MASS SPECTROMETRIC PROTEIN FOOTPRINTING OF CONFORMATIONAL
CHANGES INDUCED BY CHROMATOGRAPHIC SURFACES

by

Kimberly Nicole Focke

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Chemistry

Charlotte

2022

Approved by:

_____
Dr. Brian T. Cooper

_____
Dr. Jordan Poler

_____
Dr. Jerry Troutman

_____
Dr. Yuri Nesmelov

ACKNOWLEDGEMENTS

ABSTRACT

KIMBERLY NICOLE FOCKE. Mass Spectrometric Protein Footprinting of Surface-
Induced Conformational Changes. (Under the direction of Dr. BRIAN T. COOPER)

Reversed-phase liquid chromatography (RPLC) of intact proteins induces

conformational changes due to the denaturing effects of both the hydrophobic stationary

phase and the high-organic mobile phase, which can lead to poor recovery and badly

distorted or artifactual chromatographic peaks. A surprising additional consequence is

that the length of the column does not matter. We are investigating these structural

changes using mass spectrometric protein footprinting, which gives information about

changes in the solvent accessibility of modifiable sites. Differences in site accessibility

following adsorption to and desorption from chromatographic surfaces indicate regions of

structural alteration in proteins. We seek to determine if bovine serum albumin (BSA)

adsorbs in a preferred binding orientation, if denaturation increases with surface

residence time, and if the conformation in high-organic solvent depends on whether the

protein was first adsorbed. Glutamate and aspartate groups of BSA are amidated with

methylamine and EDAC. Tryptic digests of the modified proteins are analyzed using LC-

MS/MS. Peptides are identified with tandem mass spectrometry, and modification sites

are located through mass shifts in the tandem mass spectra. Quantification is performed

by comparing relative peak areas of modified and unmodified peptides in extracted ion

chromatograms for their respective precursors. While results remain somewhat

ambiguous in the absence of instrumental replicates of individual samples and modeling,

our research presents a method for footprinting proteins adsorbed to chromatographic

surfaces and confirms the apparent contribution of both the hydrophobic stationary phase

and the high-organic mobile phase to conformational changes observed during RPLC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACN   Acetonitrile

CE   Capillary electrophoresis

CZE   Capillary zone electrophoresis

DTT   Dithiothreitol

EDAC   1-Ethyl-3(3-dimethylaminopropyl)carbodiimide

ESI   Electrospray ionization

IAM   Iodoacetamide

LC   Liquid chromatography

LC-MS/MS   Liquid chromatography-tandem mass spectrometry

LTQ   Linear quadrupole ion trap

MA   Methylamine

MALDI   Matrix-assisted laser desorption/ionization

MeOH   Methanol

m/z   mass-to-charge ratio

MS   Mass spectrometry

MS/MS   Tandem mass spectrometry

RPLC   Reversed-phase liquid chromatography

RT   Retention time

SPE   Solid phase extraction

TIC   Total ion chromatogram

XIC   Extracted ion chromatogram

CHAPTER 1: PROTEIN ADSORPTION DURING LIQUID CHROMATOGRAPHY

1.1 Reversed-Phase Liquid Chromatography (RPLC)

Reversed-phase liquid chromatography (RPLC) is a surface-mediated separation technique that is widely used to separate analytes that are too polar to be separated by gas chromatography. This technique is also extensively used for both small molecule applications and in proteomics for protein and peptide separations. RPLC utilizes a nonpolar stationary phase (C_4, C_8, or C_{18}) covalently bonded to a silica support and a polar mobile phase typically composed of a mixture of water and acetonitrile (ACN) or methanol (MeOH). Mobile phase strength increases with increasing organic content. Elution is in order of increasing hydrophobicity.

The mechanism of retention is vastly different between small molecules and large polymers such as proteins. In the case of traditional reversed-phase chromatography of small molecules, the analytes equilibrate between the stationary and the mobile phase for the full length of the column. Interactions between analytes and the stationary phase due to intermolecular forces such as hydrogen bonding, dipole-dipole interactions, and Debye forces result in differing levels of retention. The longer the alkyl chain of the stationary phase, the more strongly retained nonpolar compounds will be. Highly nonpolar compounds will be strongly retained by the nonpolar stationary phase and elute later than more polar analytes, which have more preferential interactions with the polar mobile phase than with the stationary phase and thus travel through the column at a much faster rate. Separations of small molecules may be performed under either isocratic or gradient conditions. In isocratic separations, the mobile phase is held at a fixed composition

throughout the run. However, for a realistically complex sample, there is often no ideal set of isocratic conditions which can both adequately resolve early-eluting peaks while also eluting more strongly retained compounds in a reasonable experimental timeframe. This problem, commonly known as the general elution problem, may be overcome by performing a gradient separation. In gradient separations, the organic content of the mobile phase is increased throughout the run in order to more quickly elute the strongly retained compounds.

Column selection in RPLC is of particular importance—variations in column length, particle size, and stationary phase can all affect the retention and resolution of analytes. Longer columns tend to have more resolving power than shorter columns: for small molecule separations, an increased distance of travel between the head of the column and the detector provide more opportunities for interactions with the stationary phase. For otherwise unresolvable chromatographic peaks, choosing a longer column can help improve resolution. This is not the case for protein separations, however.

1.2 RPLC of Intact Proteins

Protein separations differ significantly from small molecule separations in both the mechanism of retention and elution. In contrast with small molecules, proteins adsorb strongly to the chromatographic surface and do not readily partition back into the mobile phase. Hydrophobic regions of the protein surface bind strongly to the stationary phase at the head of the column, and the protein begins to unfold to make more favorable connections with the chromatographic surface. Environmental conditions such as temperature, pH, ionic strength, and solvent composition strongly influence protein unfolding and surface adsorption.[1-11] This is reflected in one of the most unusual

characteristics of proteins adsorbed to RPLC surfaces: after adsorption, the solvent range over which the protein changes from completely adsorbed to completely desorbed is high, but very narrow. Desorption and elution can thus be achieved through slight changes in the organic content of the mobile phase.[4, 6, 9] The sensitivity of proteins to change in mobile phase composition is pronounced, and extends even to extremely large proteins (>500 kDa), which desorb from a chromatographic surface following increases in organic content of only a few percent.[12] An interesting consequence of this is that for intact-protein RPLC separations, column length does not matter.[9, 13] Although elution times increase with column length, resolution is negligibly improved, if at all—for example, the time between the elution of bovine serum albumin and ovalbumin was found to increase by only 13% upon a fivefold increase in column length.[13] Furthermore, recovery of surface hydrophobic proteins has previously been improved without loss of resolution by reducing column length from 25 cm to 1 cm.[9] This surprising chromatographic behavior is indicative of large structural changes occurring during the chromatographic process which reduce the affinity of the protein for the stationary phase post-desorption. These changes have been subject to limited research and are poorly understood; however, we speculate that they are due to a combination of the denaturing effects of the high-organic mobile phase and desorption from the hydrophobic stationary phase.

Although RPLC is capable of achieving very high selectivity—notably, it can be used to separate proteins differing by a single amino acid[1, 2, 14-16]—poor protein recoveries and aberrant retention behavior arise from structural changes which occur during the chromatographic process. These structural changes are largely due to

denaturing effects of both the hydrophobic stationary phase and the high-organic mobile

phase. Low protein recovery during RPLC has also been attributed to electrostatic

interactions between the protein and residual free silanol groups; non-specific interactions

with stainless steel tubing, porous frits, and metal ion impurities in the silica support; and

irreversible adsorption to the stationary phase itself.[9] This has limited the commercial use

of RPLC as a preparative method to relatively small, stable proteins such as recombinant

insulin.[2] Additionally, *a priori* retention predictions are made difficult by the presence of

protein conformers on-column: native and unfolded protein conformations can differ

significantly in relative binding strength—by factors as large as 10^{12}—under the same

fixed mobile phase composition.[8] This often leads to the elution of broadened,

shouldered, or multiple peaks concomitant with the presence of multiple protein

conformers,[2, 3, 7, 8, 11, 14, 17, 18] even from a pure protein. This may lead researchers to make

specious assumptions about sample purity or column quality and contributes to the

difficulty in making accurate predictions of protein retention behavior under RPLC

conditions.
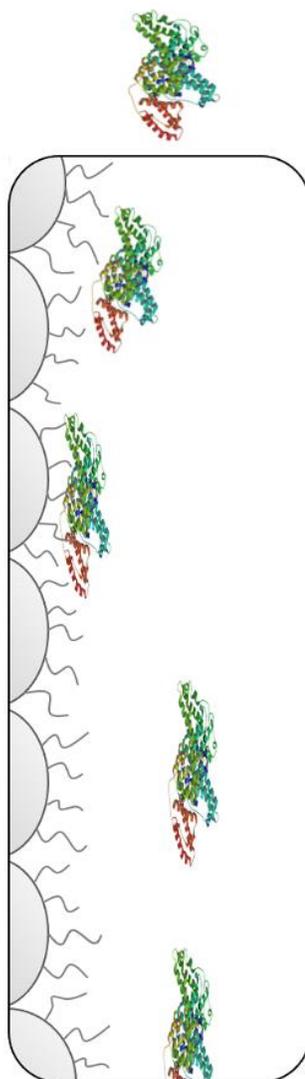
Figure 1.1: Illustration of column cross-section demonstrating reversed-phase separation of an intact protein. Following adsorption at the head of the column and desorption back into the high-organic mobile phase, the protein interacts minimally with the stationary phase for the remainder of the separation. Not shown to scale.

1.3 Conformer Formation of Adsorbed Proteins

The tendency of proteins toward surface adsorption is well-established. The effects of protein-surface interactions on adsorption are extremely complex due to the variety of contributions to the total interaction. Van der Waals forces, electrostatic interactions, hydrogen bonding, and the hydrophobic effect all drive the dynamic adsorption process, which is dependent on characteristics of both the surface and the adsorbing protein.[19-22] The unusual retention and elution behavior of proteins has been attributed to the presence of secondary and tertiary structure in the folded protein[1, 3, 5] which yields a heterogenous surface with patches characterized by hydrophobic/hydrophilic or charged/neutral properties.[6, 19-22] These heterogeneous surfaces initiate elaborate protein-surface interactions which have been shown to give rise to preferred binding orientations,[6, 22, 23] and interactions with adjacent protein molecules on the surface have been shown to regulate binding density and rearrangement of adsorbed proteins.[6, 22] Proteins generally demonstrate strong and extensive binding to hydrophobic surfaces,[6, 15, 20-24] where reduction of native structure is most frequently observed and increases with contact time as the protein undergoes rearrangements to form more favorable points of contact with the surface.[6, 22, 23] This is particularly relevant to RPLC protein separations, as surface affinity has been well-correlated with extent of structural change.[6, 22, 23] Structural changes are most pronounced in less stable proteins,[6, 10, 23, 24] with timescales comparable to chromatographic retention times.[3, 6, 17] This is illustrated by the aberrant chromatographic behavior of the less stable "soft" proteins prone to conformational changes (such as BPTI[14]) compared with more stable "hard" proteins (such as α-chymotrypsinogen[25]) that display no obvious unfolding.[24]

Although it is clear that surface interactions contribute significantly to structural changes during RPLC, solvent effects confer additional complexity to protein retention and cannot be neglected. The three-dimensional protein structure is highly sensitive to slight changes in non-physiological environments,[1] and the secondary and tertiary structure of a protein is dependent on several variables. Protein unfolding and surface adsorption can be further influenced by the ionic strength and pH of the environment.[1-10]

1.4 Previous Work Involving Adsorbed Proteins

Although limited, some research has been conducted on the effects of different reversed-phase chromatographic conditions on protein conformation. Much of this work has focused on soft proteins susceptible to undergoing conformational changes, particularly bovine pancreatic trypsin inhibitor (BPTI)—one of the simplest and smallest globular proteins.

Several techniques have been employed to evaluate conformational changes in surface-adsorbed proteins, including Raman[26] and Fourier-transform infrared (FTIR) spectroscopy,[27] which assess changes in secondary structure through comparison of amide band intensities; and fluorescence[28] of tryptophan residues or fluorescent labels, which is a sensitive indicator of changes in tertiary structure but does not give structural information directly. More recently, hydrogen-deuterium exchange (HDX) detected by nuclear magnetic resonance (NMR)[2, 14, 29] or mass spectrometry (MS)[29-31] and limited proteolysis detected by MS[21, 30] have been exploited to measure changes in solvent accessibility between protein conformers. Although NMR is capable of atomic-level resolution,[32] its use is limited to smaller proteins ($< \sim 50$ kDa) due to unresolvable overlapping peaks in the NMR spectra of large proteins.[30, 33] The requirement of high

analyte concentrations further limits its potential.[34] McNay et al. explored the role of salt type and ionic strength on conformational changes using HDX and NMR. It was found that for BPTI, increasing the ionic strength of a salt present with reversed-phase surface-adsorbed protein caused a decrease in the degree of hydrogen exchange on the surface. This resulted in an exchange pattern similar to that found on the native protein, and was attributed to stabilization of the native structure due to the salts. This stabilization of native structure by salts is well-known and most pronounced for salts at the top of the Hofmeister series and for salts present in high concentrations. However, non-volatile salts are not amenable to mass spectrometry, limiting the possibility of their use for stabilizing proteins analyzed by LC-MS.

Structural MS, which is amenable to larger proteins and can be used to obtain sequential and structural information, has been enlisted to overcome these limitations. Popular approaches including HDX and covalent labeling.[32] HDX is a sensitive structural probe which relies on the exchange of backbone amide protons with deuterium, causing observable mass shifts in the MS spectrum. This technique provides valuable site-specific information about backbone solvent accessibility; however, it is not without limitations: the reversibility of deuterium labelling makes back-exchange ineluctable, reducing the dynamic range of observable exchanges.[29, 32] Moreover, conditions required for quenching and reduction of back-exchange are incompatible with many proteolytic enzymes, limiting those which can be used to acid proteases such as pepsin.[29] In contrast, covalent labeling techniques are amenable to large proteins and an array of proteases, and are rooted in historically well-established protein footprinting methods for covalent

modification of proteins.[32]

1.5 Bovine Serum Albumin

The high stability and availability of bovine serum albumin (BSA), combined with its low cost and well-characterized native structure, make it well-suited and widely used for adsorption studies.[10] It also has strong similarities to human serum albumin (HSA)[35] and relevance to biofouling of medical devices.[19] Serum albumin (SA), the most abundant mammalian plasma protein, is a multifunctional globular "soft" protein with exceptional binding capacity. Following synthesis in the liver, it is transported as a non-glycosylated protein to the plasma. Once in the plasma, SA reaches high concentrations (approximately 0.6 mM) and regulates the colloidal osmotic pressure of blood. The high binding capacity of serum albumin makes it a critical protein for the transport of metabolites, nutrients, drugs, and other small molecules and ions. Notably, it is a primary circulatory protein responsible for moderating levels of ionized Zn^{2+}, Ca^{2+}, and Mg^{2+} in the blood. These binding properties contribute to the use and study of serum albumin for myriad clinical and biochemical applications.

