PROTEIN STRUCTURAL INVESTIGATION USING HIGH SPATIAL RESOLUTION HYDROXYL RADICAL PROTEIN FOOTPRINTING COUPLED WITH LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY

by

ZIXUAN LI

(Under the Direction of Joshua S. Sharp)

ABSTRACT

Hydroxyl radical protein footprinting (HRPF) coupled with liquid chromatography and tandem mass spectrometry (LC-MS/MS) has become a very promising technique to study intact protein, protein conformational change and protein – ligand interactions in solution. HRPF oxidizes amino acid side chains at a rate based on a combination of the inherent reactivity and the accessibility to the hydroxyl radical diffusing in solution, giving its advantages of fast time frame for protein modification, high labeling efficiency and stability of labeled analyte. By monitoring the change in the amount of oxidation under different experimental conditions, changes in protein structure can be determined by how the structural changes affect the solvent accessibility of the different regions of the protein. The work presented describes a newly developed method using high spatial resolution HRPF coupled with electron-transfer dissociation (ETD)-based LC-MS/MS to investigate complex molecular interaction interfaces at residue level. Protein-GAG interaction and chemokine association binding interfaces were successfully determined by this method. An improvement of the method by adding a supercharging reagent *m*-NBA to increase sequence coverage and spatial resolution for protein quantitation is also presented.

INDEX WORDS: Mass Spectrometry, Tandem Mass Spectrometry, High Performance Liquid Chromatography, Hydroxyl Radicals, Protein Footprinting, Protein Structure, Electron-transfer dissociation, Fast Photochemical Oxidation of Proteins, *m*-NBA, Robo1, Heparin, CXCL4, Glycosaminoglycans

PROTEIN STRUCTURAL INVESTIGATION USING HIGH SPATIAL RESOLUTION HYDROXYL RADICAL PROTEIN FOOTPRINTING COUPLED WITH LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY

by

ZIXUAN LI

B.S., Northwest University, China, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

PROTEIN STRUCTURAL INVESTIGATION USING HIGH SPATIAL RESOLUTION HYDROXYL RADICAL PROTEIN FOOTPRINTING COUPLED WITH LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY

by

ZIXUAN LI

Major Professor:	Joshua S. Sharp
Committee:	Ron Orlando I. Jonathan Amster

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

DEDICATION

Mom and Dad, this is for you! Thanks for always being there for me.

ACKNOWLEDGEMENTS

I cannot express enough thanks to my advisor, Joshua Sharp for the continuous support of my study and research. It is his immense knowledge, patience, and excellent guidance that make this dissertation possible. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. I could not have imagined having a better advisor and mentor for my Ph.D. study. I also thank my committee members, Ron Orlando and Jon Amster, for their time and advices throughout my study in graduate school.

My completion of this dissertation could not have been accomplished without my collaborators. I'd like to acknowledge Dr. Kelley Moremen, Dr. Shuo Wang and Heather Moniz for continually providing Robo1 protein and the mutants. Thanks to Dr. Lianchun Wang for providing porcine intestinal heparin. Thanks to Dr. Chau-wen Chou at the Proteomic and Mass Spectrometry Facility, Dr. Lance Wells and Rob Bridger for their instrumentation and assistance of mass spectral data acquisition. Thanks to Dr. Fuming Zhang and Dr. Robert J. Linhardt at Rensselaer Polytechnic Institute for performing SPR data. Thanks to Dr. Gregory Wylie and Dr. Chuck Mobley for assistance with dynamic light scattering. Thanks to Dr. Tracy Handel and Dr. Douglas Dyer at University of California, San Diego for providing CXCL4 protein. Thanks to Qi Gao for teaching me protein expression and purification. Thanks to all my former and present group members and our instrumentation managers, for their assistance, advices, and encouragement.

Last but not least, a special thanks to my dear family and sweet friends for their unending love, support and company.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Protein Structural Analysis and Techniques

Proteins are macromolecules that perform a cell's myriad processes and functions in all living organisms via dynamic interactions with other proteins, small molecules and cellular components. All proteins' functions are dependent on their structure. Protein tertiary and quaternary structure are fundamental to determining mechanisms of protein function. The interactions of protein and their assembly into large complexes are crucial to almost all biological processes, such as enzymatic catalysis, transport and storage, and immune protection. Knowledge of the details of protein structure and protein molecular interactions in macromolecule assemblies is therefore significant to understand protein function, mechanisms in biological systems as well as their roles in disease and potentials for drug development. The primary experimental methods used in the high-resolution study of protein structure include X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and electron microscopy (EM). The methods available for studying proteins provide different types of structural information with its advantages and disadvantages.

X-ray crystallography is a tool used for identifying the atomic and molecular structure of a crystaland obtaining structural information, and accounts for almost 90% of the structures entered into the protein databank (www.pdb.org) [1]. For this method, the protein is purified and crystallized, then subjected to an intense beam of X-rays. The proteins in the crystal diffract the X-ray beam into one or another characteristic pattern of spots, which are then analyzed to determine the distribution of electrons in the protein. By measuring the angles and intensities of these diffracted beams, a crystallographer can produce a three-dimensional picture of the density of electrons within the crystal. From this electron density, the mean positions of the atoms in the crystal can be determined, as well as their chemical bonds, their disorder and various other

information [2]. X-ray crystallography can provide very detailed atomic information, showing every atom in a protein or nucleic acid along with atomic details of ligands, inhibitors, ions, and other molecules that are incorporated into the crystal[3]. However, proteins must be examined in crystalline form; therefore, the behavior of molecules in solutions cannot be studied and only one conformation of the protein can be observed [4]. The process of crystallization is difficult and can impose limitations on the types of proteins that may be studied by this method. Flexible proteins, for instance, are far more difficult to study by this method, and flexible portions of protein will often be invisible in crystallographic electron density maps.

NMR is another commonly used tool for obtaining high resolution protein structure, dynamics, and interaction information by studying molecules in solutions using a variety of experiments [5]. The protein is purified, placed in a strong magnetic field, and then probed with radio waves. A distinctive set of observed resonances may be analyzed to give a list of atomic nuclei that are close to one another, and to characterize the local conformation of atoms that are bonded together. This list of restraints is then used to build a model of the protein that shows the location of each atom. A major advantage of NMR spectroscopy is that it provides information on proteins in solution, as opposed to those locked in a crystal or bound to a microscope grid, and thus, NMR spectroscopy is the premier method for studying the atomic structures of flexible proteins. Even though NMR doesn't require crystals and studies proteins under more physiological conditions, disadvantages to NMR include protein size limitation of ~35 kDa [6], difficulty in manual data interpretation [4], and requirement of milligrams of isotopically labeled proteins for comprehensive studies. Both NMR and X-ray crystallography are high resolution techniques used to examine protein structures; however, both methods are low throughput, labor intensive and are not suitable for all proteins or protein complexes.

Electron microscopy is used to determine structures of macromolecular complexes that are either too large or too heterogeneous to be investigated by conventional X-ray crystallography or NMR [7]. A beam of electrons is used to image the molecule directly. For a few particularly well-behaved systems, electron diffraction produces atomic-level data, but typically, electron micrographic experiments do not allow the researcher to see each atom. Electron micrographic studies often combine information from X-ray crystallography or NMR spectroscopy to sort out the atomic details. Atomic structures are docked into the electron density map to yield a model of the complex. The disadvantages include large radiation damage and complexes artifacts [8].

Other techniques with lower resolution, such as absorption, Raman, fluorescence, and circular dichroism (CD) spectroscopy are used routinely to evaluate the structural integrity of proteins and other biomolecules. These techniques do not yield all the information the atomic coordinates would, but can be easily interpreted, which provide information on structural reproducibility in bioproduct manufacture, overall secondary structure, and real-time refolding.

Over the last 20 years, mass spectrometry (MS) has become an important emerging method used to determine both the primary and higher order structures of proteins and dynamics. The basis for these investigations lies in the ability of mass analysis techniques to detect changes in protein conformation under differing conditions [9]. The experiments can be conducted on protein alone or in combination with proteolytic digestion or chemical modification [10, 11]. In addition to primary structure determination, proteases and chemical modification have long been used as probes of higher order structure, an approach that has been recently rejuvenated with the emergence of highly sensitive and accurate mass analysis techniques. MS-based technologies are among the most versatile tools for monitoring protein folding, structure, dynamics, and function.

Protein Footprinting

"Footprinting" refers to assays that examine ligand binding and conformational changes by determining the solvent accessibility of portions of macromolecules through their sensitivity to chemical or enzymatic cleavage or modification reactions. Protein footprinting methods are used to examine protein structure and conformational changes by monitoring solvent accessibility using either modification or cleavage reactions [12]. The standard approach to protein footprinting is to determine which sites or areas of the protein reveal a change in solvent accessibility between two or more states of the protein. A variety of approaches have been used to map macromolecules, including hydrogen-deuterium exchange (HDX) [13-15], intra- and intermolecular cross-linking[16, 17], noncovalent structural probing[18], and protein structural mapping via covalent labeling[19, 20].

Hydrogen-deuterium exchange (HDX) in conjunction with mass spectrometry (MS) is commonly used to study protein structure, dynamics, and protein-ligand interactions [15, 21-24]. Amide hydrogen on the backbone of a biomolecule exchanges with deuterium in solution at different rates partly dependent on conformation. These changes can be measured by MS and therefore the degree of activity at various sites can be inferred. The location of deuteriums can be roughly determined by digesting the deuterated protein with a non-specific acid protease such as pepsin, and fast HPLC separation at 0ºC to decrease the rate of back-exchange prior to MS. Relative folding and dynamics can be determined from the different rates of uptake at different locations [24]. Disadvantages of HDX include the difficulties of accurate data acquisition, the use of pepsin for digestion which introduces a complex mixture of overlapping partial digestion products, and the exchange process' reversibility which complicates data analysis [25].

Cross-linking methods create new intramolecular or intermolecular bonds that impose distance constraints on the location of two amino acid side chains, which allow one to deduce information about the three-dimensional structure of a single protein or protein complex [16]. In general, chemical cross-linking experiments are carried out by first linking the interacting proteins through covalent bonds followed by a series of well-established protocols, such as SDS-PAGE, in-gel digestion, and shotgun LC/MS/MS for identification of the cross-linked proteins [26]. General limitations of any cross-linking method result from the limited spatial resolution that this technique provides and from the sometimes low number of cross-links obtained from a complex.

Covalent labeling approaches use reagents that either modify specific amino acids or react generally with many different amino acids. In comparison to H/D exchange methods that use deuterium as a reagent, the possibilities of back-exchange and scrambling are essentially non-existent with covalent labeling reagents. In addition, covalent labeling approaches provide structural information about amino acid side chains that is generally not available with H/D exchange methods [19].Covalent labeling is usually carried out with either non-specific labels or amino acid-specific labels. Approaches that use non-specific labels rely much on protein reactions with radicals [12, 20, 27, 28].

The development of the hydroxyl radical as a modification reagent for protein footprinting and its application in conjunction with mass spectrometry were directly inspired by hydroxyl radical nucleic acid footprinting. Initially, hydroxyl radicals were used to nonspecifically cleave the protein backbone in conjunction with SDS-PAGE to separate cleavage products [29-33]. Hydrogen abstraction from the Cα carbon and subsequent radical reaction with oxygen leads to backbone cleavage [34, 35]. However, the side chain reactions with hydroxyl radicals occur at rates 10 to 1000 times faster than hydrogen abstraction from the Cα carbon [34, 36, 37]. Therefore,

amino acid side chains are preferable as probes for protein structural studies. Analyses of quantitative mass spectrometry of the modified protein fragments provide a footprinting approach used to probe protein structure, dynamics and protein-ligand interactions [38-45].

Mass Spectrometry-based Hydroxyl Radical Protein Footprinting (HRPF)

Due to the current technical limitations with NMR and crystallographic methods, mass spectrometry plays an ever-increasing role in protein structure determination given its speed, sensitivity, and specificity. HRPF coupled with MS has become increasingly popular as a labeling technique to probe intact protein structure [46-52], protein-protein interactions [39, 53, 54], protein folding [55, 56], and protein-small molecule ligand interactions [57] in solution. Hydroxyl radicals have significant advantages for labeling over other footprinting reagents for several reasons. First, the van der Waals area and solvent properties are similar to water molecules, making hydroxyl radicals ideal as solvent accessibility probes. Second, they have high reactivity with well understood chemical selectivity. Third, they can be generated safely and conveniently under multiple solution conditions. Finally, their non-specificity allows many residues to be covalently probed with one experiment. Covalently labeling the protein using HRPF overcomes the shortcomings of HDX and facilitates sample analysis making it easy to denature, digest and analyze the peptides by MS.

Generally, in an HRPF experiment the protein solution is exposed to hydroxyl radicals and the radicals covalently react with the solvent accessible side chains of residues [36]. Following the reaction the modified protein can be frozen for later analysis or immediately subjected to specific proteases to maximize sequence coverage and simplify data analysis. The stable modifications to the protein side chains are analyzed by LC-MS, relatively quantifying modified and unmodified peptides, and the MS signal of the oxidized version(s) of each peptide is compared to the MS signal

of the unoxidized version of the same peptide to quantify oxidation. By comparing the same protein under two different structural states, changes in solvent accessibility caused by the structural differences are detected.

Generation of Hydroxyl Radicals in Solution

The common generation of hydroxyl radicals for footprinting includes Fenton chemistry [52, 58, 59], electron pulse radiolysis [60], synchrotron radiolysis of water [39, 42, 61-63], gamma (γ) radiolysis of water [64-67], laser photolysis of H_2O_2 [44, 45, 68] and UV photolysis of peroxide [45, 48, 68, 69]. Upon oxidation, the native protein structure can be altered due to the oxidative modifications; even a single oxidation event can induce protein unfolding which causes a rapid increase in oxidation as compared to the folded conformation [70-72]. In order for hydroxyl radical protein footprinting to be a reliable method for protein structural determination, hydroxyl radicals must react with the protein exclusively in its native conformation. The time scale for large scale protein motions, like helix coiling/uncoiling, ranges from long microseconds to milliseconds [73]; however, most hydroxyl radical generation techniques, including water radiolysis by gamma (γ) rays and X-rays, are typically performed on a millisecond to minute time scale, which increases the likelihood that the protein of interest will experience unfolding due to oxidative modifications [70, 71]. Previous strategies for ensuring native protein probing include circular dichroism to monitor conformational changes as oxidation is increased [46], limiting the amount of oxidation to where the vast majority of protein molecules had one or no oxidations [51, 52], or monitoring the reaction kinetics of the protein [74] or constituent peptides [39, 53, 75] to be oxidation follows the an expected pseudo-first order rate law. However, these strategies all require that the overall amount of oxidation be strictly limited to prevent probing oxidatively-unfolded structures, which limits the amount of oxidation sites resulting in lower structural resolution.

Fast photochemical oxidation of proteins (FPOP) [47, 48] and pulsed electron beam water radiolysis [76] are successful generation methods of hydroxyl radicals for footprinting as they produce hydroxyl radicals in less than a microsecond, shorter than large scale protein motions. In FPOP experiments, addition of a scavenger controls radical exposure time to the protein to prevent labeling of the oxidatively unfolded conformation [70, 71, 74]. The hydroxyl radical extinction coefficient with Beer-Lambert Law was used to calculate the expected lifetime of hydroxyl radicals in solution [68]. Calculations show that addition of twenty millimolar glutamine is sufficient to consume the hydroxyl radicals in around one microsecond under the described conditions [77]; this does not consider hydroxyl radical reactions with buffer components or protein analyte, both of which will further decrease the half-life of the radical. Immediately following irradiation, methionine amide and catalase are added to devour any secondary oxidation products formed in the solution. By using glutamine to control radical lifetime and minimizing exposure to peroxide post FPOP, proteins can be labeled without excessively perturbing their conformation [77].

The FPOP method for producing hydroxyl radicals is employed by using a 248 nm pulsed laser beam to photodissociate H_2O_2 into two hydroxyl radicals (·OH). The diffusing radical is dispersed through the protein solution reacting with the analyte as well as buffer components and glutamine scavenger. A flow system is designed such that the laser produces a small window of high flux light designed to maximize radical exposure to a small volume of protein solution (**Figure 1.1**). The syringe pump flow rate and the frequency of the laser are set to allow for 90% of the sample to be oxidized; this prevents any portion of the sample from being irradiated more than once and accounts for laminar flow. Following irradiation the sample is collected in a tube containing methionine amide and catalase to immediately quench secondary oxidants.

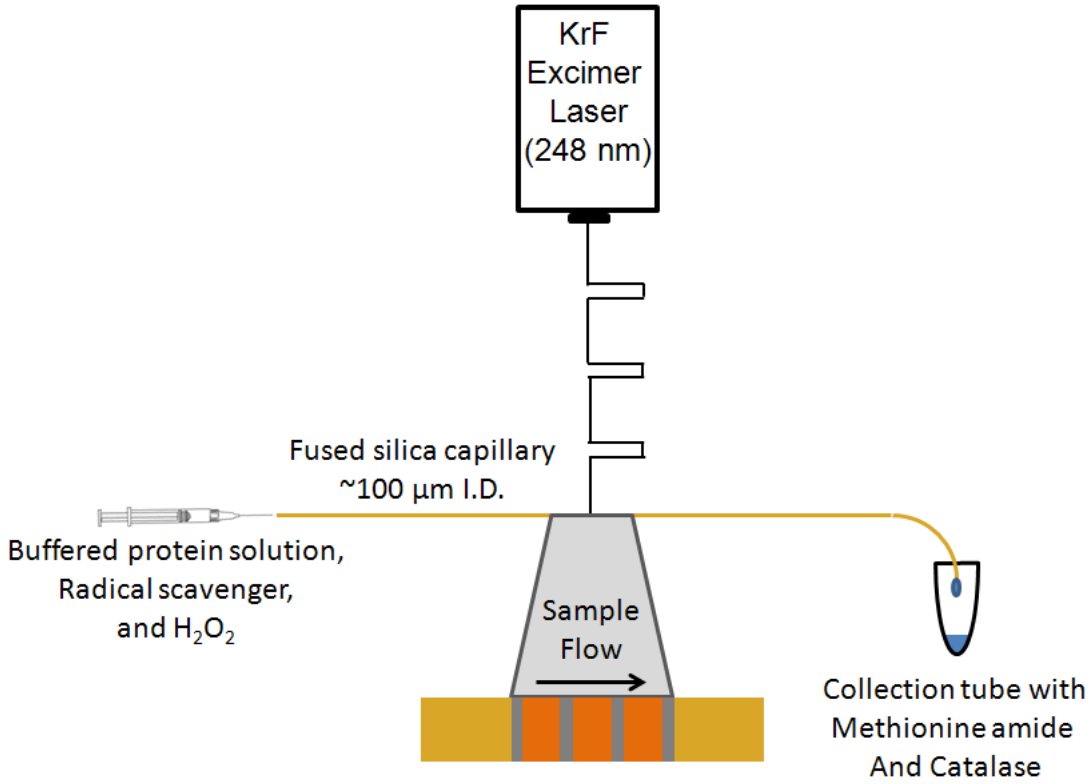

Figure 1.1. FPOP Schematic. Schematic of typical FPOP setup and conditions.

Reactions of Hydroxyl Radicals with Amino Acid Side Chains

The extent of oxidation of amino acid side chain residues in proteins in aqueous media is determined by two factors: inherent reactivity and solvent accessibility of the exposed residues to the hydroxyl radicals diffusing in solution [20, 52, 53, 78]; this observation allows us to directly relate the relative rate of oxidation with the relative solvent accessibility of the residue in the structure [55, 78]. The usefulness of a side chain in footprinting experiments depends on both its ability to react with hydroxyl radicals and the ease of detection of the reaction products in MS experiments [79]. **Table 1.1** lists the rate constants for the reaction of each free amino acid with hydroxyl radical; the relative reactivity of the side chains with hydroxyl radicals is as follows: Cys>Met>Trp>Tyr>Phe>Cysteine>His>Leu , Ile>Arg, Lys, Val>Ser, Thr, Pro>Gln, Glu>Asp,

Asn>Ala>Gly [37]. Oxidation of amino acid side chains often leads to a variety of products via multiple competing reaction mechanisms [37, 80]. When measuring oxidation products, the potential of an amino acid residue to serve as a footprinting probe is determined not only by its reactivity but also by the ability to detect stable oxidation products by mass spectroscopy. In addition, footprinting experiments are normally carried out in the micromolar concentration range, unlike those traditional studies that were generally performed at millimolar concentrations of analyte. A detailed comparison of all 20 amino acids and their products has been carried out using mass spectrometric analysis, and side chain modifications and mass shift observed in hydroxyl radical mediated protein footprinting experiments have been identified [36, 37, 80-85] as shown in **Table 1.1**. Most commonly, all reactive residues experience a net addition of oxygen, i.e. +16, +32, +48 Da, etc. [36, 37, 80-86]; the sulfur-containing residues are highly reactive toward hydroxyl radicals; the aliphatic residues have additional +14 mass shifts due to the formation of carbonyl groups at the aliphatic hydrocarbon side chains. Some side chains give rise to characteristic oxidation products. Arginine has a characteristic mass shift of -43 Da, acidic residues have a -30 Da mass shift due to the decarboxylation of the C-terminal carboxyl, and oxidation of histidine residues leads to very complex products including +16, -22, -23, +5, and -10 Da. Some major and minor products are shown in **Figure 1.2**. Multiple oxidation products complicate the spectrum analysis; however, the uniqueness of the chemical products can in some cases identify the oxidation sites on the basis of characteristic mass signatures.

While hydroxyl radicals are considered non-specific reactants, they do not label side chains with equal efficiency. The second order rate constants listed in **Table 1.1** pertain to amino acids free in solution and not as residues in a protein sequence. The rates reflect the inherent residue reactivity in a fully exposed protein structural content [37].

Table 1.1. Initial ·OH-amino Acid Side Chain Reaction Rates and Common Mass Spectrometry Observed Products of ·OH-mediated Protein Footprinting

amino acid	$k_{\cdot OH}(M^{-1}\ sec^{-1})$ [87]	common modifications and mass shift (Da) [88, 89]
Cys	3.5×10^{10}	Sulfonic acid (+48), sulfinic acid (+32),
Trp	1.3×10^{10}	Tryptophan Hydroxy (+16, +32, +48 and so on)
Tyr	1.3×10^{10}	Hydroxy (+16, +32 and so on)
Met	8.5×10^{9}	Sulfoxide (+16), aldehydes (-32), sulfone (+32)
Phe	6.9×10^{9}	Hydroxy (+16, +32 and so on)
His	4.8×10^{9}	Oxo (+16), pyrrole-ring opening (-22, -10, +5)
Arg	3.5×10^{9}	Deguanidination (-43), hydroxy (+16), carbonyl (+14)
Ile	1.8×10^{9}	Hydroxy (+16), carbonyl (+14)
Leu	1.7×10^{9}	Hydroxy (+16), carbonyl (+14)
Val	8.5×10^{8}	Hydroxy (+16), carbonyl (+14)
Pro	6.5×10^{8}	Hydroxy (+16), carbonyl (+14)
Gln	5.4×10^{8}	Hydroxy (+16), carbonyl (+14)
Thr	5.1×10^{8}	Hydroxy (+16), -2(+16 and loss of H2O)
Lys	3.5×10^{8}	Hydroxy (+16), carbonyl (+14)
Ser	3.2×10^{8}	Hydroxy (+16), -2(+16 and loss of H2O)
Glu	2.3×10^{8}	Decarboxylation (-30), hydroxy (+16), carbonyl (+14)
Ala	7.7×10^{7}	Hydroxy (+16)
Asp	7.5×10^{7}	Decarboxylation (-30), hydroxy (+16)
Asn	4.9×10^{7}	Hydroxy (+16)
Gly	1.7×10^{7}	Hydroxy (+16)

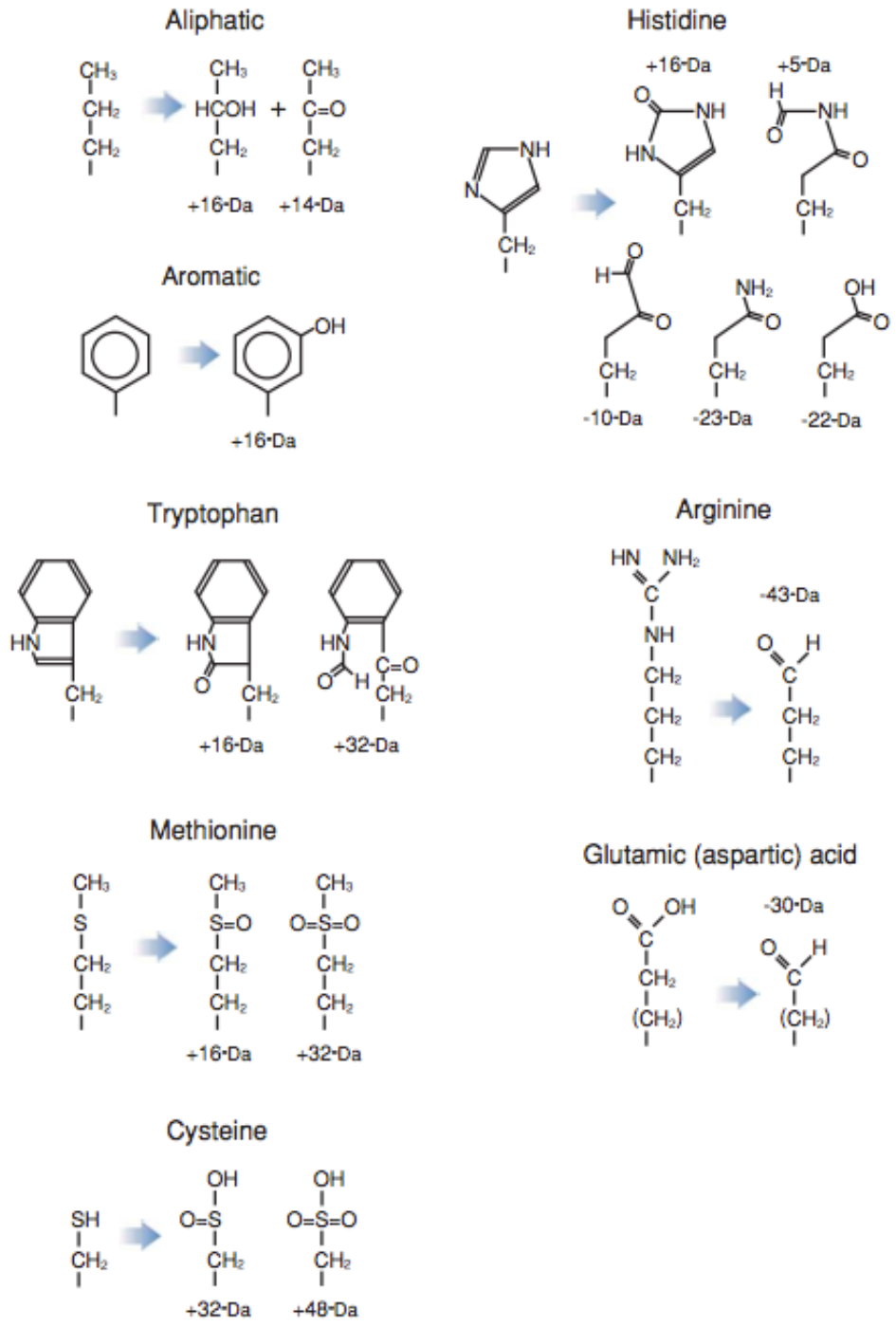

Figure 1.2. Major Oxidation Products. Major reaction products of the most reactive amino acid side chains [20].

High Resolution HRPF Quantitation of residue oxidation with LC-MS/MS

Traditional HRPF quantitation was performed on the peptide level using LC-MS allowing for structural information such as monitoring large-scale protein conformational changes or large complex interaction interfaces [90].The digested protein is typically subjected to reverse-phase HPLC, which separates the digested peptides and radiolytically modified peptides from their unmodified parent peptides. The modified peptides are generally eluted earlier than the unmodified peptide for simple alcohol or keto additions. The abundance of unmodified and modified peptides can be determined using the selected ion chromatogram peak area or the intensity of the peaks in the mass spectra. The average oxidation of peptide is calculated by the sum of the modified peptide peak intensities (areas) divided by the sum of the unmodified and modified peptide peak intensities (areas).

If multiple amino acids on the same peptide were oxidized (a common occurrence in HRPF), however, this peptide level quantitation method yielded structural information with a spatial resolution limited by the size of the peptide, which limits its use for studying smaller protein-ligand interaction interfaces. High spatial resolution HRPF methods were therefore developed to enable quantification of changes in solvent accessibility at the amino acid level [91, 92]. High spatial resolution HRPF quantiation is performed by LC-MS/MS analysis. In our study, peptide fragmentation is done using Electron-transfer dissociation (ETD). The validation of ETD-based fragmentation method used for oxidized peptide quantiation is described in **Chapter 2**. By investigating ETD fragmentation spectra corresponding to a peptide oxidation product, the c- and z-ion series can be used to map and quantitate the site(s) of oxidation to specific amino acids. **Figure 1.3** depicts a representative ETD spectrum of singly-oxidized peptide.

Changes in the solvent accessibility of each residue can be determined by quantitating change in oxidation amount of the residue. The fractional oxidation of a given sequence ion is

defined as the ratio between the oxidized sequence ion intensity to the sum of the intensity of the corresponding oxidized and unoxidized sequence ion based on MS/MS specturm. The absolute level of oxidation for a given amino acid residue is based on both the average oxidation event of peptide and the fractional oxidiation of the corresponding sequence ions. A High spatial resolution HRPF workflow is shown in **Fig 1.4**. High spatial resolution HRPF necessitates not only the identification of the sites of oxidation on a residue level, but also the quantitation of the level of oxidation at each site within the peptide, providing structural information with higher spatial resolution and more accuracy.

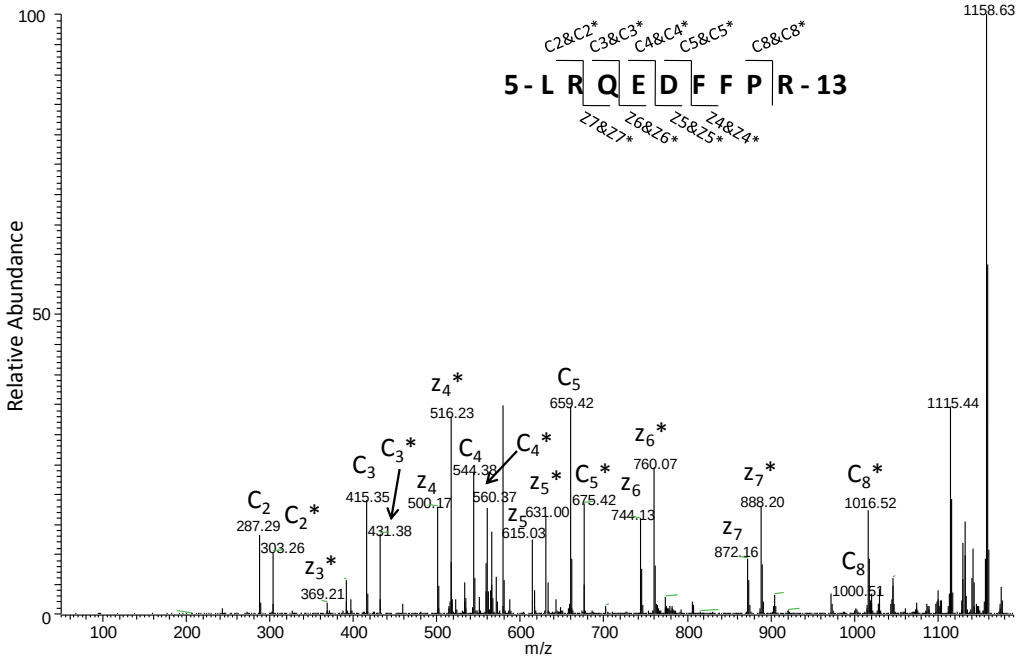

Figure 1.3. Representative ETD spectrum of singly-oxidized peptide LRQEDFFPR. The asterisks indicate the product ions that are oxidized.

