-target" oxidation with performic acid directly on a MALDI sample tip, although the details of the procedure utilized are not disclosed in the manuscript.

RESULTS AND DISCUSSION

Positive and negative ion spectra of the thermally-denatured tryptic digests of unoxidized BSA and TF were similar in appearance to previous samples, and after submission of the observed peaks to Mascot, the percent coverage fell within the standard deviation of the average of the respective proteins in Chapter IV. After ejecting the sample stage from the instrument, various ratios of oxidizing solution were deposited onto the sample spot and allowed to evaporate. The extent to which the α-cyano-4-hydroxycinnamic acid crystals dissolved was visual inspected. In every case where performic acid was mixed with the organic solution and directly deposited on top of the dried-droplet, the matrix crystals did not appear to dissolve completely. After evaporation, the peaks corresponding to peptides with oxidizable residues had not shifted in mass, indicating that no oxidation occurred. When the methanol alone was applied, the lack of surface tension caused the methanol droplet to spread rapidly into other sample spots on the MALDI plate causing cross-contamination. An equal volume of acetonitrile was mixed with the methanol to increase the surface tension and prevent "spreading". After applying the acetonitrile/methanol mixture, the matrix crystals dissolved, and 1 μL of performic acid was added to the sample droplet. The matrix crystals were significantly smaller after the oxidant mixture had evaporated. The sample was analyzed by MALDI in the same manner as described previously.

Compared to in-solution performic acid oxidation, applying the oxidant solution directly on top of a MALDI sample spot resulted in degradation of spectral quality. Average signal-to-noise ratio (S/N) decreased by a factor of 5 upon oxidation (Figure

10). Only peptide ion signals that typically have high signal intensity when oxidized in-solution were detected. It is possible that residual oxidant may have become encapsulated inside the matrix crystals, leading to an increase in chemical noise. The solvent composition of matrix/analyte solution can influence MALDI spectral quality. For example, recent results suggest that the solvent-dependant conformation of peptides can influence MALDI ionization efficiency.[23]

Despite the relatively poor signal quality, the method utilized successfully oxidized the peptide sample. Masses corresponding to unoxidized peptides with cysteine, methionine, and/or tryptophan were not detected. No methionine sulfoxide-containing peptides were observed indicating that oxidation was complete. In the case of tryptophan-containing peptides, the dominant oxidized form was that of N-formylkynurenine, the doubly-oxidized species (Figure 11). This differs from in-solution performic acid oxidation where β-(3-oxyindolyl)alanine-containing (+O) peptide peaks are dominant. The performic acid concentration used for on-target oxidation and the presence of matrix can be ruled out as causes for the observation. In preliminary studies with single peptides, the 1:32 ratio of performic acid was used with no matrix present, and the $+O_2$ oxidized tryptophan species was also the most dominant. The probable cause for the shift in the dominant oxidized species is due to the higher temperature allowed in on-target oxidation.

Using the complete, doubly-oxidized tryptophan modification in ProFound, positive and negative ion data from the on-target oxidation of BSA and TF were queried. The PMF identification of positive ion data for both BSA and TF had percent coverages