Serum albumins are negatively charged, relatively large (approximate molecular weights of 66 kDa) globular proteins with a high degree of sequential and structural similarity across organisms. Both HAS and BSA are comprised of 583 amino acids in their mature form, neglecting the signal sequence and propeptide which are cleaved to form the mature protein. Compared with HSA, the sequence of BSA has a similarity of 75.6%.[35]

Three helical domains form 74% of the secondary structure of BSA and arrange into a heart-shaped three-dimensional structure—these domains are responsible for the binding

and transport of fatty acids, metal ions, and nucleotides. The 583-residue amino acid

sequence of the negatively-charged mature protein (average molecular weight: 66 kDa)

contains 99 acidic side chains (40 aspartic acid and 59 glutamic acid residues comprise

17% of the mature sequence). Thirty-five cysteine residues form 17 intrachain disulfide

bridges, leaving a free thiol at Cys34. While the native structure of BSA has been

extensively investigated, BSA conformers resulting from reversed-phase separations are

less well characterized.[35]

Figure 1.2: Ribbon diagram of bovine serum albumin structure. RCSB PDB 3V03.

1.6 Our Research

The purpose of this research is to investigate the conformational changes that occur when a protein interacts with a hydrophobic chromatographic surface. Bovine serum albumin (BSA) was selected as a model protein due to its prevalence in serum, its extensively characterized native structure, and its potential relevance to biofouling of medical devices. Following the adsorption of BSA to chromatographic packings taken from C_8 solid phase extraction (SPE) cartridges, adsorbed protein is treated with reagents which can react with solvent-accessible glutamic acid (Glu, E) and aspartic acid (Asp, D) residues. Mass spectrometry is used to measure the extent of modification at each site. Each "footprinting" reaction is carried out under different conditions reflective of typical reversed-phase separations. In doing so, we seek to do the following:

- Investigate the role of reversed-phase surface adsorption in conformational changes of BSA.

- Investigate the role of mobile phase organic-content in conformational changes of BSA.

- Determine the preferred binding orientation of BSA on a hydrophobic reversed-phase surface.

- Evaluate the impact of surface dwell time on conformational changes of BSA.

The modified protein is desorbed from the surface and proteolyzed. Peptide fragments are identified with tandem mass spectrometry (MS/MS) and quantified with liquid chromatography-mass spectrometry (LC-MS) by comparing relative peak areas of modified and unmodified peptides. The extent of modification can be used to track solvent-accessibility of residues under different environmental conditions typical of

reversed-phase separations. This exploration of surface-induced conformational changes under reversed-phase chromatographic conditions will provide insight to the nature of protein denaturation on reversed-phase columns, which is pertinent to the development of chromatographic surfaces for intact-protein separations and for the improved prediction of retentive behavior for proteins separated using reversed-phase liquid chromatography.

CHAPTER 2: PROTEIN FOOTPRINTING

2.1 Introduction

Protein footprinting involves irreversibly "marking" solvent-accessible amino acid side chains with chemical reagents prior to mass spectrometric analysis, yielding a stably covalently labeled protein.[32, 36] The modification sites are identified by the resultant mass shift, and the extent of modification can be used to track site accessibility under different environmental conditions.[32] Covalent footprinting modifications serve as a map of solvent-accessible surface area, facilitating elucidation of tertiary and secondary structure. Quaternary structure can also be examined by identifying contacting regions of bound and unbound proteins—a decrease in the extent of modification for the ensemble at specific surface patches indicates oligomerization with other proteins. The same principle can be applied to the study of protein adsorption at surfaces: a preferred surface-binding orientation can be identified through a decrease in modification extent of surface residues for adsorbed proteins, indicating that the target residue is involved in the binding process and no longer exposed to the solvent.

Following modification and proteolysis—typically with site-specific proteases such as trypsin, which cleaves at lysine and arginine—the digested peptide mixture is analyzed by LC-MS/MS. Liquid chromatography is used to separate the peptide mixture, reducing the complexity of the sample entering the mass spectrometer. Data-dependent MS/MS is used for identification of peptides and verification of modification sites through observable expected mass shifts. If the site of modification is ambiguous, MS/MS can also be used to determine the specific site of modification. Once peptides have been

identified in the MS^2 spectra, MS^1 chromatographic peak areas can be used for quantification.[29, 32, 36]
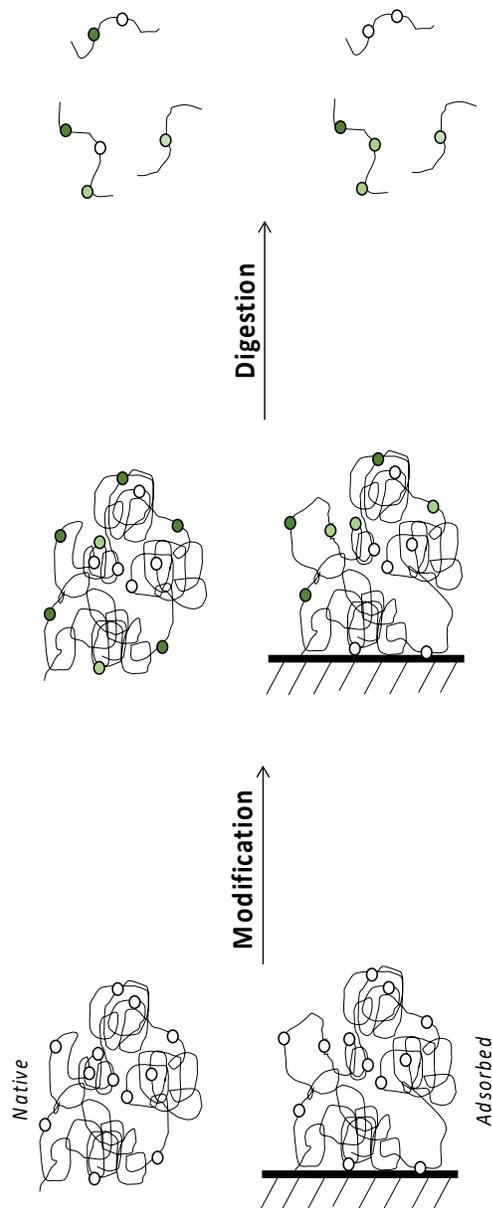
Figure 2.1: Protein footprinting illustration. Open circles indicate potential sites of modification on the native and surface-adsorbed protein. Light and dark green filled circles indicate residues with low and high extents of modification, respectively. The digested peptide mixture is analyzed with LC-MS/MS.

2.2 Modification Reaction

The charged residues of folded proteins in their native state tend to reside on or near the surface, as there is an energetic consequence to burial within the hydrophobic core of the protein. Besides the C- and N- termini, there are seven potentially modifiable amino acids in proteins. Two of these are glutamic and aspartic acid, which possess carboxylic acid containing R-groups that can be modified by amidation.

Carbodiimides mediate the formation of amide bonds through the conjugation of carboxylic acids and amines. The most popular carbodiimide used for this purpose is 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDAC), which has high water solubility and which can be easily removed along with its isourea byproduct formed during the crosslinking reaction. However, EDAC is also highly labile in water, particularly under acidic conditions. In the presence of carboxylic acids, N-substituted carbodiimides such as EDAC form a highly reactive O-acylisourea intermediate which will react with a primary amine nucleophile to generate an amide bond. Although carboxylate activation with EDAC is most effective between pH 3.5 – 4.5 and the highest yield amide formation occurs between pH 4 – 6, strong competition from hydrolysis occurs under acidic conditions below about pH 6.5. The most effective pH range for EDAC-mediated amide formation has been experimentally found to be between 4.5 – 7.5.[37] This means that it is possible to amidate proteins under physiological pH conditions.
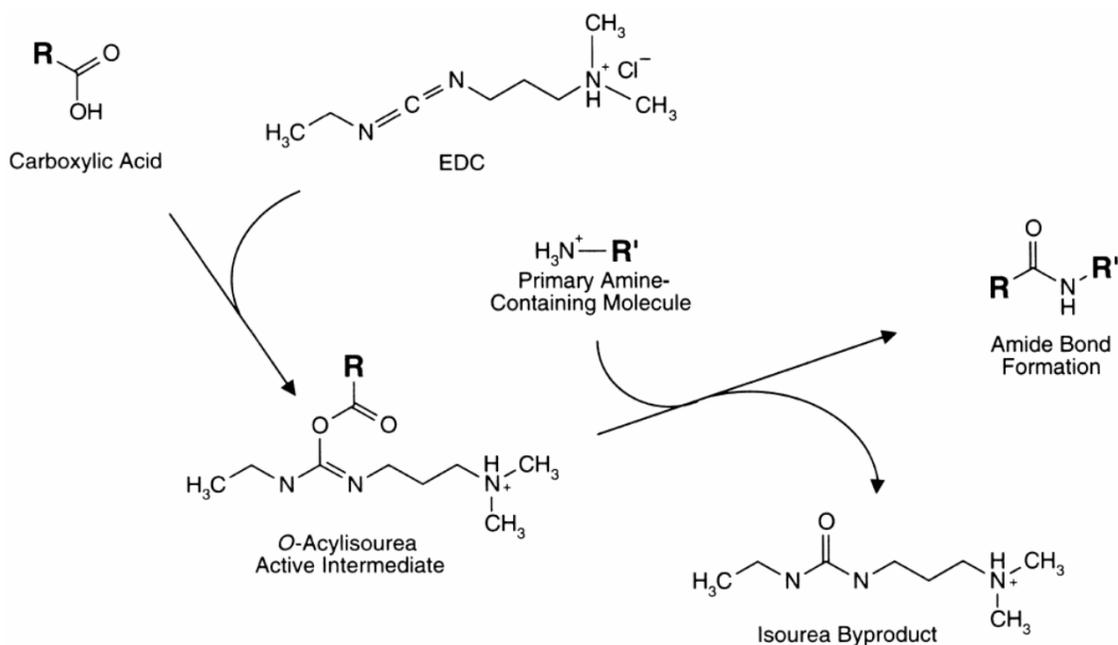
Figure 2.2: Mechanism for the EDAC-mediated amidation of aspartic and glutamic acid residues using methylamine. Used without permission.[38]

Several potential side reactions may occur, with hydrolysis potentially inhibiting desired product formation at two separate points in the reaction mechanism. Thiol ester linkages may occur if the active species is in the presence of sulfhydryl groups. The oxygen atoms in water may also act as an attacking nucleophile, rendering hydrolysis in water the major competing reaction: in this case, the activated ester intermediate is cleaved off, regenerating the carboxylic acid group and forming an isourea.[37]

In the absence of hydrolysis, an anhydride intermediate may be formed through the reaction between a neighboring carboxylate group close in proximity to the O-acylisourea ester intermediate. This is prevalent in polymers with repeating carboxylate groups—anhydrides are also reactive with amines and can still proceed to form an amide bond through one of the carboxylates contained in the anhydride. Although this may also

occur on regions of the protein with close neighboring modification sites, the end product

is the same desired amide formation, and the mechanism is inconsequential.

If EDAC is added in large excess of the number of carboxylate groups present,

the intermediate O-acylisourea ester may rearrange and react with the neighboring

secondary amine on the carbodiimide. This reaction forms an inactive N-acylisourea

derivative permanently attached to the carboxylate containing compound.[37] While the

amidating reaction tends to work best with higher concentrations of EDAC, care should

be taken to avoid adding excess and generating bulky N-acylisourea tags, which can

complicate analysis post-mass spectral acquisition.
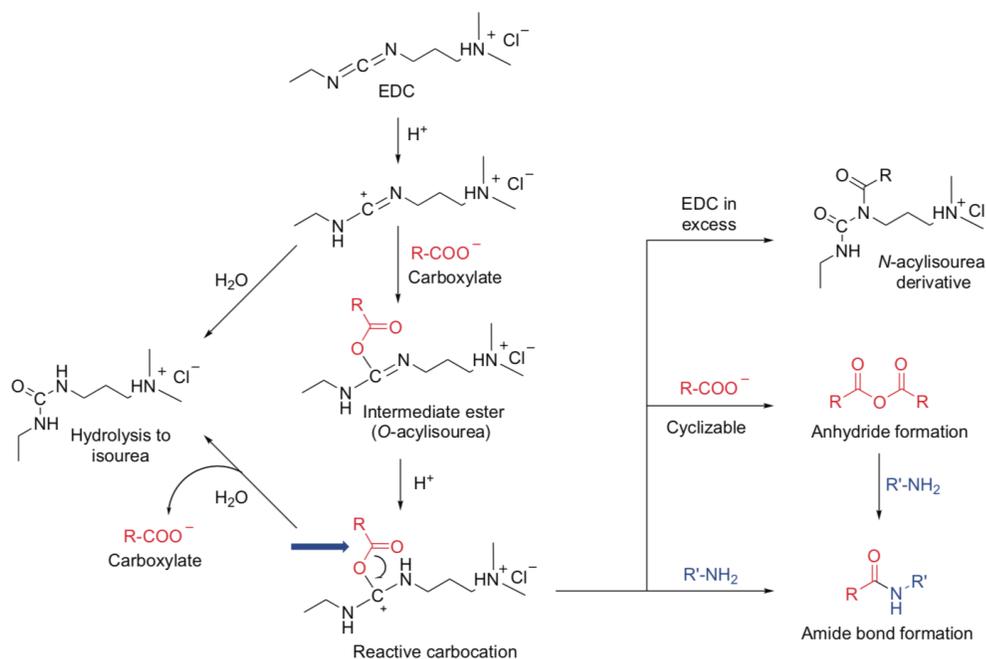
Figure 2.3: Potential side reactions associated with EDAC. Used without permission.[38]

2.3 Capillary Zone Electrophoresis

We used capillary electrophoresis to estimate the extent of amidation. Capillary zone electrophoresis (CZE) is an analytical separation technique based on differences in the migration velocity of charged species in solution under the influence of an electric field. The electrophoretic velocity of an ion is the product of its mobility (μ_{ep}) and the applied field (E).

$$v_{ep} = \mu_{ep}E$$

The electrophoretic mobility of an ion is directly proportional to its charge (q) and approximately inversely proportional to its hydrodynamic radius (r) and the solvent viscosity (η):

$$\mu_{ep} = \frac{q}{f} = \frac{q}{6\pi\eta r}$$

The observed velocity (v_{net}) in CZE is the vector sum of the electrophoretic velocity (v_{ep}) and the velocity of the bulk solution (v_{eof}) due to "electroosmotic flow" (EOF). Like v_{ep}, v_{eof} is the product of a mobility (μ_{eof}) and the applied electric field:

$$v_{eof} = \mu_{eof}E$$

Electroosmotic mobility depends on the charge density on the inner walls of the capillary and on the ionic strength and viscosity of the run buffer. Under typical conditions (bare fused silica capillaries and buffers with low ionic strength and near-physiological pH), the electroosmotic mobility will be positive (in the direction of

cations) and larger than the electrophoretic mobility of any analyte ion. EOF can be measured by injecting a neutral marker and measuring its elution time. Anionic analytes (like most proteins at physiological pH) migrate against this stronger EOF, and thus elute after the neutral marker.