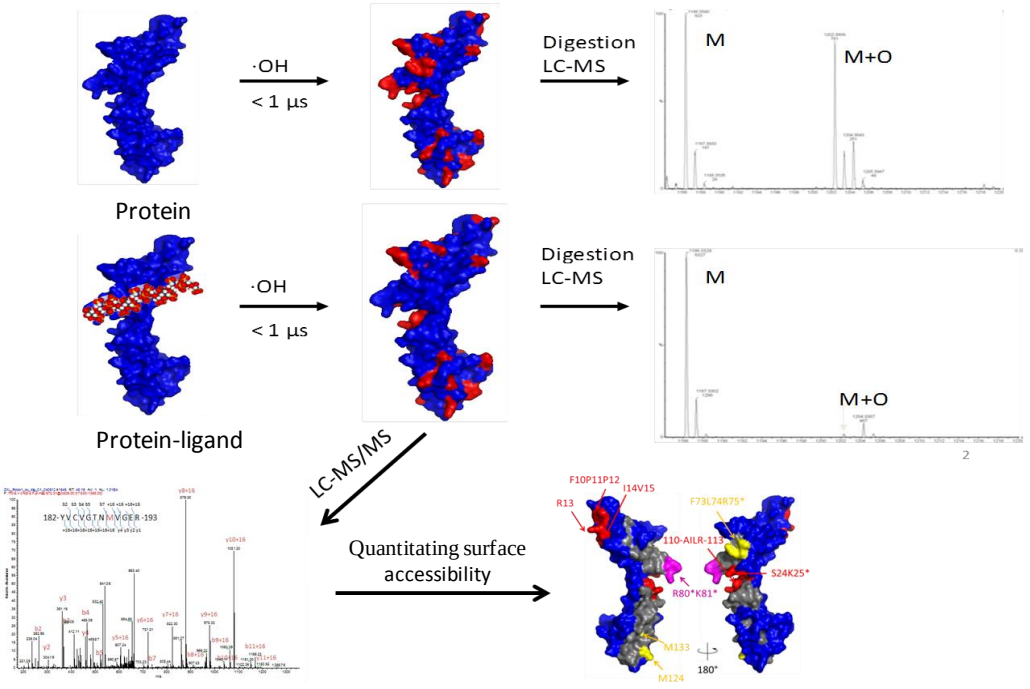

Figure 1.4. High Spatial Resolution HRPF Workflow. Sample of protein and protein with ligand were oxidatively labeled with hydroxyl radicals. The resulting peptides were analyzed by LC/MS and LC-MS/MS for peptide and residue level oxidation quantitation, respectively. Residues with significant change in surface accessibility are used to map the protein-ligand binding interface.

In **Chapter 2**, we compared the ability of ETD, collision-induced dissociation (CID) and nano-ultra high performance liquid chromatography (nano UPLC) separation to quantiate oxidation isomers with one oxidation at multiple adjacent sites in mixtures of peptide. We demonstrated the accuracy of ETD-based methods for quantifying multiple adjacent sites of isomeric oxidation products, as well as the general applicability of ETD-based dissociation to quantify oxidized peptide isomers with multiple sites of oxidation on both a series of synthetic peptide oxidation isomers and peptides from Robo1 protein sample generated by HRPF.

Efficient ETD requires high peptides charge states, which limits its applicability for HRPF. An improvement for ETD-based quantiation of HRPF is described in **Chapter 3**. We demonstrated that the addition of *m*-NBA to both a defined mixture of synthetic isomeric oxidized peptides and Robo1 protein subjected to HRPF increased the abundance of higher charge state ions, improving our ability to perform efficient ETD of the mixture, which indicates the utility of *m*-NBA for residue-level quantification of peptide oxidation in HRPF and other applications.

Interaction of transmembrane receptors of the Robo family and the secreted protein Slit provides important signals in the development of the central nervous system and regulation of axonal midline crossing. Heparan sulfate (HS), a sulfated linear polysaccharide modified in a complex variety of ways, serves as an essential co-receptor in Slit-Robo signaling. For the first time, we utilized ETD-based high spatial resolution HRPF to identify two separate binding sites for heparin interaction with Robo1 as presented in **Chapter 4**. We have proposed a model for heparin/heparan sulfate binding and activation of the Slit-Robo complex. This work marks the first use of ETD-based high spatial resolution hydroxyl radical protein footprinting, which shows great utility for the characterization of protein-carbohydrate complexes.

Platelet factor 4 (PF4, CXCL4) is a 7.8 kDa chemotactic cytokine (chemokine) that is known to be involved in numerous biological processes. Many of the biological activities of CXCL4 appear to be related to its interaction with cell-surface glycosaminoglycans via a protein surface binding domain. The structure of CXCL4 and CXCL4 K50E mutant in solution were firstly characterized at residue level by ETD-based HRPF as described in **Chapter 5**. By comparing residues with significant change of solvent accessibility between wild type CXCL4 in tetramer and CXCL4 K50E in dimer form, we have proposed a model of the dimer-dimer interface of CXCL4

and its disruption by the K50E mutation, providing more detailed structural information of the dimer-dimer interface of CXCL4 tetramer in solution.

References

1. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

2. Smyth, M.S. and J.H.J. Martin, *x Ray crystallography.* Journal of Clinical Pathology-Molecular Pathology, 2000. **53**(1): p. 8-14.

3. Ilari, A. and C. Savino, *Protein structure determination by x-ray crystallography.* Methods Mol Biol, 2008. **452**: p. 63-87.

4. Edwards, A.M., et al., *Protein production: feeding the crystallographers and NMR spectroscopists.* Nature Structural Biology, 2000. **7**: p. 970-972.

5. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.

6. Yu, H., *Extending the size limit of protein nuclear magnetic resonance*. 1999, National Acad Sciences.

7. Jonic, S. and C. Venien-Bryan, *Protein structure determination by electron cryo-microscopy.* Current Opinion in Pharmacology, 2009. **9**(5): p. 636-642.

8. Henderson, R., *The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic-Resolution Microscopy of Unstained Biological Molecules.* Quarterly Reviews of Biophysics, 1995. **28**(2): p. 171-193.

9. Walzthoeni, T., et al., *Mass spectrometry supported determination of protein complex structure.* Current Opinion in Structural Biology, 2013. **23**(2): p. 252-260.

10. Konermann, L., X. Tong, and Y. Pan, *Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches.* Journal of Mass Spectrometry, 2008. **43**(8): p. 1021-1036.

11. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.

12. Xu, G.H. and M.R. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics.* Chemical Reviews, 2007. **107**(8): p. 3514-3543.

13. Englander, J.J., et al., *Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 7057-62.

14. Campbell, S., et al., *Deuterium exchange reactions as a probe of biomolecule structure. Fundamental studies of cas phase H/D exchange reactions of protonated glycine oligomers with D2O, CD3OD, CD3CO2D, and ND3.* Journal of the American Chemical Society, 1995. **117**(51): p. 12840-12854.

15. Wales, T.E. and J.R. Engen, *Hydrogen exchange mass spectrometry for the analysis of protein dynamics.* Mass Spectrometry Reviews, 2006. **25**(1): p. 158-170.

16. Sinz, A., *Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions.* Mass Spectrom Rev, 2006. **25**(4): p. 663-82.

17. Back, J.W., et al., *Chemical cross-linking and mass spectrometry for protein structural modeling.* J Mol Biol, 2003. **331**(2): p. 303-13.

18. Ly, T. and R.R. Julian, *Using ESI-MS to probe protein structure by site-specific noncovalent attachment of 18-crown-6.* Journal of the American Society for Mass Spectrometry, 2006. **17**(9): p. 1209-1215.

19. Mendoza, V.L. and R.W. Vachet, *Probing protein structure by amino acid-specific covalent labeling and mass spectrometry.* Mass Spectrom Rev, 2009. **28**(5): p. 785-815.

20. Takamoto, K. and M.R. Chance, *Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 251-76.

21. Englander, S.W. and N.R. Kallenbach, *Hydrogen exchange and structural dynamics of proteins and nucleic acids.* Q Rev Biophys, 1983. **16**(4): p. 521-655.

22. Garcia, R.A., D. Pantazatos, and F.J. Villarreal, *Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions.* Assay Drug Dev Technol, 2004. **2**(1): p. 81-91.

23. Engen, J.R. and D.L. Smith, *Investigating protein structure and dynamics by hydrogen exchange MS.* Anal Chem, 2001. **73**(9): p. 256A-265A.

24. Zhu, M.M., et al., *Quantification of protein-ligand interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX.* J Am Chem Soc, 2003. **125**(18): p. 5252-3.

25. Chalmers, M.J., et al., *Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions.* Expert Rev Proteomics, 2011. **8**(1): p. 43-59.

26. Tang, X. and J.E. Bruce, *Chemical cross-linking for protein-protein interaction studies.* Methods Mol Biol, 2009. **492**: p. 283-93.

27. Maleknia, S.D. and K. Downard, *Radical approaches to probe protein structure, folding, and interactions by mass spectrometry.* Mass Spectrom Rev, 2001. **20**(6): p. 388-401.

28. Guan, J.Q. and M.R. Chance, *Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry.* Trends in Biochemical Sciences, 2005. **30**(10): p. 583-592.

29. Baichoo, N. and T. Heyduk, *Mapping conformational changes in a protein: application of a protein footprinting technique to cAMP-induced conformational changes in cAMP receptor protein.* Biochemistry, 1997. **36**(36): p. 10830-6.

30. Baichoo, N. and T. Heyduk, *Mapping cyclic nucleotide induced conformational changes in cyclicAMP receptor protein by a protein footprinting technique using different chemical proteases.* Protein Science, 1999. **8**(3): p. 518-528.

31. Heyduk, E. and T. Heyduk, *Mapping protein domains involved in macromolecular interactions: a novel protein footprinting approach.* Biochemistry, 1994. **33**(32): p. 9643-50.

32. Heyduk, T., et al., *Determinants of RNA polymerase alpha subunit for interaction with beta, beta', and sigma subunits: hydroxyl-radical protein footprinting.* Proceedings of the National Academy of Sciences, 1996. **93**(19): p. 10162.

33. Rana, T.M. and C.F. Meares, *Specific Cleavage of a Protein by an Attached Iron Chelate.* Journal of the American Chemical Society, 1990. **112**(6): p. 2457-2458.

34. Garrison, W.M., *Reaction-Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins.* Chemical Reviews, 1987. **87**(2): p. 381-398.

35. Garrison, W.M., Kland-English, M., Sokol, H. A., Jayko, M. E., *Radiolytic degradation of the peptide main chain in dilute aqueous solution containing oxygen.* J. Phys. Chem., 1970. **74**(26): p. 4506-4509.

36. Maleknia, S.D., M. Brenowitz, and M.R. Chance, *Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry.* Anal Chem, 1999. **71**(18): p. 3965-73.

37. Xu, G. and M.R. Chance, *Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting.* Anal Chem, 2005. **77**(14): p. 4549-55.

38. Guan, J.Q., S.C. Almo, and M.R. Chance, *Synchrotron radiolysis and mass spectrometry: A new approach to research on the actin cytoskeleton.* Accounts of Chemical Research, 2004. **37**(4): p. 221-229.

39. Guan, J.Q., et al., *Mapping the G-actin binding surface of cofilin using synchrotron protein footprinting.* Biochemistry, 2002. **41**(18): p. 5765-5775.

40. Kiselar, J.G., et al., *Structural analysis of gelsolin using synchrotron protein footprinting.* Molecular & Cellular Proteomics, 2003. **2**(10): p. 1120-1132.

41. Kiselar, J.G., et al., *Visualizing the Ca2+-dependent activation of gelsolin by using synchrotron footprinting.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(7): p. 3942-3947.

42. Kiselar, J.G., et al., *Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry.* International Journal of Radiation Biology, 2002. **78**(2): p. 101-114.

43. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry.* Analytical Biochemistry, 2003. **313**(2): p. 216-225.

44. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry.* Analytical Chemistry, 2004. **76**(3): p. 672-683.

45. Sharp, J.S., et al., *Photochemical surface mapping of C14S-Sml1p for constrained computational modeling of protein structure.* Analytical Biochemistry, 2005. **340**(2): p. 201-212.

46. Aye, T.T., T.Y. Low, and S.K. Sze, *Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry.* Anal Chem, 2005. **77**(18): p. 5814-22.

47. Hambly, D.M. and M.L. Gross, *Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale.* J Am Soc Mass Spectrom, 2005. **16**(12): p. 2057-63.

48. Hambly, D. and M. Gross, *Laser flash photochemical oxidation to locate heme binding and conformational changes in myroglobin.* International Journal of Mass Spectrometry, 2007. **259**(1-3): p. 124-129.

49. Maleknia, S.D., M.R. Chance, and K.M. Downard, *Electrospray-assisted modification of proteins: a radical probe of protein structure.* Rapid Commun Mass Spectrom, 1999. **13**(23): p. 2352-8.

50. Sharp, J.S., et al., *Photochemical surface mapping of C14S-Sml1p for constrained computational modeling of protein structure.* Anal Biochem, 2005. **340**(2): p. 201-12.

51. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry.* Anal Chem, 2004. **76**(3): p. 672-83.

52. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Protein surface mapping by chemical oxidation: structural analysis by mass spectrometry.* Anal Biochem, 2003. **313**(2): p. 216-25.

53. Goldsmith, S.C., et al., *Synchrotron protein footprinting: a technique to investigate protein-protein interactions.* J Biomol Struct Dyn, 2001. **19**(3): p. 405-18.

54. Liu, R., et al., *Structural reorganization of the transferrin C-lobe and transferrin receptor upon complex formation: the C-lobe binds to the receptor helical domain.* Biochemistry, 2003. **42**(43): p. 12447-54.

55. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochem Biophys Res Commun, 2001. **287**(3): p. 614-21.

56. Maleknia, S.D., et al., *Determination of macromolecular folding and structure by synchrotron x-ray radiolysis techniques.* Anal Biochem, 2001. **289**(2): p. 103-15.

57. Kiselar, J.G., et al., *Visualizing the Ca2+-dependent activation of gelsolin by using synchrotron footprinting.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3942-7.

58. Schantz, J.T., et al., *Repair of calvarial defects with customised tissue-engineered bone grafts - II. Evaluation of cellular efficiency and efficacy in vivo.* Tissue Engineering, 2003. **9**: p. S127-S139.

59. Berlett, B.S. and E.R. Stadtman, *Protein oxidation in aging, disease, and oxidative stress.* Journal of Biological Chemistry, 1997. **272**(33): p. 20313-20316.

60. Vonsonntag, C. and H.P. Schuchmann, *Suppression of Hydroxyl Radical Reactions in Biological-Systems - Considerations Based on Competition Kinetics.* Oxygen Radicals in Biological Systems, Pt C, 1994. **233**: p. 47-56.

61. Adilakshmi, T., R.A. Lease, and S.A. Woodson, *Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation.* Nucleic Acids Research, 2006. **34**(8).

62. Goldsmith, S.C., et al., *Synchrotron protein footprinting: A technique to investigate protein-protein interactions.* Journal of Biomolecular Structure & Dynamics, 2001. **19**(3): p. 405-418.

63. Rashidzadeh, H., et al., *Solution structure and interdomain interactions of the Saccharomyces cerevisiae "TATA binding protein" (TBP) probed by radiolytic protein footprinting.* Biochemistry, 2003. **42**(13): p. 3655-3665.

64. Franchetbeuzit, J., et al., *Radiolytic Footprinting - Beta-Rays, Gamma-Photons, and Fast-Neutrons Probe DNA-Protein Interactions.* Biochemistry, 1993. **32**(8): p. 2104-2110.

65. Armstron.Rc and A.J. Swallow, *Pulse-Radiolysis and Gamma-Radiolysis of Aqueous Solutions of Tryptophan.* Radiation Research, 1969. **40**(3): p. 563-579.

66. Kopoldova, J. and S. Hrncir, *Gamma-Radiolysis of Aqueous-Solution of Histidine.* Zeitschrift Fur Naturforschung C-a Journal of Biosciences, 1977. **32**(7-8): p. 482-487.

67. Winchest.Rv and K.R. Lynn, *X-Radiolysis and Gamma-Radiolysis of Some Tryptophan Dipeptides.* International Journal of Radiation Biology and Related Studies in Physics Chemistry and Medicine, 1970. **17**(6): p. 541-548.

68. Hambly, D.M. and M.L. Gross, *Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale.* Journal of the American Society for Mass Spectrometry, 2005. **16**(12): p. 2057-2063.

69. Aye, T.T., T.Y. Low, and S.K. Sze, *Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry.* Analytical Chemistry, 2005. **77**(18): p. 5814-5822.

70. Sharp, J.S. and K.B. Tomer, *Analysis of the oxidative damage-induced conformational changes of apo- and holocalmodulin by dose-dependent protein oxidative surface mapping.* Biophys J, 2007. **92**(5): p. 1682-92.

71. Sharp, J.S., et al., *Measurement of multisite oxidation kinetics reveals an active site conformational change in Spo0F as a result of protein oxidation.* Biochemistry, 2006. **45**(20): p. 6260-6266.

72. Venkatesh, S., K.B. Tomer, and J.S. Sharp, *Rapid identification of oxidation-induced conformational changes by kinetic analysis.* Rapid Communications in Mass Spectrometry, 2007. **21**(23): p. 3927-3936.

73. Buxton, G.V., F.C.R. Cattell, and F.S. Dainton, *Application of Time-Dependent Rate Constant Theory to Reactions of Solvated Electrons - Reaction Distances, Rate Constants and Diffusion-Coefficients in Concentrated Aqueous-Solutions.* Journal of the Chemical Society-Faraday Transactions I, 1975. **71**(1): p. 115-122.

74. Venkatesh, S., K.B. Tomer, and J.S. Sharp, *Rapid identification of oxidation-induced conformational changes by kinetic analysis.* Rapid Commun Mass Spectrom, 2007. **21**(23): p. 3927-36.

75. Winchester, R.V., Lynn, K. R., *X- and gamma-radiolysis of some tryptophan dipeptides.* Int J Radiat Biol, 1970. **17**(6): p. 541-548.

76. Watson, C., et al., *Pulsed Electron Beam Water Radiolysis for Submicrosecond Hydroxyl Radical Protein Footprinting.* Anal. Chem, 2009. **81**(7): p. 2496-2505.

77. Gau, B.C., et al., *Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding.* Analytical Chemistry, 2009. **81**(16): p. 6563-6571.

78. Charvátová, O., et al., *Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1.* J. Am. Soc. Mass Spectr., 2008. **19**: p. 1692-1705.

79. Xu, G.Z. and M.R. Chance, *Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting.* Analytical Chemistry, 2005. **77**(14): p. 4549-4555.

80. Garrison, W.M., *Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins.* Chemical Reviews, 1987. **87**(2): p. 381-398.

81. Xu, G., K. Takamoto, and M.R. Chance, *Radiolytic modification of basic amino acid residues in peptides: probes for examining protein-protein interactions.* Anal Chem, 2003. **75(**24): p. 6995-7007.

82. Xu, G. and M.R. Chance, *Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions.* Anal Chem, 2004. **76(**5): p. 1213-21.

83. Xu, G., et al., *Secondary reactions and strategies to improve quantitative protein footprinting.* Anal Chem, 2005. **77**(10): p. 3029-37.

84. Xu, G. and M.R. Chance, *Radiolytic modification of sulfur-containing amino acid residues in model peptides: fundamental studies for protein footprinting.* Anal Chem, 2005. **77**(8): p. 2437-49.

85. Nukuna, B.N., M.B. Goshe, and V.E. Anderson, *Sites of hydroxyl radical reaction with amino acids identified by (2)H NMR detection of induced (1)H/(2)H exchange.* J. Am. Chem. Soc., 2001. **123**: p. 1208–14.

86. Goshe, M.B., Y.H. Chen, and V.E. Anderson, *Identification of the sites of hydroxyl radical reaction with peptides by hydrogen/deuterium exchange: prevalence of reactions with the side chains.* Biochemistry, 2000. **39**: p. 1761–1770.

87. Buxton, G.V., et al., *Critical-Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen-Atoms and Hydroxyl Radicals (.Oh/.O-) in Aqueous-Solution.* Journal of Physical and Chemical Reference Data, 1988. **17**(2): p. 513-886.

88. Xu, G. and M. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics.* Chem. Rev, 2007. **107**(8): p. 3514-3543.

89. Charvatova, O., et al., *Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1*. Journal of the American Society for Mass Spectrometry, 2008. **19**(11): p. 1692-1705.

90. Watson, C. and J.S. Sharp, *Conformational Analysis of Therapeutic Proteins by Hydroxyl Radical Protein Footprinting.* Aaps Journal, 2012. **14**(2): p. 206-217.

91. Chen, J.W., et al., *Fast Photochemical Oxidation of Proteins and Mass Spectrometry Follow Submillisecond Protein Folding at the Amino-Acid Level.* Journal of the American Chemical Society, 2012. **134**(45): p. 18724-18731.

92. Kaur, P., et al., *Quantitative Protein Topography Analysis and High-Resolution Structure Prediction Using Hydroxyl Radical Labeling and Tandem-Ion Mass Spectrometry (MS).* Molecular & Cellular Proteomics, 2015. **14**(4): p. 1159-1168.

CHAPTER 2

IMPROVED IDENTIFICATION AND RELATIVE QUANTIFICATION OF SITES OF PEPTIDE AND PROTEIN OXIDATION FOR HYDROXYL RADICAL FOOTPRINTING[1]

[1]Li, X.; Li, Z.; Xie, B.; Sharp, J. S. *Journal of the American Society for Mass Spectrometry*, **2013**. *24(*11): p. 1767-1776. Reprinted here with permission of publisher.

ABSTRACT

Protein oxidation is typically associated with oxidative stress and aging and affects protein function in normal and pathological processes. Additionally, deliberate oxidative labeling is used to probe protein structure and protein–ligand interactions in hydroxyl radical protein footprinting (HRPF). Oxidation often occurs at multiple sites, leading to mixtures of oxidation isomers that differ only by the site of modification. We utilized sets of synthetic, isomeric "oxidized" peptides to test and compare the ability of electron-transfer dissociation (ETD) and collision induced dissociation (CID), as well as nano-ultra high performance liquid chromatography (nanoUPLC) separation, to quantitate oxidation isomers with one oxidation at multiple adjacent sites in mixtures of peptides. Tandem mass spectrometry by ETD generates fragment ion ratios that accurately report on relative oxidative modification extent on specific sites, regardless of the charge state of the precursor ion. Conversely, CID was found to generate quantitative MS/MS product ions only at the higher precursor charge state. Oxidized isomers having multiple sites of oxidation in each of two peptide sequences in HRPF product of protein Robo-1 Ig1-2, a protein involved in nervous system axon guidance, were also identified and the oxidation extent at each residue was quantified by ETD without prior liquid chromatography (LC) separation. ETD has proven to be are liable technique for simultaneous identification and relative quantification of a variety of functionally different oxidation isomers, and is a valuable tool for the study of oxidative stress, as well as for improving spatial resolution for HRPF studies.

INTRODUCTION

Oxidative modification of proteins occurs in vivo during aging and in certain disease conditions. It can also be induced experimentally by a wide array of oxidizing agents [1, 2]. In vivo oxidative modification to proteins can lead to diverse functional consequences, such as inhibition of enzymatic and binding activities, increased susceptibility to aggregation and proteolysis, increased or decreased uptake by cells, and altered immunogenicity. Quantitatively profiling changes in modification levels and characterizing the exact amino acid sites of modification have become essential parts of trying to elucidate the mechanisms of a variety of important biological process [3–5]. Apart from this biological role, oxidation of proteins and peptides often occurs during sample processing and preparation. It is important to monitor the degree of oxidative modification of therapeutic proteins manufactured for commercial use. Oxidative protein modification occurring during both manufacture and storage of the protein can affect drug efficacy and potential side effects [6–8]. The most commonly used marker of protein oxidative modification in manufactured proteins is methionine sulfoxide [1, 9]. In some cases, the effect of oxidative modification on bioactivity of the protein is site-dependent, as is seen following oxidation of Escherichia coli-derived stem-cell factor [9]. It would be helpful to have an accurate and reliable analytical method to determine oxidation in proteins in a residue-specific manner in order to assess their biological significance.

Furthermore, oxidative labeling of proteins and peptides is used to probe protein structure and protein-ligand interactions in hydroxyl radical protein footprinting (HRPF) coupled with mass spectrometry (MS) analysis [10, 11], in which the protein of interest is exposed to diffusing hydroxyl radicals and the amounts of oxidation of solvent-exposed residues are monitored by tandem mass spectrometry. HRPF takes advantage of the fact that the rate of oxidation of various

amino acid side chains varies directly with their accessibility to solvent [12, 13], allowing the researcher to monitor conformational changes and ligand binding surfaces. The level of accuracy and structural detail generated from the HRPF data relies on improved methods for detecting and quantifying at the residue level adjacent sites of oxidation in the protein sequence. MS is a powerful tool for characterization of oxidative protein modification because of its sensitivity, speed, and specificity of detection. It is routine to differentiate oxidized peptides from their unoxidized counterparts either on the basis of their mass difference and/or liquid chromatographic separation, then the liquid chromatography-mass spectrometry (LC-MS) peak intensity or selected reaction monitoring (SRM)-based methods will be utilized for relative quantification [12, 14–19].

The likelihood of multiple significant oxidation sites existing on a single peptide increases significantly with the development of sub-microsecond HRPF techniques. Earlier methods of HRPF including water radiolysis by gamma (γ) rays [20] and X-rays [13, 21, 22], worked on the millisecond-to-minutes timescales, and required careful control of the extent of oxidation to ensure that the protein maintained the native structure during the HRPF process [23, 24]. In 2005, Hambly and Gross reported Fast Photochemical Oxidation of Proteins (FPOP), an HRPF method that allows extensive labeling of proteins (and therefore more sites of oxidation) during a short time scale of oxidation (sub-microsecond) while retaining their native structure [15–19, 25, 26]. In 2009, Watson et al. developed a novel method for hydroxyl radical protein footprinting using a Van de Graaf electron accelerator-based method that achieved HRPF by water radiolysis on a similar sub-microsecond timescale [27]. The FPOP method enjoys the advantage of using a much more readily available radiation source and offering more flexibility in the photoactivated chemistry, whereas the electron accelerator has the advantage of not requiring the addition of any exogenous radical precursors, forming the hydroxyl radicals directly from water. Both methods

complete labeling of the protein in a time scale that is faster than the protein can unfold. Therefore, unlike the previous slower methods, these methods are unrestrained by the need to dose-control the radical exposure to maintain native structure. Dose-control of the radical exposure for these sub-microsecond methods is required only to maintain the covalent integrity of the protein in order to yield identifiable fragments, which allows for much higher radical dosages. At these higher radical doses, it is common to label amino acids that are insufficiently reactive to label when native structure must be maintained for longer periods of time. This more extensive labeling results in a higher number of isomeric oxidation products that are co-localized to one tryptic peptide, resulting in an analytical challenge to quantify how much oxidation occurs at each site.

Traditionally, HRPF quantification was performed on the peptide level. Slower radical generation techniques that required lower oxidation amounts would often only oxidize a single amino acid in a given tryptic peptide to an appreciable extent, making this peptide-level quantification of oxidation acceptable. However, as the sub-microsecond oxidation techniques have been developed, and as FPOP in particular has been more widely used for HRPF experiments, the more heavily oxidized analytes have resulted in a much more complex array of isomeric oxidized peptides. While traditional oxidation quantification based on measurements of levels of oxidized peptides allows for structural information, the spatial resolution of such information is strictly limited by the size of the peptide, limiting the applicability of these data to high resolution structural analysis, including constrained modeling approaches. Improved methods for detecting and quantifying oxidation at the residue level, even for adjacent sites of oxidative modification in the protein sequence, will greatly increase the accuracy and level of structural detail of any model generated from the data.

The quantification of the apparent rate of oxidation on any specific residue requires specific quantification of oxidation at that amino acid residue from within a pool of multiple isomers that differ only by the site of oxidation. However, it is still a very challenging task to distinguish and accurately quantify peptide isomers with oxidation at multiple adjacent sites in a single sequence that results in an identical mass shift and cannot be easily separated. Previous research by Woods and co-workers [12] relied upon LC separation of oxidation isomers, combined with semi-automated MS/MS identification of the major oxidation product in each MS/MS spectrum. The spectral intensities from the MS/MS spectrum would be assigned to the major oxidation product for quantification. Similar work with superior chromatography has been reported by Gross and coworkers, using the signal intensities from the separated oxidation isomer for quantification [17–19]. Spectral counting in the MS/MS mode for partially separated oxidation isomers has also been used for quantitative analysis of peptide isomers [16]. However, because of a lack of rigorously known oxidation levels for any of these systems, the accuracy of these methods of analysis remains unknown.

Recently, Vachet and coworkers have reported the success of using electron transfer dissociation (ETD) fragment ion intensity to accurately quantify the residue oxidation level in isomeric mixtures of methionine sulfoxide-containing peptides [28] and peptides that contain both oxidized histidine residues and other oxidized residues [29]. Similarly, Jumper et al. recently reported use of ETD in quantitating mixtures of carbene footprinting products [30]. In contrast, collision-induced dissociation (CID) MS/MS spectra of these oxidized products rarely reflect the accurate oxidation extent at each given residue. This occurs because the dissociation chemistry of each oxidized isomer can be very different during CID, which gives rise to complicated spectra that are easy to misinterpret [28, 31, 32]. However, insufficient work has been done in this area to

demonstrate the accuracy of ETD-based methods for quantifying multiple adjacent sites of oxidation with the various classes of oxidation products, as well as the general applicability of ETD-based dissociation to quantify oxidized peptide isomers having more than one site of oxidation on a single peptide generated in real HRPF samples.

Here we utilized sets of synthetic, isomeric "oxidized" peptides from a model tryptic peptide with a sequence RPMFAIWK to test and compare the ability of different fragmentation methods (both ETD and CID), as well as nano-ultra high performance liquid chromatography (nanoUPLC) separation, to quantify isomers with one oxidation at multiple adjacent sites.. Oxidized isomers having multiple sites of oxidation in each of two peptide sequences in the HRPF product of the protein Robo-1 Ig1-2, a protein involved in nervous system axon guidance, were also identified, and the oxidation extent at each residue was also quantified by ETD without prior LC separation.

EXPERIMENTAL

Materials

Hydrogen peroxide (30%) was obtainedfrom J.T. Baker (Phillipsburg, NJ). Dithiothreitol (DTT), HPLC-grade acetonitrilewere purchased from Fisher Scientific (Fair Lawn, NJ). Methionine amide was purchased from Bachem (Torrace, CA, USA). Catalase, formic acid and L-glutamine were obtained from the Sigma-Aldrich Corporation (St. Louis, MO). Sequencing-grade modified Lys-C was purchased from Promega Corporation (Madison, WI, USA). All reagents were used as provided. Purified water (18 MΩ) was obtained froman in-house Milli-Q Synthesis system (Millipore, Billerica, MA). The peptide RPMFAIWK and its oxidized isomers: RP*MFAIWK (Pro2→hydroxyproline), RPMF*AIWK (Phe4→tyrosine), and RPMFA*IWK (Ala5→serine) were synthesized by GenScript USA Inc. (Piscatanay, NJ). Peptide RPM*FAIWK

(Met3→methionine sulfoxide) was prepared in house by adding 90 μM unmodified peptide RPMFAIWK to a solution of hydrogen peroxide (3%) and incubated in the dark for 1 h, followed by C18 reverse phase HPLC purification of the methionine sulfoxide-containing product. The amount of these synthetic peptides was individually quantified by amino acid analysis and was nominally in 50 μM range. Working stocks of each peptide were prepared by diluting solutions down to a concentration of 2 μM in acetonitrile:water:formic acid (50:50:0.1, vol/vol/vol).

Construction, Expression and Purification of Robo 1Ig1-2

The protein sequence of the translation of the human ROBO1 gene was analyzed for domain boundaries using the UniProt database and a truncated protein sequence comprised of Ig domains 1 and 2 (plus flanking regions) was chosen for gene synthesis. Optimized codon utilization for human protein expression was employed in the synthesis of the corresponding coding region and the resulting DNA fragment was cloned into our mammalian expression vector (pGEn2) using restriction digestion and ligation into corresponding restriction sites in the vector. Large scale DNA preparations were prepared and transiently transfected into HEK293S GnTI- suspension culture cells and recombinant protein was harvested after 6 days of transfection. The recombinant protein was purified from the culture supernatant using Ni^{2+}-NTA chromatography and concentrated to ~1mg/mL. The resulting protein preparation was digested with recombinant TEV-GFP and EndoF1-GFP and then further purified by Ni^{2+}-NTA chromatography and gel filtration.

Sample Preparation

The working stock solutions of the four peptides were mixed in five different volume ratios (1:1:1:1, 1:2:3:4, 4:3:2:1, 1:1:9:9, 9:9:1:1) in triplicate at a final total molar concentration of 2 μM to prepare the mimetic oxidative isomeric mixtures generated under different conditions.