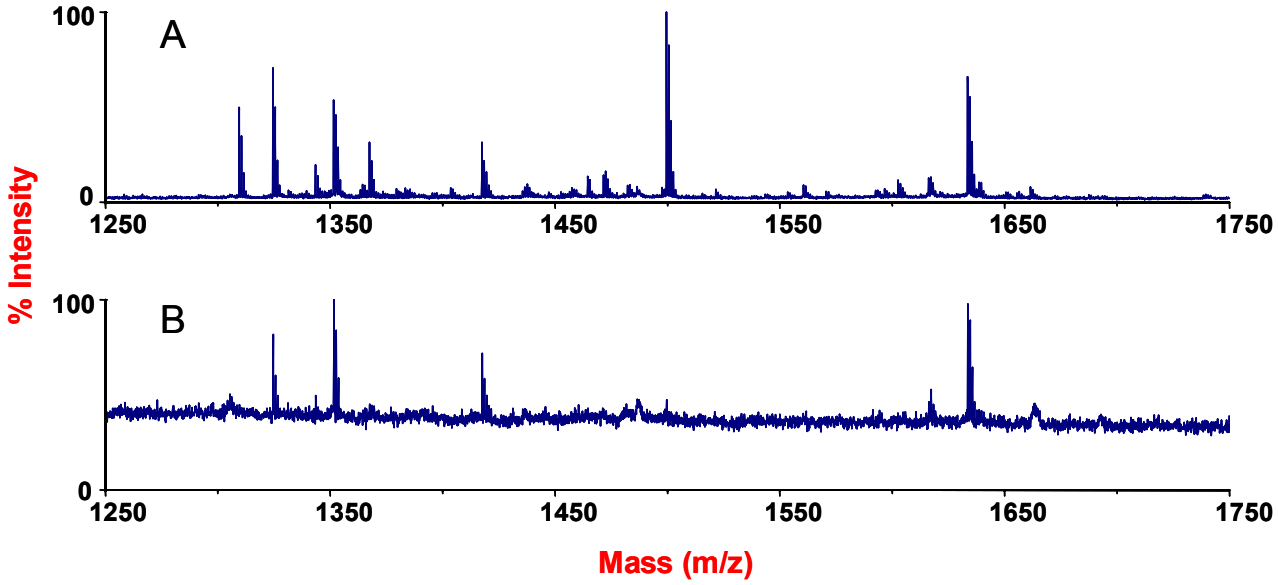

Figure 10. Representative negative mode MALDI spectra of TF when oxidized (A) in-solution and (B) "on-target".

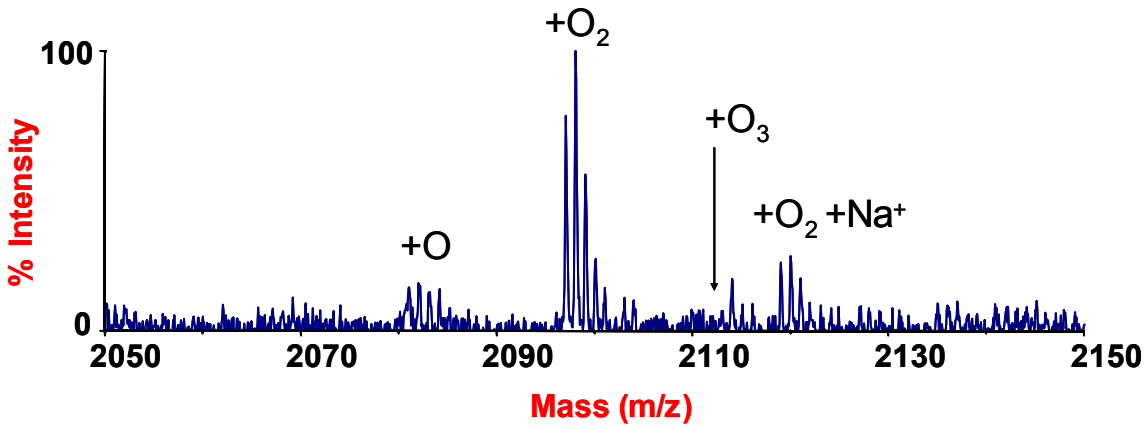

Figure 11. Positive mode MALDI spectra of the tryptic peptide EDVIWELLNHAQEHFGK after "on-target" performic acid oxidation.

that fell lower than the average range for oxidized, positive ion percent coverage of the proteins oxidized in-solution in Chapter IV. Percent coverage obtained from oxidized, negative ion data fell to approximately half of the percentages obtained from in-solution oxidization. When oxidized positive and negative ion MALDI data were combined, BSA and TF percent coverage was also below the average range for the same combined data obtained under in-solution oxidation (Chapter IV).

The total-combined percent coverage obtained from combining the four sets of data (unoxidized positive/negative and oxidized positive/negative) for BSA was ~20% lower than the total-combined percent coverage of BSA in Chapter IV. The percent coverage observed from the combination of TF data was within the average range for total-combined percent coverage of TF reported in Chapter IV.