The ends of a fused-silica capillary are submerged in source and destination vials containing run buffer and electrodes connected to a high-voltage power supply. Both fluorescence and UV-Vis absorption can be used for detection, with a detection window made on the capillary just prior to the destination vial. The total length of the capillary is denoted as L_{tot}, and the length from the beginning of the capillary to the detection window is denoted L_{det}. The effective separation length is given by L_{det}. In practice, μ_{ep} can be measured by simultaneously injecting a neutral marker with the analyte of interest. The migration time (taken at the peak centroid) of the analyte (t) can be converted to μ_{ep} using the following equation, in which t_{eof} is the elution time of the neutral marker and V is the run voltage.

$$\mu_{ep} = \mu_{net} - \mu_{eof} = \frac{L_{tot}L_{det}}{V}\left(\frac{1}{t} - \frac{1}{t_{eof}}\right)$$

Proteins typically have an overall negative charge. Protein modification leading to a change in the charge of the targeted residue will change the net charge of the protein, leading to a different electrophoretic mobility between unmodified and modified species. This difference caused by modification can be observed using CZE to confirm successful modification. Furthermore, the extent of modification may also be estimated through CZE and the determination of the modified protein's overall net charge.[39, 40]
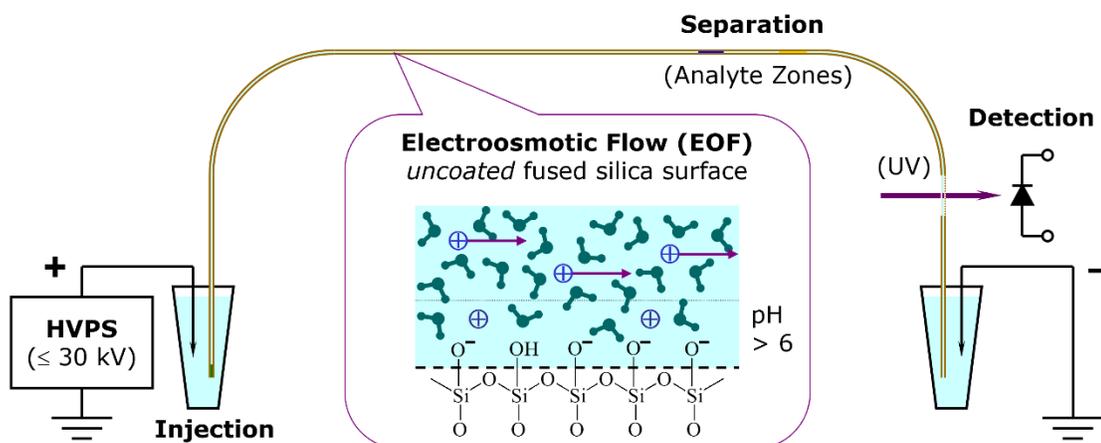
Figure 2.4: Capillary zone electrophoresis instrumental setup. An electrode in contact with the solution to be injected and run buffer is connected to a high voltage power supply. A second electrode in the destination vial generates a current across the capillary. A UV-Vis detector is placed before the destination vial in line with the detection window burned into the capillary coating. Used with permission from Dr. Brian T. Cooper.
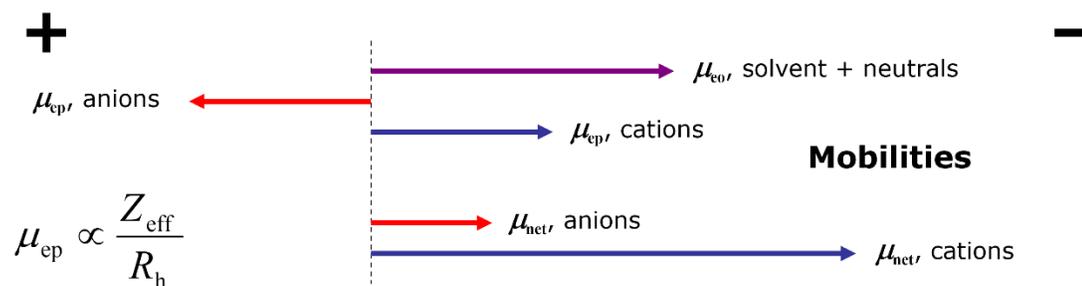

Figure 2.5: Mobility diagram demonstrating migration of analytes through a capillary. Used with permission from Dr. Brian T. Cooper.

2.4: Matrix-Assisted Laser Desorption/Ionization (MALDI)

The overall extent of covalent modification can also be monitored by mass spectrometry of the intact protein. We used matrix-assisted laser desorption/ionization (MALDI) for this purpose. The process begins with the deposition of a sample onto a slide followed by application of a suitable matrix. As the solvent evaporates, the mixture of sample and matrix is dried, resulting in a crystalline matrix structure doped with analyte. Portions of the resulting solid are irradiated with an intense, short, pulsed laser which causes sublimation of the matrix, ablation of the crystal surface, and expansion of the matrix into a gas phase containing intact analyte within the plume.[41] Ionization can occur under vacuum conditions at any time during this process, but the origin of ions produced in MALDI is not yet fully understood: the most widely accepted mechanism involves protonation within the solid phase *prior* to desorption, or proton transfer within the gas phase of photoionized matrix molecules. Ions within the gas phase are then accelerated through an electric field out of the ion source into the mass analyzer.[4]

Matrix selection is especially important for this technique. There must be a large excess of matrix compared to analyte in order to separate the analyte and prevent cluster formation which would inhibit the appearance of molecular ions. Most of the incident energy from the laser is absorbed by the matrix, which both minimizes sample damage and increases the efficiency of energy transfer to the analyte. Because it is the matrix that absorbs the laser pulse, the wavelength of the laser does not need to be adjusted to match the absorption frequency of the analyte.[41]

MALDI matrices must meet several criteria: they must be low enough in molecular mass to be sublimated, vacuum stable, and chemically inert; they must also have the ability to ionize the analyte and strong absorbance at the wavelength of the laser.

Nitrogen lasers (337 nm) and frequency tripled and quadrupled Nd:YAG lasers (355 nm

and 266 nm, respectively) are commonly used. Standard pulse widths range in the tenths

of a nanosecond with a spot diameter between 5 to 200 micrometers. Matrix molecules

are class dependent; two matrices used for proteomic analyses are 2,5-dihydroxy benzoic

acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA). DHB is compatible with

small organic molecules, and both are suitable for use with peptides, proteins, and

carbohydrates.[41] We used MALDI whenever the amount of recovered modified protein

was too small to permit analysis by CZE.

2.5 Mass Spectrometry

Mass spectrometry (MS) is a powerful tool in proteomics which enables the

identification of proteins in complex mixtures as well as detecting and characterizing a

broad assortment of possible post-translational modifications. For our purposes in the

analysis of a single, modified protein as opposed to a complex mixture of proteins, we

need only consider the ability of MS to identify peptides and modified sites. The extent of

modification must be high enough to enable mass spectrometric detection, but not so high

as to distort the protein structure. For "bottom-up" protein analysis, the protein must first

be enzymatically digested prior to mass spectrometric analysis. One of the most

commonly used proteolytic enzymes for this purpose is trypsin, which cleaves the protein

backbone from the N-terminal end after lysine and arginine residues. The presence of

proline after a lysine or arginine residue will prevent this cleavage from occurring due to

the kink in the protein backbone caused by the proline residue. Provided the protein

sequence is known, the expected masses of peptides resulting from an enzymatic

digestion can be calculated and used for peptide identification. Additionally, the masses

of modified peptides can also be calculated and used to quantify the extent of modification.

Separating peptides prior to MS simplifies the peptide mixture resulting from enzymatic digestion that reaches the instrument at any given time. Liquid chromatography (LC) is the most commonly used technique for this purpose; proteins and peptides are too large and too polar to be analyzed by gas chromatography. Reversed-phase liquid chromatography (RPLC) is the most frequently used form of LC employed for this purpose.

As discussed previously, RPLC utilizes a nonpolar, hydrophobic stationary phase and a polar mobile phase composed of aqueous and organic solvents. Mobile phase strength increases with organic content, and the concentration of organic solvent is increased throughout the run to cause elution as the analytes increasingly favor the mobile phase over the stationary phase. Elution is in order of decreasing polarity and increasing hydrophobicity. Typical stationary phases for reversed phase separations of peptides are C_{18}. Typical solvents include methanol, acetonitrile, and isopropyl alcohol, often mixed with trifluoracetic acid to improve ionization efficiency. Following separation, aqueous peptides must be ionized prior to mass spectral analysis. The most frequently employed ionization source for proteomic investigations is electrospray ionization, which can generate multiply charged ions. This enables the analysis of larger molecules that may otherwise be outside the mass range of the mass spectrometer.

2.5.1 Electrospray Ionization (ESI)

Electrospray ionization (ESI) is an atmospheric pressure ionization source that is especially well-suited for reversed-phase liquid chromatography detection. Ions are

produced through electrochemical processes which enable the formation of multiply

charged ions. Ions are produced in ESI through the application of an electric field to a

liquid traversing a capillary with weak flux: a potential difference of 3-6 kV is applied

between the capillary and a counter electrode separated by 0.3-2 cm, generating electric

fields on the order of 10^6 Vm^{-1} which cause charge accumulation on the liquid surface as

it exits the capillary. This ultimately breaks to yield highly charged droplets, which are

subsequently subjected to a coaxially injected inert gas with a low flow rate, limiting the

spatial dispersion of the resulting spray. The surface tension of the solvent in use dictates

the required onset voltage for the sprayer tip: at low voltages, a spherical droplet is

formed at the tip of the capillary; as the voltage increases, the drop becomes elongated

due to the pressure resulting from charge accumulation in the stronger electric field.

When the onset voltage is reached, the surface tension of the droplet is exceeded by the

pressure of the accumulated charges and the droplet takes the form of a Taylor cone.

Small droplets are released from this Taylor cone which quickly divide and decompose to

form a spray. These droplets further decompose as the solvent evaporates and causes the

charge per unit volume to increase with the shrinking droplet size. As the solvent

evaporates and the electrostatic repulsion of like charges on the surface increases, the

droplet becomes increasingly unstable as it approaches its Rayleigh limit—the point at

which the electrostatic repulsion is greater than the surface tension of the droplet—and

eventually undergoes Coulombic fission, exploding into smaller, more stable droplets.

About 20 smaller droplets are released from this second Taylor cone, possessing

approximately 2% of the volume of the precursor droplet and 15% of the charge of the

precursor. While the initial droplet released from the capillary carries about 50,000

elementary charges (about 10^{-4} C) and has a diameter of roughly 1.5 µm, the secondary

droplets carry 300-400 elementary charges with a diameter of approximately 0.1 µm.

These smaller droplets can undergo further desolvation and Coulombic fission. The initial

droplet size can be reduced by adding components to the solvent that increase

conductivity, such as formic acid.[42]

Several different theories attempt to explain the final production of gas-phase ions

from these liquid droplets: for low molecular weight species, the ion evaporation model

(IEM) is proposed. The IEM describes the ejection of small solvated ions from the

droplet surface as a result of the electric field emanating from the Rayleigh-charged

droplet; the departing ion remains connected to the initial droplet through an extended

solvent bridge that ruptures upon release of the solvated ion. This results in a gas-phase

cluster of the ion and a few associated solvent molecules which are lost through

background gas collisions in the sampling interface of the mass spectrometer. Analytes

residing in the interior of the droplet experience reduced sensitivity to ionization

compared with those closer to the surface—the practical significance of this is that

surfactants and nonvolatile salts must be avoided in spray solvents to avoid signal

suppression.[43]

Another mechanism for ion release, the charged residue model (CRM), describes

the release of large globular analytes into the gas phase. In the CRM, the Rayleigh-

charged droplet contains a single analyte to which the charge of the desolvating droplet is

transferred. As the solvent evaporates, the CRM droplet remains close to the Rayleigh

limit, indicating a charge reduction along with the decreasing radius. This charge

reduction can be explained by IEM ejection of solvated protons, although IEM ejection of

large analytes is typically not kinetically viable. In the case of peptides, both IEM and

CRM can be viable ESI mechanisms. The prevalence of one occurrence over the other is

dependent on droplet size: IEM ejection is favorable only for larger droplets with radii

between 4-5 nm, while CRM is favored for smaller droplets.[43, 44]

2.5.2 Linear Ion Traps

Following ionization, ions are transferred to a mass analyzer in order to obtain a

spectrum of their mass-to-charge ratios (*m/z*). The mass spectrometer utilized in this

research is a Thermo Velos Pro LTQ, in which the mass analyzer is a dual-pressure linear

quadrupole ion trap. Ions are accumulated in the higher-pressure cell (5.0 mtorr), then

transferred to the lower-pressure cell (0.35 mtorr) for mass analysis. A linear ion trap

consists of a set of four parallel segmented electrodes with a hyperbolic cross section. An

oscillating radio frequency (RF; 1.2 MHz) potential is applied across opposing pairs of

electrodes. This quadrupolar field confines ions in the radial (*xy*) direction, and a DC

potential is applied to the end segments to confine ions in the axial (*z*) direction. A

positive potential is required to trap positive ions, and a negative potential is required to

trap negative ions. Upon entry into the higher-pressure cell, collisions with He gas cool

the ions as they traverse the *z*-axis between the end electrodes while oscillating in the *xy*-

plane as a function of the RF potential applied to the rods. After accumulation and

cooling, trapped ions are transferred to the analyzer cell. Mass-dependent ejection from

low to high is accomplished by increasing the main RF voltage while simultaneously

applying a supplemental AC (alternating current) resonance excitation potential across

opposing *x*-electrodes. The AC excitation potential has lower amplitude and frequencies

than the RF trapping potential. Ejected ions exit the trap in the *x*-direction through narrow

slits in the electrodes, then strike an electron multiplier for detection. Resonance ejection can also be used to isolate ions in a narrow m/z range for further analysis, as in "tandem" mass spectrometry (MS/MS).[42]
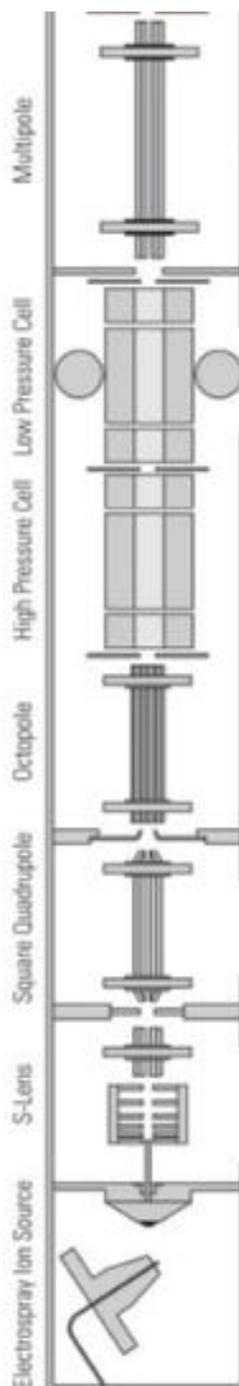
Figure 2.6: Schematic of Velos LTQ Pro MS instrument from LTQ Hardware Manual. Used without permission.