Protein Robo1 Ig1-2 was oxidized using laser flash photolysis of hydrogen peroxide as described previously [12, 18]. Briefly, a 248 nm GAM EX100 KrF excimer laser operating at 50 mJ per pulse output was used to photolyze H_2O_2. Protein Robo1 Ig1-2 was reconstituted in an ammonium bicarbonate buffered solution (50 mM, pH=7.8) at a concentration of 30 μM. Each 20μL reaction buffer (50 mM ammonium bicarbonate) was prepared with a final concentration of 6 μM Robo1 Ig1-2, 3% hydrogen peroxide (freshly prepared and added right prior to irradiation) and 20 mM glutamine. The solution was then loaded into a 100 μL syringe and introduced via a syringe pump coupled to the 100 μm i.d. fused silica tubing. The flow rate was set to 12.19 μL/min, with the laser pulse repetition rate set to illuminate each volume of sample with a single focused laser pulse as it passed through the laser path, with 10% of the total volume remaining unirradiated to correct for diffusion and laminar flow effects. The capillary outflow from a sufficient number of laser shots was collected in a microcentrifuge tube, to which was added immediately 20μL of water washed agarose-immobilized catalase (0.5 mg/mL) and methionine amide (0.5 mg/mL) to remove any remaining H_2O_2. Following irradiation, the sample (40μL) was mixed with ammonium bicarbonate buffer (40μL, 50 mM) and DTT (20 μL, 50 mM) and then incubated at 65°C for 30 min to denature and reduce the protein. After cooling to room temperature, a 1:100 weight ratio of Lys-C (enzyme to protein, w to w ratio of trypsin) was added to the protein sample and incubated at 37 °C overnight while rotating. The digested samples were stored at -20°C for LC-MS/MS analysis.

Instrumentation

Mass spectral analyses were performed on the Waters Synapt G2 Q-TOF mass spectrometer configured with both ETD and CID capabilities, and with a nano-ESI source that was coupled to either a nanoACQUITY (Waters) UPLC system, or operated in infusion mode. For direct injection experiments, the sample was delivered at 0.4μL/min using a syringe pump. For

nanoUPLC-MS/MS analysis, the samples were separated on a Symmetry C18 180 μm x 20 mm, 5μm trapping column and a BEH130 C18 75 μm x 150 mm, 1.7μmanalytical column. The gradient elution was performed from 2–40% acetonitrile in 0.1% formic acid over 50 min at a flow rate of 0.4μL/min, and then increased to 80% acetonitrile in 0.1% formic acid for 5 min followed by a 5 min re-equilibration step. Nitrosobenzene was used as ETD reagent. Trap wave height was set at 0.2 V to allow the ETD reagent to mix with the sample and fragment. Supplemental transfer CE of around 5V was used to aid the ETD of doubly-charged ions. For CID, the collision energy was set at 25V for doubly-charged ion and 15 V for triply-charged ion, respectively. All mass spectra were acquired in positive ion mode, with spray voltage of 3 kV and source temperature 100ºC.

Quantitative Analysis

For direct injection and co-eluted samples, the fragment ion intensities from ETD and/or CID are used for the calculation of oxidation rate at specific residue site using a similar approach reported by Jumper et al. [30]. Briefly, peptide isomers with oxidation distributed throughout the sequence can generate both oxidized and unoxidized sequence ions in its MS/MS spectrum. The actual fractional oxidation of a given sequence ion is defined as the ratio between the oxidized sequence ion intensity to the sum of the intensity of the corresponding oxidized and unoxidized sequence ion. For example, for ETD spectra generating a c-ion series, the actual fractional oxidation is defined as Equation 1:

$$f(Ci) = \frac{I(Ci)oxidized}{I(Ci)unoxidized + I(Ci)oxidized} \quad (1)$$

where f(Ci) actual denotes the fractional oxidation of c-ion i. I(c_i) denotes the intensity of the c ion i, whether the oxidized and unoxidized form. The sum of the molar fraction made up of peptides that contain an oxidized residue within sequence ion i is defined as the theoretical value of f(Ci) . This is shown in Equation 2,

$$f(Ci)theoretical = \sum_{i=1}^{i} f(i)oxidized \quad (2)$$

where the f(Ci) oxidized is the molar fraction of the peptide containing an oxidation at residue i. The relative oxidation rate for a specific residue i is calculated as the difference between the fractional oxidation of adjacent residues. This is shown in Equation 3:

$$oxidation\ rate/residue\ i = f(Ci)actual - f(Ci-1) (3)$$

A similar approach is followed for CID spectra generation b-ions. For UPLC peaks of separated isomers, the selected ion monitoring (SIM) peak area is used for relative quantification.

RESEULT AND DISCUSSION

Quantification of Synthetic Peptide Isomers using CID and ETD without Prior LC Separation

It has been demonstrated that side chain oxidation of the methionine residue often drastically affects the fragmentation pattern during CID, preventing correct peptide identification, and as a result, relative quantification [28, 33]. To investigate the possible fragmentation pattern effect caused by CID on the quantification of oxidized peptide isomers, the tandem mass spectra of both CID and ETD for each of the four oxidation isomers of the model peptide RPMFAIWK was obtained by direct infusion of individual isomer. The sequence RPMFAIWK was chosen in order to meet a variety of criteria. First, we wanted to cover each of the biophysical categories of major oxidation products that we could cover with synthetic peptides: large aliphatics (Pro→hydroxyproline), sulfur-containing residues (Met→methionine sulfoxide), aromatics (Phe→tyrosine), and small aliphatics (Ala→serine). We were unable to cover the generally less reactive ionic amino acids, as their major oxidation products are not readily available for peptide synthesis, and oxidation of ionic amino acids can easily change their ionization efficiencies in dramatic ways [34]. Second, we wanted to be able to probe multiple charge states by both CID and ETD to determine the effect of charge state on quantification. This required an additional basic

site to ensure the presence of the +3 charge state. Third, we wanted to ensure that the peptide had a reasonable retention on a UPLC column to test quantification by UPLC. This caused us to include the hydrophobic Ile and Trp residues near the C-terminus. Finally, we wanted our peptide to be a possible tryptic peptide, as most HRPF experiments are performed from tryptic digests of the protein. This caused us to place the extra basic site N-terminal to a Pro, which prevents tryptic cleavage. Having the basic site at the N-terminus also made the presence of abundant +2 and +3 charge states more likely due to the charging of the N-terminus and the N-terminal Arg side chain interfering with one another, preventing the peptide from existing entirely as a +3 ion. The sequence RPMFAIWK was analyzed by a protein BLAST search against the non-redundant protein sequences database [35] and found to be strongly homologous to a large number of sequences present in biology, indicating that this peptide is not a wholly artificial and unreasonable model sequence for testing analytical methods (data not shown).

Sequence isomers were confirmed straightforwardly on the basis of their sequence-specific fragment ions. For example, the CID spectrum of RP*MFAIWK shows oxidized b2–b8 ions, indicating Pro2 is oxidized, while the CID spectrum of RPM*FAIWK shows an unoxidized b2 ion and oxidized forms of the b3–b8 ions, indicating thatMet3 is oxidized. Both ETD and CID gave confident sequence information of the site of oxidative modification on each individual isomer even for the labile side chain modification methionine sulfoxide (data not shown).

The mixtures of the four isomers, RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK, at five varying ratios: 1:1:1:1, 1:2:3:4, 4:3:2:1, 9:9:1:1, and 1:1:9:9, respectively, were then subjected to tandem mass spectrometry analysis by both ETD and CID. The representative ETD spectra of both doubly- and triply-charged ions of mixture at ratio of 1:1:1:1 are shown in **Figure 2.1**. The representative CID spectra of both charged-state ions of mixture at

ratio 1:1:1:1 are presented in **Figure 2.2**. The representative ETD spectra of both charged-state ions of mixtures at other ratios are presented in Supplementary **Figures S2.1 and S2.2**. The representative CID spectra of triply-charged ion of mixtures at other ratios are presented in Supplementary **Figure S2.3**. The mere presence or absence of an oxidized fragment ion is only sufficient to determine, at best, the N-terminal and C-terminal sites of oxidation. In order to determine internal oxidations, as well as to quantify any mixture of isomeric oxidations, the ratio of product ion intensities of the oxidized fragment to the sum of the unoxidized and oxidized version of that fragment ion must be compared. The measured percentages of the oxidized form based on c-type ions ($f(ci)_{actual}$) of ETD and b-type ions ($f(bi)_{actual}$) of CID are plotted against the theoretical percentages oxidized ($f(ci)_{theoretical}$ and $f(bi)_{theoretical}$**, Figure 2.3a, b, c, and d**), respectively. The reader should note that the experimental and theoretical percentages oxidized are for the fragment ion, not the amino acid residue. For example, a c3 ion's fractional oxidation would represent the sum of all fractional oxidation occurring on the first three amino acids of the peptide. As shown in **Figure 2.3a and b**, the molar ratio of each isomer in the mixture calculated based on abundance of the c-type product ion fragmented by ETD at both doubly- and triply-charged states correlates well to the theoretical values. Except a few outliers, the quantification method based on ETD fragmentation ion intensity at doubly- and triply-charged states exhibited an accuracy (relative error) ranging from −18 % to 24 % based on the average of triplicates. The data suggest that oxidized residues present in a mixture at 5 % relative abundance can be accurately quantitated by ETD fragmentation.

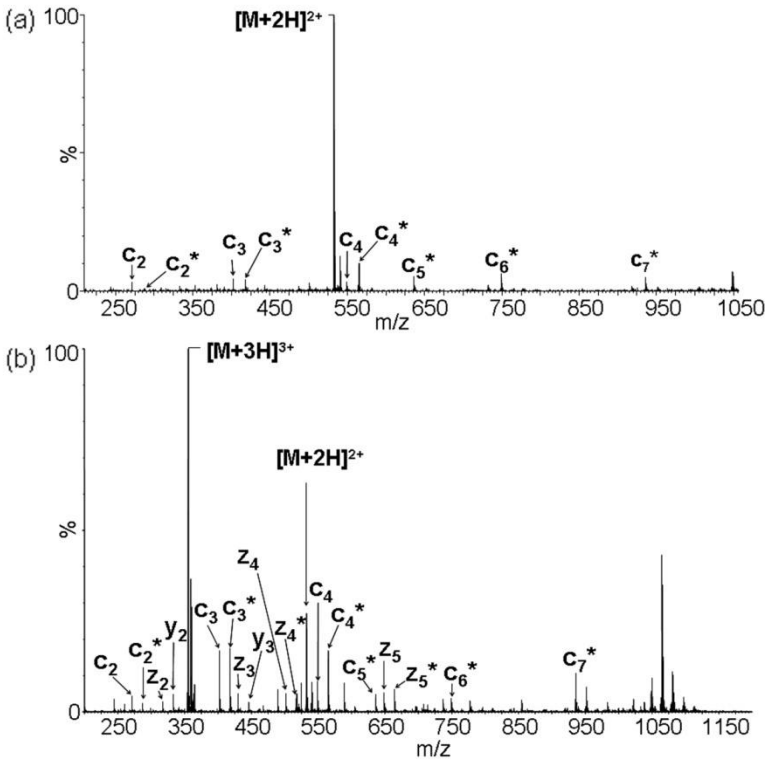

Figure 2.1. Representative ETD spectrum of mixture of four oxidation isomers of the model peptide RPMFAIWK at (a) doubly-charged state and (b) triply-charged state by direct infusion. The mixture contains RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at a molar ratio 1:1:1:1. The asterisks indicate the product ions that are oxidized.

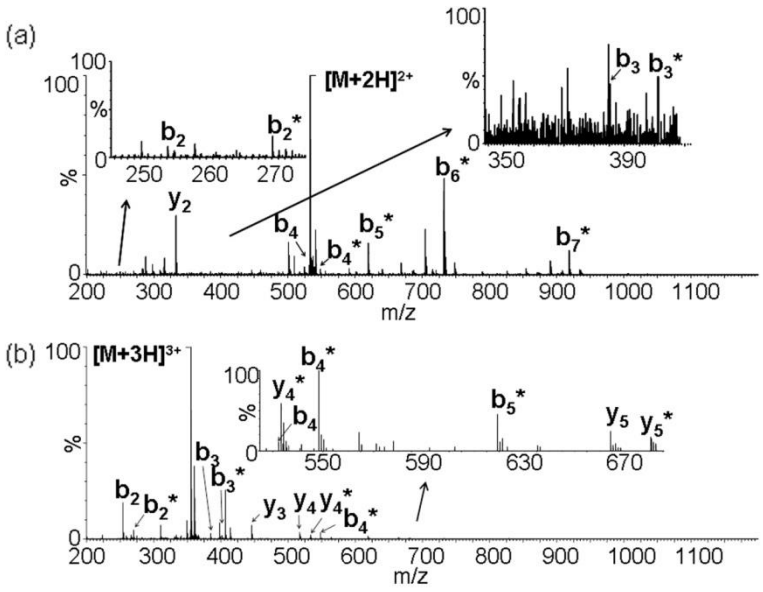

Figure 2.2. Representative CID spectrum of mixture of four oxidation isomers of the model peptide RPMFAIWK at (a) doubly-charged state and (b) triply-charged state by direct infusion. The mixture contains RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at a molar ratio 1:1:1:1. The asterisks indicate the product ions that are oxidized.

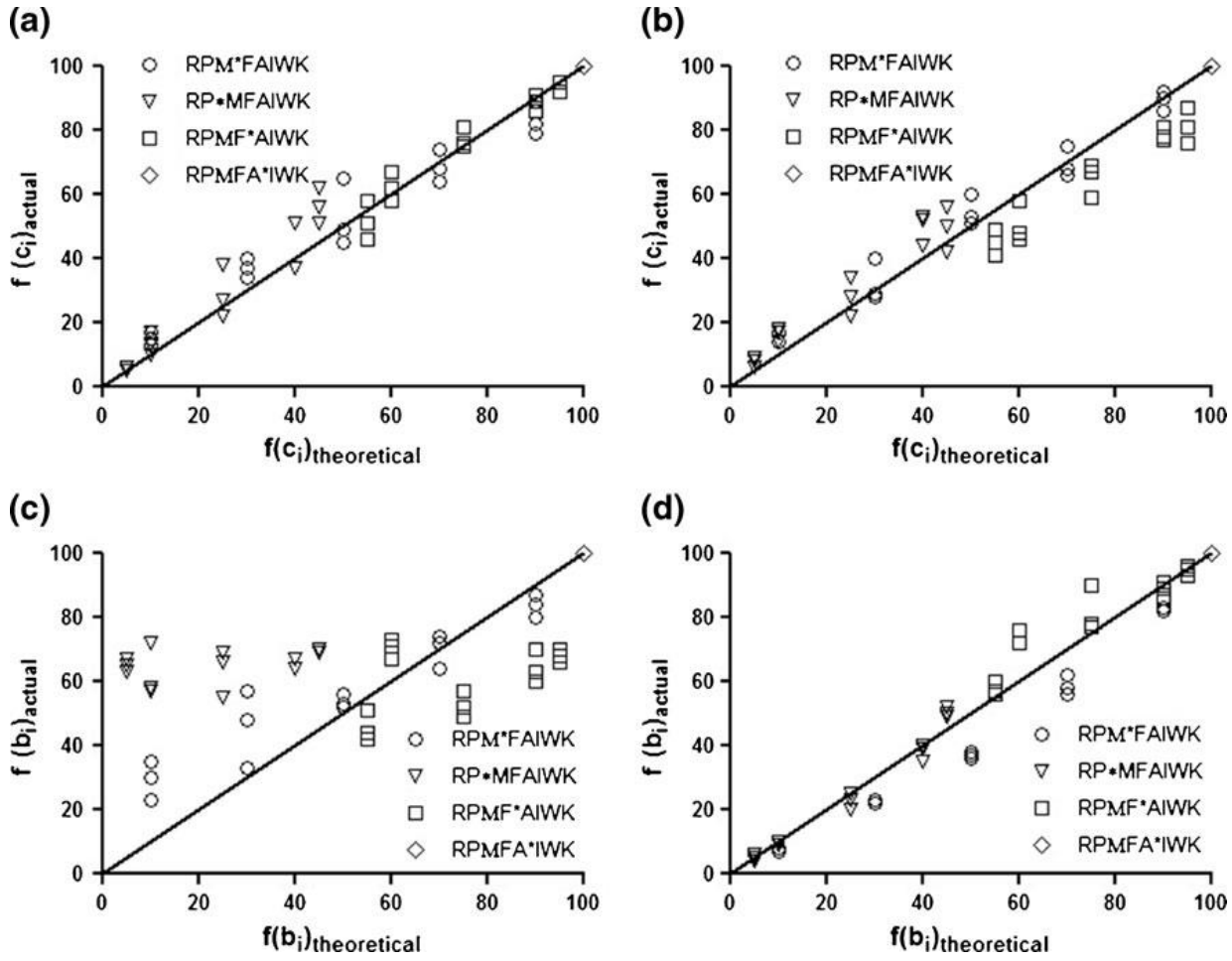

Figure 2.3. Comparison of theoretical and experimental quantification of mixtures of four oxidation isomers of the model peptide RPMFAIWK based on fragment c-type ions intensity by ETD at (a) doubly-charged state and (b) triply-charged state, and on fragment b-type ions intensity by CID at (c) doubly-charged state and (d) triply-charged state, respectively. The mixtures contain: RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at five varying molar ratios of 1:1:1:1, 1:2:3:4, 4:3:2:1, 9:9:1:1, and 1:1:9:9 (n = 3). Each replicate of the triplicate analysis is plotted as an individual data point.

As shown in **Figure 2. 3c and d**, the precursor charge state plays a major role in CID quantification; only spectra (**Figure 2.3d**) of the triply-charged ions reveals a good correlation

between the abundance of the oxidized versus unoxidized b-type product ion and the abundance of the appropriate oxidation isomers. The accuracy for quantification based on CID mass spectra of doubly-charged precursor ion ranged from −29 % to 1200 %. By contrast, the accuracy for quantification based on CID mass spectra of triply-charged precursor ion ranges from −26 % to 22 % and most of data points (over 85 %) present accuracy within ±20 %.We attempted CID of a three-component peptide mixture excluding the methionine sulfoxide-containing peptide, and found that the +2 charge state still did not yield quantitative fragmentation results (data not shown), indicating that the presence of methionine sulfoxide is not the only confounding issue in CID-based quantification at this charge state. These results lend support to the hypothesis that oxidation-influenced fragmentation is a charge-remote process that can be suppressed in the presence of sufficiently mobile protons. Further work would be required to test this hypothesis.

In addition to the successful determination of the N-terminal sequence (b and c ions), several C-terminal fragment ions were also observed in the CID and/or ETD spectra of triply-charged ions (Supplemental **Table S2.1** and **Table S2.2**). The molar ratios of isomers RPMF*AIWK and RPMFA*IWK calculated based on abundance of C-terminal fragment ions (y or z ions) are close to expected, in concert with the ratios calculated based on abundance of N-terminal fragment ions (b and/or c ions). It should be noted that some y ions were also formed in the ETD spectra of triply-charged ion, possibly due to in-source dissociation or supplemental energy activation. The mass of observed y ion series is 16.0187 da higher than the z ion series of the same amino acid residue, and the oxidized z ion is 15.9994 da higher than unoxidized, which means there is only 0.0193 da mass differences between the oxidized z ion and y ion of the same amino acid residue. High resolution MS allows these two peaks to be distinguished. For example, the oxidized z5 ion at m/z 664.343 could be resolved from the y ion at m/z 664.363 (Supplementary

Figure S2.4), which makes the quantification of oxidized isomers based on z ion intensity reliable by excluding the interference from y ions. However, in mass spectrometers with lower resolution, the interference between unoxidized y-ions and oxidized z-ions prevents accurate quantification based on C-terminal ions in ETD.

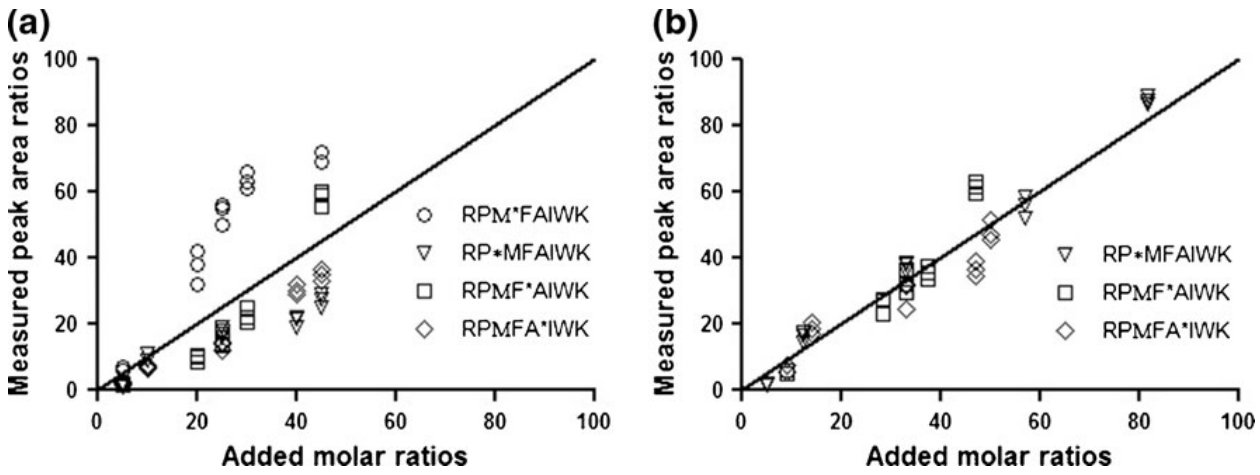

Figure 2.4. Comparison of added molar ratio and experimental quantification calculated with RPM*FAIWK included (a) and excluded (b) of mixtures of four oxidation isomers of the model peptide RPMFAIWK based on SIM peak area after UPLC separation (n = 3). Each replicate of the triplicate analysis is plotted as an individual data point.

Given the observed fact that reliability of CID-based quantification of oxidation isomers depends heavily on charge state, while ETD-based quantification is robust across both +2 and +3 charge states, it stands to reason that ETD-based quantification will be more reliable in cases where the isomeric state is unknown and the charge state sufficient for quantitative CID fragmentation is not determined beforehand.

Relative Quantification of Synthetic Peptide Isomers by nanoUPLC-MS in Selected Ion Monitoring Mode

Due to similar properties, separation of isomers by liquid chromatography (LC) requires intensive screening of various columns and mobile phase conditions. In cases of hydrophilic oxidized peptides, this task becomes even more challenging due to limited retention on traditional reversed-phase HPLC columns. During the process of the optimization of nanoUPLC conditions, two different columns: BEH130 C18 column (75 μm × 100 mm, 1.7 μm) and porous shell Halo C18 column were tested. With an optimized LC gradient, 3 of the 4 peptide isomers were separated from each other, while RP*MFAIWK and RPMFA*IWK were co-eluted (data not shown). As shown in Supplementary **Figure S2.5**, a longer BEH130 C18 column (75 μm × 150 mm, 1.7 μm) is able to produce excellent peak shapes for all four oxidation isomers of the model peptide RPMFAIWK. Baseline separation within 60 min was achieved with linear gradient elution. It should be noted that gradient separation conditions are necessary for complete baseline separation of the four isomers on BEH130 C18 column. Different isocratic separation conditions were tested on the columns mentioned above, under the best of circumstance, only three peaks could be observed and peaks for RP*MFAIWK and RPMFA*IWK were co-eluted.

The four isomeric peptides analyzed were of synthetic origin, so that their individual assignment in the UPLC-MS/MS runs was straightforward. The retention times for the four isomers: RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK, are 32.8 min, 30.5 min, 29.5 min, and 33.4 min, respectively. The molar ratio of each isomer was determined by integration of the SIM peak. The peak area of each isomer correlated well with the defined amount of the isomer in solution when the data was calculated with methionine sulfoxide containing isomer (RPM*FAIWK) excluded (**Figure 2.4**). The amounts of methionine sulfoxide-containing

isomer are overestimated by peak area in all tested samples. This variation was not observed when the isomers were individually injected and eluted under a step-gradient to strong elution conditions using a mobile phase of 90 % acetonitrile (UPLC peak areas are used for quantification, data not shown). Under these step gradient conditions, no separation of the methionine amide diastereomers can occur, eliminating peak shape irregularities at the cost of any UPLC resolving power. These data, combined with the successful quantification by ETD fragmentation, suggests that the discrepancy between the UPLC peak areas and the theoretical abundance of each isomer is not due to differences in ionization efficiency between the isomers.

Based on the results above, UPLC-based quantification may be unreliable, even in cases where the oxidation isomers can be cleanly separated to baseline. This may be due to the fact that in our sample, the MetSO-containing peptide is the only peptide that is present as a mixture of two stereoisomers, whereas the other oxidation isomers are stereochemically pure. Analysis of the MetSO peak in Supplementary **Figure S2.5** indicates peak broadening, shouldering, or splitting, which may be due to partial separation of these stereoisomers. This peak asymmetry may be causing the overestimation of MetSO quantity by UPLC. Unfortunately, in HRPF and most oxidative stress situations, oxidative modification of peptides will not result in a stereochemically pure sample. Therefore, this peak asymmetry may occur rampantly in any optimized UPLC protocol, further complicating quantification by nanoUPLC analysis.

Relative Quantification of Mixtures of OxidizedIsomers Having Multiple Sites of Oxidationin HRPF Product of Protein Robo-1 Ig1-2 by ETD

To demonstrate the ability of ETD-based dissociation to quantify oxidized peptide isomers having more than one site of oxidation on a single peptide generated in real HRPF samples, we performed a HRPF study of the protein Robo-1 Ig1-2 and identified a series of oxidized products

with multiple oxidation sites on a single peptide sequence. An example is the identification and quantification of sites of oxidation of an identified peptide LMITYTRK by ETD.

Protein Robo-1 Ig1-2 was subjected to hydroxyl radical footprinting and Lyc-C digestion, and the peptide LMITYTRK produced by Lys-C digestion was sequenced based on the tandem mass spectra by ETD (Supplementary **Figure S2.6a**). A singly-oxidized product of this peptide LMITYTRK generated by HRPF was also identified and sequenced as LM*ITYTRK based on its ETD spectrum (Supplementary **Figure S2.6b**). The ETD spectrum of LM*ITYTRK, $[M + O + 3H]^{3+}$ contains unoxidized c1 ion, oxidized c2–c7 ions, and unoxidized z2–z3ions, which indicate that Met2 is the sole site of oxidation.

We also observed a doubly-oxidized product of the same peptide sequence LMITYTRK by comparing the full mass spectrum of unoxidized Robo-1 Ig1-2 sample with oxidized sample. The SIM chromatogram shows only one single peak corresponding to this doubly-oxidized product with optimized nanoUPLC separation on the same column used for the synthetic peptide isomers. However, the ETD spectrum shows unoxidized c1 and singly-oxidized c2 ions, and singly- and doubly-oxidized forms of c3–c7 product ions (**Figure 2.5**). The presence of only singly-oxidized c2 ion and singly-oxidized and doubly-oxidized c3–c7 product ions indicate that 100 % Met2 was singly-oxidized. In addition, the increasing percentages of these doubly-oxidized forms of c3–c6 product ions provide clear evidence that Ile3, Thr4, Tyr5, Thr6, and Lys8 are all partially oxidized (**Table 2.1**). Therefore, this single SIM chromatographic peak at m/z 353 ($[M + 2O + 3H]^{3+}$ species of the peptide LMITYTRK actually represents a mixture of five doubly-oxidized isomers differing by the site of second oxidation. Indeed, the percentages of these doubly-oxidized forms of c3–c8 product ions indicate that the relative molar ratio of the five doubly-oxidized isomers with

oxidation on Met2/Iso3, Met2/Thr4, Met2/Tyr5, Met2/Thr6, and Met2/Lys8 is approximately 15:32:20:19:14.

Table 2.1. Percentage (mean ± SD, n = 3) of Oxidized Product Ions Observed in the ETD Spectrum of the Doubly Oxidized Peptide LMITYTRK

Product Ion	m/z unoxidized	m/z singly oxidized	m/z doubly oxidized	Ratio of singly oxidized to unoxidized product ion	Percentage of doubly oxidized product ion
c1	131	147	163	0	0.0
c2	262	278	294	100	0.0
c3	375	391	407	100	15.1±2.5
c4	476	492	508	100	46.8±2.0
c5	639	655	671	100	66.7±3.5
c6	740	756	772	100	85.9±3.2
c7	896	912	928	100	89.6±6.3
c8	---	---	---	100	100

Another example in which the ETD spectrum has demonstrated the capability of simultaneously identifying and quantifying the oxidation sites of peptide isomers involves another HRPF product of protein Robo-1 Ig1-2: peptide IVEHPSDLIVSK. The ETD spectrum of the triply-charged oxidized peptide from an unseparated sample not only identifies three oxidation sites but also provides the degree of oxidation at each residue (**Figure 2.6** and **Table 2.2**). As shown in **Table 2.2**, the percentages of oxidized c1–c3 andc5 ions are consistent, indicating only the N-terminal amino acid is oxidized to a significant extent. Meanwhile, it is clearly shown that c6 ion is significantly oxidized, as the percentage increases substantially at this residue. The analysis of the percentage of the oxidized forms indicates that the degree of oxidation at Ile1, Ser6, and Lys12 is about 40 %, 20 %, 40 %, respectively. Examination of **Figure 2.6** indicates that ETD fragmentation efficiency of this peptide was not particularly robust, which has obvious consequences when examining the increased standard deviations of oxidation measurement shown in **Table 2.2**. However, even with the poor signal-to-noise, we were clearly able to identify and

quantify the major oxidation isomers of this peptide, which were impossible to resolve by reverse phase UPLC separation.

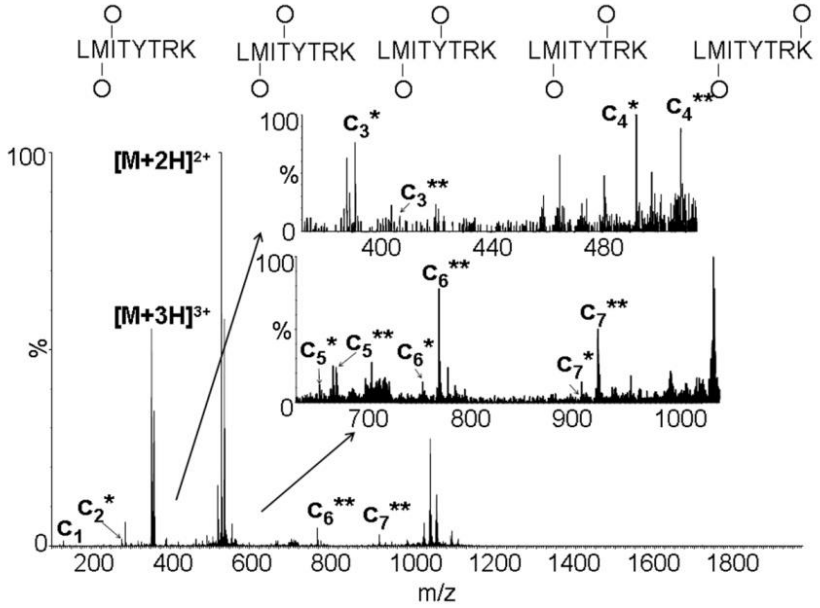

Figure 2.5. The ETD spectrum of doubly-oxidized peptide LMITYTRK $[M + 2O + 3H]^{3+}$. The single asterisks indicate the product ions that are singly oxidized. The double asterisks indicate the product ions that are doubly oxidized.

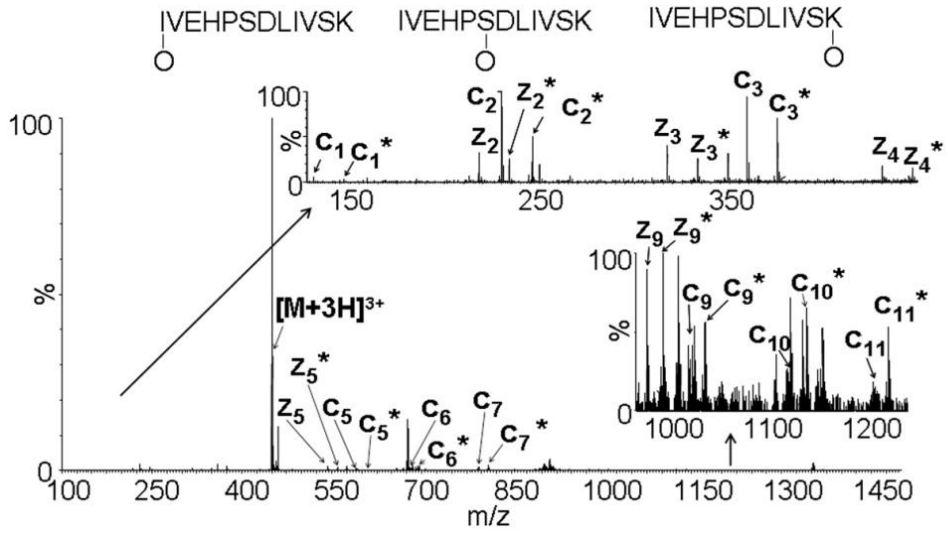

Figure 2.6. The ETD spectrum of singly oxidized peptide isomers IVEHPSDLIVSK $[M + O + 3H]^{3+}$. The asterisks indicate the product ions that are oxidized.