CHAPTER VII

CONCLUSIONS

Performic acid oxidation of tryptic protein digests alters the chemical nature of peptides that contain cysteine, methionine, and/or tryptophan residues. Dependant on the amino acid composition of a protein, oxidation can aid in increasing sequence coverage when the digest is analyzed by MALDI in positive and negative polarity and the data is submitted for PMF identification. Oxidation influences sequence coverage by enhancing ionization efficiency of some peptides and reducing mass spectral overlap for oxidizable peptide ion signals that are eclipsed by an unoxidizable peptide. High confidence identification of a protein by PMF can be obtained by observing mass shifts in peaks corresponding to a peptide with an oxidizable residue when comparing spectra from before and after oxidation due to the low frequency of abundance of cysteine, methionine, and tryptophan residues in proteins. A shift of +32 Da indicates the presence of methionine, whereas a mass shift of +48 Da confirms the presence of cysteine. Despite the detection of fewer tryptophan-containing peaks in MALDI after performic acid oxidation, observing the emergence of a triplet of peaks with 16 Da gaps, confirms the presence of tryptophan. Thus, the mass shifts of cysteine, methionine, and/or tryptophan can greatly enhance confidence when performing PMF protein identifications.

After detailed analyses of MALDI spectra of oxidized protein digests and protein digest mixtures, peptides with methionine sulfone are best detected in positive mode,

and those with cysteic acid are best detected in negative mode. Methionine oxidation may add to the basicity of the peptide to enhance the probability of protonation. It is also important to note that high m/z peaks, corresponding to either an oxidized peptide or not, were more readily detected in positive mode MALDI of oxidized digests. Due to the high acidity of cysteic acid, cysteic acid-containing peptides more readily loose a proton, and thus are more easily detected in negative mode MALDI. The majority of cysteine-containing peptides are not observed prior to oxidation in the proteins used here (Table 1).

Performic acid oxidation can be performed "on-target" by directly applying an organic solution onto a dried-droplet MALDI sample immediately followed by the oxidant solution. This approach to oxidation allows rapid data acquisition of the same sample before and after oxidation, and allows micro scale oxidation when sample volume is limited. To correct for the degradation of spectral quality after "on-target" oxidation, acetone redeposition[46,49] or applying additional matrix solution may be used as a third application after complete evaporation of the oxidant solution. These additional applications may free any residual oxidant that may become incorporated into matrix crystals.

After an optimized protocol is determined, performic acid oxidation can be adapted to high-throughput, automated proteomics experiments. For example, samples analyzed by electrospray ionization mass spectrometry (ESI-MS) could be mixed using a flow-injection (FI) method where a long reaction capillary is used to perform on-line

oxidation prior to ESI. Also, MALDI sample oxidation can be automated using commercially-available robotic sample handling devices.

Future experiments will involve time trials of in-solution oxidation to determine if the 180 min incubation period of tryptic digests is truly necessary for complete oxidation. When performic acid oxidation is performed "on-target", complete oxidation can be performed in ~10 min although the temperature environment is ~ 30°C higher than when oxidation is performed in-solution. Temperature should also be varied when conducting in-solution oxidation since the $+O_2$ tryptophan species becomes the most dominant when oxidation is performed "on-target". At temperatures above 25°C, the triply-oxidized tryptophan may become more dominant. The best case scenario for tryptophan oxidation would be restricting oxidation to only the singly-oxidized species when dealing with a true unknown; then, the amount of additional oxygen atoms would correlate only to tryptophan, methionine, or cysteine.

Performic acid oxidation is extremely useful in an additional manner concerning PMF identification of proteins. In instances where PMF experiments return a number of highly-scored protein matches that are theoretically possible as components in the biological sample, performic acid oxidation can be used to identify the true protein(s) present. Comparing the proposed peptide sequence reported by the search engine as matching with the mass shift, if any, of the m/z after performic acid oxidation will either enhance or refute the validity of the PMF protein identification due to the ability of cysteine, methionine, and tryptophan to incorporate a specific number of oxygen atoms. Observing any mass shifts of masses corresponding to purported peptides, coupled with

the low frequency of occurrence of cysteine, methionine, and tryptophan residues in proteins, adds an even higher amount of confidence in PMF identifications of proteins. Therefore, performic acid oxidation is of great utility when used in conjunction with peptide mass fingerprinting protein identification.