2.4.3 Tandem Mass Spectrometry

Tandem mass spectrometry entails multiple stages of mass spectrometric analysis, the first of which is the acquisition of a full scan of all intact peptide ions eluting from the column at a given point in time. While this precursor spectrum can give information on the m/z value of intact peptides, it is insufficient for peptide identification and sequential analysis. For this purpose, tandem mass spectrometry (MS/MS or MS^2) can be used. In MS/MS, the first mass analyzer stage is used to select a specific precursor ion. The mass-selected precursor is sent to a collision cell for fragmentation, and a spectrum of the fragments is obtained by the second mass analyzer stage.[29, 32, 36, 37] The Thermo Velos LTQ Pro utilized in this research performs tandem-in-time mass spectrometry, in which both the precursor ion selection and subsequent fragmentation (and possibly product-ion mass analysis) are performed in the same physical mass analyzer.

Peptides tend to fragment along the backbone, with the amide linkages along the peptide backbone being the preferred protonation sites during ionization. These protonated amide linkages are thus weakened and susceptible to breakage upon collisions with inert gas during the collision-induced dissociation (CID) fragmentation step. Energy imparted from the collisions is redistributed throughout the peptide, allowing it to cleave at various peptide bonds along the backbone. This generates fragments in the MS/MS spectra that provide sequence information directly. Fragments in which the charge resides on the N-terminal fragment are referred to as b-ions, and fragments in which the charge resides on the C-terminal fragment are referred to as y-ions. Additional "satellite" ions may also be produced from these ion types through further loss of NH_3 or H_2O.[41] For

singly-modified peptides with multiple possible modification sites, the fragment spectrum of the precursor can be used to identify the specific site of modification.

For our investigations into the surface-induced conformational changes of BSA, we will use mass spectrometric protein footprinting to evaluate solvent-accessibility of residues in native and adsorbed BSA conformers under typical RPLC conditions.
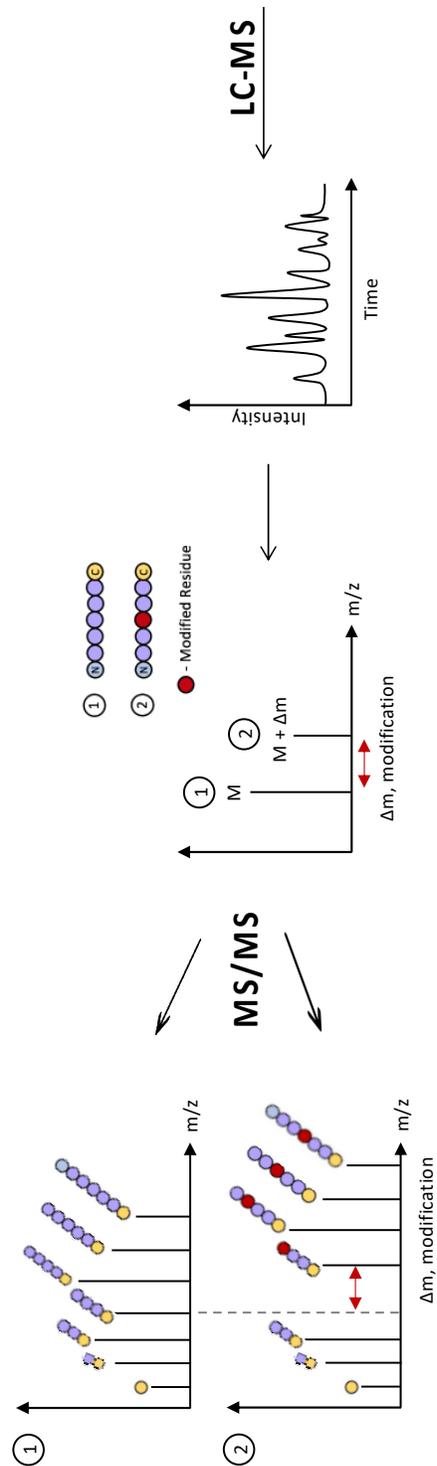
Figure 2.7: Tandem mass spectrometry illustration. For modified peptides with ambiguous amidation sites, MS/MS can be used to determine the exact location of the modification.

CHAPTER 3: MATERIALS AND METHODS

3.1 Capillary Zone Electrophoresis

Capillary zone electrophoresis experiments were conducted on a custom-made instrument. A Spellman (Hauppauge, NY) CZE 1000R power supply was used to provide high voltage (± 30 kV) and a Linear Instruments Model 200 UV-Vis detector was for analyte detection. Fused silica capillaries (50 µm i.d., 375 µm o.d.) from Polymicro Technologies (Phoenix, AZ) were cut to a length of 50.0 cm, and a detection window was made 5.0 cm from the outlet end of the capillary for a separation length of 45.0 cm. The detection window was made by removing the polyimide capillary coating with several drops of concentrated sulfuric acid boiled with the tip of a soldering iron. Samples were injected electrokinetically

Electrokinetic sample injection was performed at 20 kV for 2 sec for protein samples and 1 sec for injections of a 0.01% v/v mesityl oxide (MO) neutral marker. Run voltages were set to 20 kV, resulting in currents ranging from 15-25 uA. Runs were performed at ambient temperatures. The run buffer used was 60 mM TES, pH 7.40, and the capillary was rinsed with run buffer at the beginning and end of the day and between samples. Mesityl oxide was co-injected with each protein sample to generate a clean neutral marker peak in the electropherograms for comparison. The capillary was rinsed with water at the end of day and held in water between uses.

3.2 Bradford Assay

Eight standard BSA solutions with concentrations ranging from 0-2 mg/mL were prepared by serial dilution. A 10 µL aliquot of each standard was added to 300 µL of room temperature Coomassie Plus protein assay reagent. A blank was prepared by adding

35 µL of deionized water to 1050 µL of Coomassie Plus reagent. All unknown BSA samples were treated in the same manner as the standards: 10 µL of sample was added to 300 µL of the Coomassie Plus reagent. For volume-limited samples, only 2 µL of the protein sample was added to 60 µL of Coomassie Plus reagent. Replicate absorbances were collected for each test solution at 595 nm using a NanoDrop UV Vis spectrophotometer. A calibration curve was constructed using the average absorbances of each standard solution. The linear region was determined by examining a residual plot for the curve and used to determine unknown concentrations of BSA.

For all samples modified on-column, 2 µL of the eluted protein solution were added to 60 µL of Coomassie Plus reagent without vortexing. A sharp color change of the Coomassie Plus reagent from purple to blue as the low-concentration protein sample diffused into the reagent indicated successful protein recovery. This color change was used to confirm the elution of the protein prior to tryptic digestion.

3.3 Development of Adsorption Experiments

3.3.1 Protein Modification in Solution

Phosphate buffer (10 mM, pH 7.40) was prepared from sodium phosphate monobasic dihydrate obtained from Fluka (Buchs, Switzerland) and adjusted to the desired pH with NaOH. Methylamine hydrochloride (MA) was obtained from Aldrich Chemical Company (Milwaukee, WI) and EDAC (1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride, 98+%) was obtained from Acros Organics (Geel, Belgium). Bovine serum albumin (BSA, min. 99% fatty acid and globulin free) was obtained from Sigma (Burlington, MA).

Modification experiments were first performed on proteins in solution to validate our method for using EDAC and MA to amidate carboxylate-containing residues on BSA. Phosphate buffer at physiological pH was selected as the reaction medium in order to most closely reflect the solvent conditions of a typical RPLC separation of BSA. Fresh BSA stock solutions ranging in concentration from 0.5-2 mg/mL were freshly prepared daily for modification experiments. For experiments using low concentrations of EDAC (>10 mM, 1.92 mg/mL) to modify the protein in phosphate buffer at physiological pH, no measurable extent of modification was noted.

It was found that by increasing the concentration of EDAC to ~52 mM (10.00 mg/mL), the buffered protein solution could be modified in the presence of ~70 mM (13.23 mg/mL) MA. Capillary electrophoresis was used to confirm success of the reaction and estimate the extent of modification.
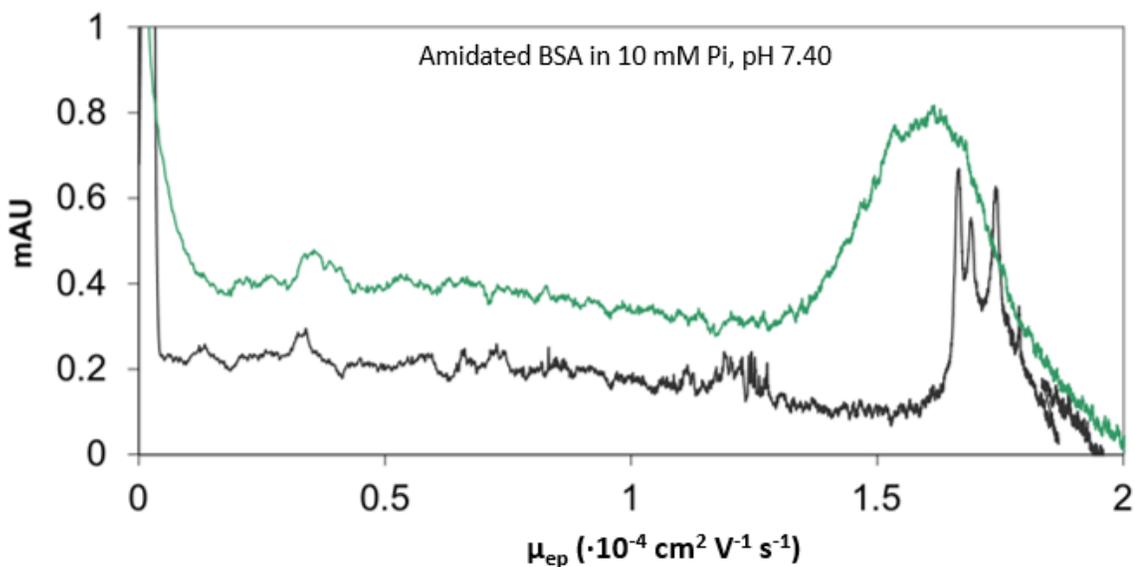
Figure 3.1: Overlay of electropherograms for modified BSA (top, green) and unmodified BSA (bottom, black), both 2.0 mg/mL in 10 mM Pi, pH 7.40.

Figure 3.1 shows an overlay of unmodified BSA (bottom, black) with BSA modified in phosphate buffer at pH 7.40 (top, green). The structure of the unmodified BSA peak in the electropherogram is a result of the microheterogeneity of the protein. Each subpeak within the broad protein peak corresponds to an oxidation state of the free Cys34 residue and were identified in previous group research conducted by Xu Ying. The leftmost peak corresponds to reversibly reduced BSA, with a charge state between -1 and 0. The middle peak corresponds to BSA-S-S-Cys and potentially a small fraction BSA-S-S-homocysteine, both with a charge contribution of 0. The rightmost peak corresponds to the irreversibly oxidized forms of BSA: $BSA-SO_2^-$ and $BSA-SO_3^-$, sulfinic and sulfonic acid, respectively. A small amount of BSA-S-S-glutathione also contributes. All irreversibly oxidized forms have a charge contribution of -1. Because the separation of these peaks is a result of the ± 1 contribution, the distance separating them yields a 'charge ladder' that can be used to determine the change in net charge of a protein.

The conversion of negatively charged carboxylates to neutral amide bonds upon modification lowers the net negative charge of the protein, reducing its electrophoretic mobility. This causes a shift in the elution time of the modified protein, as seen in Figure 3.1. The difference in electrophoretic mobility between the two peaks (measured at the centroid of the peak maximum) is used to estimate the extent of modification using the previously described 'charge ladder'. The figure shown demonstrates an early version of the modification experiments using low concentrations of EDAC, resulting in an estimated charge difference of +2 for the modified protein corresponding to two amidated residues per protein. This technique was used as to guide the development of ideal reaction conditions to be used for adsorption experiments. In order to achieve an extent of

modification sufficiently high to be detected by mass spectrometry, the final

concentrations of reactants for adsorption experiments were held at 222 mM EDAC and

395 mM MA. The possibility of performing the modification in a solution of 50%

acetonitrile was also checked and confirmed.

The reaction between EDAC and carboxylates to form the previously discussed

O-acylisourea intermediate is the rate-determining step. The presence of a primary amine

results in rapid formation of the amide bond, on the order of a few milliseconds. Due to

competition with hydrolysis, timing of the amidation reaction is of particular importance.

Although some protocols advise an incubation period prior to the addition of the primary

amine in order to active carboxylates, it was not found to be necessary. Simultaneous

addition of both reagents in the form of a pre-combined reaction mixture was sufficient

for modifying the proteins, and this approach was used for adsorption experiments.

3.3.2 Column Preconditioning and Protein Adsorption

Packing material was removed from HyperSep C_8 Solid Phase Extraction (SPE)

cartridges obtained from Thermo Electron Corporation (Waltham, MA) for use in

adsorption experiments. Empty 0.8 mL Pierce centrifuge columns with a polyethylene

filter (~30 µm pore size) were purchased from Thermo Scientific (Rockford, IL).

Packing material (>300 mg) was added to empty centrifuge spin columns and

preconditioned in a multi-step process with centrifugation steps between each solvent

addition. A 400 µL aliquot of acetonitrile (ACN) was added to the column and

centrifuged at 100 x g for 2 minutes in a Sorvall Legend Micro 21R centrifuge to

thoroughly saturate the packing material. This preconditioning step was followed by five

200 µL aliquots (1 mL total) of phosphate reaction buffer (10 mM, pH 7.40). The first

four aliquots were centrifuged for 2 minutes at 100 x *g*. The final aliquot was centrifuged

at 100 x *g* in 30 second intervals until the supernatant had fully soaked into the column,

checking between each interval to ensure the column remained saturated. The

equilibrated column was treated with 250 µL of 2.0 mg/mL BSA prepared in phosphate

reaction buffer. The protein solution was centrifuged into the packing material at 100 x *g*

in 30 second intervals until there was no remaining supernatant.

Adsorption of added protein was confirmed using both CZE and a Bradford assay.