Table 2.2. Percentage (mean ± SD, n = 3) of Oxidized Product Ions Observed in the ETD Spectrum of the Singly Oxidized Peptide Isomers IVEHPSDLIVSK

Product ion	$m/z_{unoxidized}$	$m/z_{oxidized}$	Percentage oxidized
c1	131	147	40.3±8.0
c2	230	246	33.3±9.3
c3	359	375	42.9±3.5
c4	---	---	---
c5	593	609	32.7±8.6
c6	680	696	60.7±5.5
c7	795	811	61.1±12.2
c8	---	---	---
c9	1022	1038	72.6±5.6
c10	1121	1137	69.3±6.5
c11	1208	1224	73.6±15.3
c12	---	---	100
z2	218	234	42.9±14.6
z3	317	333	39.4±5.3
z4	430	446	46.3±3.4
z5	543	559	44.7±2.5
z9	980	996	53.7±7.0
z12	---	---	100

CONCLUSION

In this work, we have demonstrated the accuracy of quantification of mixtures of synthetic oxidation isomers based on ETD fragment ion intensities. The amount of oxidation occurring at given residues could be accurately determined from ETD spectra of peptides in both doubly- and triply-charged states. In contrast, CID spectra of doubly-charged precursor ions for such peptides rarely reflect the accurate oxidation extent, though CID spectra of the triply-charged ion allows accurate determination for such peptides. Based on the current limited data set (only one set of four synthetic peptide sequences was tested), we are unable to predict with certainty what charge state would result in quantitative CID fragmentation for a given peptide sequence, although initial

results from our model peptide lead to the hypothesis that each basic site in the peptide must be protonated for quantitative CID.

LC separation is one of major techniques currently used for relative quantification of mixtures of oxidation isomers [12, 17–19]. But carefully optimized LC conditions are needed to fully separate the oxidized isomers. In addition, even though these isomers could often (but not always) be separated successfully after intensive optimization of LC conditions, the relative quantification of the isomers based on individual chromatographic peak area shows significant variation for each individual isomer. The only isomer that gave aberrant readings in our model UPLC separation-based quantification was the methionine sulfoxide-containing peptide. This error cannot be explained by differences in ionization efficiency, as ionization efficiency differences would be revealed in any ESI-MS-based method, be it LCMS based or fragment ion intensity-based. The methionine sulfoxide peptide is the only model peptide that existed as a mixture of two structures: the two stereoisomers of the sulfoxide. The broader peak shape and shouldering of the oxidized methionine peak shown in Supplementary **Figure S2.5** (probably attributable to these diastereomers) compared with the other isomeric oxidation products may result in poor peak area calculations, leading to the errors in quantification. Similarly, when isomers were injected individually and eluted with a strong step-gradient to eliminate peak shape aberrations, accurate quantification of all four isomers could be achieved, although this method sacrifices the resolving power of the UPLC and is impractical for mixtures. Unfortunately, in HRPF experiments, these oxidation products almost always exist as multiple isomers within a single amino acid residue. In our experience, these HRPF-based mixtures often have UPLC peaks with apparent partial separation of these in-residue isomers, leading to peak broadening and poor peak shapes. If these peak shapes lead to errors in peak area calculations, LCMS- based quantification of these oxidation

isomers is likely to be rife with errors. Further work comparing ETD-based and UPLC-based quantification of complex HRPF-derived samples should be pursued to probe the extent of this potential issue.

The application of ETD in simultaneously identifying and quantifying the oxidation sites has also been demonstrated through analyzing a series of oxidized products with multiple oxidation sites on two peptide sequences in HRPF product of protein Robo-1 Ig1-2. As discussed above, it is quite challenging to separate these isomers by LC because of their similar properties and high hydrophilic character, even using the nanoUPLC system, which has demonstrated the capability to separate four synthetic peptide isomers with adjacent oxidation. In such a situation, ETD-based tandem mass spectrometry provides a fast, convenient, and accurate method that allows simultaneous identification and relative quantification of the oxidation of each individual amino acid in proteins. Its advantages in the ability to obtain structural details on modified peptides present in complex mixtures, as well as its accuracy for quantitative analysis, make it a powerful tool in the study of oxidative stress, commercial protein oxidative modification, as well as for improving spatial resolution for HRPF studies. The latter capability should further advance the HRPF power for the determination of protein structure and protein–ligand interactions. While the well-known limitations of ETD (the need for higher charge states, inability to cleave N-terminal to proline) may limit its applicability in some cases, in most cases it is an important tool in the quantification of peptide oxidation isomers.

ACKNOWLEDGEMENTS

The authors acknowledge support of this research by the National Institute of General Medical Sciences (1R01GM096049-01A1) and in part by the National Institute of General Medical Sciences funded “Research Resource for Integrated Glycotechnology”(P41 GM103390)

from the National Institutes of Health. The authors thank Professor Kelley Moremen for the expression and purification of the Robo-1 Ig1-2 protein.

SUPPLEMENTARY INFORMATION

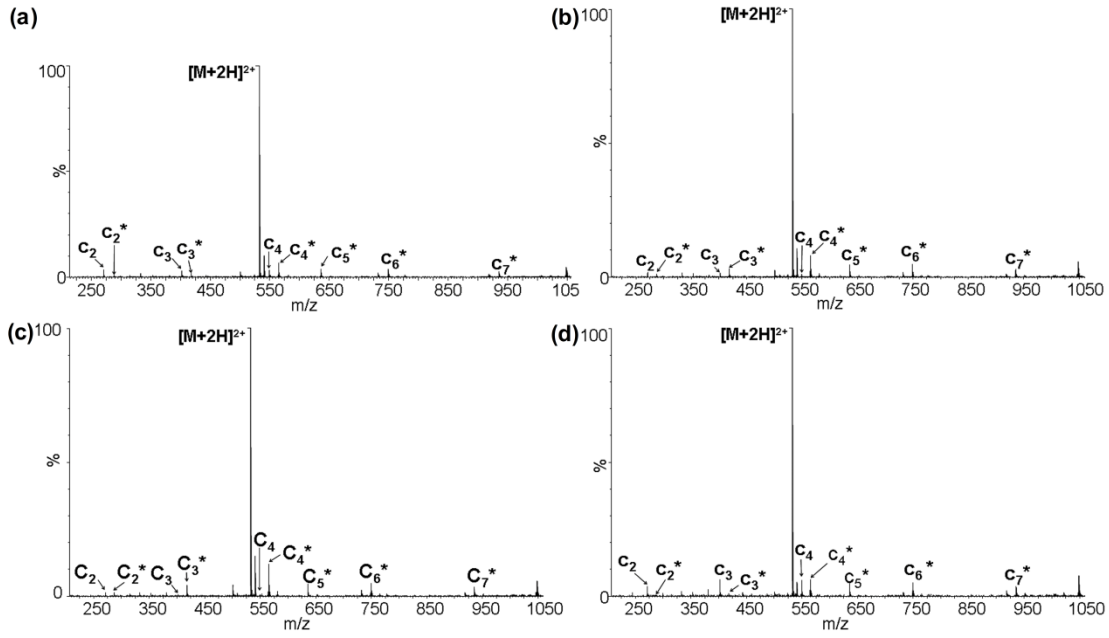

Figure S2.1. Representative ETD spectrum of doubly-charged ion of mixture of four oxidation isomers of the model peptide RPMFAIWK. The mixtures contain RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at varying ratios: **(a)**1:2:3:4; **(b)** 4:3:2:1; **(c)** 9:9:1:1; **(d)** 1:1:9:9. The asterisks indicate the product ions that are oxidized.

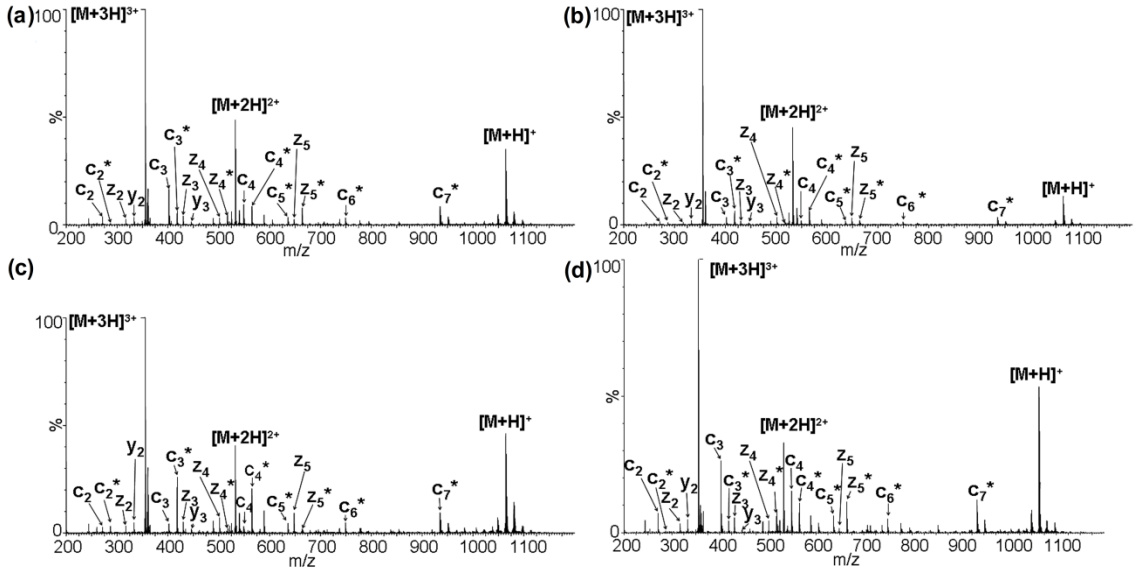

Figure S2.2. Representative ETD spectrum of triply-charged ion of mixture of four oxidation isomers of the model peptide RPMFAIWK. The mixtures contain RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at varying ratios: **(a)**1:2:3:4; **(b)** 4:3:2:1; **(c)** 9:9:1:1; **(d)** 1:1:9:9. The asterisks indicate the product ions that are oxidized.

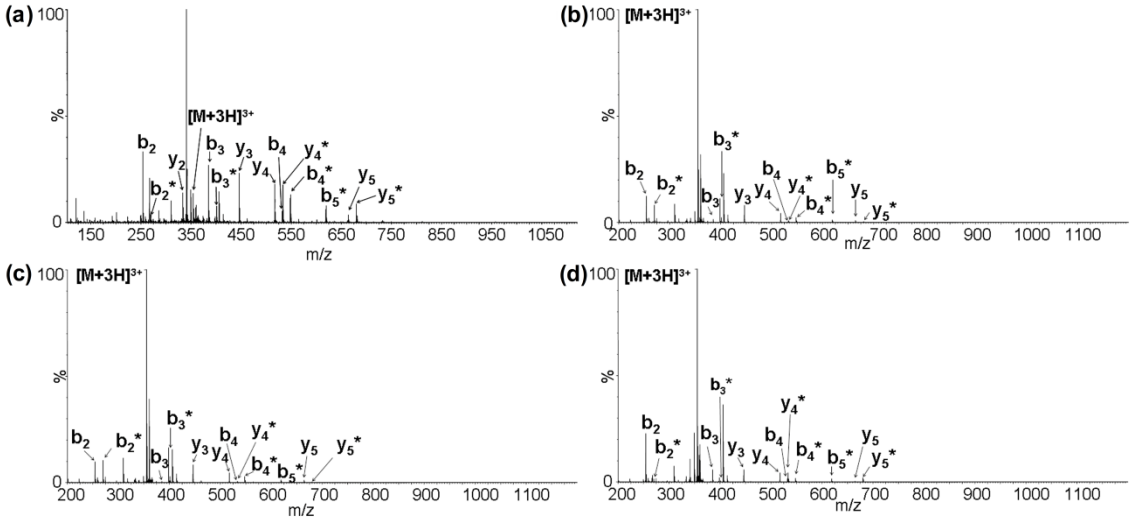

Figure S2.3. Representative CID spectrum of triply-charged ion of mixture of four oxidation isomers of the model peptide RPMFAIWK. The mixtures contain RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at varying ratios: **(a)** 1:2:3:4; **(b)** 4:3:2:1; **(c)** 9:9:1:1; **(d)** 1:1:9:9. The asterisks indicate the product ions that are oxidized.

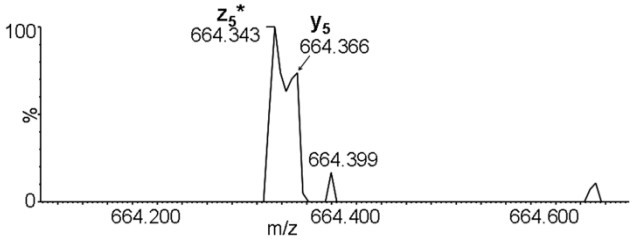

Figure S2.4. Identification of oxidized z5 ion and y5 ion on high resolution ETD spectrum of triply-charged ion of mixture of four oxidation isomers of the model peptide RPMFAIWK.

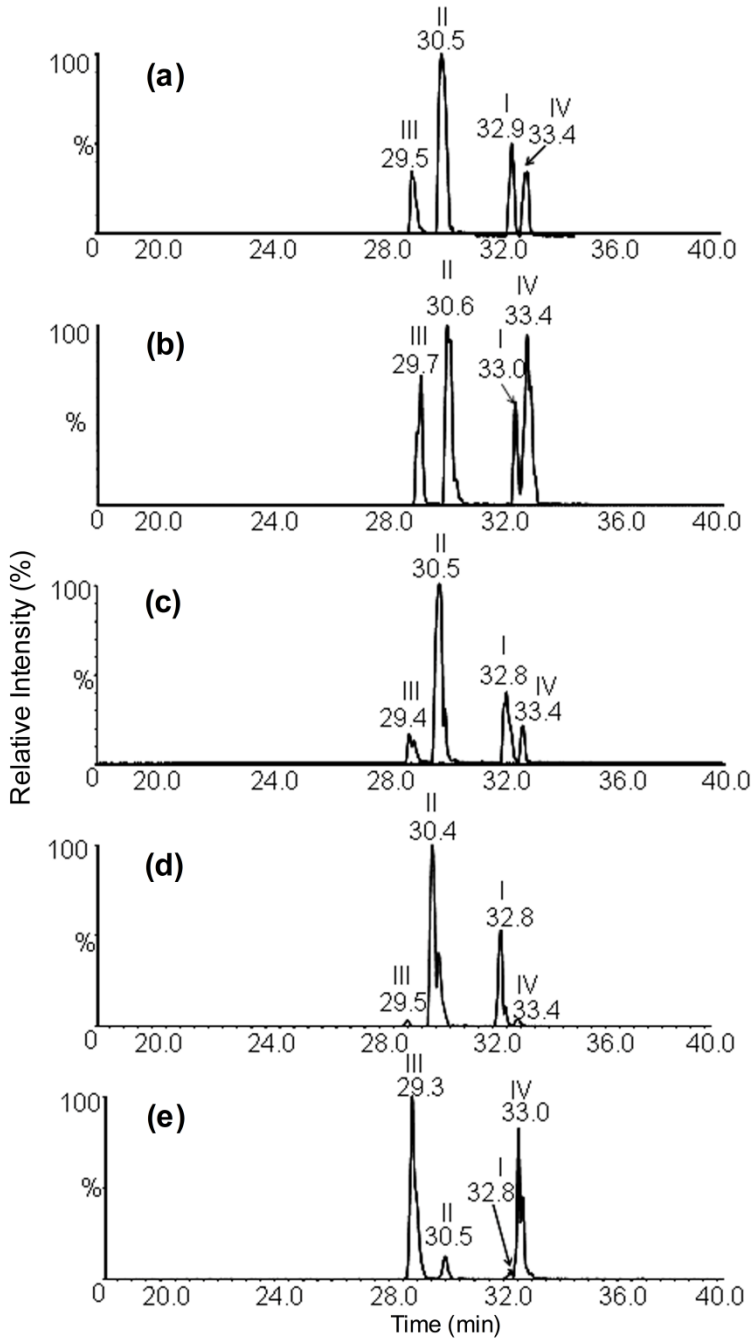

Figure S2.5. Representative selected ion chromatogram of UPLC separation of four oxidation isomers of the model peptide RPMFAIWK. The mixtures contain RP*MFAIWK, RPM*FAIWK, RPMF*AIWK, and RPMFA*IWK at five varying molar ratios of I:II:III:IV. **(a)** 1:1:1:1;**(b)** 1:2:3:4;**(c)** 4:3:2:1;**(d)** 9:9:1:1;**(e)** 1:1:9:9

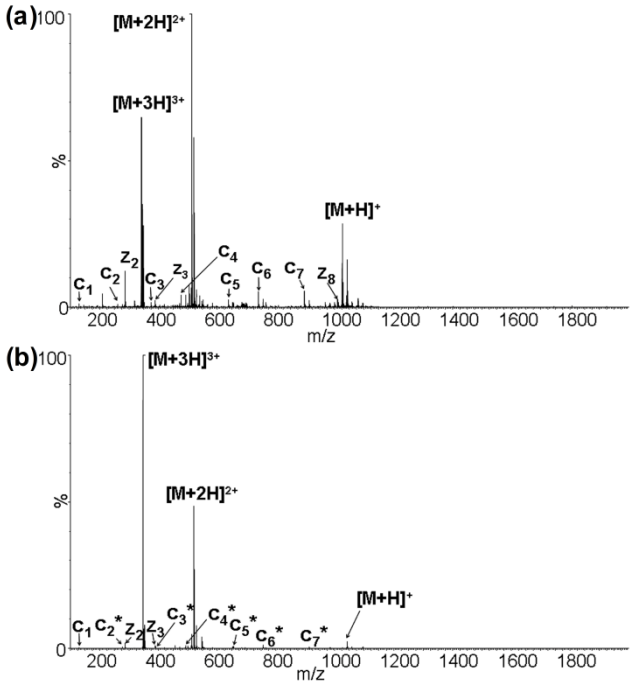

Figure S2.6. The ETD spectrum of unoxidized peptide LMITYTRK $[M+3H]^{3+}$ **(a)** and singly oxidized peptide LM*ITYTRK $[M+O+3H]^{3+}$ **(b)**. The asterisks indicate the product ions that are oxidized.

Table S2.1. Percentage (Mean ± SD, n = 3) of oxidized product ions observed in the ETD spectrum of the doubly-charged ion and triply-charged ion of four oxidation isomers of the model peptide RPMFAIWK

Product ion	$m/z_{unoxidized}$	$m/z_{oxidized}$	Theoretical percentage oxidized	Percentage oxidized based on fragmentation of doubly-charged ion	Percentage oxidized based on fragmentation of triply-charged ion
c2	271	287	25	29±8.3	28±6.2
c3	402	418	50	53±10	55±4.9
c4	549	565	75	77±3.4	65±4.9
c5	620	636	100	100	100
c2	271	287	10	13±3.5	16±2.2
c3	402	418	30	37±3.2	32±67
c4	549	565	60	62±4.6	51±6.2
c5	620	636	100	100	100
c2	271	287	40	42±8.2	50±4.9
c3	402	418	70	69±5.2	70±4.9
c4	549	565	90	89±2.5	79±2.1
c5	620	636	100	100	100
c2	271	287	5	5.4±0.8	7.3±1.6
c3	402	418	10	15±1.9	15±1.7
c4	549	565	55	52±6.4	45±3.9
c5	620	636	100	100	100
c2	271	287	45	56±5.2	49±6.6
c3	402	418	90	84±5.1	90±3.2
c4	549	565	95	94±1.5	81±5.5
c5	620	636	100	100	100
z5	648	664	50		47±7.6
z4	501	517	25		36±3.0
z5	648	664	70		74±5.0
z4	501	517	40		50±6.1
z5	648	664	30		38±0.8
z4	501	517	10		23±3.2
z5	648	664	10		17±1.9
z4	501	517	5		8±1.2
z5	648	664	90		88±1.7
z4	501	517	45		54±6.5

REFERENCES

1. Shacter, E., *Quantification and significance of protein oxidation in biological samples.* Drug Metab Rev, 2000. **32**(3-4): p. 307-26.

2. Stadtman, E.R. and B.S. Berlett, *Reactive oxygen-mediated protein oxidation in aging and disease.* Drug Metab Rev, 1998. **30**(2): p. 225-43.

3. Marondedze, C., et al., *Structural and functional characteristics of cGMP-dependent methionine oxidation in Arabidopsis thaliana proteins.* Cell Commun Signal, 2013. **11**(1): p. 1.

4. Kuhns, L.G., et al., *Role of Helicobacter pylori methionine sulfoxide reductase in urease maturation.* Biochem J, 2013. **450**(1): p. 141-8.

5. Chen, H.J. and Y.C. Chen, *Reactive nitrogen oxide species-induced post-translational modifications in human hemoglobin and the association with cigarette smoking.* Anal Chem, 2012. **84**(18): p. 7881-90.

6. Gitlin, G., et al., *Isolation and characterization of a monomethioninesulfoxide variant of interferon alpha-2b.* Pharm Res, 1996. **13**(5): p. 762-9.

7. Chen, W., et al., *Modification of cysteine residues in vitro and in vivo affects the immunogenicity and antigenicity of major histocompatibility complex class I-restricted viral determinants.* J Exp Med, 1999. **189**(11): p. 1757-64.

8. Berti, P.J., et al., *Affinity purification and elimination of methionine oxidation in recombinant human cystatin C.* Protein Expr Purif, 1997. **11**(1): p. 111-8.

9. Hsu, Y.R., et al., *In vitro methionine oxidation of Escherichia coli-derived human stem cell factor: effects on the molecular structure, biological activity, and dimerization.* Protein Sci, 1996. **5**(6): p. 1165-73.

10. Xu, G. and M.R. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics.* Chem Rev, 2007. **107**(8): p. 3514-43.

11. Wang, L. and M.R. Chance, *Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting.* Anal Chem, 2011. **83**(19): p. 7234-41.

12. Charvatova, O., et al., *Quantifying protein interface footprinting by hydroxyl radical oxidation and molecular dynamics simulation: application to galectin-1.* J Am Soc Mass Spectrom, 2008. **19**(11): p. 1692-705.

13. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochem Biophys Res Commun, 2001. **287**(3): p. 614-21.

14. Kiselar, J.G. and M.R. Chance, *Future directions of structural mass spectrometry using hydroxyl radical footprinting.* J Mass Spectrom, 2010. **45**(12): p. 1373-82.

15. Zhang, H., et al., *Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system.* Anal Chem, 2011. **83**(1): p. 311-8.

16. Jones, L.M., et al., *Fast photochemical oxidation of proteins for epitope mapping.* Anal Chem, 2011. **83**(20): p. 7657-61.

17. Gau, B., et al.,*Mass spectrometry-based protein footprinting characterizes the structures of oligomeric apolipoprotein E2, E3, and E4.* Biochemistry, 2011.**50**: p. 8117–8126.

18. Chen, J., et al.,*Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level.* J. Am. Chem. Soc., 2012. **134**: 18724–18731.

19. Gau, B.C., et al., *Fast photochemical oxidation of protein footprints faster than protein unfolding.* Anal Chem, 2009. **81**(16): p. 6563-71.

20. Smedley, J.G., et al.,*Probing the pH-dependent prepore to pore transition of Bacillus anthracis protective antigen with differential oxidative protein footprinting.* Biochemistry, 2008. **47**: p.10694–10704

21. Maleknia, S.D., M. Brenowitz, and M.R. Chance, *Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry.* Analytical Chemistry, 1999. **71**(18): p. 3965-3973.

22. Goldsmith, S.C., et al., *Synchrotron protein footprinting: A technique to investigate protein-protein interactions.* Journal of Biomolecular Structure & Dynamics, 2001. **19**(3): p. 405-418.

23. Sharp, J.S., et al., *Measurement of multisite oxidation kinetics reveals an active site conformational change in Spo0F as a result of protein oxidation.* Biochemistry, 2006. **45**(20): p. 6260-6266.

24. Sharp, J.S. and K.B. Tomer, *Analysis of the oxidative damage-induced conformational changes of apo- and holocalmodulin by dose-dependent protein oxidative surface mapping.* Biophys J, 2007. **92**(5): p. 1682-92.

25. Hambly, D.M. and M.L. Gross, *Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale.* Journal of the American Society for Mass Spectrometry, 2005. **16**(12): p. 2057-2063.

26. Gau, B.C., et al., *Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding.* Analytical Chemistry, 2009. **81**(16): p. 6563-6571.

27. Watson, C. and J.S. Sharp, *Conformational Analysis of Therapeutic Proteins by Hydroxyl Radical Protein Footprinting.* Aaps Journal, 2012. **14**(2): p. 206-217.

28. Srikanth, R., et al., *Improved sequencing of oxidized cysteine and methionine containing peptides using electron transfer dissociation.* Journal of the American Society for Mass Spectrometry, 2007. **18**(8): p. 1499-1506.

29. Srikanth, R., J. Wilson, and R.W. Vachet, *Correct identification of oxidized histidine residues using electron-transfer dissociation.* Journal of Mass Spectrometry, 2009. **44**(5): p. 755-762.

30. Jumper, C.C., et al., *High-Resolution Mapping of Carbene-Based Protein Footprints.* Analytical Chemistry, 2012. **84**(10): p. 4411-4418.

31. Wiesner, J., T. Premsler, and A. Sickmann, *Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications.* Proteomics, 2008. **8**(21): p. 4466-4483.

32. Zhou, Y.P., J. Dong, and R.W. Vachet, *Electron Transfer Dissociation of Modified Peptides and Proteins.* Current Pharmaceutical Biotechnology, 2011. **12**(10): p. 1558-1567.

33. Lagerwerf, F.M., et al., *Identification of oxidized methionine in peptides.* Rapid Communications in Mass Spectrometry, 1996. **10**(15): p. 1905-1910.

34. Xu, G.Z., K. Takamoto, and M.R. Chance, *Radiolytic modification of basic amino acid residues in peptides: Probes for examining protein-protein interactions.* Analytical Chemistry, 2003. **75**(24): p. 6995-7007.

35. Altschul, S.F., et al., *Basic Local Alignment Search Tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-410.

CHAPTER 3

SUPERCHARGING BY M-NBA IMPROVES ETD-BASED QUANTIFICATION OF HYDROXYL RADICAL PROTIEN FOORPRINTING[2]

[2] Li, X.; Li, Z. (co-first author); Xie, B.; Sharp, J. S. *Journal of the American Society for MassSpectrometry*, **2015**:p. 1-4. Reprinted here with permission of publisher.

ABSTRACT

Hydroxyl radical protein footprinting (HRPF) is an MS-based technique for analyzing protein structure based on measuring the oxidation of amino acid side chains by hydroxyl radicals diffusing in solution. Spatial resolution of HRPF is limited by the smallest portion of the protein for which oxidation amounts can be accurately quantitated. Previous work has shown electron transfer dissociation (ETD) to be the most reliable method for quantifying the amount of oxidation of each amino acid side chain in a mixture of peptide oxidation isomers, but efficient ETD requires high peptide charge states, which limits its applicability for HRPF. Supercharging reagents have been used to enhance peptide charge state for ETD analysis, but previous work have shown supercharging reagents to enhance charge state differently for different peptides sequences; it is currently unknown if different oxidation isomers will experience different charge enhancement effects. Here,we report the effect of m-nitrobenzyl alcohol (m-NBA) on the ETD-based quantification of peptide oxidation. The addition of m-NBA to both a defined mixture of synthetic isomeric oxidized peptides andRobo1 protein subjected to HRPF increased the abundance of higher charge state ions, improving our ability to perform efficient ETD of the mixture. No differences in the reported quantitation by ETD were noted in the presence of absence of m-NBA, indicating that all oxidation isomers were charge-enhanced to a similar extent. These results indicate the utility ofm-NBA for residue-level quantification of peptideoxidationin HRPF and other applications.

INTRODUCTION

Protein tertiary and quaternary structure are fundamental todetermining mechanisms of protein function. Understandingthe structure and function of proteins and their interactions inmacromolecular assemblies is critical to achieve an overallunderstanding of biological systems. Hydroxyl radical proteinfootprinting (HRPF) is a relatively recent covalent labelingapproach coupled with mass spectrometry, and has been developedover the last decade to a powerful method for analyzingprotein structure and dynamics. HRPF has several advantagesthat recommend it for the analysis of protein structure, particularlyfor difficult systems such as large, heterogeneous proteincomplexes, membrane proteins, and flexible protein systems [1–3]. HRPF takes advantage of the fact that the rate of oxidation ofeach amino acid varies directly with the solvent accessibility ofthat amino acid [4, 5]. This relationship allows for changes inprotein structure to be monitored by monitoring the apparent rateof oxidation of a particular amino acid side chain [6, 7].

Initial uses of HRPF were limited in spatial resolution to the size of a proteolytic peptide, as the amount of oxidation of any individual amino acid within the peptide could not be accurately quantified by CID [8-10]. As sub-microsecond HRPF technologies such as Fast Photochemical Oxidation of Proteins (FPOP) [3] and pulsed electron beam radiolysis [11] began to allow for heavier oxidation of proteins, the need to quantitate isomeric peptide oxidation products became even more pronounced. Reports from Gross and coworkers have used UPLC to separate isomeric peptide products and quantify based on peak area in a selected ion chromatogram [12]; however, the only attempt to use UPLC separation coupled with peak area quantification using known oxidized peptide standards found this method to be inaccurate in some cases, while electron transfer dissociation (ETD) provided an accurate and reliable quantification of oxidation at the residue level for isomeric mixtures [13].

Although ETD gave reliable results for residue-level quantification of oxidation, ETD is widely known for having poor fragmentation efficiency for doubly-charged peptides, which are commonly observed for tryptic digestion products. This poor fragmentation efficiency limits both the sensitivity of ETD-based quantification as well as the spatial resolution of HRPF information, as cleavage of each peptide bond in the peptide is required for true residue-level resolution. One approach to improve ETD fragmentation is based on addition of supercharging reagent into electrospray solution to increase the charge state of tryptic peptide ions [14, 15]. As the ability to quantify oxidation by ETD depends upon the ability of m-NBA to equally alter the charge state of each oxidation isomer of a given peptide sequence, as well as the ETD fragmentation process remaining transparent to the site of oxidation in the presence of m-NBA, the applicability of supercharging to ETD-based HRPF remains in question. In this study, we test the effect of the charge-enhancing reagent m-NBA on the ability to accurately quantify the amount of oxidation on each amino acid by ETD, as well as the ability of m-NBA to positively impact actual HRPF studies of an oxidized protein.

EXPERIMENTAL

Materials

m-Nitrobenzyl alcohol (m-NBA), catalase, formic acid and L-glutamine was purchased from the Sigma-Aldrich Corporation (St. Louis, MO).Hydrogen peroxide (30%) was obtainedfrom J.T. Baker (Phillipsburg, NJ). Dithiothreitol (DTT) andHPLC-grade acetonitrile (ACN) were purchased from Fisher Scientific(Fair Lawn, NJ). Methionine amide was purchased from Bachem (Torrace, CA).Sequencing-grade modified trypsinwas purchased from Promega Corporation (Madison, WI, USA).All reagents were used as provided. Purified water (18 MΩ) was obtained from an in-house Milli-Q Synthesis system (Millipore, Billerica, MA).Peptide oxidation analog

standards of the peptides RPMFAIWK and MLLPSGSLFFLR were synthesized and purified as previously described [13]. Robo1 Ig1-2 protein was a gift from Prof. Kelley Moremen (University of Georgia) and expressed and purified.