REFERENCES

(1) Hood L. *J Proteome Res.* **2002**, *1*, 399-409.

(2) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.

(3) Farwig, Z. N.; Campbell, A. V.; MacFarlane, R. D. *Anal. Chem.* **2003**, *75*, 3823-3830.

(4) Kim, Y. J.; Freas, A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 1544-1548.

(5) Apuy, J. L.; Chen, X.; Russell, D. H.; Baldwin, T. O.; Giedroc, D. P. *Biochemistry* **2001**, *40*, 15164-15175.

(6) Henzel, W.J.; Billeci, T.M.; Stults, J.T.; Wong, S.C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011-5015.

(7) Pappin, D.J.C.; Hojrup, P.; Bleasby, A.J. *Curr. Bio.* **1993**, *3*, 327-332.

(8) Jensen, O.N.; Podtelejnikov, A.V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741-4750.

(9) *Mascot: Peptide Mass Fingerprint*; Matrix Science Ltd. http://www.matrixscience.com/cgi/index.pl?page=../home.html (accessed August 2003).

(10) *ProFound-Peptide Mapping*; Rockefeller University, Version 4.10.5. http://129.85.19.192/profound_bin/WebProFound.exe (accessed August 2003).

(11) Zhang, W.; Chait, B.T. *Anal. Chem.* **2000**, *72*, 2482-2489.

(12) *MS-Fit*; UCSF Mass Spectrometry Facility. http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm (accessed August 2003).

(13) *Swiss-Prot*; ExPASy Molecular Biology Server. Swiss Institute for Bioinformatics. http://us.expasy.org/sprot/ (accessed August 2003).

(14) *NCBI*; National Center for Biotechnology Information. National Library of Medicine, National Institutes of Health. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein (accessed August 2003).

(15) Beavis R.C.; Chait B T. *Methods Enzymol.* **1996**, *270*, 519-51.

(16) Knochenmuss, R.; Zenobi, R. *Chem. Rev.* **2003**, *103*, 441-452.

(17) Gimon, M. E.; Preston, L. M.; Solouki, T.; White, M. A.; Russell, D. H. *Org. Mass Spectrom.* **1992**, *27*, 827-830.

(18) Stults, J.T.; Billeci, T.M. *Anal. Chem.* **1993**, *65*, 1709-1716.

(19) Eriksson, J.; Chait, B.T.; Fenyo, D. *Anal. Chem.* **2000**, *72*, 999-1005.

(20) Eriksson, J.; Fenyo, D. *Proteomics* **2002**, *2*, 262-270.

(21) Kratzer, R.; Eckerskorn, C.; Karas, M.; Lottspeich, F. *Electrophoresis* **1998**, *19*, 1910-1919.

(22) Knochenmuss, R. *Anal. Chem.* **2003**, *75*, 2199-2207.

(23) Figueroa, I. D.; Torres, O.; Russell, D. H. *Anal. Chem.* **1998**, *70*, 4527-4533.

(24) Park, Z.-Y. Ph.D. Dissertation, Texas A&M University, College Station, May 2001.

(25) Krause, E.; Wenschuh, H.; Jungblut, P.R. *Anal. Chem.* **1999**, *71*, 4160-4165.

(26) Karty, J.A.; Ireland, M.M.E.; Brun, Y.V.; Reilly, J.P. *J. Proteome Res.* **2002**, *1*, 325-335.

(27) Cohen, S.L.; Chait, B.T. *Anal. Chem.* **1996**, *68*, 31-37.

(28) Campo, K. K. M.S. Thesis, Texas A&M University, College Station, December 1996.

(29) Sanger, F. *Biochem. J.* **1949**, *44*, 126-128.

(30) Benassi, C.A.; Scoffone, E.; Veronese, F.M. *Tetrahedron Lett.* **1965**, *49*, 4389-4393.