Eluent from the addition and centrifugation of BSA through the column was collected in

a clean microcentrifuge tube. A 2 µL aliquot of the stock protein solution was deposited

on the surface of 60 µL of Coomassie Plus protein assay reagent and compared with

aliquots of the recovered flow-through treated identically. In the absence of mixing,

diffusion of the protein-containing solutions into the Coomassie Plus reagent rendered a

color change from purple to blue that was limited to a narrow band at the surface of the

solution. A less intense shade of blue for flow-through samples compared with that of the

stock solution indicated successful adsorption.

3.3.3 Modification and Elution of Adsorbed Proteins

A reaction mixture was made by mixing 150 µL of 222 mM EDAC and 200 µL of

MA prepared in phosphate buffer. This 350 µL mixture was added to the column and

centrifuged into the packing material in 30 second increments at 100 x *g*. Columns were

incubated at 30 °C and the reaction was allowed to proceed for at least an hour. The

reaction mixture was then removed by centrifuging the column at 1,500 x *g* for 1 minute,

at which point further centrifugation yielded no eluent. To elute the modified protein, 200

µL of acetonitrile was added and gently centrifuged into a 2 mL microcentrifuge tube for

5 minutes at 100 x *g*. This setting is insufficient to fully remove solvent from the packing material, however. To ensure all eluted protein had been collected, 200 µL of digestion buffer (100 mM ammonium bicarbonate, pH 7.98) was added to the column and centrifuged into the collection vial for 1 minute at 1,500 x *g*.

Because of the low volume and concentration of protein recovered from these experiments (approximately 20 µL of 0.5 mg/mL BSA, on average), the use of an intact SPE cartridge (500 mg) was explored as a possibility for improving recovery. A 3-mL syringe was attached to a Luer fitting positioned at the outlet of the cartridge and used to pull solvents and reagents through the intact cartridge. Previously described solvent and reagent volumes were commensurately increased to account for the larger mass of packing material and ensure sufficient wetting. While this did result in slightly higher protein recovery, this approach was less suitable than the described adsorption experiments performed in miniature spin columns. Specifically, the larger volume of eluent (7.5 mL) proved unwieldy—the required process of concentrating the large volume (in 500 µL iterations) was inefficient and not justified by the negligible increase in protein recovery.

Additionally, a one-vial approach in which the packing material, protein solution, reaction mixture, and eluting solvents are added to a single microcentrifuge tube was considered. The inability to effectively remove excess protein and reagents, in conjunction with the limited volume of the vial rendered this approach inferior to the simulated-column experiments described here.

The limited volume and low concentration of recovered proteins prevented the use of CZE for estimating the extent of modification of eluted, modified BSA. Matrix-

assisted laser desorption/ionization (MALDI) mass spectrometry was considered as an

alternative technique for estimating the extent of modification for adsorbed proteins.
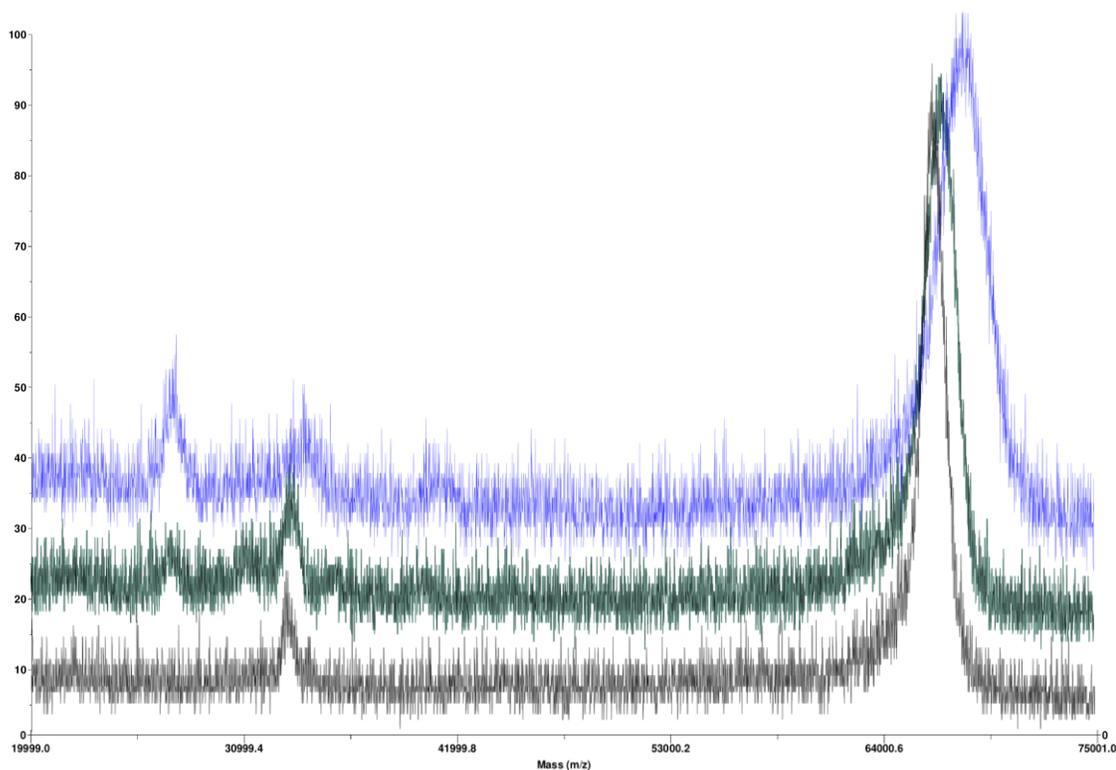
Figure 3.2: MALDI-MS spectral overlay of unmodified BSA (m/z = 66538; black, bottom), BSA amidated in-solution (m/z = 67059; 10 mM Pi, pH 7.40; green, middle), and BSA amidated following adsorption to C_8 SPE packing material (m/z = 68165; top, blue).

A sinapinic acid (10 mg/mL) matrix was dotted onto the sample plate of a

Voyager MALDI-MS along with protein sample. Unmodified BSA and BSA modified in

phosphate buffer were used as controls. MALDI-MS of the unmodified protein yielded a

peak with the expected molecular weight of BSA. Both of the modified protein samples

showed an expected increase in m/z value, with the protein modified on-column

demonstrating a greater m/z value than that of the protein modified in-solution. However,

as can be seen in Figure 3.2, the resolution of the instrument was insufficient for

quantifying an extent of modification and this technique was used sparingly.

During protocol development for adsorption experiments, CZE was repurposed to

confirm successful adsorption of protein to the stationary phase of loose packing

material. Loose packing material was weighed into a microcentrifuge tube and 200 µL of

10 mg/mL BSA in 10 mM Pi, pH 7.40 was added and vortexed to form a slurry.

Supernatant remaining on the packing material surface was injected immediately (less

than 5 minutes post-addition), and the slurry was incubated at 30 °C for 45 minutes

before the supernatant was injected again. Overlaying the electropherograms

demonstrated a reduction in the concentration of BSA in the supernatant, confirming

adsorption to the stationary phase and loss of protein in free-solution.
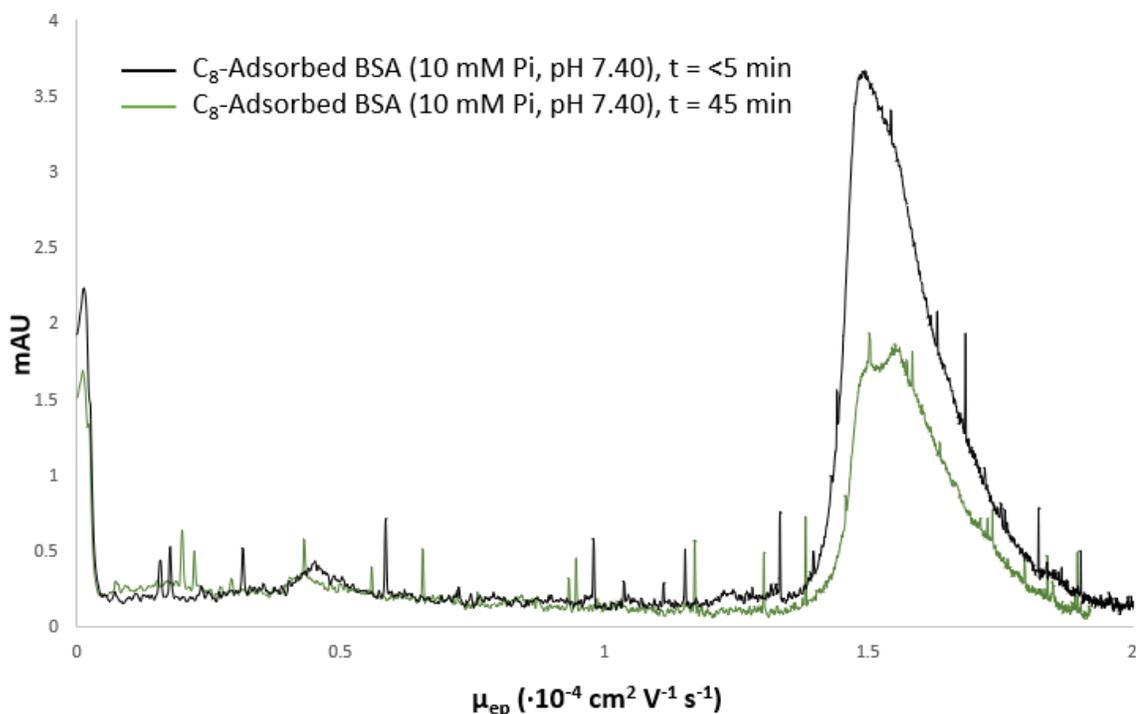
Figure 3.3: Electropherogram of protein at different time points following addition of C_8 SPE packing material. The decrease in absorbance over time indicates successful adsorption of protein to the stationary phase. The detection wavelength was 200 nm, and the CZE run buffer was 60 mM TES, pH 7.98.

3.3.4 Simulated RPLC Experiments

Proteins were modified at various incubation intervals (5 min, 30 min, 60 min, >24 hrs) and post-desorption, as well as under organic conditions ranging from 0-40% ACN. To investigate the retention of conformational changes following desorption into the mobile phase, the unmodified protein was eluted from the column with acetonitrile and concentrated using Amicon Ultra-0.5 centrifugal filters. The concentrated protein was diluted in phosphate reaction buffer and the amidation reaction was performed in the same manner as a typical in-solution protein modification.

To simulate the effect of increasing the organic content of the mobile phase throughout the run, proteins were allowed to incubate for approximately 15 minutes on the column in phosphate buffer. They were then treated with 200 µL of a mixture of acetonitrile and phosphate buffer combined at the concentration of interest. This simulated mobile phase was centrifuged into the packing material in 30 second intervals at 100 x g. The adsorbed protein was incubated under these simulated high-organic conditions at 30 °C for 15 minutes prior to modification.

Control samples were made by first incubating 200 µL of 2.0 mg/mL BSA for 15 minutes at 30 °C. A reaction mixture was made by combining 150 µL of 222 mM EDAC with 200 µL of MA. The reaction mixture was added to the protein solution and briefly vortexed. The reaction was allowed to proceed for one hour at 30 °C, and excess reagents were removed from the solution using Amicon Ultra-0.5 centrifugal filters with a 30 kDa molecular weight cutoff from Sigma Aldrich.

3.3.5 Buffer Exchange

A buffer exchange was performed on modified proteins prior to tryptic digestion. Phosphate buffer from the modification reaction was exchanged with digest buffer (100 mM ammonium bicarbonate, pH 7.98). Eluted proteins modified on-column were preconcentrated and excess reagents removed from solution using Amicon Ultra-0.5 centrifugal filters (30 kDa MWCO). Protein samples eluted with pure acetonitrile were diluted with digest buffer to 20% acetonitrile (1 mL final volume) prior to preconcentration in order to achieve a solvent composition amenable to the filter devices. Diluted protein samples were added to the 500 µL centrifugal filters and centrifuged for 10 minutes at 14,000 x g. The filter was then inverted into a new microcentrifuge tube and the concentrated protein was recovered by centrifuging the vial at 1,000 x g for 2 minutes. The approximate volume recovered from each 500 µL aliquot was 20 µL of 0.5-2 mg/mL BSA, depending on the amount of packing material used. Elution of the protein was confirmed with Coomassie Plus assay reagent as described above prior to digestion. The presence of a blue color change was a reliable indicator of eluted protein.

3.4 Tryptic Digestion of Modified BSA

A buffer exchange was performed on modified proteins prior to tryptic digestion. Phosphate buffer from the modification reaction was exchanged with digest buffer (100 mM ammonium bicarbonate, pH 7.98). Eluted proteins modified on-column were preconcentrated and excess reagents removed from solution using Amicon Ultra-0.5 centrifugal filters with a 30 kDa molecular weight cutoff from Sigma Aldrich. Protein samples eluted with pure acetonitrile were diluted with digest buffer to 20% acetonitrile prior to preconcentration in order to achieve a solvent composition amenable to the filter

devices. Diluted protein samples were added to the 500 µL centrifugal filters and centrifuged for 10 minutes at 14,000 x g. The filter was then inverted into a new microcentrifuge tube and the concentrated protein was recovered by centrifuging the vial at 1,000 x g for 2 minutes. The approximate volume recovered from each 500 µL aliquot was 20 uL of 0.5-2 mg/mL BSA, depending on the amount of packing material used.

Digest buffer was added to bring the concentrated sample volume to approximately 100 µL. The protein was reduced by adding 5 µL of 200 mM dithiothreitol (DTT) and placing the sample in a hot water bath for 30 minutes. To prevent disulfide bonds from reforming, 5 µL of 1 M iodoacetamide (IAM) was added to the BSA sample and incubated in the dark at 37 °C for 30 minutes. This alkylation reaction was quenched by adding 20 µL of 200 mM DTT and incubating at 37 °C for 45 minutes. A 1 mg/mL stock solution of trypsin was made in 1 mM HCl and frozen between uses. This stock solution was used to perform a two-step tryptic digestion. Two microliters of 1 mg/mL trypsin in 1 mM HCl were added to each BSA sample and incubated at 37 °C. A second 2 µL aliquot of trypsin was added to each sample after approximately an hour, and the digestion mixture was incubated overnight (>18 hours) at 37 °C. Digested BSA samples were frozen until ready for analysis.

The 200 mM DTT and 1 M IAM were prepared in 100 mM ammonium bicarbonate, pH 7.98. DTT and IAM were obtained from Thermo Scientific (Waltham, MA). Vials containing 20 µg of trypsin were purchased from Thermo Scientific (Rockford, IL), and 20 µL of 1 mM HCl was added to prepare the 1 mg/mL trypsin-HCl solution.