Synthetic peptide oxidation analog standards

The peptide RPMFAIWK, and its oxidized isomers: RP*MFAIWK (Pro2→hydroxyproline), RPMF*AIWK (Phe4→tyrosine), and RPMFA*IWK (Ala5→serine) as well as the peptide MLLPSGSLFFLR (Robo1 64-75) and its oxidized isomers: MLLP*SGSLFFLR (Pro2→hydroxyproline) and MLLRSGSLF*FLR (Phe9→tyrosine) were synthesized by GenScript USA Inc. (Piscataway, NJ), and their purity and quantity were validated in-house by HPLC analysis. Peptides RPM*FAIWK and M*LLRSGSLFFLR (Met→methionine sulfoxide) were prepared in house by adding 90 µM unmodified peptideto a solution of hydrogen peroxide (3%) and incubated in the dark for 1 h, followed by C18 reverse phase HPLC purification of the stable methionine sulfoxide-containing product.

Robo 1Ig1-2Expression and Purification

The protein sequence of the translation of the human ROBO1 gene was analyzed for domain boundaries using the UniProt database and a truncated protein sequence comprised of Ig domains 1 and 2 (plus flanking regions) was chosen for gene synthesis. Optimized codon utilization for human protein expression was employed in the synthesis of the corresponding coding region and the resulting DNA fragment was cloned into our mammalian expression vector (pGEn2) using restriction digestion and ligation into corresponding restriction sites in the vector. Large scale DNA preparations were prepared and transiently transfected into HEK293S GnTI- suspension culture cells and recombinant protein was harvested after 6 days of transfection. The recombinant protein was purified from the culture supernatant using Ni2+-NTA chromatography

and concentrated to ~1 mg/mL. The resulting protein preparation was digested with recombinant TEV-GFP and EndoF1-GFP and then further purified by Ni2+-NTA chromatography and gel filtration.

Robo1 FPOP

A 248 nm EX100 KrF excimer laser (GAM Laser Inc., Orlando FL, USA) operating at 50 mJ per pulse output was used to photolyze H2O2. Robo1 Ig1-2 was reconstituted in an ammonium bicarbonate buffer (50 mM, pH=7.5) at a concentration of 50 μM. Each 20 μL reaction buffer (50 mM ammonium bicarbonate) was prepared with a final concentration of 20 μM Robo1 Ig1-2, 3% hydrogen peroxide (freshly prepared and added immediately prior to irradiation) and 20 mM glutamine. The solution was then loaded into a 100 μL syringe and introduced via a syringe pump coupled to the 100 μm i.d. fused silica tubing. The flow rate was set to 12.19 μL/min, with the laser pulse repetition rate set to illuminate each volume of sample with a single focused laser pulse as it passed through the laser path, with 10% of the total volume remaining unirradiated to correct for diffusion and laminar flow effects. The capillary outflow was collected in a microcentrifuge tube containing catalase (0.5 mg/mL) and methionine amide (0.5 mg/mL) with ammonium bicarbonate buffer (50 mM, pH 7.5) to remove any remaining H_2O_2, as well as scavenge secondary oxidants. Following irradiation, samples were reduced with 5mM DTT and incubated at 65°C for 30 min. After cooling to room temperature, a 1:20 weight ratio of trypsin (enzyme to protein, w to w ratio of trypsin) was added to the protein sample and incubated at 37°C overnight while rotating.

LC-MS/MS and data analysis

The working stock solutions of the four synthetic peptides were prepared in 50% ACN or in 50% ACN containing 0.1% m-NBA. The working solutions with or without m-NBA were mixed

in 1:1:1:1 volume ratios, respectively. The final molar concentration for each peptide in the mixture is 2 μM.

Mass spectral analyses of the RPMFAIWK peptide series and the Robo1 samples were performed on the Waters Synapt G2 Q-TOF mass spectrometer configured with both ETD and CID capabilities, and with a nano-ESI source that was coupled to either a nanoACQUITY (Waters) UPLC system, or operated in infusion mode. For direct injection experiments, the sample was delivered at 1μL/min using a syringe pump. For nanoUPLC-MS/MS analysis, the samples were separated on a Symmetry C18 180 μm x 20 mm, 5μm trapping column and a BEH130 C18 75 μm x 150 mm, 1.7μmanalytical column. Mobile phase A consistedof 0.1% formic acid (FA) in water/ acetonitrile (1:99, v/v). Mobile phase B consistedof 0.1% FA in acetonitrile.In m-NBA involved experiments, appropriate amount of m-NBA was added to mobile phases A and B to obtain 0.1% m-NBA (v/v) in both mobile phases. For synthetic peptide analysis, the gradient elution was performed in 60% mobile phase B over 15 min. For Robo 1 digestion sample analysis, the gradient elution was performed from 2–40% mobile phase B over 50 min at a flow rate of 0.4μL/min, and then increased to 80% mobile phase B for 5 min followed by a 5 min re-equilibration step.Nitrosobenzenewas used as ETD reagent. Trap wave height was set at 0.2 V to allow the ETD reagent to mix with the sample and fragment. Supplemental transfer CE of around 5V was used to aid the ETD of doubly-charged ions. For CID, the collision energy was set at 25 V for doubly-charged ion and 15 V for triply-charged ion, respectively. All mass spectra were acquired in positive ion mode, with a spray voltage of 3 kV and source temperature 100ºC.

Mass spectral analyses of the MLLPSGSLFFLR peptide series were carried out using an Orbitrap Elite mass spectrometer with ETD coupled to a nano-ESI source operating in direct infusion mode (Thermo Scientific) with a spray voltage of 2.5 kV and a source temperature of

275ºC. For CID analyses, the CID collision voltage was set at 25V. For ETD analysis, the ETD reaction time was set to 100 ms, and the ETD reagent ion used was fluoranthene.

The fragment ion intensities from ETD and/or CID are used for the calculation of oxidation rate at specific residue site using a similar approach reported previously [13, 18]. The actual fractional oxidation of a given sequence ion is defined as the ratio between the oxidized sequence ion intensity to the sum of the intensity of the corresponding oxidized and unoxidized sequence ion. This is shown in Equation 1:

$$f(c_i)_{actual} = \frac{I(c_i)_{oxidized}}{I(c_i)_{oxidized} + I(c_i)_{unoxidized}} (1)$$

where f $(ci)_{actual}$ denotes the fractional oxidation of c-ion i. I(ci) denotes the intensity of the c ion i, whether the oxidized and unoxidized form. The relative oxidation rate for a specific residue i is calculated as the difference between the fractional oxidation of adjacent residues. This is shown in Equation 2:

$$\text{oxidation rate/residue}_i = f(c_i)_{actual} - f(c_{i-1})_{actual} (2)$$

RESULTS AND DISCUSSION

Effect of 0.1% m-NBA on relative quantitation of synthetic peptide isomers by ETD and CID

The effect of different concentrations of m-NBA on the charge states of the oxidized peptide RPMFA*IWK (where the alanine is oxidized) is shown in **Table S3.1**. The absolute intensity of the triply-charged ion stopped increasing appreciably above a concentration of 0.1% m-NBA, which was used in the following studies. A representative ESI mass spectra of the peptide RPMFA*IWK with 0% and 0.1% m-NBA is shown in **Figure S3.1**, where it can be seen that 0.1% m-NBA did not appreciably add to the background of the spectrum. Mobile phase containing 0.1% m-NBA was not found to have a negative effect on either the chromatographic peak ion intensity or the peak shape of peptide RPMFA*IWK in LC-MS (data not shown).

In order for m-NBA to be useful for ETD-based high resolution HRPF, it is necessary that both the supercharging effect and the subsequent ETD fragmentation of the peptide be uninfluenced by the position of the site of oxidation in a mixture of isomers. Therefore, the effect of 0.1% m-NBA on the relative quantification of oxidation isomers was evaluated by direct infusion of a mixture of peptide oxidation isomers of either the peptide RPMFAIWK or MLLPSGSLFFLR, the Robo1 tryptic peptide 64-75, present at anequimolar ratio in 50% ACN solutions containing 0% and 0.1% m-NBA. The amount of oxidation at each specific residue site was determined based on the CID and ETD fragment ion intensity at different charge states and calculated by Equations 1 and 2. The measured average oxidation percentages from triplicate samples based on product ions from ETD are plotted against the theoretical percentages oxidized in the mixture (**Figure 3.1**). The percentage of oxidation in the mixture calculated based on abundance of the c-type and z-type product ion fragmented by ETD with and without m-NBA correlates very well to the theoretical values for both peptide oxidation isomer series, indicating m-NBA has no effects on ETD-based quantification of oxidation isomers with multiple oxidation sites at adjacentsites. For comparison, the same analysis for CID ion from thesesamples is shown in **Figure S3.2**. As reported previously [13], the precursor charge state and oxidized side chain identity plays amajor role in CID quantification, with CID-based quantification remaining unreliable. UPLC analysis was previouslyshown to be unable to quantitate the RPMFAIWK methionine oxidation isomer [13], and was unable to separate the oxidized methionine and oxidized proline isomers in the MLLPSGSLFFLR peptide after gradient optimization, preventing quantification by UPLC (data not shown)

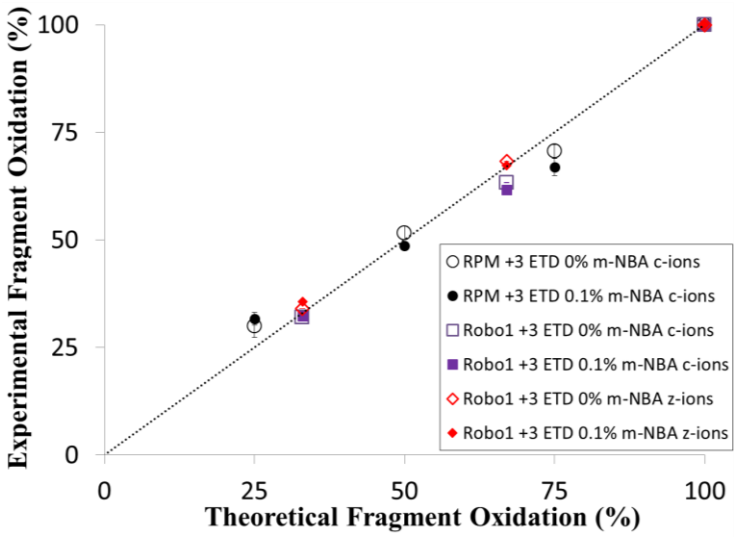

Figure 3.1. Comparison of theoretical and experimental quantification of equimolar mixtures of four oxidation isomers of the peptide RPMFAIWK based and three oxidation isomers of the peptide MLLPSGSLFFLR on product ion intensities by ETD of triply-charged ions in the presence of 0% or 0.1% *m*-NBA. The line represents the ideal 1:1 relationship between theoretical and measured oxidation. Error bars represent one standard deviation from a triplicate dataset.

Application of m-NBA in protein structural characterization by HRPF

The effect of m-NBA on ETD-based analysis of Robo1 subjected to HRPF was also studied. As shown in **Table 3.1**, addition of 0.1% m-NBA in mobile phase increases the relative abundance of the +3 and higher charge states for all unoxidized peptides measured (64-75 was only detected in the oxidized state after HRPF). m-NBA was found to increase the abundance of the higher charge states not only for the unoxidized peptides, but also for the oxidized products. The practical effect of the increased charge states of the oxidized peptides in the presence of 0.1% m-NBA in high resolution HRPF can be easily observed by comparing the ETD MS/MS spectrum of the highest charge state precursor detected for the oxidation product of peptide 194-205, shown

in Supplementary Information **Figure S3.3**. In the absence of m-NBA, no +3 charge state precursor for the oxidized peptide was detected; the resulting ETD fragmentation of the +2 charge precursor resulted in poor ETD-based sequence coverage and lower signal-to-noise ratios. Upon addition of m-NBA, a sufficient abundance of the +3 charge state precursor ion could be detected for ETD fragmentation, and the resulting spectrum yielded many more product ions and superior signal-to-noise ratios.

Table 3.1. Charge State Distribution of peptides from Robo 1 trypsin digestion sample in mobile phaseswith or without *m*-NBA

	0% *m*-NBA					0.1% *m*-NBA				
Peptide	**Charge-State Distribution**				**AV**	**Charge-State Distribution**				**AV**
	1+	2+	3+	4+		1+	2+	3+	4+	
14-25	0.00	96.53	3.47	0.00	2.03	0.17	22.52	77.31	0.00	2.77
35-47	0.00	52.76	47.24	0.00	2.47	0.00	6.05	92.99	0.96	2.95
64-75 (M+O)	0.00	100.00	0.00	0.00	2.00	0.00	95.15	4.85	0.00	2.05
140-149	0.00	95.13	4.87	0.00	2.05	0.00	52.44	47.56	0.00	2.48
151-161	0.00	93.23	6.77	0.00	2.07	0.00	47.38	52.62	0.00	2.53
194-205	12.00	88.00	0.00	0.00	1.88	0.00	82.33	17.67	0.00	2.18

To ensure that the measured amount of oxidation observed from the sample analyzed in the presence of m-NBA yielded identical quantification results as the sample analyzed without m-NBA, we compared the +3 charge states of the +16 oxidation products of other peptides from Robo1. A representative MS/MS spectra of oxidized peptide 140-149 are shown in Supplementary Information, **Figure S3.4**. An obvious increase of the intensity of the triply-charged ion of this oxidized peptide was found in 0.1% m-NBA sample compared to 0% m-NBA (data not shown). In both the sample with 0% m-NBA and with 0.1% m-NBA, the ETD spectra of the oxidized peptide 140-GHPEPTISWK-149, $[M+O+3H]^{3+}$ reveals no oxidation of the c3-c7 ions, and both oxidized and unoxidized c8 and c9 product ions, indicating residue S, W and K were oxidized in this mixture by HRPF. Furthermore, the quantification of oxidation extent based on ETD

fragmentation intensity clearly indicates that residue S, W and K were oxidized 14%, 59%, 27% respectively in the sample with 0% m-NBA. In the sample with 0.1% m-NBA, 18% of S, 52% of W and 30% of K were oxidized, indicating that 0.1% m-NBA has no substantial effect on quantitating site-specific oxidation in HRPF samples. Similar consistency in the measured oxidation amounts were found for other oxidized peptides of Robo1 in the presence and absence of 0.1% m-NBA (data not shown).

CONCLUSION

In this work, we have demonstrated the ability of m-NBA in increasing charge state distribution and thus improve sequence coverage of ETD spectra without affecting the ETD-based quantification of site-specific oxidation. Both synthetic peptide mixtures and actual tryptic peptides from an HRPF experiment of Robo1 gave a robust increase in the abundance of higher charge states without negatively impacting the ETD-based quantification, indicating the oxidation isomers all have their charge states affected similarly by m-NBA. These results indicate that the use of m-NBA is a workable method for increasing sequence coverage and spatial resolution in HRPF quantification as well as quantification of protein oxidation in general by ETD-based LC-MS/MS.

ACKNOWLEDGMENTS

This research is supported by the National Institute of General Medical Sciences-funded "Research Resource for Integrated Glycotechnology" (P41 GM103390), and in part by the National Institute of General Medical Sciences (1R01GM096049-01A1) from the National Institutes of Health. The authors would like to thank Prof. Kelley Moremen for the expression and purification of the Robo-1 Ig1-2 protein.

SUPPLEMENTAL INFORMATION

Table S3.1. Effect of different concentrations of *m*-NBA on the charge state of synthetic RPMFA*IWK

*m*NBA%	Intensity of doubly charged ion (532.79)	Intensity of triply charged ion (355.5)	Ratio of $[M+3H]^{3+}/[M+2H]^{2+}$
0.00	5.83E+05	7.19E+03	0.012
0.01	3.59E+05	3.38E+04	0.094
0.10	2.84E+05	4.30E+04	0.151
0.50	1.96E+05	3.71E+04	0.189
1.00	1.23E+05	4.36E+04	0.354

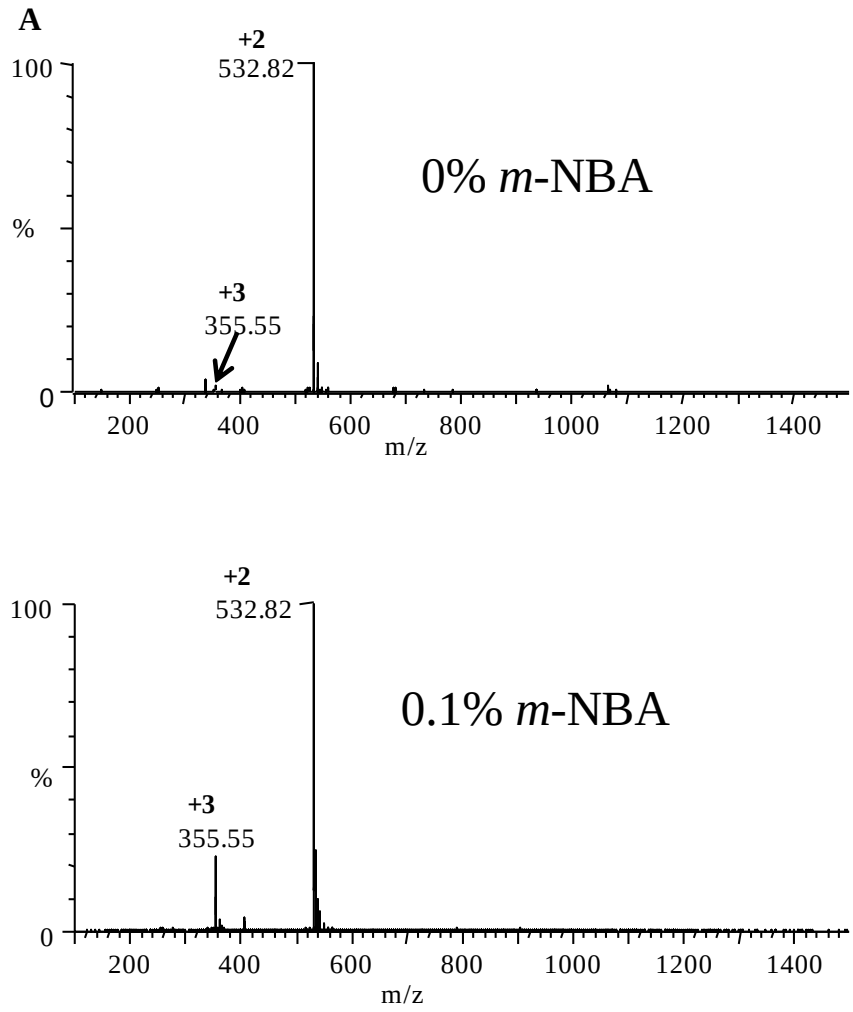

Figure S3.1. Full Mass Spectra of Peptide RPMFA*IWK in infusion solution (A) without *m*-NBA and (B) with 0.1% *m*-NBA.

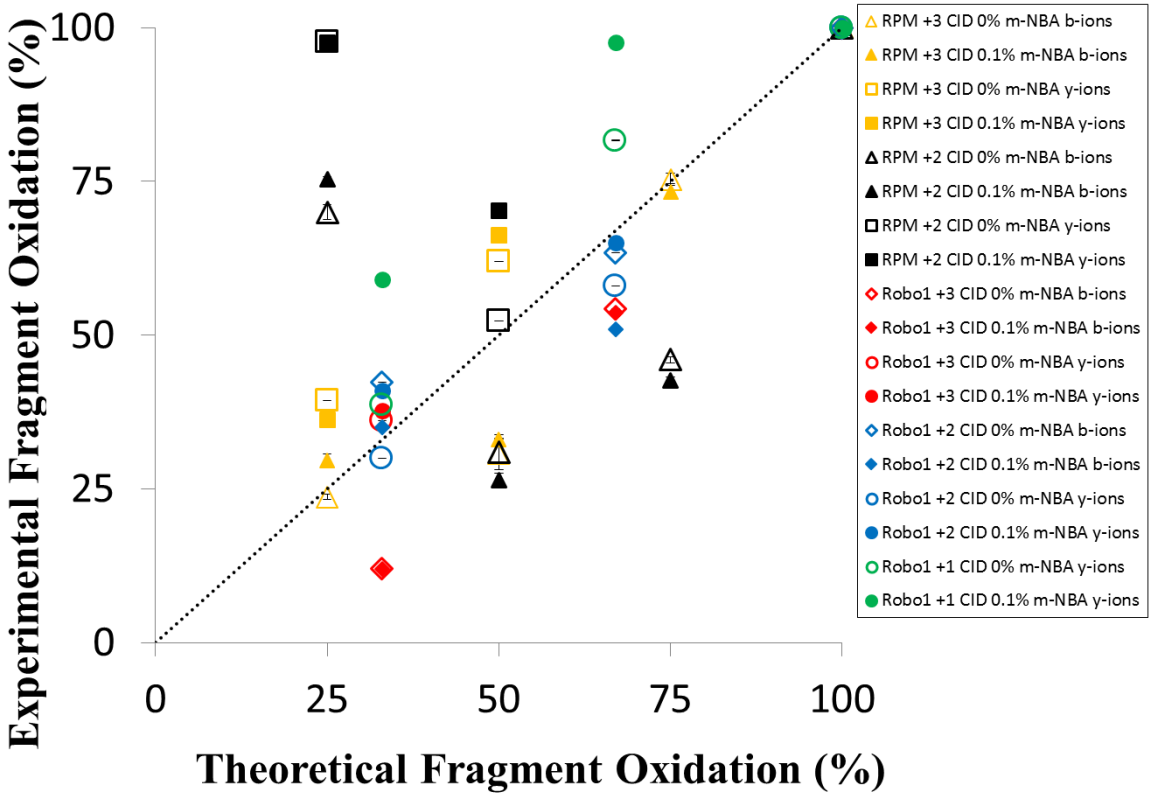

Figure S3.2. Comparison of theoretical and experimental quantification of equimolar mixtures of four oxidation isomers of the peptide RPMFAIWK (triangles and squares) and three oxidation isomers of the Robo1 tryptic peptide MLLPSGSLFFLR (diamonds and circles). The dotted line represents the relationship that should be shown by a quantitative technique. Error bars represent one standard deviation from a triplicate dataset. For both peptides, two factors were found to contribute most strongly to CID inaccuracies. If the charge state was not greater than than the number of basic side chains (two for the RPM peptide series, one for the Robo1 peptide series), or if the product ion included the methionine sulfoxide residue, the product ion abundances tended to be less quantitative. Not all product ion series were present or complete; product ions that were not detected at least 5:1 (S:N) were not represented in the chart.

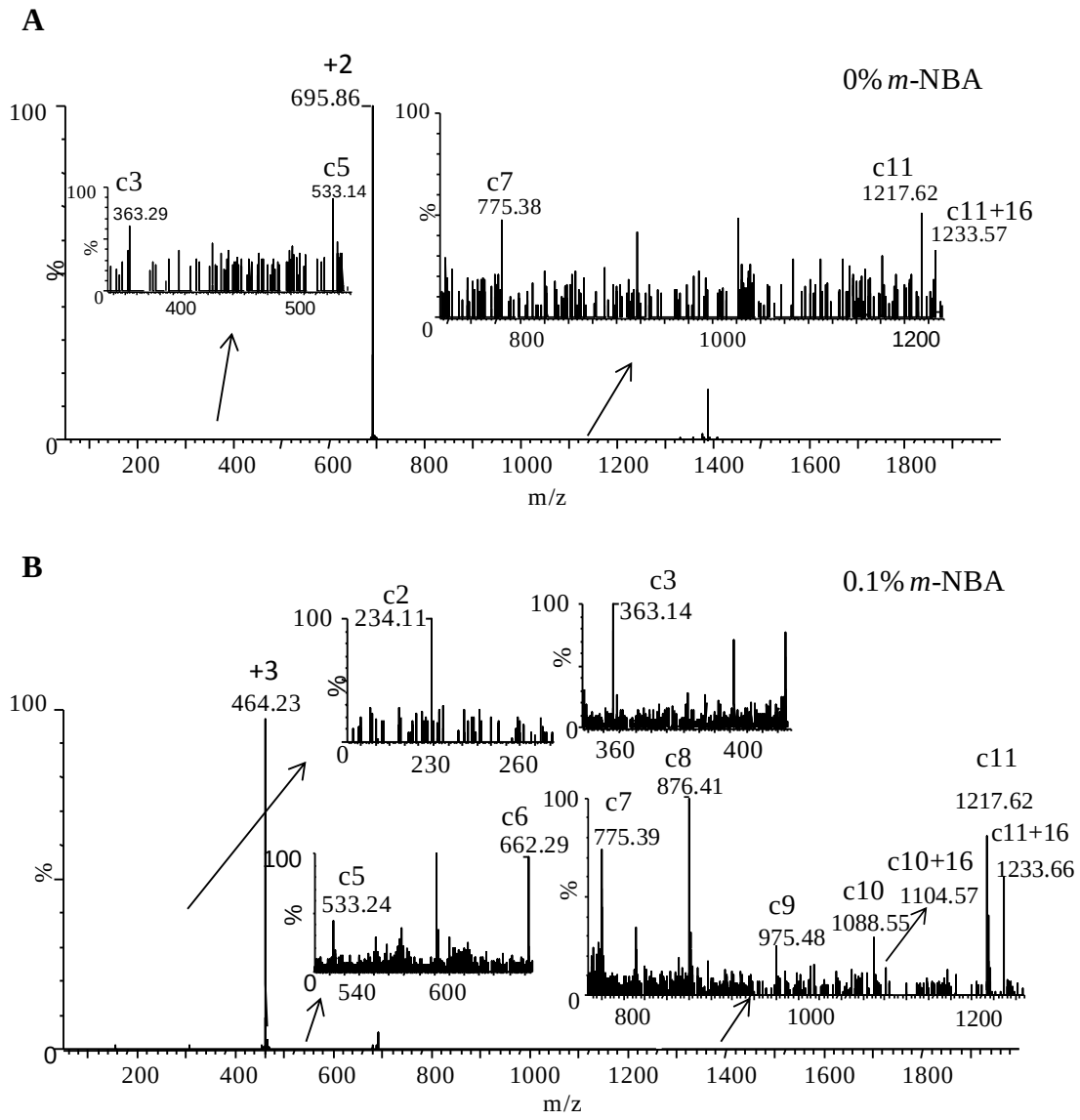

Figure S3.3. ETD spectra of the oxidized Robo1 peptide 194-ESEVAELTVLER-205 (+16) from (A) doubly charged precursor ion with 0% *m*-NBA and (B)triply charged precursor ion with 0.1% *m*-NBA. The +16 designates the product ions that are oxidized.

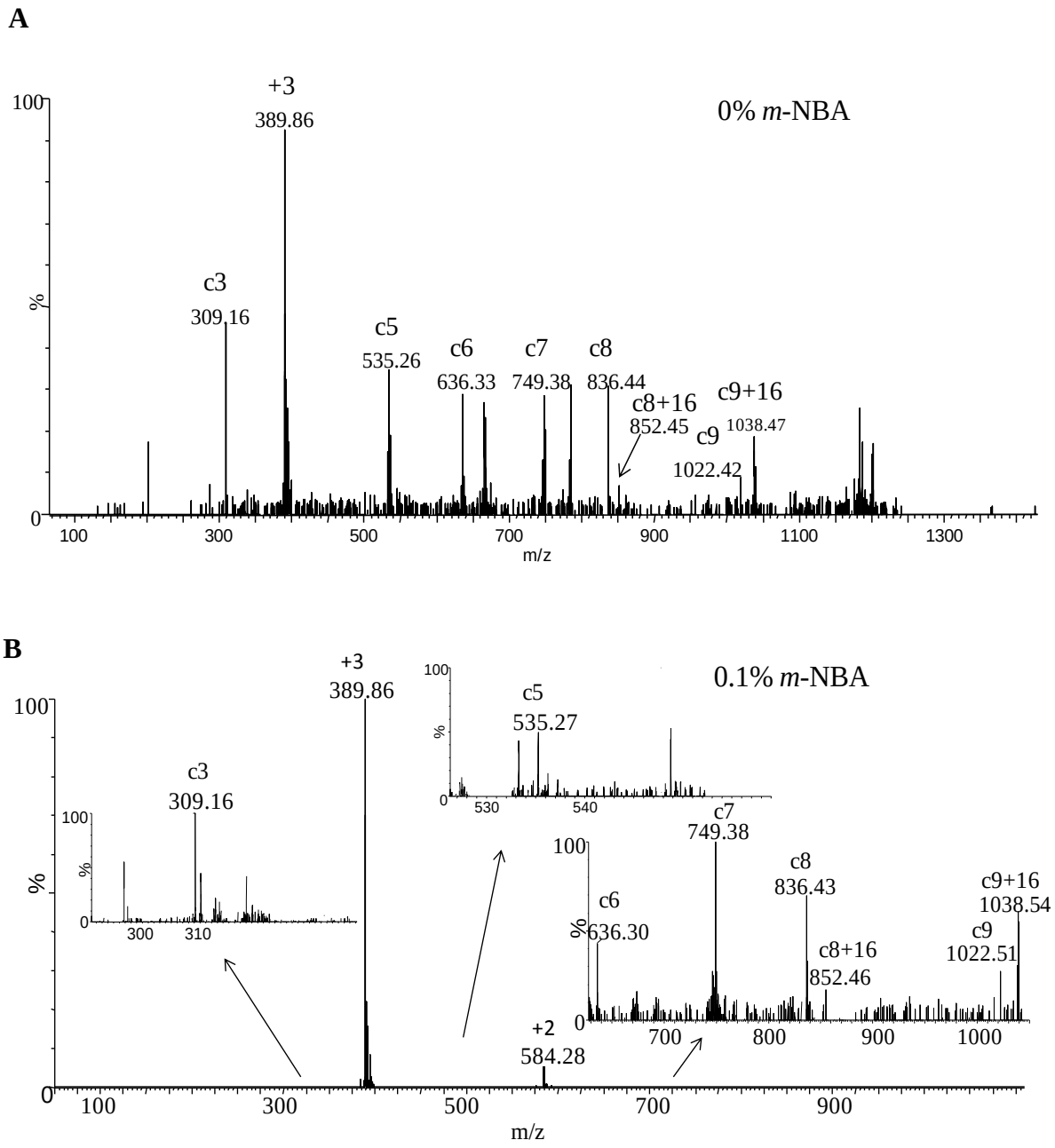

Figure S3.4. ETD spectra of the oxidized Robo1 peptide 140-GHPEPTISWK-149 (+16) from triply charged precursor ion (A) with 0% *m*-NBA and (B) with 0.1% *m*-NBA. The +16 designates the product ions that are oxidized.

REFERENCES

1. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry.* Analytical Biochemistry, 2003. **313**(2): p. 216-225.

2. Maleknia, S.D., M. Brenowitz, and M.R. Chance, *Synchrotron radiolysis studies of peptides & proteins by mass spectrometry.* Biophysical Journal, 1999. **76**(1): p. A172-A172.

3. Hambly, D.M. and M.L. Gross, *Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale.* J Am Soc Mass Spectrom, 2005. **16**(12): p. 2057-63.

4. Xu, G.H. and M.R. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics.* Chemical Reviews, 2007. **107**(8): p. 3514-3543.

5. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry.* Analytical Chemistry, 2004. **76**(3): p. 672-683.

6. Charvatova, O., et al., *Quantifying protein interface footprinting by hydroxyl radical oxidation and molecular dynamics simulation: application to galectin-1.* J Am Soc Mass Spectrom, 2008. **19**(11): p. 1692-705.

7. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochem Biophys Res Commun, 2001. **287**(3): p. 614-21.

8. Bridgewater, J.D., et al., *The effect of histidine oxidation on the dissociation patterns of peptide ions.* J Am Soc Mass Spectr, 2007. **18**(3): p. 553-562.

9. Srikanth, R., et al., *Improved sequencing of oxidized cysteine and methionine containing peptides using electron transfer dissociation.* J Am Soc Mass Spectrom, 2007. **18**(8): p. 1499-506.

10. Srikanth, R., J. Wilson, and R.W. Vachet, *Correct identification of oxidized histidine residues using electron-transfer dissociation.* Journal of Mass Spectrometry, 2009. **44**(5): p. 755-762.

11. Watson, C., et al., *Pulsed electron beam water radiolysis for submicrosecond hydroxyl radical protein footprinting.* Anal Chem, 2009. **81**(7): p. 2496-505.

12. Yan, Y., et al., *Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin.* J Am Soc Mass Spectrom, 2014.

13. Li, X., et al., *Improved identification and relative quantification of sites of peptide and protein oxidation for hydroxyl radical footprinting.* J Am Soc Mass Spectrom, 2013. **24**(11): p. 1767-76.

14. Iavarone, A.T., J.C. Jurchen, and E.R. Williams, *Supercharged protein and peptide ions formed by electrospray ionization.* Anal Chem, 2001. **73**(7): p. 1455-60.

15. Kjeldsen, F., et al., *Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry.* Anal Chem, 2007. **79**(24): p. 9243-52.