(31) Hirs, C.H.W. *Methods Enzymol.* **1967**, *11*, 197-199.

(32) Bondarenko, P.V. "On-Target Performic Acid Oxidation for MALDI Analysis." Manuscript; Department of Chemistry, Texas A&M University, College Station, 1998.

(33) Wagner, R.M.; Fraser, B.A. *Biomed. Environ. Mass Spectrom.* **1987**, *14*, 69-72.

(34) Sun, Y.; Smith, D.L. *Anal. Biochem.* **1988**, *172*, 130-138.

(35) Van de Weert, M.; Lagerwerf, F.M.; Haverkamp, J.; Heerma, W. *J. Mass Spectrom.* **1998**, *33*, 884-891.

(36) *Merck Index*, 12th ed.; Merck & Co.: Rahway, N.J., 1996.

(37) Gevaert, K.; Van Damme, J.; Goethals, M.; Thomas, G.R.; Hoorelbeke, B.; Demol, H.; Martens, L.; Puype, M.; Staes, A.; Vandekerckhove, J. *Mol. Cell. Proteomics* **2002**, *1*, 896-903.

(38) Simat, T.J.; Steinhart, H. *J. Agric. Food Chem.* **1998**, *46*, 490-498.

(39) Villar, H.O.; Koehler, R.T. *Biopolymers* **2000**, *53*, 226-232.

(40) Park, Z.-Y.; Russell, D.H. *Anal. Chem.* **2000**, *72*, 2667-2670.

(41) Sharp, J.S.; Becker, J.M.; Hettich, R.L. *Anal. Biochem.* **2003**, *313*, 216-225.

(42) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotech.* **1999**, *17*, 994-999.

(43) Russell, D.H.; Park, Z.-Y. In *Advances in Nucleic Acid and Protein Analyses, Manipulation, and Sequencing*; Limbach, P.A., Owicki, J.C., Raghavachari, R., Tan, W., Eds.; Proceedings of SPIE, SPIE Press: Bellingham, WA, 2000; Vol. 3926, 48-55.

(44) Hill, H. H.; Hill, C. H.; Asbury, G. R.; Wu, C.; Matz, L. M.; Ichiye, T. *Int. J. Mass Spectrom.* **2002**, *219*, 23-37.

(45) Smith, B.J.; Wheeler, C. In *Protein Sequencing Protocols*; Smith, B.J., Ed.; Methods in Molecular Biology, Humana Press: Totowa, N. J., 1997; Vol. 64, Chapter 5.

(46) Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, *66*, 3281-3287.

(47) Anderson, N.L.; Anderson, N.G. *Mol. Cell. Proteomics* **2002**, *1*, 845-867.

(48) Jensen, O.N.; Podtelejnikov, A.V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741-4750.

(49) Yao, J.; Scott, J.R.; Young, M.K.; Wilkins, C.L. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 805-813.

VITA

JESSICA ELAINE LUCAS

Home: P.O. Box 814
Alpine, TX 79831
catfish623@cs.com

Work: Texas A&M University
Department of Chemistry
MS 3255
College Station, TX 77843
(979) 845-0613
lucas@mail.chem.tamu.edu

Education

- M.S., Chemistry (December2003), Texas A&M University, College Station.
  Major: Analytical Chemistry, Biological Mass Spectrometry
  Research Director: David H. Russell
- B.S., Chemistry (May 2000), University of Texas at Dallas, Richardson, Texas.
  Major: Synthesis of Mesoporous Silica
  Research Director: Kenneth J. Balkus, Jr.

Honors and Awards

- Regents' Fellowship Award for Excellence in Chemistry, 2000-2001
- First Place Poster at the 5th Annual Biology and Chemistry Graduate Student Symposium at the University of Texas at Dallas, 2000

Professional Affiliations

- American Society for Mass Spectrometry (ASMS)
- American Chemical Society (ACS)
- Phi Lambda Upsilon (PLU)
- Certified Pharmacy Technician (CPhT) 1999-2001, 2003-2005