3.5 LC-MS/MS

Tryptic digests of modified BSA were chromatographically separated using a
Thermo Vanquish LC system and detected using a Thermo Velos Pro LTQ mass
spectrometer. A Dionex Acclaim 120 C18 column (2.1 x 100 mm, 3 µm particle size)
was used to separate the tryptic peptides. Reservoir A held LC-MS grade water with
0.1% formic acid, and Reservoir B contained LC-MS grade acetonitrile with 0.1% formic
acid. The column was rinsed with 95% B prior to each set of runs. A gradient LC method
was developed in which the column was pre-equilibrated with 5.0% B for 10 minutes
prior to sample injection. The flow rate was held at 250 µL/min, and the injection volume
for each sample was 10.0 µL. After column equilibration to 5.0% B and sample injection,
the solvent composition was held at 5.0% B for 2.50 minutes. A linear solvent gradient
from 5.0 to 45.0% B over 52.50 minutes was followed by a faster cleaning ramp from
45.0 to 95.0% B over 2.00 minutes. The column was held at 95.0% B for 10.00 minutes
to rinse off strongly retained compounds, and returned to 5.0% B over 0.50 minutes to
prepare for the next run. The total run length was 68 minutes. The MS detector was
turned off during the pre-equilibration step and the initial 2.50 minute hold at 5.0% B.
Mass spectrometry data was collected from the start of the gradient until the 95.0% B
wash step.
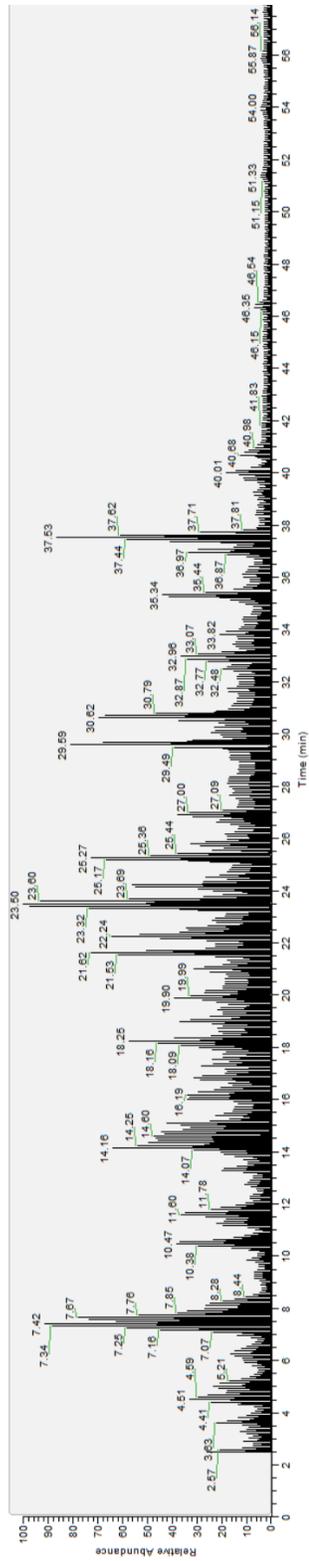
Figure 3.4: Example total ion chromatogram (TIC) of a tryptic digest for a modified BSA sample.

3.6 Data Dependent Acquisition

Tryptic digests were analyzed using a variety of "Big n" experiments, with n = 5, 10, 15, etc., on a Thermo Velos LTQ Pro mass spectrometer. Big 5 refers to an experimental method in which a full-scan MS^1 spectrum is followed by five successive MS^2 scans of the top of five most abundant peaks in the preceding full scan. Big 10 and Big 15 experiments were also performed. Large contaminant peaks in the background can complicate these analyses and prevent the acquisition of useful data, so an exclusion list of potential or observed contaminants is added to the method to allow contaminants to be ignored. Additionally, modified peptides are not expected to be as abundant as unmodified peptides due to the low extent of protein modification. A precursor mass list with expected m/z values for modified peptides was generated and included in the search parameters: if a precursor on the mass list was seen by the instrument in the full scan, it would be preferentially selected for fragmentation over other otherwise more abundant peaks.

"Dynamic exclusion" was also used. After a tandem spectrum has successfully been obtained for a particular precursor, the precursor m/z is added to an exclusion list for approximately the width of the chromatographic peak, typically 30-90 seconds. This greatly improves peptide coverage by allowing lower-abundance precursors to be selected for fragmentation, even in the presence of dominant high-abundance ions.

3.7 Post-Acquisition Data Analysis

Following MS/MS analysis, peptides were identified by the MS-GF+ search engine, using SearchGUI and Peptide Shaker as graphical user interfaces for running the search and viewing the results. Sequence searching—in which the search engine iterates

over a FASTA file of the protein sequence(s) and generates theoretical MS/MS spectra for comparison with observed spectra—can be used to identify potential peptide matches.[45, 46] Theoretical spectra generated by the search engine consist of b- and y- ions in addition to their satellite ions. MS-GF+ is a broadly useful tool for identifying peptides that scores experimental MS/MS spectra against those generated from a known protein sequence.[47]

This research utilizes trypsin, a proteolytic enzyme that cleaves on the C-terminal side of lysine and arginine residues, unless followed by proline. Because both the sequence of BSA and the cleavage points are known, the sequences and masses of the peptides generated from a tryptic digestion can be calculated. The m/z values for likely charge states expected to be observed in the mass spectra can also be calculated. Furthermore, because the modification sites are restricted to aspartic and glutamic acid residues, the expected masses and m/z values of modified peptides can be calculated in addition to those of the unmodified peptides. This provides an additional benefit to using sequence searching over a library search, in which the MS/MS spectra are compared with those in a library of previously acquired spectra: MS-GF+ is capable of handling a variety of digestion conditions, experimental protocols, instruments and fragmentation methods that might generate spectra unlikely to be in a spectral library. The search engine was only supplied with the FASTA sequence file for BSA, as it is the only [non-contaminant] protein in our samples. PeptideShaker is then used to interpret the proteomics identification results generated from search engines.

Theoretical spectra for peptides with specified modifications can also be generated. Modifications are defined as either "static" or "dynamic." Static modifications

are expected to be present invariably, meaning the search engine will only search for

peptides containing that modification and theoretical spectra will not be generated for an

unmodified version of the peptide. Dynamic or variable modifications are expected to be

present inconsistently: theoretical spectra will be generated for both the unmodified and

the modified versions of the peptide, and both will be sought in the experimental data.

Search parameters included treating alkylation of cysteine with iodoacetamide as

a static modification. Amidation of glutamic and aspartic acid residues, as well as the C-

terminal, was considered a variable modification. Mono- dioxidation of methionine were

also included as potential variable modifications. The n-acylisourea side product was also

included as a possible static modification, however, no peptides containing this

modification were confidently identified under the final reaction conditions. Other search

parameters included limiting the number of missed tryptic cleavages to two, and

searching exclusively for +1, +2, and +3 charge states. The precursor and fragment mass

tolerances were set to 0.3 Da.

Table 3.1: Modifications included in MS-GF+ search parameters.

Modification	Target Residue	Mass Shift (Da)
Fixed		
Carbamidomethylation	Cys	+ 57.02
Variable		
Oxidation	Met	+ 15.99
Dioxidation	Met	+ 31.99
Amidation	Glu, Asp, C-terminal	+ 13.03
N-acylisoureation	Glu, Asp, C-terminal	+ 155.14

Peaks at specific retention times in the LC-MS precursor scan are assigned to corresponding peptides identified from the MS/MS spectra. Relative response factors are used for quantification. This is achieved by integrating extracted ion chromatograms (XIC) of the LC-MS spectra and comparing the ratio of modified and unmodified peptides.[29] Xcalibur (the LTQ instrument software) was used for viewing and integrating chromatographic and MS/MS data. PeptideShaker was used to visualize potential peptide matches generated from the sequence search. Modified and unmodified peptides identified in the experimental data were presented with the associated spectrum number and chromatographic retention time.

CHAPTER 4: FOOTPRINTING RESULTS AND DISCUSSION

4.1 Peptide Identification

Manual inspection of the spectra was required to confirm each potential peptide identification: data collected in profile mode was used to confirm the correct charge state had been identified and assigned, and XICs were generated for each match to confirm that the result corresponded to a real chromatographic peak. We do not have a high-resolution instrument so some of the peak areas are probably biased by the inclusion of nearly isobaric co-eluting peaks. But since we are analyzing a purified protein and not a complex mixture of proteins, this is less likely to be a problem.

The m/z values of observed charge states for unmodified peptides and their associated retention times are presented in Table 4.1. Peptides containing a single modification were identified, and their associated retention times and m/z values of observed charge states are presented in Table 4.2. Retention times and m/z values observed for peptides with multiple modified sites are similarly collected in Table 4.3.

Table 4.1: Unmodified peptides identified in the chromatographic separation of tryptically digested BSA. Retention times are given in a range of typical peak locations and identified charge states are listed. The potentially modifiable residue(s) associated with each peptide are listed.

Modifiable Site	Tn-m	Sequence	RT (min)	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$
D13, E16, E17	T3-4	FKDLGEEHFK	17.10	1249.62	625.31	417.21
D13, E16, E17	T4	DLGEEHFK	15.20	974.46	487.73	352.49
D37, E38	T5	GLVLIAFSQYLQQCPFDEHVK	40.05		1246.64	831.43
E45, E48	T6	LVNELTEFAK	26.75	1163.63	582.32	-
D56, E57, E63	T7	TCVADESHAGCEK	5.91	1463.59	732.30	488.53
D72, E73	T8	SLHTLFGDELCK	26.90	1419.69	710.35	473.90
E82, D86, D89, E92	T10	ETYGDMADCCEK	15.60	1478.52	739.77	493.51
E100	T12	NECFLSHK	14.35	1034.47	517.74	345.50
E100, D107, D108, D111	T12-13	NECFLSHK DDSPDLPK	19.70-19.80	1901.87	951.44	634.63
D107, D108, D111	T13	DDSPDLPK	14.40	886.42	443.71	300.48
D118, D124, E125	T14	LKPDPNTLCDEFK	23.20-23.50	1576.77	788.89	526.26
E152	T19-20	RHPYFYAPELLYYANK	30.56	-	1023.02	682.35
E152	T20	HPYFYAPELLYYANK	33.10-33.40	1888.93	944.97	630.31

E166, E171	T21	YNGVFQECCQAEDK	19.40-19.60	1747.71	874.36	583.23
E226, E229	T34	AEFVEVTK	18.0-18.25	922.49	461.75	-
D236	T35	LVTDLTK	16.75	789.47	395.24	-
E243, D248, E251, D254	T37	ECCHGDLLECADDR	18.65	1749.66	875.33	583.89
D265, D268	T39	YICDNQDTISSK	15.10	1443.64	722.32	481.84
E276, D279, E284	T40-41	LKECCDKPLLEK	14.62	-	652.82	435.55
E276, D279, E284	T41	ECCDKPLLEK	14.55	-	766.89	511.60
E291, E293	T42	SHCIAEVEK	11.10	1072.51	536.76	358.17
D295, E299, D307, E310	T43	DAIPENLPPLTADFAEDK	32.47	1955.96	978.48	652.99
E320	T45	NYQEAK	17.40	752.36	376.68	-
E320, D323, E332	T45-47	NYQEAKDAFLGSFLYEYSRR	30.97	-	1229.10	819.73
D323, E332	T46	DAFLGSFLYEYSR	37.36	-	784.38	-
E339	T47-48	RHPEYAVSVLLR	23.95	1439.81	720.41	480.61
E339	T48	HPEYAVSVLLR	26.68	1283.71	642.36	428.58
E351, E353, E357, E358	T50	EYEATLEECCAK	17.52	1502.61	751.81	506.54
D363, D364, D374	T51	DDPHACYSTVFDK	20.88	1554.65	777.83	518.89
D381, E382	T53	HLVDEPQNLIK	21.53	1305.72	653.36	435.91

D392, E395	T54	QNCDQFEK	9.73	1068.44	534.72	-
E399	T55	LGEYGFQNALIVR	30.65	1479.80	740.40	-
E424	T57-58	KVPQVSTPTLVEVSR	23.08-23.77	1639.94	820.47	547.32
E424	T58	VPQVSTPTLVEVSR	24.87-25.59	1511.84	756.43	-
E449, D450	T62	MPCTEDYLSLILNR	37.55	1724.83	862.92	-
E464	T63	LCVLHEK	14.13	898.48	449.74	300.17
E470	T64	TPVSEK	5.18	660.36	330.68	-
E478	T65-66	VTKCCTESLVNR	25.07	1466.71	733.86	489.57
E478	T66	CCTESLVNR	14.60	1138.50	569.75	-
D493, E494	T67	RPCFSALTPDETYVPK	25.10		940.96	627.65
D511, D517, E519	T69	LFTFHADICTLPDTEK	29.60	1907.92	954.46	631.99
E530	T71-72	KQTALVELLK	25.71	1142.71	571.86	381.58
E530	T72	QTALVELLK	29.66-29.72	1014.62	507.81	-
E540, E541	T74	ATEEQLK	4.88-5.14	818.43	409.72	-
E548, D555	T75	TVMENFVAFVDK	35.33	1399.69	700.35	467.24
D561, D562, E564, E570	T76-77	CCAADDKEACFAVEGPK	18.90	1927.80	964.40	643.27
E564, E570	T77	EACFAVEGPK	18.93	1107.51	554.26	-
C-term	T78	LVVSTQTALA_OH	23.50	1002.58	501.79	-

Table 4.2: Singly modified peptides identified in the chromatographic separation of tryptically digested BSA. Retention times are given in a range of typical peak locations and identified charge states are listed. The potentially modifiable residue(s) associated with each peptide are listed. Residues at which confirmed modifications were observed are highlighted in boldface.