16. Watson, C. and J.S. Sharp, *Conformational analysis of therapeutic proteins by hydroxyl radical protein footprinting.* AAPS J, 2012. **14**(2): p. 206-17.

17. Gau, B.C., et al., *Fast photochemical oxidation of protein footprints faster than protein unfolding.* Anal Chem, 2009. **81**(16): p. 6563-71.

18. Jumper, C.C., et al., *High-resolution mapping of carbene-based protein footprints.* Anal Chem, 2012. **84**(10): p. 4411-8.

CHAPTER 4

HIGH STRUCTURAL RESOLUTION HYDROXYL RADICAL PROTEIN FOOTPRINTING REVEALS AN EXTENDED ROBO1-HEPARIN BINDING INTERFACE[3]

[3]Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K..; Linhardt, R.; Sharp, J. *Journal of Biological Chemistry*, 2015. **290**(17): p. 10729-10740. Reprinted here with permission of publisher.

ABSTRACT

Interaction of transmembrane receptors of the Robo family and the secreted protein Slit provides important signals in the development of the central nervous system and regulation of axonal midline crossing. Heparan sulfate (HS), a sulfated linear polysaccharide modified in a complex variety of ways, serves as an essential co-receptor in Slit-Robo signaling. Previous studies have shown that closely-related heparin octasaccharides bind to *Drosophila* Robo directly, and SPR analysis revealed that Robo1 binds more tightly to full-length unfractionated heparin. For the first time, we utilized ETD-based high spatial resolution hydroxyl radical protein footprinting to identify two separate binding sites for heparin interaction with Robo1: one binding site at the previously-identified site for heparin dp8, and a second binding site at the N-terminus of Robo1 that is disordered in the X-ray crystal structure. Mutagenesis of the identified N-terminal binding site exhibited a decrease in binding affinity as measured by SPR and heparin affinity chromatography. Footprinting also indicated heparin binding induces a minor change in the conformation and/or dynamics of the Ig2 domain, but no major conformational changes were detected. These results indicate a second low-affinity binding site in the Robo-Slit complex, as well as suggesting the role of the Ig2 domain of Robo1 in heparin-mediated signal transduction. This study also marks the first use of ETD-based high spatial resolution hydroxyl radical protein footprinting, which shows great utility for the characterization of protein-carbohydrate complexes.

INTRODUCTION

During neuronal development, growing axons are initially attracted to the midline by following specific pathways and encountering a number of guidance cues in their extracellular environment [1]. A ligand-receptor system known to be very important in axon guidance involves the interaction of Slit and Robo at the CNS midline [2]. Members of the Roundabout (Robo) family of proteins are single-pass transmembrane cell adhesion molecules that are highly conserved across many species, from C. elegans to humans [3]. Four Robo proteins (Robo1, Robo2, Robo3/Rig-1, and Robo4/magic Roundabout) have been identified in mammals [4, 5]. Robo1-3 are expressed on the surface of growing axons [1], while Robo4 is specifically expressed in vascular endothelial cells and hematopoietic stem cells [6]. In all identified Robo receptors except for vertebrate Robo4, the Ig1 and Ig2 domains have been evolutionarily conserved and are crucial for forming the Robo-Slit protein-protein complex. Slit-Robo interactions are required in the development of the central nervous system and regulation of axonal midline crossing as well as influencing dendrite repulsion and self-avoidance [7]. Slit-Robo signaling is also essential for many neurodevelopmental processes including formation of the olfactory tract, the optic nerve, and motor axon fasciculation [8, 9]. In addition, Slit-Robo signaling contributes to cell migration [10], leukocyte chemotaxis [11] and the development of other tissues such as the lung, kidney, muscle and breast [12, 13]. Mutations in Robo genes have been linked to multiple human disorders such as dysplasia [14].

Previous studies have shown that heparan sulfate is required for Slit-Robo signaling [15, 16], and acts at least in part by promoting the formation of a ternary Slit-Robo-HS signaling complex [16-18]. Heparan sulfate (HS) is a member of the glycosaminoglycan (GAG) family, consisting of linear polysaccharides bearing repeating disaccharide units of (-4GlcNAcα1,

4GlcUAβ1-) that are modified by variable epimerization of the uronic acid and both N-sulfation or deacetylation of the GlcNAc and O-sulfation of both monosaccharides. Heparin is a closely related GAG typically stored in the secretory vesicles of mast cells that tends to be more heavily sulfated. HS and heparin chains are attached to extracellular matrix proteins to form HS proteoglycans [19, 20]. HS-protein interactions have been shown to modulate protein activity by various mechanisms, such as stabilizing proteins at their sites of production and in the matrix for future mobilization, protection of proteins against degradation, binding ligands to other receptors, and regulating enzyme activity [20]. Glycosaminoglycan (GAG) - protein interactions are essential to many biological processes, including those involving cell-cell communication, cell migration, and stabilization of cellular structure [20]. However, the large size, heterogeneous structure, and inherent flexibility of GAG chains make the detailed structural interrogation of protein-GAG interactions difficult, particularly for full-length GAG chains. Previous studies have shown that heparin octasaccharides bind to *Drosophila* Robo directly and the heparin dp8 binding sites have been identified by mutagenesis and partial direct crystallographic observation of a bound heparin oligosaccharide [21]. These observations indicate that HS/heparin plays a direct and essential role in Slit-Robo signaling. However, these studies of Robo1-heparin binding are all performed on relatively small oligosaccharides. Previous surface plasmon resonance (SPR) analysis of Robo1-heparin interactions revealed that heparin dp8 bound to Robo1 with considerably lower affinity than full-length unfractionated heparin [22]. Additionally, heparin dp8 SPR binding curves fit a single state binding model, while full-length heparin-Robo1 SPR binding curves fit a two-state binding model, including a fast binding event similar to the single state model for heparin dp8 and an additional slow binding event that is not observed with the dp8 heparin ligands [22]. These data indicate that the dp8 ligand does not fully engage the Robo1 binding site, and a structural

characterization of the Robo1-unfractionated heparin complex is required to understand the interaction with extended HS/heparin ligands. However, unfractionated heparin is a polydisperse mixture of large, flexible chains (~10-40 kDa) with a variety of modification patterns. Standard high-resolution structural technologies are unable to handle such a diverse mixture of highly-dynamic complexes.

Hydroxyl radical protein footprinting (HRPF) coupled with mass spectrometry (MS) has become an increasingly popular structural analysis technique to probe molecular interaction interfaces given its fast time frame for protein modification, high labeling efficiency and stability of labeled analyte [23]. HRPF oxidizes amino acid side chains at a rate based on a combination of the inherent reactivity and the accessibility to the hydroxyl radical diffusing in solution. This rate dependence on solvent accessibility makes HRPF a very promising technique to probe intact protein structure [24], protein conformational change [25], protein-protein interactions [26], and protein-ligand binding surfaces in solution [27]. Generally, in an HRPF experiment the protein of interest is exposed to diffusing hydroxyl radicals and the radicals covalently react with the solvent accessible side chains of residues [28]. Certain methods of radical generation for HRPF, including Fast Photochemical Oxidation of Proteins (FPOP) [29] and electron accelerator radiolysis [30] are able to complete the labeling process faster than the protein can undergo labeling-induced conformational changes, ensuring that the HRPF is not probing an artifactual structure. The stable modifications to the protein side chains are analyzed by liquid chromatography-mass spectrometry (LC-MS), relatively quantifying modified and unmodified peptides, and the MS signal of the oxidized version(s) of each peptide is compared to the MS signal of the unoxidized version of the same peptide to quantify oxidation. By comparing the same protein under two different states (e.g. ligand-bound versus ligand-free), changes in solvent accessibility caused by the structural

differences are detected. Traditional HRPF quantitation was performed on the peptide level allowing for structural information. If multiple amino acids on the same peptide were oxidized (a common occurrence in HRPF), this method yielded structural information with a spatial resolution limited by the size of the peptide. While such information is useful for monitoring large-scale protein conformational changes or large protein-protein interaction interfaces, they are much less useful for studying smaller protein-ligand interaction interfaces. High spatial resolution HRPF necessitates not only the identification of the sites of oxidation on a residue level, but also the quantitation of the level of oxidation at each site within the peptide, providing structural information with higher spatial resolution and more accuracy. Moreover, as HRPF is measuring the changes in the solvent accessibility of the protein, it can be carried out in the presence of a complex mixture of ligands, only some of which bind tightly to the protein.

In our previous work, we have successfully demonstrated the accuracy of electron transfer dissociation (ETD)-based methods for quantifying multiple adjacent sites of isomeric oxidation products, as well as the general applicability of ETD-based dissociation to quantify oxidized peptide isomers with multiple sites of oxidation on both a series of synthetic peptide oxidation isomers and peptides from Robo1 Ig1-2 protein generated by HRPF [31]. Here, this technology is applied to characterizing the interaction of Robo1 Ig1-2 with unfractionated heparin. We have identified basic residues in and around the previously identified site for high affinity heparin binding [21], as well as a previously unidentified second binding site near the N-terminus of Robo1 that is evolutionarily conserved and influence heparin binding as determined by site-directed mutagenesis, SPR and heparin affinity chromatography. We have also identified residues in the hydrophobic core of the Ig2 domain that show altered solvent accessibility upon binding of unfractionated heparin, suggesting a heparin binding-induced conformational change in the Ig2

domain that may be involved in heparin-induced signal transduction. We propose a model of Robo-Slit-heparin/HS binding and signaling involving the previously identified high-affinity heparin binding site, and a second low affinity binding site near the N-terminus that is present in the Slit-Robo complex, which is important for triggering the conformational changes required for signal transduction. This work also marks the first determination of protein-GAG interaction by high-resolution HRPF, providing binding interface information down to the amino-acid residue level in some cases.

EXPERIMENTAL PROCEDURES

Protein expression and purification

Human Robo1 (Ig domains 1 and 2 encoding amino acids 58-161 from UniProtQ9Y6N7) was generated as a soluble secreted fusion protein in HEK293 suspension cultures as previously described [22]. Briefly, the Robo1 protein coding region was designed and synthesized for in-frame insertion into the pGEn2 mammalian expression vector [32]as a fusion protein expression product. The vector encodes a 25-amino acid signal sequence is followed by a His8tag, AviTag, and the "superfolder" GFP coding region. The synthetic Robo1 cassette was subcloned downstream of the vector sequence as a EcoRI-HindIII restriction fragment containing an NH2-terminal 7-amino acid recognition sequence of the tobacco etch virus (TEV) protease and residues 1958-2276665 of human Robo1 (UniProt Q9Y6N7). For generation of mutants that eliminate the primary heparin sulfate binding site (Mutant I: R136→A, K137→A) and the secondary heparin sulfate binding site (Mutant II: R62→A, R69→A) the final expression constructs were subjected to site-directed mutagenesis using the QuikChange^{TM}mutagenesis kit (Stratagene, La Jolla, CA). The recombinant product was expressed in suspension culture mutant HEK293S GnTI- cells (ATCC catalog no. CRL-3022) by transient transfection as a PEI-DNA complex as previously

described [33].Following expression for 6 days, the cultures were harvested, clarified by centrifugation at 3500 rpm for 15 min, and the recombinant product was purified by Ni-NTA Superflow (Qiagen, Valencia, CA) chromatography as described [22]. The eluted GFP-Robo1 fusion protein was concentrated to ~1 mg/ml by ultrafiltration with a 10-kDa molecular mass cutoff membrane (Millipore, Billerica, MA) and tags and glycans were excised by treatment with recombinant TEV protease and endoglycosidase F1 (EndoF1) and further purified as described [33].For surface plasmon resonance binding studies wild type and mutant forms of Robo1 were expressed in wild type HEK293 cells (FreeStyle 293-F cells, Invitrogen)and purified by Ni-NTA chromatography and subjected to TEV protease cleavage without addition of EndoF1 and further purified as described [33]. The protein preparation was concentrated by ultrafiltration to ~3 ml and was further purified on a Superdex 75 column (GE Healthcare) preconditioned with a buffer containing 137 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4, 2 mM KH_2PO_4, pH 7.4. Peak fractions of Robo1 were collected and concentrated. Later MS analysis indicated that the final expressed products are lacking the C-terminal lysine, resulting in an expressed protein of residues 58-265.

HRPF of proteins

HRPF of Robo1 with and without a heparin ligand was performed by FPOP as previously described [29, 34]. Robo1 (20 μM) and unfractionated heparin (estimated at 40 μM based on average mass of heparin chain) were mixed together and incubated for 1h. Both Robo1 alone and Robo1 with heparin were combined with glutamine and nanopure water to give a final protein concentration of 20 μM Robo1, ~40 μM unfractionated heparin in 20 μL final volume. Glutamine was added as a radical scavenger at a concentration of 20mM. Angiotensin peptide (10 μM) was added to each protein solution as a radical dosimeter to monitor the available radical dose in each protein solution. Concentrated hydrogen peroxide (1 M) was added to each replicate of a triplicate

set to give a final concentration of 100 mM just prior to flow through the laser beam path for oxidative modification. The protein samples were loaded into a 100-μL syringe and introduced via a syringe pump coupled to the 100-μm inner diameter fused silica tubing. Samples flowed through the beam path of an EX100 KrF excimer laser at 248 nm with a laser power of 50 mJ/pulse (GAM Laser Inc., Orlando, FL). The flow rate of the syringe pump was set at 12.19 μl/min, with the laser pulse repetition rate set such that each segment of protein sample was irradiated with a single ~ 20-ns UV pulse with a 10% unirradiated buffer region between irradiated segments to help account for sample diffusion and laminar flow [29, 35]. Samples were collected in a microcentrifuge tube containing methionine amide (0.5 μg/μL) and catalase (0.5 μg/μL) with sodium phosphate (50 mM, pH 7.4) to immediately quench oxidation by destroying excess hydrogen peroxide and other reactive oxygen species. In addition, control samples were made with hydrogen peroxide under same conditions, but they were not subjected to laser irradiation.

Following irradiation, each sample was lyophilized and resuspended in 20 μL of 50mM sodium phosphate. Samples were reduced with 5mM DTT and incubated at 65°C for 30 min. After cooling to room temperature, samples were digested overnight with one of three proteases: trypsin at 37℃ at an enzyme: protein ratio of 1:20, Lys-C orGlu-C at the ratio of 1:100. The digested samples were stored at -20° C for LC-MS/MS analysis.

LC-MS/MS acquisition and data analysis

Samples were analyzed on a LTQ Orbitrap XL (Thermo Fisher Scientific) controlled by Xcalibur 2.0.7 software (Thermo Fisher, San Jose, CA). Samples were loaded on a 75μm (i.d.105 mm) C18 fritless reverse phase column (packed in house, YMC GEL ODS-AQ120 ÅS-5, Waters) by nitrogen pressure bomb. Samples were eluted directly into the nanospray source of an LTQ Orbitrap XL with a 160min linear gradient consisting of 5-100% acetonitrile with a constant

concentration of 0.1% formic acid over 100 min at a flow rate of 250 nL/min. The spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to 200℃. Full MS scans were acquired from m/z 200 to 2000 at a nominal resolution of 60 000 for ions of m/z 400 with the FT mass spectrometer. To obtain MS/MS data of oxidized peptides, ions were fragmented by ETD with the corresponding parent ion mass list. ETD-based precursor activation was carried out for 100ms, including charge-state-dependent supplemental activation. Precursor ions were isolated with a width of 3 m/z units.

Unoxidized Robo1 controls, oxidized Robo1 and oxidized Robo1-unfractionated heparin complex peptide sequences were identified using MASCOT 2.2.2 (Matrix Science, London, UK) and ByOnic V1.2-250 (Protein Metrics, CA, USA). All tandem mass spectra assignments and sites of oxidation were verified manually due to peptide modification complexity. The tryptic and Lys-C digested peptides and corresponding oxidation products were identified from the LC-MS runs manually to calculate the average oxidation events per peptide in the sample. Peptide level quantitation of the average oxidation events per peptide is calculated by summing the ion intensities of all the oxidized peptide masses multiplied by the number of oxidation events required for the mass shift (e.g., one event for +16, two events for +32), then divided by the sum of the ion intensities of all unoxidized and oxidized peptide masses (eq 1):

$$P = \frac{I(+16)oxidized*1+I(+32)oxidized*2+I(+48)oxidized*3+\cdots}{Iunoxidized+I(+16)oxidized+I(+32)oxidized+I(+48)oxidized+\cdots} (1)$$

where P denotes the average oxidation events per peptide and I donates to the intensities of oxidized and unoxidized peptides.

Residue level quantitation was calculated by the fragment ion intensities from ETD to determine oxidation extent at specific residue sites based on our previous study [31] and a similar

approach reported by Jumper et al for carbene-based footprinting [36].Briefly, oxidized peptide with oxidation at multiple sites can generate both oxidized and unoxidized sequence ions in its MS/MS spectrum. The fractional oxidation of a given sequence ion is defined as the ratio between the oxidized sequence ion intensity to the sum of the intensity of the corresponding oxidized and unoxidized sequence ion (eq 2).

$$f(Ci) = \frac{I(Ci)oxidized}{I(Ci)unoxidized + I(Ci)oxidized} (2)$$

Where $f(Ci)$ denotes the fractional oxidization of sequence ion i (e.g., oxidized c3 ions generated by ETD), and $I(Ci)$ denotes the intensity of the oxidized and unoxidized ion of interest (e.g. oxidized c_3 ion plus unoxidized c_3 ion). The absolute level of oxidation for a given amino acid residue i is based on both the average oxidation event of peptide and the fractional oxidiation of the corresponding sequence ions, as shown in eq 3:

$$\text{oxidation/residue}i = \text{P}[f(Ci) - f(Ci-1)] (3)$$

where P is the average oxidation events/peptide as derived from Equation 1, and the term in square brackets is the fractional oxidation difference of two adjacent sequence ions, Ci and Ci-1. Multiplying the average number of oxidations/peptide by the fraction of that oxidation that occurs on a given amino acid residue yields the average oxidation events per residue. For ETD MS/MS spectra where ETD fragmentation ions are not adjacent in sequence, fractional oxidation for multiple contiguous residues within the peptide can be calculated by using non-adjacent ETD fragments in Equation 3, resulting in a spatial resolution of HRPF that is limited by the distance between ETD fragmentation sites.

Dynamic light scattering

A DynaProNanoStar system (Wyatt Technology Corporation, Santa Barbara, CA) was used to conduct all dynamic light scattering (DLS) measurements. Laser wavelength is 825 nm. Samples of different concentration were held at 25 °C by a temperature controlled sample holder and allowed to equilibrate for 120 seconds prior to analysis. Each size measurement was determined from 20 scans, 5 seconds per scan. All DLS data were collected and analyzed using DYNAMIC V6™ Software – Version 6.3.01. Hydrodynamic radius calculations based on the Robo1 XRC structure were performed using WINHYDROPRO10 [37].

Heparin affinity chromatography

TEV protease digested, purified Robo1 proteins devoid of the GFP fusion tags (500 µg each of wild type, mutant 1, or mutant 2 Robo1 protein) were injected onto a 5-ml HiTrap Heparin HP column (GE Healthcare) equilibrated in 0.05 M Tris-HCl, pH 7.5. Bound proteins were eluted with a linear NaCl gradient of 0–1 M NaCl over 5 column volumes at a flow rate of 1 ml/min.

Surface plasmon resonance

The biotinylated heparin was prepared by reaction of sulfo-N-hydroxysuccinimide long-chain biotin (Thermo Fisher, Rockford, IL) with free amino groups of unsubstituted glucosamine residues in the polysaccharide chain following a published procedure [38]. The SPR analysis was conducted using a BIAcore 3000 (GE Healthcare, Uppsala, Sweden) SPR instrument. The biotinylated heparin was immobilized to streptavidin (SA) chip (GE Healthcare,) based on the manufacturer's protocol. In brief, 20 µl solution of the heparin-biotin conjugate (0.1 mg/mL) in HBS-EP running buffer was injected over flow cell 2 (FC2) of the SA chip at a flow rate of 10 µL/min. The successful immobilization of heparin was confirmed by the observation of a 100 to 200 resonance unit (RU) increase in the sensor chip. The control flow cell (FC1) was prepared by

1 min injection with saturated biotin. To measure the interaction between heparin and Robo1, protein samples were diluted in HBS-EP buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% surfactant P20, pH 7.4). Different dilutions of protein samples were injected at a flow rate of 50 µL/min. At the end of the sample injection, the same buffer was flowed over the sensor surface to facilitate dissociation. After a 2 min dissociation time, the sensor surface was regenerated by injecting with 50 µL of 0.25% SDS to get fully regenerated surface. The response was monitored as a function of time (sensorgram) at 25 °C. Data were analyzed using BIA evaluate version 4.0.1 using a 1:1 Langmuir model (corrected for bulk signal); Robo1 (A) binds to heparin (B) to form complex (AB),

$$A + B \underset{kd}{\overset{ka}{\rightleftharpoons}} AB$$

REACTION 1

$$d[AB]/dt = k_a[A]\,[B] - k_d\,[AB] \quad (4)$$

$$d[RU]/dt = k_a\,C\,(RU_{max} - RU) - K_d\,RU \quad (5)$$

where ka is the apparent association rate constant [$M^{-1}s^{-1}$] and kd is the apparent dissociation rate constant (s^{-1}). The apparent dissociation equilibrium constant K_D = kd / ka.

RESULTS

The interaction of Robo1and heparin previously described by SPR analysis suggests that Robo1 binds more tightly to full-length unfractionated heparin, and the binding event with unfractionated heparin contains a slow component that is not detected in the Robo1-heparin dp8 interaction [22]. To investigate the binding site between Robo1 and unfractionated heparin at a high spatial resolution, Robo1 and Robo1-heparin samples were aliquoted into three replicates, and each sample was oxidized for HRPF by FPOP. As the protein is exposed to the short burst of

hydroxyl radicals, the radicals react with the amino acid side chains with an apparent rate that is a function of the amino acid sequence, which is constant between samples, and the exposure of the side chain to the solvent, which changes at the binding interface as the heparin binds to the protein, as well as changing in the vicinity of any heparin binding-induced conformational changes. Immediately after FPOP oxidation, the samples were quenched to eliminate excess hydrogen peroxide and any other long-lived secondary oxidants, digested with multiple enzymes to achieve full sequence coverage of the protein, and analyzed by LC-MS/MS.

High spatial resolution HRPF of Robo1-heparin complex

The Robo1 protein analyzed in this study consists of the two domains Ig1-2 with 209 amino acids and an N-glycan site (**Fig. 4.1**). To study the interaction of Robo1 with heparin, a Robo-heparin complex was prepared by mixing Robo1 with unfractionated heparin at an approximate 1:2 molar ratio. The samples were oxidized under the same experimental conditions with a peptide radical dosimeter used to ensure that all samples received the same amount of available hydroxyl radical. Peptide level quantitation of oxidation for Robo1-heparin was performed and compared with the extent of oxidation of Robo1alone (**Fig. 4.2**). Chromatography of the oxidized residues was intentionally performed so that peptide oxidation isomers did not resolve. A residue level quantitation was employed by measuring the intensity of unoxidized and oxidized c/z ions generated by ETD and calculating the relative average oxidation for each residue as described under "Experimental Procedures."In order to determine the statistical difference in the amount of oxidation for these residues, two-tailed independent Student's t tests were used for statistical analysis, with an of 0.05 selected for statistical significance. Residues exhibiting statistically significant differences upon heparin binding are shown in **Table 4.1**. The amounts of oxidation of Robo1 and Robo1-heparin and protection for these residues are also shown. As can be seen in

Table 4.1, residues R62, F66-P67-P68, R69, I70-V71, S80-K81, and A166-I167-L168-R169 have heparin-bound oxidation levels less than 50% of the Robo1 without heparin, indicating strong protection upon heparin binding. F129-L130-R131, C147, M180, and M189 have heparin-bound oxidation levels greater than 50% of Robo1 without heparin, although statistically significant protection is observed.

```
58  GSRLRQEDF  PPRIVEHPSD  LIVSKGEPAT  LNCKAEGRPT  PTIEWYKGGER
108 VETDKDDPR  SHRMLLPSGS  LFFLRIVHGR  KSRPDEGVYV  CVARNYLGEAV
158 SHNASLEVA  ILRDDFRQNP  SDVMVAVGEP  AVMECQPPRG  HPEPTISWKKD
208 GSPLDDKDE  RITIRGGKLM  ITYTRKSDAG  KYVCVGTNMV  GERESEVAELT
258 VLERPSFV
```

Figure 4.1.Amino acid sequence of Robo1Ig1-2 domain. Identified oxidized amino acids are colored red. The underline indicates the N-glycan site. Unoxidized residues in the Ig1 domain are colored green; unoxidized residues in theIg2 domain are colored blue.

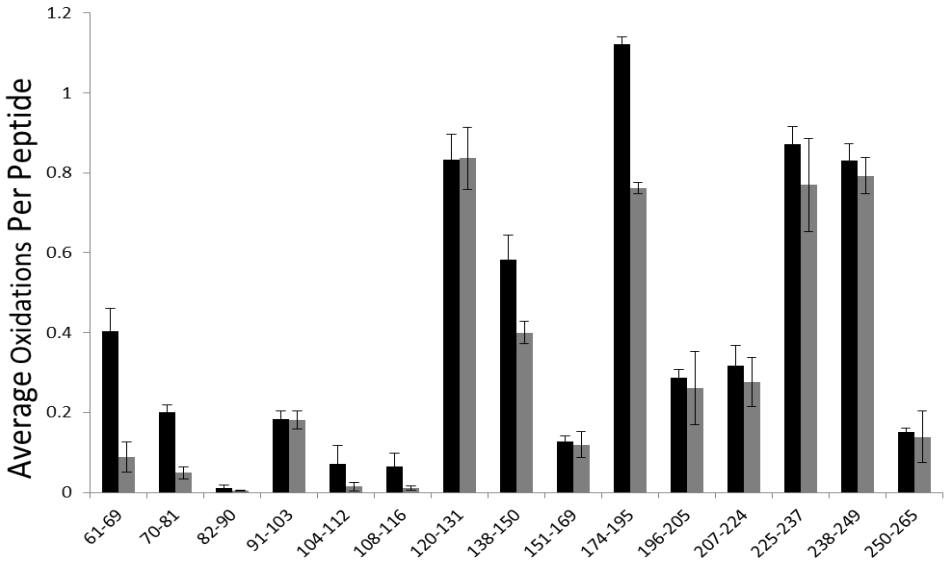

Figure 4.2. Hydroxyl radical protein footprinting of Robo1-Heparin. Extent of oxidation of Robo1 with and without unfractionated heparin at peptide level is shown. *Error bars*, S.D.from a triplicate set of FPOP oxidations and analyses.

Table 4.1. Independent Two-Tailed Student's T-Test of HRPF between Robo1 and Robo1-heparin

Protected Residues	Avg. Oxid./Residue Robo1 (±SD)	Avg. Oxid./Residue Robo1-heparin (±SD)	*p*-value	%Protection upon heparin binding
R62	0.198 ± 0.035	0.048 ± 0.023	0.02	75.8
F66P67P68	0.185 ± 0.011	0.040 ± 0.014	<0.01	78.4
R69	0.012 ± 0.003	0.003 ± 0.001	0.03	75.0
I70V71	0.056 ± 0.001	0.015 ± 0.005	<0.01	73.2
S80K81	0.062 ± 0.016	0.016 ± 0.003	0.04	74.2
F129L130R131	0.433 ± 0.024	0.364 ± 0.002	<0.01	15.9
C147	0.582 ± 0.063	0.399 ± 0.028	0.03	31.4
A166I167L168R169	0.057 ± 0.001	0.028 ± 0.003	<0.01	50.9
M180	1.078 ± 0.028	0.726 ± 0.021	<0.01	32.6
M189	0.967 ± 0.030	0.683 ± 0.020	0.01	29.4

The oxidation of Robo1 and Robo1-heparin at residue level are shown as the average number of oxidation events per residue. S.D. values were calculated from triplicate samples.

Identification of heparin binding sites

Heparin-protein binding interactions are most often mediated by interactions between basic residues and the negative charges present on heparin. Previous mutagenesis studies of Robo1 of D. melanogaster found four basic residues that affect elution from a heparin column: K81, R131, R136 and K137 [21]. These residues are conserved within human Robo1.To evaluate the definition of the binding site for full-length heparin, we examined the residue-level HRPF protection information at or around basic residues in the structure. Residue level quantitation by ETD for peptide 61-69 was employed to determine oxidation extent at specific residues site (**Figure 4.3a**). An example product-ion ETD spectrum for singly-oxidized peptide 61-69 is shown in **Figure 4.4**. Residues L61-R62, F66-P67-P68, and R69 all show significant protection upon binding to heparin. The average number of oxidation events per residue for L61-R62decreases from 0.198 to 0.048. Residues F66-P67-P68 are not typical for a binding site for heparin as heparin interacts with positively charged amino acids in most GAG-protein interaction systems. These residues, which show decreased oxidation from 0.185 to 0.040, are adjacent to the basic residue R69shown in **Figure 4.3b**, which is also a minor oxidation target that shows an large oxidation decrease from 0.012 to 0.003, which is a 75% reduction in oxidation and is statistically significant. Protection of the F66-P67-P68 region and the I70-V71 region before and after the R69 site (discussed below) confirms the protection observed in the R69 residue through binding to heparin. These residues were not previously detected as heparin-binding, and are distal to the previously identified binding site [21].

(a) **(b)**

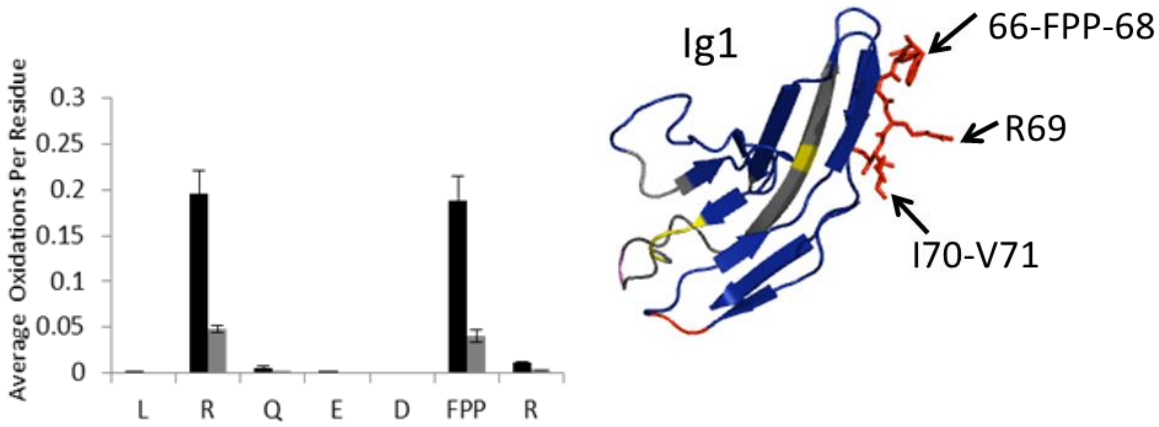

Figure 4.3. Residues involved in binding with unfractionated heparin. **(a)** Extent of oxidation of Robo1 alone (*black bars*) compared to heparin-bound Robo1 (*gray bars*) at the residue level. Error bars S.D. from a triplicate set of experiments. **(b)** Protected residues F66, P67, P68, I70 and V71 surround protected basic residue R69 in three-dimensional space, suggesting that R69 interacts with heparin, which shields neighboring residues. Protected residues L61and R62 are not shown in the crystal structure.