	Tn-m	Sequence	RT (min)	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$
D37, E38	T5	GLVLIAFSQYLQQCPF**D**EHVK	41.40	-	1246.64	831.43
E45, E48	T6	LVN**E**LTEFAK	25.90-26.10	1176.66	588.83	-
D56, E57, E63	T7	TCVA**D**ESHAGCEK	5.37	-	738.81	492.88
D56, **E57**, E63	T7	TCVAD**E**SHAGCEK	8.15	1476.62	738.81	492.88
D56, E57, **E63**	T7	TCVADESHAGC**E**K	7.47	1476.62	738.81	492.88
D72, **E73**	T8	SLHTLFGD**E**LCK	26.40-26.50	1432.73	716.87	478.25
E82, D86, D89, **E92**	T10	ETYGDMADCC**E**K	15.80-16.00	1491.55	746.28	-
E100	T12	N**E**CFLSHK	14.62	1047.50	524.26	349.84
E100, D107,	T12-13	N**E**CFLSHK DDSPDLPK	19.60-19.70	-	957.95	638.97

D108, D111						
D107, D108, D111	T13	**D**DSPDLPK	14.40-14.60	899.45	450.23	-
D107, **D108**, D111		DD**S**PDLPK	13.70-14.30	899.45	450.23	300.48
D107, D108, **D111**		DDSP**D**LPK	14.00-14.20	899.45	450.23	300.48
D118, **D124**, E125	T14	LKPDPNTLC**D**EFK	22.85-23.15	-	795.40	530.60
E152	T20	HPYFYAP**E**LLYYANK	33.60-34.00		951.48	634.66
E166, E171	T21	YNGVFQ**E**CCQAEDK	20.40-20.60	1760.74	880.87	587.58
E166, **E171**		YNGVFQECCQA**E**DK	19.90-20.10	-	880.87	587.58
E226, E229	T34	A**E**FVEVTK	18.15-18.35	935.52	468.26	-
E226, **E229**	T34	AEFV**E**VTK	18.15-18.35	935.52	468.26	-
D236	T35	LVT**D**LTK	16.65-16.55	802.50	401.76	-

E243, **D248**, E251, D254	T37	ECCHG**D**LLECADDR	19.30-19.50	-	881.85	588.24
E243, D248, **E251**, D254	T37	ECCHGDLL**E**CADDR	18.30-18.50	1762.69	881.85	588.24
D265, D268	T39	YIC**D**NQDTISSK	15.15-15.35	1456.67	728.84	486.20
D265, **D268**	T39	YICDNQ**D**TISSK	15.65-15.85	1456.67	728.84	486.20
E276, **D279**, E284	T40-41	LKECC**D**KPLLEK	14.35-14.50	1545.81	773.41	515.94
E276, D279, **E284**	T40-41	LKECCDKPLL**E**K	17.73-17.93	-	773.41	515.94
E276, **D279**, E284	T41	ECC**D**KPLLEK	13.60-13.90	130	652.82	435.55
E291, E293	T42	SHCIA**E**VEK	10.74-10.94	1085.54	543.27	362.52
E291, **E293**	T42	SHCIAV**E**K	10.90-11.10	1085.54	543.27	362.52
D295, E299,	T43	DAIPENLPPLTA**D**FAEDK	32.01-32.21	-	9585.00	657.02

D307, E310						
E320	T45	NYQ**E**AK	22.00-22.20	765.39	383.20	-
E320, D323, E332	T45-47	NYQ**E**AKDAFLGSFLYEYSRR	40.80-41.00	-	1235.61	824.08
D323, E332	T46	**D**AFLGSFLYEYSR	37.64-37.84	1580.77	790.89	527.60
E339	T47-48	RHP**E**YAVSVLLR	23.60-23.80	1452.84	726.93	484.95
E351, E353, E357, E358	T50	EYEATLEECCAK	17.40-17.60	1515.65	758.33	506.54
D381, E382	T53	HLV**D**EPQNLIK	20.61-20.81	1318.75	659.88	440.25
D381, **E382**	T53	HLVD**E**PQNLIK	21.07-21.27	1318.75	659.88	440.25
D392, E395	T54	QNC**D**QFEK	12.21-12.41	1081.47	541.24	-
D392, **E395**	T54	QNCDQF**E**K	10.49-10.69	1081.47	541.24	-
E399	T55	LG**E**YGFQNALIVR	30.35-30.55	-	746.92	498.27
E424	T58-59	VPQVSTPTLV**E**VSRLGK	22.37	1910.11	955.56	637.37

E449, **D450**	T62	MPCTE**D**YLSLILNR	38.73	-	869.44	579.95
E464	T63	LCVLH**E**K	19.76	911.51	456.26	
E470	T64	TPVS**E**KVTK	11.40	1001.60	501.30	334.54
E478, D493, E494	T66-67	CCT**E**SLVNRRPCFSALTPDETYVPK	33.71	-	-	1005.15
D493, E494	T67	RPCFSALTP**D**ETYVPK	25.20	-	947.48	631.
D511, D517, **E519**	T69	LFTFHA**D**ICTLPD**T**EK	29.23	-	960.98	640.99
E530	T72	QTALV**E**LLK	28.72	1027.65	514.33	-
E548, D555	T75	TVM**E**NFVAFVDK	35.52	1412.72	706.87	471.58
E548, **D555**	T75	TVMENFVAFV**D**K	35.09	1412.72	706.87	471.58
C-term	T78	LVVSTQTALA_Am	23.98	1015.61	508.31	-

Table 4.3: Multiply modified peptides identified in the chromatographic separation of tryptically digested BSA. Retention times are given in a range of typical peak locations and identified charge states are listed. The potentially modifiable residue(s) associated with each peptide are listed. Residues at which confirmed modifications were observed are highlighted in boldface.

Modifiable Site	Tn-m	Sequence	RT (min)	[M+H]⁺	[M+2H]²⁺	[M+3H]³⁺
D13, E16, E17	T4	**D**LGEE**H**FK	29.60-29.80	1013.55	507.28	338.52
D56, E57, E63	T7	TCVA**DE**SHAGCEK	15.17-15.37	1463.59	732.30	488.53
E82, D86, D89, E92	T10	ETYG**D**MADCCEK	15.90-16.10	1504.59	752.80	-
E82, **D86**, D89, **E92**	T10	**E**TYG**D**MADCCEK	16.35-16.55	-	759.31	506.54
D107, D108, **D111**	T13	**D**DSP**D**LPK	14.25-14.45	912.48	456.74	-
E226, E229	T34	AEFV**E**VTK	25.10-25.30	948.55	474.78	-
D295, E299, D307, E310	T43	DAIPENLPPLTA**D**FAEDK	31.89-32.09		991.52	661.35
E320, D323, E332	T45-47	NYQ**E**AK**D**AFLGSFLYEYSRR	41.10-41.30	1242.12	828.42	621.57
D381, E382	T53	HLV**DE**PQNLIK	26.15-26.35	1331.78	666.39	444.60
E548, D555	T75	TVM**E**NFVAFV**D**K	34.38-34.58	1425.76	713.38	475.92

4.2 Footprinting Results

To automate the integration of peptide peaks in our extracted-ion chromatogram, an R script to calculate peak areas from raw data files given a list of retention times and m/z values was graciously provided by Guanghui Wang (NIST Mass Spectrometry Data Center). The extent of modification was calculated on a per-peptide basis, weighted by the number of amidation sites in each peptide (actual over possible). The overall extent of modification for the protein was estimated by dividing the total peak area for all modified peptides by the total peak area of all integrated peaks (modified and unmodified). Because the overall extent of protein modification will differ between samples due to variation in reaction conditions, apparent extents of modification at the residue level must be normalized to enable direct comparison between samples. The data presented here has been normalized to 10%. Because time and resource constraints prevented the acquisition of sufficient replicates for a meaningful analysis of all samples, a pooled comparison was performed. Table 4.4 lists the average extents of modification for experimentally observed tryptic peptides across pooled samples. Samples were divided into control (modified only in-solution), adsorbed protein modified in aqueous phosphate buffer, adsorbed protein modified in the presence of acetonitrile, protein modified post-desorption, and protein modified in high-organic without prior adsorption.

Table 4.4: Average modification extent of observed tryptic peptides for each sample type. Samples were pooled based on experimental conditions, and the extent of modification was normalized to 10%.

Modifiable Site	Tn-m	Control (%)	Adsorbed (%)	Adsorbed + Organic (%)	Post-Desorption (%)	High-Organic (%)
['D13', 'E16', 'E17']	T4	14.25	7.93	6.15	23.00	25.87
['D37', 'E38']	T5	6.56	9.61	5.06	8.92	7.33
['E45', 'E48']	T6	7.65	4.09	3.68	0.96	6.25
['D56', 'E57', 'E63']	T7	14.31	9.23	14.76	27.48	17.90
['D72', 'E73']	T8	5.26	2.01	2.90	10.65	4.29
['E82', 'D86', 'D89', 'E92']	T10	33.36	29.04	24.86	4.43	25.89
['E100']	T12	12.39	15.78	16.16	8.11	21.57
['D107', 'D108', 'D111']	T13	19.86	6.91	6.44	17.20	17.00
['D118', 'D124', 'E125']	T14	2.48	7.47	8.34	4.12	3.14
['E152']	T20	2.74	13.02	3.50	0.86	4.37
['E166', 'E171', 'D172']	T21	18.20	14.10	11.98	10.69	16.83
['E226', 'E229']	T34	19.38	17.15	18.39	7.87	20.03
['D236']	T35	7.49	11.33	9.20	12.64	1.87
['E243', 'D248', 'E251', 'D254', 'D255']	T37	6.33	5.99	5.52	5.53	8.10

['D265', 'D268']	T39	9.81	14.56	20.30	14.07	15.12
['E276', 'D279', 'E284']	T41_100	18.95	16.16	11.94	14.34	15.15
['E291', 'E293']	T42	5.36	5.79	8.33	4.54	4.89
['D295', 'E299', 'D307', 'E310', 'D311']	T43	9.78	6.42	7.68	4.87	10.35
['E320']	T45	22.29	6.54	8.33	15.08	22.09
['D323', 'E332']	T46	3.86	5.26	9.11	4.09	7.74
['E339']	T48	1.09	24.34	16.87	4.21	4.44
['E351', 'E353', 'E357', 'E358']	T50	12.92	9.92	8.74	7.92	11.66
['D363', 'D364', 'D374']	T51	0.00	0.00	0.00	0.00	0.00
['D381', 'E382']	T53	1.68	9.10	10.19	10.68	7.80
['D392', 'E395']	T54	7.91	11.37	8.56	9.93	11.33
['E399']	T55	20.12	30.16	11.36	4.82	18.35
['E424']	T58	4.95	14.44	12.46	3.12	8.64
['E449', 'D450']	T62	4.94	5.22	5.69	13.69	8.53
['E464']	T63	21.72	15.41	19.92	24.55	15.10
['E470']	T64	63.22	40.78	39.62	34.62	48.66
['E478']	T66	7.31	3.05	7.67	18.57	5.61
['D493', 'E494']	T67	27.77	28.32	15.78	19.71	21.27
['D511', 'D517', 'E519']	T69	2.92	3.91	4.74	6.82	2.22

['E530']	T72	0.16	4.58	3.41	0.86	5.49
['E540', 'E541']	T74	0.00	0.00	0.00	0.00	0.00
['E548', 'D555']	T75	5.73	5.40	5.44	12.26	9.90
['D561', 'D562', 'E564', 'E570']	T76-77	0.00	0.00	0.00	0.00	0.00
['C-term']	T78	2.60	7.80	9.26	0.42	10.63

4.3 Discussion and Future Work

Modified versions of some peptides were never observed over the course of these experiments: specifically, T_{76-77} was found as an unmodified peptide with a missed cleavage (individual T_{76} and T_{77} peptides were not identified) and T_{51} was only found in its unmodified form. It is possible that either the extent of modification at these sites was too low to be detectable by the mass spectrometer and confidently identified, or that these peptides are strongly shielded within the protein core and are not exposed to the solvent in any conditions tested. Additionally, some peptides were not confidently identified in either their modified or unmodified forms: T_1-T_3 and T_9, T_{11}, T_{15}-T_{20}, T_{21}-T_{33}, T_{36}, T_{38}, T_{40}, T_{44}, T_{47}, T_{49}, T_{52}, T_{56}-T_{57}, T_{59}-T_{61}, T_{65}, T_{68}, T_{70}-T_{71}, T_{73}. Further, no peptides containing oxidized methionine residues (a very common sample handling artifact) were confidently identified. It is possible that the occurrence of these was too infrequent to render a detectable peak, and the absence of these peptides may bias the analysis slightly for peptides containing methionine residues (T_{10}, T_{62}, and T_{75}).

Surprisingly, some potentially modifiable sites known to be solvent-exposed were either never observed or not confidently identified in these analyses: notably, T_{22}-T_{33}, T_{68}, and T_{56}-T_{57}. It is possible that the extended digestion times used in this research (>18 hours) may contribute to the loss of these, and other, peptides—recent research has shown that extended digestion times can lead to loss of sequence coverage.[48] We expect that similar tryptic peptide loss occurred in our samples and propose that decreasing the digestion time may help improve sequence coverage. Additionally, reproducibility may be improved by more carefully controlling the incubation time during tryptic digestion.

While our current results implicate some residues in binding to C_8 surfaces, improving sequence coverage will help to confirm the presence of the preferred binding orientation of BSA to these chromatographic surfaces. Tryptic peptide T_{10} (containing E82, D86, D89, and E92) is highly solvent-exposed, with site E82 (in conjunction with the nearby E97) known to be one of the primary residues involved in the dimerization of BSA.[49] Our results are consistent with the expectation that this site will be highly modified in-solution, with an average extent of modification of 33.36%. Because these samples are all normalized to a 10% overall extent of protein modification, this means the site is three times more likely to be modified than the overall average.

Following adsorption, the extent of modification decreases slightly to 29.04%. The adsorbed protein modified in the presence of acetonitrile shows a roughly similar degree of modification to that of the protein modified in acetonitrile-containing solvent: 24.86% and 25.89%, respectively. While we would expect the desorbed protein to exhibit a similar degree of modification to that of the in-solution protein samples, our results showed an extent of modification of only 4.43%. Although it is possible that the extent of modification is underestimated and expanding our search of missed-cleavage peptides may improve the accuracy of our results, it is also possible that the site is highly polar and becomes shielded on the interior of the protein following desorption into the high-organic solvent. Modeling of the conformational changes may shed light on shielding of sites due to the perturbation of adjacent regions.

In contrast, E424—found on tryptic peptide T_{58}—is known to be involved in ligand binding in the interior of the protein.[49] As expected, this site shows a relatively low degree of modification (4.95%) in-solution. During adsorption, the extent of

modification increases to 14.44% in phosphate buffer and 12.46% in the presence of acetonitrile. In acetonitrile alone, the region also appears to undergo conformational changes and become more solvent exposed, with an extent of modification of 8.92%. After desorption, it again exhibits a low degree of modification—3.12%—indicating that the site is only one-third as likely to be modified as that of the overall protein, based on the 10% normalization.

Tryptic peptide T_{50} contains several modifiable residues, including E351 and E358, which participate in the first of two parallel patches on the third and fourth alpha-helix of subdomain IIB that has been found to induce T cell proliferation.[35] A decrease in extent of modification from the in-solution protein (12.92%) to the adsorbed protein in both phosphate buffer and acetonitrile (9.92% and 8.74%) appeared to be retained following desorption into acetonitrile (7.92%). T_{50} modified in the presence of acetonitrile without prior adsorption exhibited a similar extent of modification (11.66%) to that of the control, suggesting that the stationary phase contributes to enduring perturbation of protein structure in the region surrounding the peptide. Site D363 resides on T_{51} and also contributes to this first patch;[35] however, only the unmodified form of T_{51} was identified.

The second patch contains D381, located on tryptic peptide T_{53}, which showed a low degree of modification in the control samples (1.68%). This increases significantly following adsorption, regardless of solvent composition—adsorbed protein modified in phosphate buffer exhibits a similar degree of modification (9.10%) to that modified while adsorbed under high-organic conditions (10.19%). This increase is higher than that of the protein modified in acetonitrile without previous adsorption (7.80%), and retained

following desorption (10.68%). This lends support to the theory that both the hydrophobic stationary phase and the high-organic mobile phase contribute to conformational changes observed during RPLC, and that these changes may be retained following desorption.

Residue D517, located on peptide T_{69}, is exposed to the protein surface and resides on a helix of subdomain IIIB.[35] Upon adsorption, the extent of modification at this site appears to increase slightly from 2.92% to 3.91%. For the adsorbed protein modified in the presence of organic solvent, the extent of modification continues to increase slightly to 4.74%, rising to 6.82% following desorption. In contrast, the protein modified in organic solvent without previously having been adsorbed exhibits a modification extent of 2.22%, comparable to that of the protein modified in phosphate buffer.

Tryptic peptide T_{72}, containing residue E530, is also located on subdomain IIIB.[35] Examination of the crystal structure of BSA shows that this site is on the interior of the protein, and likely to have very low solvent accessibility. In-solution, T_{72} exhibited an extent of modification of only 0.16%. Under both adsorbed and high-acetonitrile conditions, the extent of modification increases significantly to 4.58% and 5.49%, respectively. The adsorbed protein modified under organic conditions also exhibits an increase in modification extent, to 3.41%. Following desorption, the observed extent of modification was found to be 0.86%—this indicates that while the conformational changes induced in this region are not fully retained, the region also does not fully revert to its native structure. Missed cleavages were commonly noted for this peptide, however, and improving digestion conditions and sequence coverage as previously discussed may help to ameliorate our conclusions.

Other potential adjustments to the experimental protocol should be noted for future experiments: first, due to the low concentration and volume of the recovered protein, it would be advantageous to use a larger mass of packing material for adsorption experiments to allow for larger samples to be analyzed. Additionally, a washing step to remove excess BSA prior to modification should be considered. This was briefly checked during experimental development: after centrifuging the added protein solution out of the packing material at 1,500 x *g*, phosphate buffer was added to the column and collected after centrifugation. A Bradford check on this eluent indicated that no un-adsorbed BSA was present, and the wash step was omitted from future experiments. Additionally, we assume that any excess, un-adsorbed BSA remaining in the column that may be inadvertently modified will be removed with the final centrifugation step prior to the elution of the modified, adsorbed protein. However, it would be preferable to ensure that the results reflect only the extent of modification for adsorbed protein: adding a wash step following the addition of BSA or centrifuging the column to relative dryness prior to the addition of the modification reagents would facilitate this. In doing so, the extent of modification for the adsorbed protein may be increased and the quantities of the modification reagents would need to be adjusted accordingly. Further, more instrumental replicates of acquired samples are needed to enable comparison between minute changes in experimental conditions.

While the results remain somewhat ambiguous in the absence of sufficient replicate instrumental data or modeling, the analysis of tryptic peptides discussed above lends credence to our theories surrounding our research questions. To determine if BSA adsorbs with a preferred binding orientation, a consistent decrease in the extent of

modification only for the implicated residues on the protein surface would be seen following adsorption across samples and replicates. In the absence of a preferred binding orientation, all residues on the protein surface would show an equal decrease in extent of modification, reflecting the equal probability of binding in any position. While denaturation of the protein following adsorption was observed, correlation between the degree of deformation and surface residence time requires instrumental replicates of samples at different incubation times, as opposed to pooling of all adsorbed samples. In seeking to answer whether the conformation in high-organic solvent depends on the adsorption history of the protein, the apparent dissimilarity between modification extent for previously-adsorbed BSA and BSA modified in acetonitrile without previous adsorption suggests that the hydrophobic stationary phase contributes strongly to structural changes in some regions of the protein. Modeling of these changes would provide greater insight into the degree and nature of these changes, and their effect on neighboring regions of the protein.

REFERENCES

1.   Oroszlan, P.; Wicar, S.; Teshima, G.; Wu, S. L.; Hancock, W. S.; Karger, B. L., Conformational effects in the reversed-phase chromatographic behavior of recombinant human growth hormone (rhGH) and N-methionyl recombinant human growth hormone (Met-hGH). *Analytical Chemistry* **1992,** *64* (14), 1623-1631.

2.   Sokol, J. M.; Holmes, B. W.; O'Connell, J. P.; Fernandez, E. J., Aprotinin conformational distributions during reversed-phase liquid chromatography. Analysis by hydrogen-exchange mass spectrometry. *J Chromatogr A* **2003,** *1007* (1-2), 55-66.

3.   Richards, K. L.; Aguilar, M. I.; Hearn, M. T. W., Effect of protein conformation on experimental bandwidths in reversed-phase high-performance liquid chromatography. *Journal of Chromatography A* **1994,** *676* (1), 33-41.

4.   Geng, X.; Regnier, F. E., Retention model for proteins in reversed-phase liquid chromatography. *J Chromatogr* **1984,** *296*, 15-30.

5.   Katzenstein, G. E.; Vrona, S. A.; Wechsler, R. J.; Steadman, B. L.; Lewis, R. V.; Middaugh, C. R., Role of conformational changes in the elution of proteins from reversed-phase HPLC columns. *Proceedings of the National Academy of Sciences of the United States of America* **1986,** *83* (12), 4268-4272.

6.   Yu, L.; Zhang, L.; Sun, Y., Protein behavior at surfaces: Orientation, conformational transitions and transport. *Journal of Chromatography A* **2015,** *1382*, 118-134.

7.   de Frutos, M.; Cifuentes, A.; Díez-Masa, J. C.; Camafeita, E.; Méndez, E., Multiple Peaks in HPLC of Proteins: Bovine Serum Albumin Eluted in a Reversed-Phase System. *Journal of High Resolution Chromatography* **1998,** *21* (1), 18-24.

8.   Shiwen, L.; Karger, B. L., Reversed-phase chromatographic behavior of proteins in different unfolded states. *Journal of Chromatography A* **1990,** *499*, 89-102.

9.   Pearson, J. D., High-performance liquid chromatography column length designed for submicrogram scale protein isolation. *Analytical Biochemistry* **1986,** *152* (1), 189-198.

10.   Barreto, M. S. C.; Elzinga, E. J.; Alleoni, L. R. F., The molecular insights into protein adsorption on hematite surface disclosed by in-situ ATR-FTIR/2D-COS study. *Scientific Reports* **2020,** *10* (1), 13441.

11.   McNay, J. L.; Fernandez, E. J., How does a protein unfold on a reversed-phase liquid chromatography surface? *Journal of Chromatography A* **1999,** *849* (1), 135-148.

12.   Astefanei, A.; Dapic, I.; Camenzuli, M., Different Stationary Phase Selectivities and Morphologies for Intact Protein Separations. *Chromatographia* **2017,** *80* (5), 665-687.

13.   Pearson, J. D.; Lin, N. T.; Regnier, F. E., The importance of silica type for reverse-phase protein separations. *Analytical Biochemistry* **1982,** *124* (1), 217-230.

14.   McNay, J. L.; Fernandez, E. J., Protein unfolding during reversed-phase chromatography: I. Effect of surface properties and duration of adsorption. *Biotechnol Bioeng* **2001,** *76* (3), 224-32.

15.   McNay, J. L.; O'Connell, J. P.; Fernandez, E. J., Protein unfolding during reversed-phase chromatography: II. Role of salt type and ionic strength. *Biotechnol Bioeng* **2001,** *76* (3), 233-40.

16.   de Collongue-Poyet, B.; Vidal-Madjar, C.; Sebille, B.; Unger, K. K., Study of conformational effects of recombinant interferon gamma adsorbed on a non-porous reversed-phase silica support. *J Chromatogr B Biomed Appl* **1995,** *664* (1), 155-61.

17.   Hearn, M. T. W.; Grego, B., High-performance liquid chromatography of amino acids, peptides and proteins: IV. Studies on the origin of band broadening of polypeptides and proteins separated by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A* **1984,** *296*, 61-82.

18.   Place, H.; Sebille, B.; Vidal-Madjar, C., Split-peak phenomenon in nonlinear chromatography. 2. Characterization of the adsorption kinetics of proteins on reversed-phase supports. *Analytical Chemistry* **1991,** *63* (13), 1222-1227.

19.   Roach, P.; Farrar, D.; Perry, C. C., Interpretation of Protein Adsorption: Surface-Induced Conformational Changes. *Journal of the American Chemical Society* **2005,** *127* (22), 8168-8173.

20.   Jeyachandran, Y. L.; Mielczarski, E.; Rai, B.; Mielczarski, J. A., Quantitative and qualitative evaluation of adsorption/desorption of bovine serum albumin on hydrophilic and hydrophobic surfaces. *Langmuir* **2009,** *25* (19), 11614-20.

21.   Larsericsdotter, H.; Oscarsson, S.; Buijs, J., Structure, stability, and orientation of BSA adsorbed to silica. *Journal of Colloid and Interface Science* **2005,** *289* (1), 26-35.

22.   Yano, Y. F., Kinetics of protein unfolding at interfaces. *J Phys Condens Matter* **2012,** *24* (50), 503101.

23.   Rabe, M.; Verdes, D.; Seeger, S., Understanding protein adsorption phenomena at solid surfaces. *Advances in Colloid and Interface Science* **2011,** *162* (1), 87-106.

24.   Malmsten, M., Formation of Adsorbed Protein Layers. *Journal of Colloid and Interface Science* **1998,** *207* (2), 186-199.

25.   Deitcher, R. W.; O'Connell, J. P.; Fernandez, E. J., Changes in solvent exposure reveal the kinetics and equilibria of adsorbed protein unfolding in hydrophobic interaction chromatography. *Journal of Chromatography A* **2010,** *1217* (35), 5571-5583.

26.   Sane, S. U.; Cramer, S. M.; Przybycien, T. M., A holistic approach to protein secondary structure characterization using amide I band Raman spectroscopy. *Anal Biochem* **1999,** *269* (2), 255-72.

27.   Smith, J. R.; Cicerone, M. T.; Meuse, C. W., Tertiary Structure Changes in Albumin upon Surface Adsorption Observed via Fourier Transform Infrared Spectroscopy. *Langmuir* **2009,** *25* (8), 4571-4578.

28.   Benedek, K.; Dong, S.; Karger, B. L., Kinetics of unfolding of proteins on hydrophobic surfaces in reversed-phase liquid chromatography. *Journal of Chromatography A* **1984,** *317*, 227-243.

29.   Cornwell, O.; Radford, S. E.; Ashcroft, A. E.; Ault, J. R., Comparing Hydrogen Deuterium Exchange and Fast Photochemical Oxidation of Proteins: a Structural Characterisation of Wild-Type and ΔN6 β2-Microglobulin. *Journal of The American Society for Mass Spectrometry* **2018,** *29* (12), 2413-2426.

30.   Xiao, K.; Zhao, Y.; Choi, M.; Liu, H.; Blanc, A.; Qian, J.; Cahill, T. J.; Li, X.; Xiao, Y.; Clark, L. J.; Li, S., Revealing the architecture of protein complexes by an orthogonal approach combining HDXMS, CXMS, and disulfide trapping. *Nature Protocols* **2018,** *13* (6), 1403-1428.

31.   Wang, C.; Isaacson, S. C.; Wang, Y. C.; Lionti, K.; Volksen, W.; Magbitang, T. P.; Chowdhury, M.; Priestley, R. D.; Dubois, G.; Dauskardt, R. H., Surface Chemical Functionalization to Achieve Extreme Levels of Molecular Confinement in Hybrid Nanocomposites. *Advanced Functional Materials* **2019,** *29* (33).

32.   Wang, L.; Chance, M. R., Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Analytical Chemistry* **2011,** *83* (19), 7234-7241.

33.   Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R., H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for a top-down approach? *Analytical chemistry* **2009,** *81* (19), 7892-7899.

34.   Lindner, R.; Heintz, U.; Winkler, A., Applications of hydrogen deuterium exchange (HDX) for the characterization of conformational dynamics in light-activated photoreceptors. *Front Mol Biosci* **2015,** *2*, 33-33.

35.   Majorek, K. A.; Porebski, P. J.; Dayal, A.; Zimmerman, M. D.; Jablonska, K.; Stewart, A. J.; Chruszcz, M.; Minor, W., Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Mol Immunol* **2012,** *52* (3-4), 174-82.

36.   Liu, X. R.; Zhang, M. M.; Gross, M. L., Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chemical Reviews* **2020,** *120* (10), 4355-4454.

37.   Zhang, H.; Wen, J.; Huang, R. Y.; Blankenship, R. E.; Gross, M. L., Mass spectrometry-based carboxyl footprinting of proteins: method evaluation. *Int J Mass Spectrom* **2012,** *312*, 78-86.

38.   Hermanson, G. T., 3 - Zero-Length Cross-linkers. In *Bioconjugate Techniques*, Hermanson, G. T., Ed. Academic Press: San Diego, 1996; pp 169-186.

39.   Sanzgiri, R. D.; McKinnon, T. A.; Cooper, B. T., Intrinsic charge ladders of a monoclonal antibody in hydroxypropylcellulose-coated capillaries. *Analyst* **2006,** *131* (9), 1034-43.

40.   Cooper, B. T.; Sanzgiri, R. D.; Maxey, S. B., Probing the conformational behavior of a monoclonal antibody with surfactant affinity capillary electrophoresis (SurfACE). *Analyst* **2012,** *137* (24), 5777-84.

41.   Guthals, A.; Bandeira, N., Peptide Identification by Tandem Mass Spectrometry with Alternate Fragmentation Modes*. *Molecular & Cellular Proteomics* **2012,** *11* (9), 550-557.

42.   Hoffmann, E. d. S. V., *Mass spectrometry : principles and applications*. J. Wiley: Chichester, West Sussex, England; Hoboken, NJ, 2007.

43.   Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry* **2013,** *85* (1), 2-9.

44.   Aliyari, E.; Konermann, L., Formation of Gaseous Peptide Ions from Electrospray Droplets: Competition between the Ion Evaporation Mechanism and Charged Residue Mechanism. *Analytical Chemistry* **2022,** *94* (21), 7713-7721.

45.   Schulze, W. X.; Usadel, B., Quantitation in mass-spectrometry-based proteomics. *Annu Rev Plant Biol* **2010,** *61*, 491-516.

46.   Deutsch, E. W.; Lam, H.; Aebersold, R., Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiol Genomics* **2008,** *33* (1), 18-25.

47.   Kim, S.; Pevzner, P. A., MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **2014,** *5*, 5277.

48.   Hildonen, S.; Halvorsen, T. G.; Reubsaet, L., Why less is more when generating tryptic peptides in bottom-up proteomics. *PROTEOMICS* **2014,** *14* (17-18), 2031-2041.

49.   Scognamiglio, P. L.; Vicidomini, C.; Fontanella, F.; De Stefano, C.; Palumbo, R.; Roviello, G. N., Protein Binding of Benzofuran Derivatives: A CD Spectroscopic and In Silico Comparative Study of the Effects of 4-Nitrophenyl Functionalized Benzofurans and Benzodifurans on BSA Protein Structure. *Biomolecules* **2022,** *12* (2), 262.