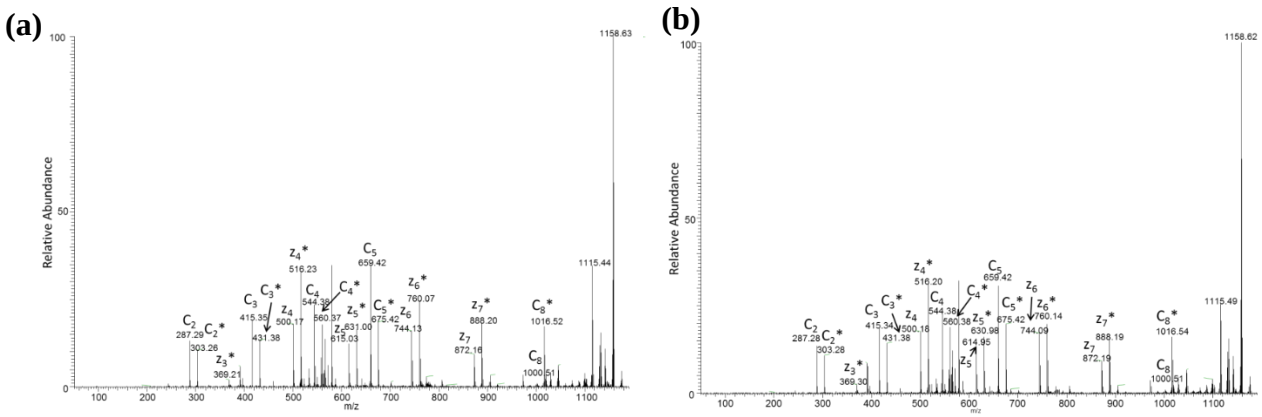

Figure 4.4. Representative ETD spectrum of singly-oxidized peptide LRQEDFFPR for (a) Robo1 alone and (b) Robo1 with heparin. The *asterisks* indicate the product ions that are oxidized. The ETD spectrum shows unoxidized and oxidized c ions as pairs from c2 to c8 while unoxidized and oxidized c1, c6 and c7 ions are absent in the spectrum. Oxidized z3, unoxidized and oxidized z4-z7 ions generated by ETD are also shown in the spectrum. The intensities of c ions are preferred to calculate for quantitation as the interference between unoxidized y-ions and oxidized z-ions prevents accurate quantification in mass spectrometers with lower resolution as demonstrated by our previous work.

Protection was also observed in and around the previously identified heparin binding site. Residues F129-L130-R131 show a decreased amount of oxidation from 0.433 to 0.364 (**Figure 4.5a**). Previous mutagenesis and X-ray crystallography studies have indicated that R131 is involved in heparin interaction [21], which is strongly supported by our high spatial resolution HRPF data (**Figure 4.5b**). As we do not have sufficient ETD fragment ions to differentiate oxidation of these three amino acids, we must quantitate oxidation of this three amino acid stretch as a whole. Taken together, these three amino acids exhibit only modest protection upon heparin binding. However, of the three amino acids in this stretch, F129 is substantially more chemically

reactive and has not been previously implicated in the heparin binding interface. All other factors being equal, the majority of oxidation of this three amino acid stretch would be expected to occur on F129. If the solvent accessibility of F129 is not substantially impacted by heparin binding, even strong protection from heparin binding at R131 would be overshadowed by the greater reactivity and negligible protection of F129 upon heparin binding.

Residue level oxidations for peptide 70-81 are shown in **Figure 4. 6a**. The amount of oxidation for residues I70-V71 in peptide 70-81 decreases from 0.056 to 0.015. As previously mentioned, these residues are directly adjacent to R69, which is heavily protected upon heparin binding. In addition to protection of I70-V71, peptide 70-81 also shows strong protection of S80-K81 upon heparin binding. K81 was shown to bind with heparin in a previous study [21], which we can confirm by high spatial resolution HRPF. The N-glycosylated peptide 151-169 shows either no oxidation or no protection on its residues except for the C-terminus (**Figure 4.6b**). Protected residues A166-I167-L168-R169 in peptide 151-169 are at a basic patch of the protein adjacent to residues previously shown to be involved in heparin dp8 binding (**Figure 4.6c**). We have identified significant protection of these residues upon binding to unfractionated heparin, revealing R169 to be involved in the extended binding site, along with the previously-identified residues involved in dp8 binding (K81, R131, R136and K137) [21]. The latter two residues were not identified by HRPF because they were not oxidized under ligand-free or ligand-bound conditions, although they were reported to be involved in binding with heparin. Our inability to detect significant oxidation of these residues in either the free or bound state made us unable to probe these residues directly. No basic residues were found to be protected by heparin binding within the Ig2 domain.

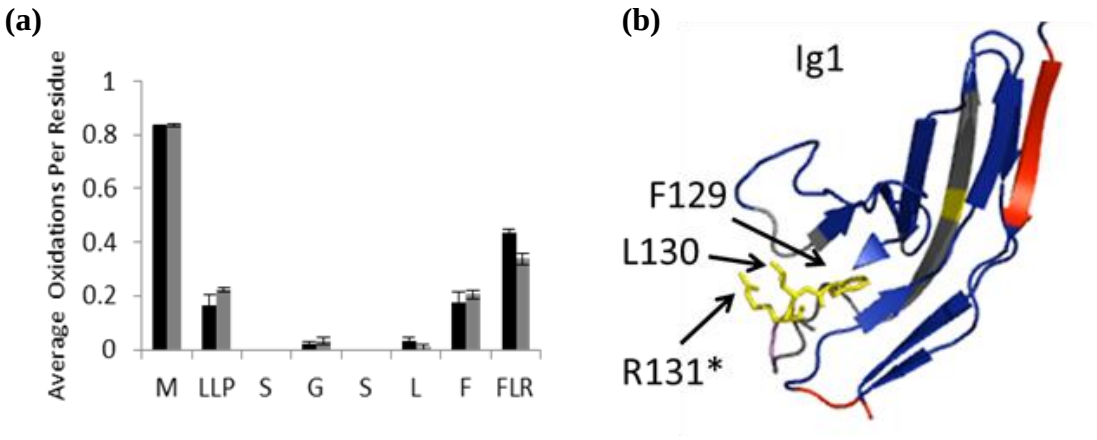

Figure 4.5. Residues involved in binding with unfractionated heparin. (a) Extent of oxidation of Robo1 alone compared to heparin-bound Robo1 at the residue level. *Error bars*, S.D. from a triplicate set of experiments. **(b)** Structure highlighting F129-L130-R131are shown in yellow. Asterisk indicates the residue involved in a heparin dp8 binding reported previously [21].

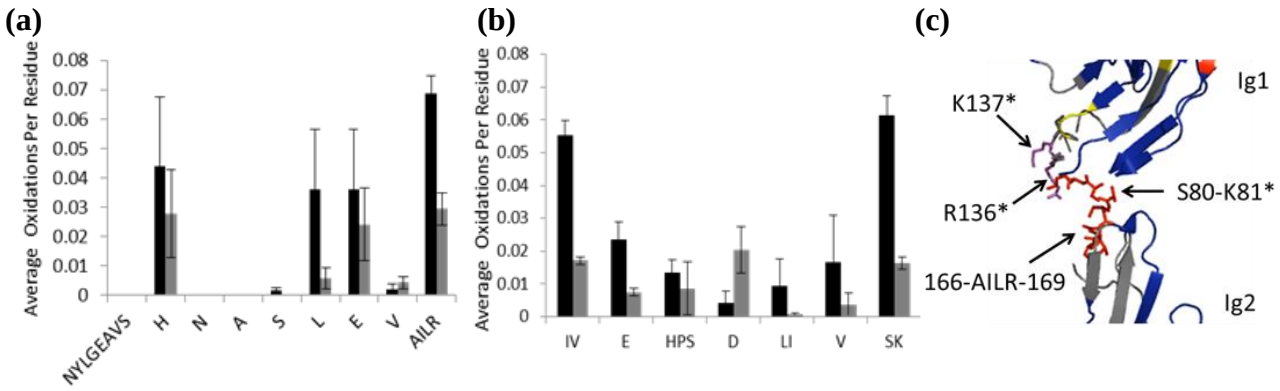

Figure 4.6. Residues identified as binding sites around basic patch. *Error bars*, S.D. from a triplicate set of experiments. **(a)** Residue-level extent of oxidation for peptide 70-81 of Robo1 (*black bars*) and Robo1-Heparin (*gray bars*). **(b)** Residue-level extent of oxidation for peptide 151-169 of Robo1 (*black bars*) and Robo1-Heparin (*gray bars*). **(c)** Structure shows the protected residues S80-K81 and A166-I167-L168-R169 (colored in red) are at a basic patch with basic residues in magenta unidentified by HRPF. *Asterisks* indicate the residue identified as interacting with heparin dp8 previously [21].

Changes of protection of non-binding residues upon heparin binding

A significant decrease of oxidation was also shown in residues C147, M180, and M189 (**Figure 4.7a**). These residues were the sole sites of oxidation in peptide 138-150 and 174-195with a decrease in oxidation from 0.582 to 0.400 for peptide 138-150 and 1.121 to 0.761 for peptide 174-195 (**Figure 4.2**). These residues are all highly reactive to hydroxyl radicals. Previous studies have indicated that residues with high inherent reactivities exhibit drastic changes to their HRPF oxidation level in response to relatively modest changes in the average solvent accessible surface area [27]; therefore, while the protection observed in these three residues is statistically significant, the fact that only these three highly reactive non-binding residues were observed to be modestly protected by heparin binding suggests that the changes in structure and/or dynamics they reflect is relatively modest. Larger conformational changes would be expected to result in measurable changes in the oxidation of a wider variety of less reactive amino acids.

C147is involved in the core disulfide bond of the Ig1 domain (**Figure 4.7b**) and is fully protected from solvent in the static X-ray crystal structure [39]. Oxidation of this cysteine in the heparin-free form despite its full burial in the X-ray structure may indicate dynamics in the Ig1 domain in solution. The modest decrease of oxidation of this highly reactive residue upon heparin binding may indicate a possible modest stabilization of the folded structure of the Ig1 domain upon ligand binding. Alternatively, heparin interactions could induce a modest conformational change not captured in the X-ray crystal structure upon binding of dp8. The change in the structure and/or dynamics of the Ig1 domain must be minor to only cause measurable changes in the HRPF footprint for C147, and not for other residues we measured as unchanged. It should also be noted that the disulfide partner of C147 (C89) remains almost wholly protected in the heparin-bound and heparin-free form, and is unaffected by heparin binding. Exposure and labeling of only one half

of a disulfide bond has been reported previously for lysozyme, where it indicated exposure of one sulfur of the disulfide bond to solvent [40].

M180 and M189 are hydrophobic residues in and around the core of the Ig2 domain (**Figure 4.7c**), and similar to C147 are highly reactive with hydroxyl radicals. Protection of these residues in the Ig2 domain distal from the extended heparin binding site suggests that this protection is likely due to either a heparin-induced multimerization event or a binding-induced change in conformation and/or dynamics of the Ig2 domain. Previous studies showed no evidence of Robo1 multimerization upon heparin binding and dynamic light scattering (DLS) of Robo1 at a concentration five times greater than that used for HRPF also showed that Robo1 exists as a monomer. DLS of the Robo1-heparin complex was inconclusive due to the polydisperse nature of the ligand, but indicated that oligomerization beyond a dimer was unlikely (**Table 4. 2**). Additionally, the lack of protection of any less reactive residues in the Ig2 domain suggests that no stable multimer exists; otherwise, we would expect to see widespread protection along the monomer-monomer interface(s). A very modest change in conformation and/or dynamics of Robo1 upon binding to heparin would be consistent with prior SPR data that indicated a two-state binding model for Robo1 interaction with unfractionated heparin [22], and with the very limited changes in HRPF oxidation observed in Ig2 here.

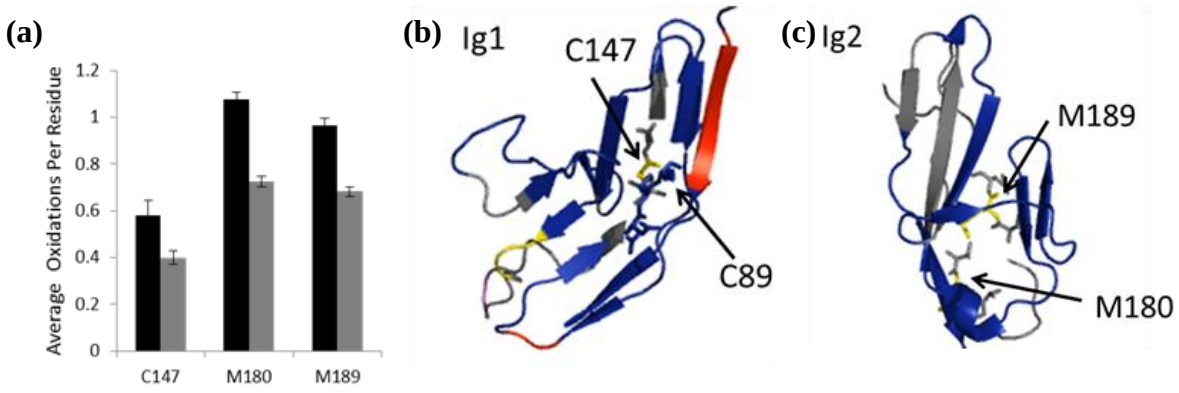

Figure 4.7. Non-binding residues show significant protections. (a) Residue-level extent of oxidation of Robo1 alone (*black bars*) compared with heparin-bound Robo1 (*gray bars*) for residues C147, M180 and M189. *Error bars*, S.D. from a triplicate set of experiments. **(b)** Structural model showing C147 (*yellow*) involved in the disulfide bond with C89 in Ig1 domain. **(c)** Hydrophobic residues M180and M189 are shown in yellow in the Ig2 domain of Robo1.

Table 4.2. Dynamic light scattering results for Robo1 and unfractionated heparin

Sample	R (nm)*
0.1 mM Robo1	2.7
1 mM heparin	2.4
0.5 mM Robo1 + 1 mM heparin	3.7

*The hydrodynamic radius for the monomer of Robo1as calculated from the X-ray crystal structure (PDB ID: 2V9Q) is 2.85 nm.

Heparin binding of site-directed mutants

The HRPF data indicate protection of two basic, conserved residues at the N-terminus of the Robo1 Ig1 domain. In order to determine if these two residues are involved in binding to heparin rather than protected due to a binding-induced structural event, site-directed point mutations were made and expressed as described in Experimental Procedures. Mutant I consists of a double mutation at the previously identified binding site near the Ig1-Ig2 domain interface, transforming two known heparin binding basic residues to alanine (R136→A, K137→A). This

double-mutant in the Drosophila Robo1 homolog (K122A, K123A) was previously shown to strongly decrease binding to a heparin column [21]. Mutant II consists of a double mutation converting both N-terminal basic residues identified as protected by heparin binding to alanine (R62→A, R69→A).

The effect of the mutations on heparin binding was tested by heparin affinity chromatography, with the results shown in **Figure 4.8**. Mutant I shows a large decrease in the NaCl concentration of elution, as previously reported [21]. Mutant II shows a much smaller, but reproducible, decrease in the NaCl concentration required for elution. These results suggest that the N-terminal binding site identified by HRPF is directly involved in heparin binding, but influences the overall affinity of Robo1 to heparin less than the previously identified binding site near the Ig1-Ig2 interface.

These Robo1 mutants were also tested for binding to heparin by SPR using immobilized heparin. Sensorgrams of the Robo1 (WT), Mutant I and Mutant II interactions with immobilized heparin chip are shown in **Figure 4.9 a-c**, respectively. Each sensorgram was globally fitted using the 1:1 Langmuir model to generate the binding kinetics data. The 1:1 Langmuir model fit the binding data of both mutants very well. The fit of the 1:1 Langmuir model to the WT data was not as good, as previously reported [22]; however, fitting of these WT data to a two-state model did not result in a substantial enough improvement to warrant use of this model, possibly due to the use of a slower association period than the previous report (data not shown).We retained the 1:1 Langmuir model for all sensorgrams to facilitate direct comparisons of the data. The SPR data show that the interaction of both mutants to heparin displayed a slower on-rate (k_a), a faster off-rate (k_d) and consequently a reduced affinity (higher K_D) to heparin as compared to the Robo1WT.The SPR data suggest that both binding sites influence binding to Robo1 similarly,

based on the similar k_a, k_d and K_D for each mutant (**Table 4.3**), with the N-terminal binding site identified here having a slightly lower binding affinity than the previously identified binding site at the Ig1-Ig2 interface. However, given the complexity of the Robo1-heparin interaction and the presence of multiple diverse binding sequences in unknown abundances, it is difficult to quantitatively determine exact binding affinities for each of the two binding sites discussed here. However, the heparin affinity column and SPR data both indicate a role for the N-terminal basic residues identified by HRPF in binding to heparin.

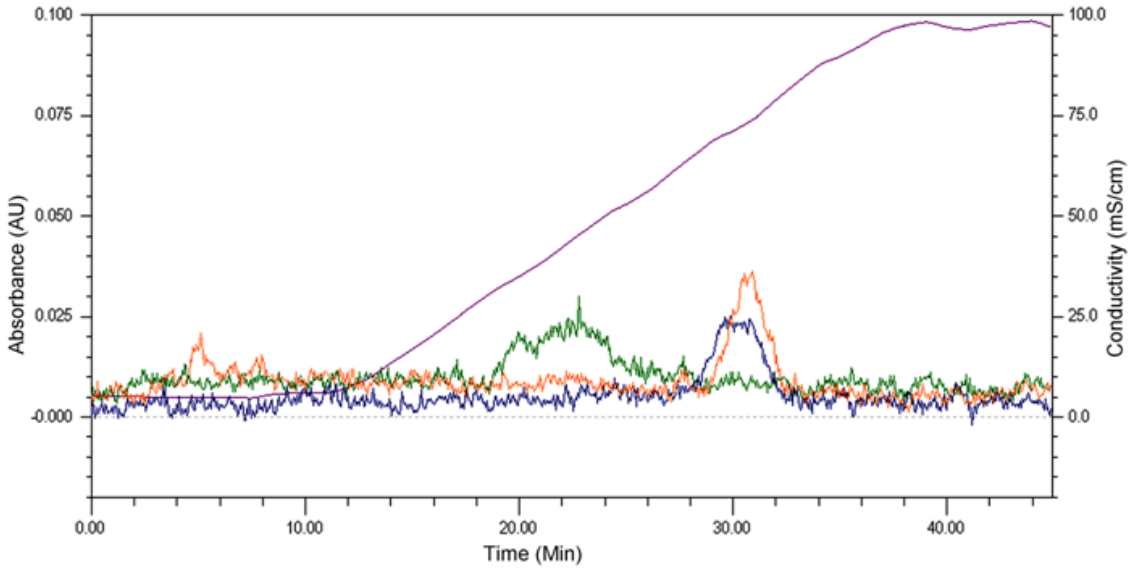

Figure 4.8. Heparin affinity chromatography of Robo1 site-directed mutants. The chromatogram is a representative replicate; the analyses were performed in duplicate and the elution profiles were reproducible. *Orange*, WT Robo1; *green*, Mutant I (R136A, K137A); *blue*, Mutant II (R62A, R69A). The *purple curve* shows the conductivity of the elution buffer (mS/cm) due to the increasing NaCl concentration, with the scale shown on the *righty* axis.

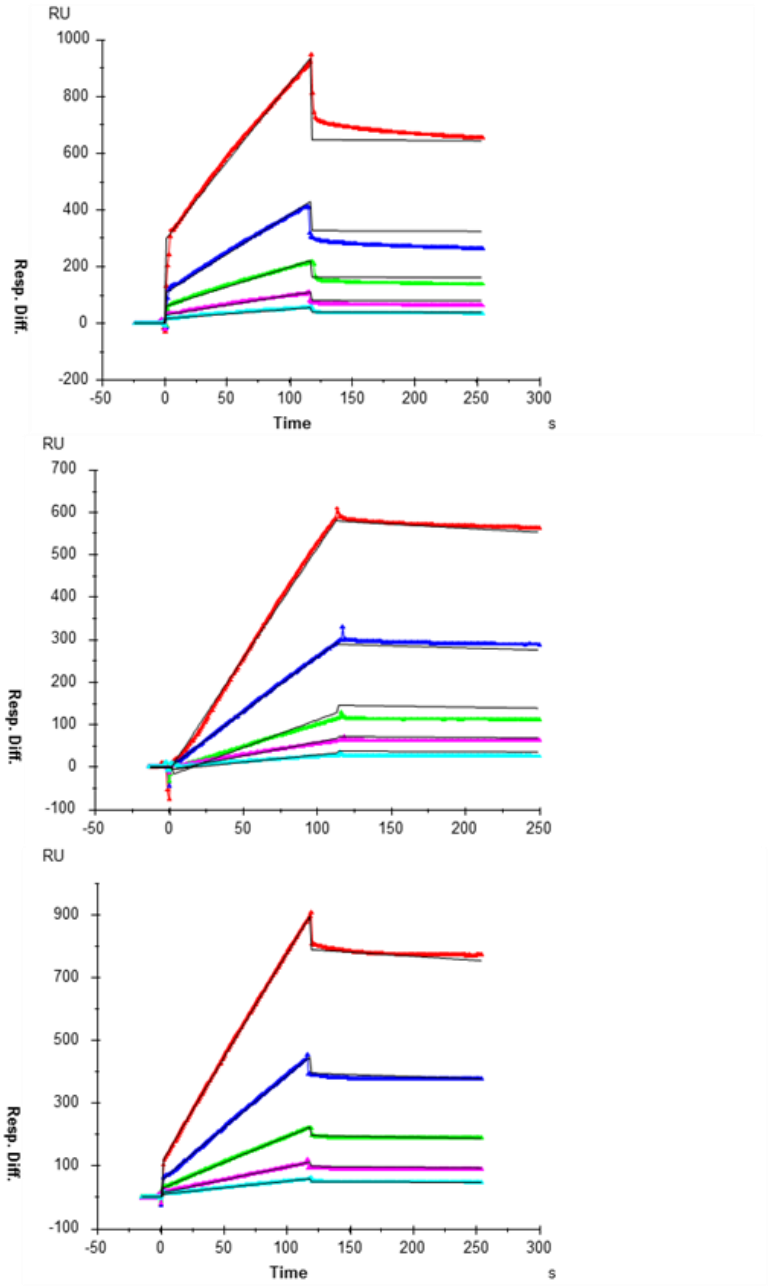

Figure 4.9. SPR sensorgrams of Robo1-heparin interaction. (a) Robo1 WT/heparin; **(b)** Mutant I (R136A, K137A)/heparin; **(c)** Mutant II (R62A, R69A)/heparin. Concentrations of Robo1 (from top to bottom) are as follows: 1000 nM (red), 500 nM (blue), 250 nM (green), 125 nM (magenta), and 63 nM (cyan). The black curves are the fitting curves from a 1:1 Langmuir model from BIAevaluate version 4.0.1. Resp. Diff., response difference.

Table 4.3 Summary of kinetic data of Robo1-heparin interactions*

Interaction	k_a *(1/Ms)*	k_d *(1/s)*	K_D *(M)*
Robo1wt /Heparin	216 (±10)	4.3×10^{-5} ($\pm 5\times 10^{-6}$)	2.0×10^{-7}
Robo1 Mutant I/ Heparin	31 (±2)	3.4×10^{-4} ($\pm 3\times 10^{-5}$)	1.1×10^{-5}
Robo1 Mutant II/ Heparin	44 (±1.4)	3.4×10^{-4} ($\pm 1.5\times 10^{-5}$)	7.8×10^{-6}

*The data with (±) in parentheses are the standard deviations (SD) from global fitting of five different concentration injections.

DISCUSSION

We have carried out a high spatial resolution hydroxyl radical protein footprinting analysis to study the structure of the complex of Robo1 and unfractionated heparin. These HRPF results revealed heparin-induced protection at and around the previously identified site of heparin dp8 interaction, as well as a second binding site containing basic residues near the N-terminus of the Ig1 domain. No protection of basic residues in the Ig2 domain was observed, indicating that heparin does not interact directly with the Ig2 domain. This is consistent with previous studies suggesting no direct interaction between heparin and the Ig2 domain [21].

Based on the results of our HRPF analysis, combined with the previously reported X-ray crystal structures of the human Slit2-Robo1 complex [39], mutagenesis studies of heparin binding to Drosophila Robo1 [21] and human Slit2 [18], and SPR studies of Robo1-heparin/HS interaction [22], we propose a model for Slit2-Robo1-heparin/HS ternary complex formation and signaling shown in **Figure 4.10**. In this model, the initial interaction between Robo1 and heparin/HS occurs at what we propose to be the high-affinity binding site centered around the previously identified residues in the Robo1 Ig1 domain (K81, R131, R136, K137, and R169) and the Slit2 D2 domain (R465, R466, K470, R471, K476, and K479), based on our heparin affinity column and SPR data

for Mutants I and II. This initial binding event serves to stabilize the Robo1-Slit2 ternary complex, and stabilizes the complex with the long heparin/HS chain. A separate binding motif on the heparin-HS chain then interacts with the novel binding site on Robo1 we report here (R62and R69), as well as with conserved basic residues spatially adjacent in Slit2 (R310, K328, K329, R331, and R332). Based on the SPR and heparin affinity column data, we propose this N-terminal binding site to be lower affinity, and serves to enhance and stabilize the overall ternary complex.

In addition to the basic residues in the Ig1 domain, HRPF also found modest heparin-induced protection at C147involved in the core disulfide bond of the Ig1 domain. Interestingly, C147 exhibits protection upon binding, but its disulfide partner C89remains strongly protected in both the heparin-free and heparin-bound states. As the reactive atom in cysteine is the sulfur, this suggests that only the C147side of the disulfide bond is exposed to any substantial degree in heparin-free Robo1. The first β-strand of the Ig1 domain sits atop this disulfide bond, and this strand and the preceding flexible region contains the two residues making up the N-terminal binding site for heparin. We hypothesize that the N-terminal tail and the first β-strand of the Robo1 Ig1 domain are inherently disordered, and heparin binding to the low-affinity binding site consisting of R62and R69stabilizes the β-strand, protecting the C147side of the disulfide bond from solvent. Mutagenesis studies would not observe a strong effect on heparin binding upon mutagenesis of this low-affinity site [18, 21], as elution from the heparin binding column would be dictated primarily by the affinity of the high-affinity binding site. Examination of the low-affinity binding site we propose indicates that, in the unbound state, the side chain orientation of the R62 and R69 residues of Robo1 are not in a geometry that would seem conducive to a shared binding site between Robo1 and Slit2 (**Figure 4.10**). However, our evidence suggests that this

region is probably flexible in the unbound state, allowing for an induced-fit binding of the heparin/HS polysaccharide.

Our results also show a modest protection of hydrophobic residues M180 and M189 in the Ig2 domain upon binding to unfractionated heparin (**Figure 4.7**). M180 is only partially protected from solvent by residues involved in the small C-terminal α-helix of the Ig2 domain, and is heavily oxidized in the absence of heparin. In the static X-ray crystal structure, M189is highly protected from solvent, yet we detect heavy oxidation of this residue as well. These results suggest that M189 may be somewhat more solvent exposed in the absence of heparin than is indicated by the crystal structure. M189 is protected from solvent largely by the β-strand containing M180, suggesting that this strand may be dynamic in solution in the heparin-free state, exposing both M180 and M189.

Previous SPR studies performed on Robo1 Ig1-2 interacting with unfractionated heparin indicated a two-state binding mechanism, including a slow event upon binding to unfractionated heparin that was not present upon binding to heparin octasaccharides [22]. While SPR data presented here were not able to conclusively verify the two-state binding model, the fit of the WT Robo1 to a standard 1:1 Langmuir binding model was not as good as the fit of either Mutant I or Mutant II, indicating that the WT Robo1 may have a more complex binding mechanism than the standard 1:1 model. Our binding model contains a second, low-affinity binding site for full-length heparin/HS chains. Our model suggests that the heparin/HS chain loops back and interacts with a distal binding site near the N-terminus of the Robo1 Ig1 domain and the Slit2 D2 domain, an interaction that would not be present for shorter heparin/HS chains. We additionally find evidence for a modest change in the conformation and/or dynamics of the Ig2 domain upon heparin binding to the Ig1 domain, but no evidence for a large conformational change under these conditions. We hypothesize that this change in conformation/dynamics may be due to heparin/HS binding at the

low affinity binding site, which would explain why previous SPR data found a two-state mechanism of binding only for full-length heparin, not for shorter oligosaccharides [22]. The model also provides a testable hypothesis for the mechanism of Slit-Robo-heparin/HS interaction in signal transduction.

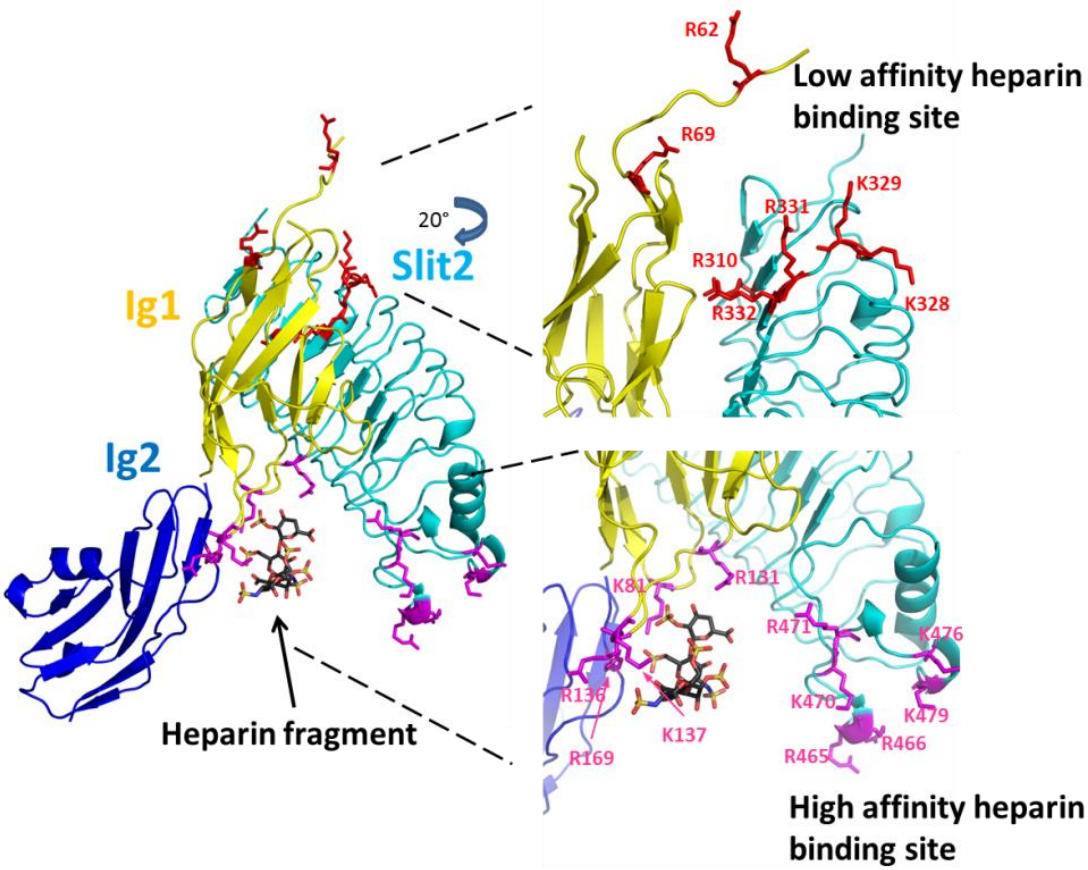

Figure 4.10. A model of the mechanism of Slit2-Robo1-heparin interactions. The Slit2-Robo1 complex has two binding sites for heparin: the previously identified high affinity binding site near the Ig1-Ig2 interface of Robo1 (*magenta*), and a novel low-affinity binding site located near the disordered N-terminus of Robo1, as well as within adjacent conserved basic residues in Slit2 (*cyan*). Full-length heparin/HS binds first to the high affinity binding site, which then allows for binding of a separate portion of the heparin/HS chain to bind the low affinity binding site. The binding to the low-affinity binding site prompts conformational changes required for signal transduction. This model was generated using the X-ray crystal structure of the second LRR domain of human Slit2 in complex with the Ig1 domain of human Robo1 (PDB ID: 2V9T). The heparin tetrasaccharide and the Ig2 domain of *Drosophila* Robo1 were aligned and joined from the X-ray crystal structure of dRobo1 bound to heparin dp8 (PDB ID: 2VRA).

ACKNOWLEDGEMENTS

We thank Professor Lianchun Wang for providing porcine intestinal heparin. We thank Dr. Chau-wen Chou of the Proteomic and Mass Spectrometry Facility at the University of Georgia for the mass spectral data acquisition assistance as well as Professor Lance Wells and Rob Bridger. We would also like to thank Dr. Gregory Wylie and Professor James Prestegard assistance with dynamic light scattering.

REFERENCES

1. Garbe, D.S. and G.J. Bashaw, *Axon guidance at the midline: From mutants to mechanisms.* Critical Reviews in Biochemistry and Molecular Biology, 2004. **39**(5-6): p. 319-341.

2. Dickson, B.J. and G.F. Gilestro, *Regulation of commissural axon pathfinding by slit and its robo receptors.* Annual Review of Cell and Developmental Biology, 2006. **22**: p. 651-675.

3. Evans, T.A. and G.J. Bashaw, *Slit/Robo-mediated axon guidance in Tribolium and Drosophila: Divergent genetic programs build insect nervous systems.* Developmental Biology, 2012. **363**(1): p. 266-278.

4. Kidd, T., et al., *Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors.* Cell, 1998. **92**(2): p. 205-215.

5. Huminiecki, L., et al., *Magic roundabout is a new member of the roundabout receptor family that is endothelial specific and expressed at sites of active angiogenesis.* Genomics, 2002. **79**(4): p. 547-552.

6. Carmeliet, P. and M. Tessier-Lavigne, *Common mechanisms of nerve and blood vessel wiring.* Nature, 2005. **436**(7048): p. 193-200.

7. Gibson, D.A., et al., *Dendrite self-avoidance requires cell-autonomous slit/robo signaling in cerebellar purkinje cells.* Neuron, 2014. **81**(5): p. 1040-56.

8. Li, H.S., et al., *Vertebrate slit, a secreted ligand for the transmembrane protein roundabout, is a repellent for olfactory bulb axons.* Cell, 1999. **96**(6): p. 807-818.

9. Fricke, C., et al., *astray, a zebrafish roundabout homolog required for retinal axon guidance.* Science, 2001. **292**(5516): p. 507-510.

10. Wang, B., et al., *Induction of tumor angiogenesis by Slit-Robo signaling and inhibition of cancer growth by blocking Robo activity.* Cancer Cell, 2003. **4**(1): p. 19-29.

11. Wu, J.Y., et al., *The neuronal repellent Slit inhibits leukocyte chemotaxis induced by chemotactic factors.* Nature, 2001. **410**(6831): p. 948-952.

12. Englund, C., et al., *Attractive and repulsive functions of Slit are mediated by different receptors in the Drosophila trachea.* Development, 2002. **129**(21): p. 4941-4951.

13. Kramer, S.G., et al., *Switching repulsion to attraction: Changing responses to slit during transition in mesoderm migration.* Science, 2001. **292**(5517): p. 737-740.

14. Jen, J.C., et al., *Mutations in a human ROBO gene disrupt hindbrain axon pathway crossing and morphogenesis.* Science, 2004. **304**(5676): p. 1509-1513.

15. Hu, H., *Cell-surface heparan sulfate is involved in the repulsive guidance activities of Slit2 protein.* Nat Neurosci, 2001. **4**(7): p. 695-701.

16. Inatani, M., et al., *Mammalian brain morphogenesis and midline axon guidance require heparan sulfate.* Science, 2003. **302**(5647): p. 1044-1046.

17. Esko, J.D. and U. Lindahl, *Molecular diversity of heparan sulfate.* Journal of Clinical Investigation, 2001. **108**(2): p. 169-173.

18. Hussain, S.A., et al., *A molecular mechanism for the heparan sulfate dependence of Slit-Robo signaling.* Journal of Biological Chemistry, 2006. **281**(51): p. 39693-39698.

19. Capila, I. and R.J. Linhardt, *Heparin - Protein interactions.* Angewandte Chemie-International Edition, 2002. **41**(3): p. 391-412.

20. Bishop, J.R., M. Schuksz, and J.D. Esko, *Heparan sulphate proteoglycans fine-tune mammalian physiology.* Nature, 2007. **446**(7139): p. 1030-1037.

21. Fukuhara, N., et al., *Structural and functional analysis of Slit and heparin binding to immunoglobulin-like domains 1 and 2 of Drosophila Robo.* Journal of Biological Chemistry, 2008. **283**(23): p. 16226-16234.

22. Zhang, F., et al., *Characterization of the interaction between Robo1 and heparin and other glycosaminoglycans.* Biochimie, 2013.

23. Xu, G.H. and M.R. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics.* Chemical Reviews, 2007. **107**(8): p. 3514-3543.

24. Aye, T.T., T.Y. Low, and S.K. Sze, *Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry.* Analytical Chemistry, 2005. **77**(18): p. 5814-5822.

25. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochemical and Biophysical Research Communications, 2001. **287**(3): p. 614-621.

26. Goldsmith, S.C., et al., *Synchrotron protein footprinting: A technique to investigate protein-protein interactions.* Journal of Biomolecular Structure & Dynamics, 2001. **19**(3): p. 405-418.

27. Charvatova, O., et al., *Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1*. Journal of the American Society for Mass Spectrometry, 2008. **19**(11): p. 1692-1705.

28. Maleknia, S.D., M. Brenowitz, and M.R. Chance, *Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry.* Analytical Chemistry, 1999. **71**(18): p. 3965-3973.

29. Gau, B.C., et al., *Fast photochemical oxidation of protein footprints faster than protein unfolding.* Anal Chem, 2009. **81**(16): p. 6563-71.

30. Watson, C., et al., *Pulsed electron beam water radiolysis for submicrosecond hydroxyl radical protein footprinting.* Anal Chem, 2009. **81**(7): p. 2496-505.

31. Li, X., et al., *Improved identification and relative quantification of sites of peptide and protein oxidation for hydroxyl radical footprinting.* J Am Soc Mass Spectrom, 2013. **24**(11): p. 1767-76.

32. Barb, A.W., et al., *NMR characterization of immunoglobulin G Fc glycan motion on enzymatic sialylation.* Biochemistry, 2012. **51**(22): p. 4618-26.

33. Meng, L., et al., *Enzymatic basis for N-glycan sialylation: structure of rat alpha2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation.* Journal of Biological Chemistry, 2013. **288**(48): p. 34680-98.

34. Hambly, D.M. and M.L. Gross, *Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale.* Journal of the American Society for Mass Spectrometry, 2005. **16**(12): p. 2057-2063.

35. Konermann, L., B.B. Stocks, and T. Czarny, *Laminar Flow Effects During Laser-Induced Oxidative Labeling for Protein Structural Studies by Mass Spectrometry.* Analytical Chemistry, 2010. **82**(15): p. 6667-6674.

36. Jumper, C.C., et al., *High-Resolution Mapping of Carbene-Based Protein Footprints.* Analytical Chemistry, 2012. **84**(10): p. 4411-4418.

37. Ortega, A., D. Amoros, and J.G. de la Torre, *Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models.* Biophysical Journal, 2011. **101**(4): p. 892-898.

38. Hernaiz, M., et al., *Enzymatic modification of heparan sulfate on a biochip promotes its interaction with antithrombin III.* Biochem Biophys Res Commun, 2000. **276**(1): p. 292-7.

39. Morlot, C., et al., *Structural insights into the Slit-Robo complex.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(38): p. 14923-14928.

40. Sharp, J.S., J.M. Becker, and R.L. Hettich, *Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry.* Anal Chem, 2004. **76**(3): p. 672-83.

CHAPTER 5

EXTENDED DIMER-DIMER INTERFACE OF CXCL4 REVEALED BY HIGH STRUCTURAL RESOLUTION HYDROXYL RADICAL PROTEIN FOOTPRINTING[5]

[5]Li, Z.; Dyer, D.; Handel, T.; Sharp, J. To be submitted to *Journal of Biological Chemistry*.

ABSTRACT

CXCL4 (also known as platelet factor 4) is a platelet α-granule protein that is identified as the first member of the C-X-C chemokine family. CXCL4 is involved in numerous biological processes including a role in heparin-induced thrombocytopenia that requires tetramer formation, and a role in endothelial cell migration that does not require the tetrameric form. Previous study on the crystal structure of recombinant human platelet factor 4 has shown that tetrameric structure of CXCL4 is stabilized by salt bridges in antiparallel β-sheets region ($E28_A/K50_C$, $E28_C/K50_A$) and hydrogen bonding interactions, but the structure of the dimeric form of CXCL4 is unknown. We utilized ETD-based high spatial resolution hydroxyl radical protein footprinting to study the binding interface of the CXCL4 tetramer by characterizing the structure of wild-type (WT) CXCL4 and its dimeric mutant CXCL4 K50E in solution for the first time. We have identified residues involved in the formation and maintenance of the dimer-dimer interface, including the N-terminal residues ($D7_A/K14_C$, $D7_B/K14_C$, $D7_C/K14_A$, $D7_D/K14_B$) that also contribute to stabilize the tetramer in addition to two previously identified salt bridges. We found no change in protection in any residue distal to the AB:CD dimer-dimer interface. Our model provides a more complete dimer-dimer interface of CXCL4 and its disruption by the K50E mutation.

INTRODUCTION

Platelet factor 4 (PF4, CXCL4) is a 7.8 kDa chemotactic cytokine (chemokine) that is synthesized by bone marrow megakaryocytes, stored in α-granules bound to sulfated glycosaminoglycans (GAGs) and released upon platelet activation [1-3]. CXCL4 is known to be involved in numerous biological processes including hematopoiesis, inhibition of angiogenesis [4, 5], interference with platelet aggregation [6], promotion of host inflammatory responses [7], granulocyte activation and migration [8], solid tumor growth [9], and vascular endothelial cell migration [10]. Many of the biological activities of CXCL4 appear to be related to its interaction with cell-surface glycosaminoglycans via a protein surface binding domain. CXCL4 binds to heparin with high affinity and to other less sulfated glycosaminoglycans with lower affinity roughly in order of their negative charge density [2].When released by activated platelets during coagulation, CXCL4 prevents the formation of ternary complexes between heparin, AT-III, and various pro-coagulant serine proteases [11]. By neutralizing heparin, CXCL4 may inhibit heparin-induced anti-thrombin III activation and permit platelets to bind to endothelial cells, thus supporting the early phases of coagulation [12, 13]. CXCL4 also plays a central role in heparin-induced thrombocytopenia (HIT), an autoimmune disorder caused by the development of large immune complexes against CXCL4 bound to heparin [14, 15].

The CXCL4 monomer is comprised of a glutamate-rich N-terminal aperiodic sequence followed by the characteristic CXC motif of CXC chemokines, the so-called "N-loop", a 3-stranded anti-parallel β-sheets, and C-terminal amphipathic α-helix. The presence of the antiparallel β-sheet structure and two cysteine disulfide bridges impose conformational constraints on CXCL4 backbone folding [16], while a repeating structure of two lysines followed by two hydrophobic amino acid residues near the C-terminus, are thought to be critical for heparin binding

[17]. CXCL4 monomers are known to form asymmetric tetramers at low concentration (50 µg/mL) under physiological conditions [18]. Furthermore, NMR studies indicate that CXCL4 exists in solution in a concentration-dependent mixture of monomers, homodimers and homotetramers [16]. With the individual subunits designated as A, B, C, and D, X-ray structure study of the tetramer of bovine CXCL4 shows that the crystal arrangement of the four monomer subunits can form AB/CD dimers, and AC/BD dimers (**Figure 5.1**) [19], and AC-type dimmers are more thermodynamically favored in solution [16]. Previous studies on the crystal structure of recombinant CXCL4 have shown that the electrostatic interactions of multiply charged amino acid side chains and hydrogen bonding interactions at the AB/CD interface serve to stabilize the tetrameric structure [20]. The most important interactions stabilizing the CXCL4 tetramer appear to involve 2 pairs of salt bridges in chains A and C between E28 and K50 ($E28_A/K50_C$, $E28_C/K50_A$) [20]. Therefore, a CXCL4 K50E mutant was created to disrupt formation of the CXCL4 tetramer, by replacing K50 with the analogous residue in the homologous CXC chemokine IL-8. IL-8 forms dimers but is not known to form tetramers or higher-ordered structures [21]. SDS-PAGE results demonstrate that CXCL4 K50E exists as dimer and monomer around concentration of 10 µg/mL. SEC-HPLC chromatography shows that CXCL4 K50E retained its capacity to bind heparin, but is unable to form tetramers or ultralarge complexes that WT CXCL4 can form with unfractionated heparin over a narrow molar range [22]. Introducing a second complementary mutation in CXCL4 (E28R- K50E) restored tetramer formation and ultralarge complexes to almost the same extent as WT CXCL4 with heparin binding [22]. Given the importance of tetrameric CXCL4 in HIT [14, 15], understanding the structural process involved in tetramer formation, as well as the structural properties of the dimeric form of CXCL4 is essential to designing treatments for HIT. Detailed structural characterization of the CXCL4 K50E mutant has never been reported, limiting its

usefulness in studying CXCL4 structure-function relationships. The primary dimer-dimer interface disrupted by the K50E mutation has not been verified, and the conformation of the dimers has not been characterized. A deeper understanding of characteristics of K50E mutant regarding its stability and dynamics as a dimer provides insights to the structural organization and nature of subunit interaction for CXCL4 tetramers in solution, and is fundamental to a complete understanding of the mechanism of action of this versatile binding protein.

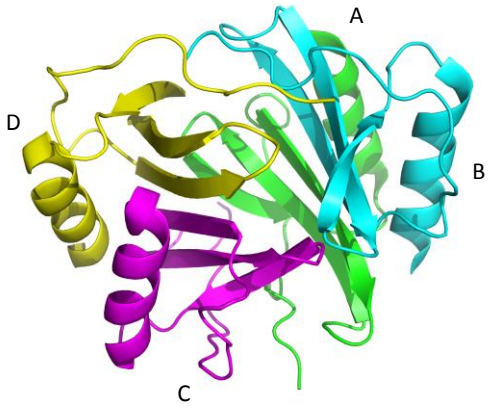

Figure 5.1. Crystal structure model of CXCL4 tetramer (PDB ID: 1RHP). Each color represents a CXCL4 monomer designated as A, B, C, and D with residue numbers 7-70. Each monomer contains one α-helical and three anti-parallel β-strands. Dimers AB/CD are maintained by hydrogen bonding of the β-sheet regions in an antiparallel fashion. Dimers AC/BD are maintained by hydrogen bonding of β-sheet like structures of N-termini in random-coil region [18].

Hydroxyl radical protein footprinting (HRPF) is a relatively recently developed covalent labeling approach that, when coupled with mass spectrometry, can be used to probe molecular interaction interfaces [23-25]. In HRPF, amino acid side chains are oxidized by hydroxyl radicals at a rate based on a combination of their inherent reactivity (based on protein sequence) and accessibility to hydroxyl radicals in solution (based on protein topology) [24]. By comparing the

same protein under different structural conditions, changes in the footprint of a side chain can be directly related to changes in the solvent accessibility of that side chain [26, 27]. The advantages of HRPF labeling include a fast time frame for protein modification, high labeling efficiency and stability of the labeled analyte [24]. HRPF has been used in various studies to probe intact protein structures [28], protein conformational changes [29], protein-protein interactions [30], protein-ligand binding surfaces in solution [27], and particularly difficult systems such as large, heterogeneous protein complexes, membrane proteins, and flexible proteins [31]. In HRPF, the protein of interest is exposed to hydroxyl radicals which react with solvent accessible side chains [32]. Due to the recent development of ultra-fast labeling processes including Fast Photochemical Oxidation of Proteins (FPOP) [28] and electron accelerator radiolysis [33] that are able to complete side chain oxidation faster than the protein can undergo labeling-induced conformational changes, researchers can now heavily label proteins while still ensuring that the HRPF is probing native structure [34]. After HRPF has been used to take a stable chemical "snapshot" of the protein surface, the protein sample can be denatured, proteolytically digested, and the proteolytic peptides are analyzed by liquid chromatography-mass spectrometry (LC-MS). The hydroxyl radical footprint of the protein is determined by comparing the amount of oxidation of a given peptide under two different structural states. Changes in oxidation of a given region of a protein under identical effective radical doses are due to changes in the solvent accessible surface area of that protein. Until very recently, HRPF quantitation was performed at the peptide level but the structural resolution provided by such quantification is low. High spatial resolution HRPF methods were therefore developed to enable quantification of changes in solvent accessibility at the amino acid level [31, 35-38].

In previous work, we demonstrated the accuracy of electron transfer dissociation (ETD) -based methods for quantifying multiple adjacent sites of isomeric oxidation products [35, 36], and successfully applied this high spatial resolution HRPF method to the identification of a protein-GAG complex binding interface [31]. Here, we report the use of this technology to characterize the dimer-dimer interface of WT CXCL4, as well as the structural consequences of the disruption of this tetramer in the K50E mutant, which forms a dimer under physiological conditions. The use of ETD-based high resolution HRPF has allowed the acquisition of footprinting data at the individual amino acid level in most cases for CXCL4, allowing for identification of specific amino acid residues impacted by disruption of the tetramer. We have additionally identified a structural role for the N-terminus in the dimer-dimer interaction, which was previously unreported.

EXPERIMENTAL PROCEDURES

Protein expression and purification

Wild type CXCL4 and the mutant CXCL4 K50E were expressed and purified using a similar method to that described previously [39]. CXCL4 was subcloned into the His-Ub fusion construct (pHUE 3D3) and the CXCL4 K50E mutant was produced with QuickChange site-directed mutagenesis (Stratagene). Both were expressed insolubly following transformation into BL21 (DE3) pLysS Escherichia coli cells, cultured in Luria broth containing 30 mg/ml kanamycin, at 37˚C. Cultures were induced when OD_{600nm} reached 0.6, using 0.5 mM isopropyl-D-1-thiogalactopyranoside, and subsequently cell pellets were collected by centrifugation, before solubilization with denaturing buffer (6M guanidine-HCl, 50 mM NaCl, 50 mM Tris, pH 8.0). Tagged protein was then extracted by Nickel-nitrilotriacetic acid (NTA) affinity chromotography and refolded in 0.5 M Arginine, 1 mM GSSG, 1 mM EDTA, 50 mM Tris, pH 8.0 (confirmed by reversed-phase high-pressure liquid chromatography (HPLC)). CXCL4 and CXCL4 K50E were

then cleaved using ubiquitinase at a 1:100 molar ratio (chemokine:enzyme) for 24 hours, progression of cleavage was monitored using HPLC. Cleavage products were applied to Ni-NTA beads, in order to remove unrequired fragments, and final purification was achieved using a C18 semi-prep column and HPLC. Chemokine size and purity was finally confirmed with SDS-PAGE and electrospray-ionization mass spectrometry analysis.

HRPF of proteins

HRPF of WT CXCL4 and the mutant CXCL4 K50E were performed by FPOP as previously described [29, 32]. Briefly, an 18μL solution of 2μM WT or mutant protein, 20mM glutamine (a hydroxyl radical scavenger), 1mM adenine (a hydroxyl radical dosimeter) and nanopure water were mixed. 2 μL of hydrogen peroxide was added to the 18 μL protein sample to a final concentration of 100 mM hydrogen peroxide immediately prior to irradiation by excimer laser. The protein samples were loaded into a 100 μL syringe and introduced via a syringe pump coupled to 100 μmi.d. fused silica tubing running through the beam path of the excimer laser, with a UV-transparent window burned in the tubing in the beam path. Samples flowed through the beam path of an EX100 KrF excimer laser (GAM Laser Inc., Orlando FL, USA) at a laser energyof 95 mJ/pulse. The flow rate of the syringe pump was set at 12.19 μL/min, with the laser pulse repetition rate set such that each segment of protein sample was irradiated with a single ~20 ns UV pulse leaving an unirradiated buffer region between irradiated segments to help account for sample diffusion and laminar flow [29]. Samples were collected in a microcentrifuge tube containing methionine amide (0.5 μg/μL) and catalase (0.5 μg/μL) with sodium phosphate (50 mM, pH 7.4) to immediately quench oxidation by destroying excess hydrogen peroxide, protein peroxides and other reactive oxygen species [40]. Negative control samples of WT and K50E mutant CXCL4 were prepared under identical conditions, except the laser was not turned on. All samples had an

adenine dosimeter absorbance reading at 260 nm of 0.451± 0.019, indicating the proteins received equivalent doses of hydroxyl radicals.

Following FPOP, each sample was lyophilized and resuspended in 20 µL of 50mM sodium phosphate. DTT was added to a final concentration of 5mM and samples were incubated at 65°C for 30 min to reduce disulfide bonds. After cooling to room temperature, samples were digested overnight with trypsin at 37˚C at an enzyme: protein ratio of 1:20. The digested samples were stored at -20°C for LC-MS/MS analysis.

LC-MS/MS acquisition and data analysis

Samples were analyzed on a LTQ Orbitrap Elite (Thermo Fisher Scientific) controlled by Xcalibur 2.0.7 software (Thermo Fisher, San Jose, CA). Samples were loaded on a 75µm (i.d.150 mm) C18 Acclaim PepMap® RSLC nanocolumn (Thermo Fisher Scientific). Samples were eluted directly into the nanospray source of the LTQ Orbitrap Elite using a 61 min linear gradient consisting of 4-100% acetonitrile with a constant concentration of 0.1% formic acid at a flow rate of 0.3 µL/min. The spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to n. Full MS scans were acquired from m/z 200 to 2000 at a nominal resolution of 120,000 using the FT mass analyzer. To obtain MS/MS data of oxidized peptides, ions were fragmented by ETD with the corresponding parent ion mass list. ETD-based precursor activation was carried out for 100ms, including charge-state-dependent supplemental activation. Precursor ions were isolated with a width of 2m/z units.

Unoxidized CXCL4 and CXCL4 K50E controls, oxidized CXCL4 and CXCL4 K50E peptide sequences were initially identified using ByOnic V1.2-250 (Protein Metrics, CA, USA). All tandem mass spectra assignments and sites of oxidation were verified manually due to peptide modification complexity. The tryptic digested peptides and corresponding oxidation products were

identified from the LC-MS runs manually to calculate the average oxidation events per peptide in the sample. Peptide level quantitation of the average oxidation events per peptide, residue level quantitation and the absolute level of oxidation for a given amino acid residue in whole protein were performed using following equations respectively in a similar approach based on our previous study [31, 35].

$$P = \frac{\mathrm{I(+16)oxidized}*1+\mathrm{I(+32)oxidized}*2+\mathrm{I(+48)oxidized}*3+\cdots}{\mathrm{Iunoxidized}+\mathrm{I(+16)oxidized}+\mathrm{I(+32)oxidized}+\mathrm{I(+48)oxidized}+\cdots} (1)$$

where P denotes the average oxidation events per peptide and I donates to the intensities of oxidized and unoxidized peptides.

$$f(Ci) = \frac{I(Ci)oxidized}{I(Ci)unoxidized+I(Ci)oxidized} (2)$$

Where $f(Ci)$ denotes the fractional oxidization of sequence ion i (e.g., oxidized c3 ions generated by ETD), and $I(Ci)$ denotes the intensity of the oxidized and unoxidized ion of interest (e.g. oxidized c3 ion plus unoxidized c3 ion). The absolute level of oxidation for a given amino acid residue i is based on both the average oxidation event of peptide and the fractional oxidiation of the corresponding sequence ions, as shown in eq 3:

$$\mathrm{oxidation/residue}\ i = \mathrm{P}[f(Ci) - f(Ci-1)] (3)$$

where P is the average oxidation events per peptide as derived from eq 1, and the term in square brackets is the fractional oxidation difference of two adjacent sequence ions $Ci$and Ci -1. In cases where ETD product ions could not differentiate between Ci and Ci -1, the next sequence ion available was used for quantitation, and the length of the quantifiable segment was noted in the results.

RESULTS

To investigate the structure of CXCL4 K50E and dimer-dimer interface of WT CXCL4 at high spatial resolution, the two protein samples were aliquoted into three replicates, and each sample was oxidized for HRPF by Fast Photochemical Oxidation of Proteins (FPOP). As the protein is exposed to short bursts of hydroxyl radicals, the radicals react with the amino acid side chains with an apparent rate that is a function of the amino acid sequence (which is constant between samples) and the exposure of the side chain to the solvent, which changes at the binding interface between dimer and tetramer. Immediately after FPOP oxidation, the samples were quenched to eliminate excess hydrogen peroxide and any other long-lived secondary oxidants, digested with trypsin, and analyzed by LC-MS/MS using ETD.

High spatial resolution HRPF of CXCL4 and CXCL4 K50E

Both CXCL4 and CXCL4 K50E consist of 70 amino acids. To study the dimer-dimer interface of WT CXCL4, WT CXCL4 and CXCL4 K50E samples were oxidized under the same experimental conditions, and peptide level quantitation (low resolution) of oxidation for WT CXCL4 and CXCL4 K50E was firstly performed. In our HRPF studies, we were able to obtain 94% sequence coverage, missing only amino acid K62 and the short peptide 47NGR49 (**Figure 5.2**). Peptide 1-14, 32-46, 50-61 show significant reductions in oxidation in WT CXCL4 compared to the mutant, which indicates more protection/less solvent exposure in the case of the WT protein (**Figure 5.3**). Peptide 23-31 and 63-66 showed no significant change in the amount of oxidation, while peptide 15-22 and 66-70 showed no oxidation in the WT or mutant protein. To achieve high spatial resolution of HRPF, residue level quantitation was conducted by measuring the intensity of unoxidized and oxidized c and z ions generated by ETD and calculating the relative average

oxidation for each residue as described in Materials and Methods [31]. Segments produced by residue level quantitation include single residue and residue stretch due to insufficient ETD fragmentation ions (**Figure 5.2**). Residues exhibiting statistically significant differences in the amount of oxidation between WT and mutant ($p \leqslant 0.05$) are shown in **Table 5.1**. The amount of oxidation of CXCL4 and CXCL4 K50E and protection for these residues are also shown. As shown in **Table 5.1**, five segments show statistically significant protection upon dimer formation: D7, C10, 28EVIK31, C36 and C52. Five segments were detected that show no protection upon tetramer formation. Other segments displayed no detectable oxidation in dimer or tetramer form.

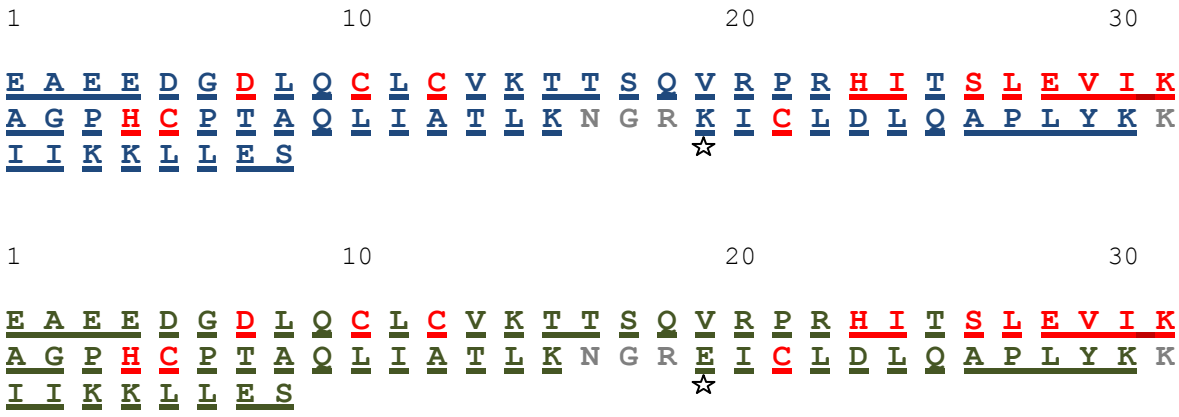

Figure 5.2. Amino acid sequence of CXCL4 (blue) and CXCL4 K50E (green). Identified oxidized amino acids were colored in red. Unoxidized residues in CXCL4 are colored blue; unoxidized residues in CXCL4 K50E are colored green. Unidentified residues are colored grey in CXCL4 and CXCL4 K50E. Mutation is indicated in star. 51 quantitative segments in HRPF samples are shown in separate underline.

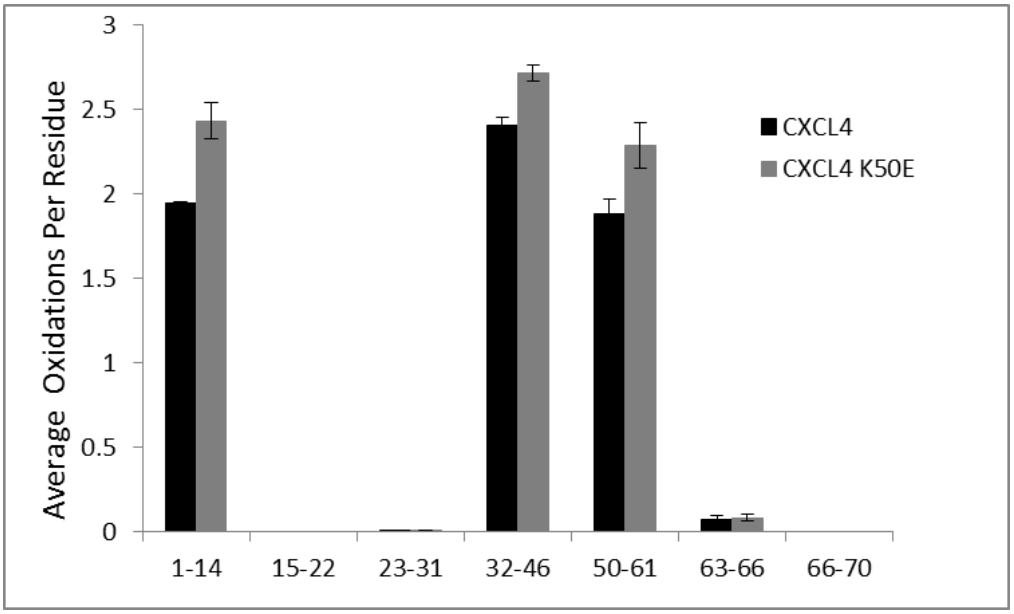

Figure 5.3. Hydroxyl radical protein footprinting of CXCL4 and CXCL4 K50E. Extent of oxidation of CXCL4 and CXCL4 K50E at peptide level is shown. Error bars represent ±1 SD from a triplicate set of FPOP oxidations and analyses.

Table 1. Residues that show significant protection upon tetramer formation

Protected Residues	Avg. Oxid./Residue CXCL4 K50E (±SD)	Avg. Oxid./Residue CXCL4 (±SD)	*p*-value	%Protection upon tetramer formation
D7	2.437 ± 0.231	0.110 ± 0.016	0.02	95.5
C10	2.181 ± 0.073	1.783 ± 0.033	<0.01	18.2
E28V29I30K31	0.005 ± 0.001	0.003 ± 0.001	0.03	32.2
C36	2.389 ± 0.051	1.725 ± 0.057	<0.01	27.8
C52	1.869 ± 0.142	1.371 ± 0.057	0.04	26.6

The oxidation of CXCL4 and CXCL4 K50E at residue level are shown as the average number of oxidation events per residue. Standard deviations were calculated from triplicate samples.

Identification of dimer-dimer interface in WT CXCL4

Previous reports of the crystal structure of human CXCL4 indicate that the two dimmers (AB/CD) associate with each other in the tetramer through surface interactions of the β-sheets with

the two pairs of α-helices located on opposite sides of the surface of the tetramer. The tetrameric structure of CXCL4 is stabilized by salt bridge interactions between dimers AB/CD. Tetramers can also be formed between AC and BD dimmers by β-sheet-type hydrogen bond interactions in the N-terminal [20]. In our study, five segments were identified showing statistically significant protection upon dimer-dimer association. Three residues were oxidized in peptide 1-14. The average number of oxidation events per residue for D7 decreases significantly showing a 95.5% reduction in oxidation between WT and mutant samples, indicating strong protection upon tetramer formation (**Figure 5.4**). A +48 Da oxidation product is observed in the product-ion ETD spectrum shown for this peptide in Figure S1. Aspartic acid residues contain acidic functional groups that normally react with the hydroxyl radical to lose CO2, resulting in an aldehyde with an overall reductionin mass of 30 Da, or show a gain of a single oxygen [25]. The observed increase of 48 Da of this aspartic acid residue is unusual, as is the observed high reactivity of this residue. However, the tandem mass spectra as exemplified in **Figure S5.1** are unambiguous. Additionally, the reduction in this product ion formation upon tetramerization is dramatic. As the protein is fully digested prior to MS/MS analysis, if the unusual +48 Da oxidation product of D7 was due to an unexpected gas-phase rearrangement of a cysteine-based oxidation product, we would expect the rearrangement in the ETD MS/MS spectrum to be similar, regardless of the solution-phase structure of the protein. Figure 4a clearly shows that, while the cysteine oxidation products show little or no reduction upon tetramer formation, the +48 of D7 is drastically reduced in the tetramer, eliminating the possibility of the +48 product of D7 oxidation being a gas-phase rearrangement of a cysteine oxidation product. D7 is the first resolved residue at the N-terminus of the protein in the crystal structure [20]. The crystal structure failed to reveal the involvement of this residue in the tetramerization process, probably due to dynamics of the N-terminal tail preventing resolution of

the N-terminal structure. However, it's suggested by HRPF results that D7 is likely to be involved in dimer-dimer interface in the full-length N-terminus structure in solution. Oxidation of residue C10 was also observed to decrease from 2.181 to 1.783. However, C12 in this peptide didn't show a significant change in oxidation extent upon tetramer formation (**Figure 5.4**).

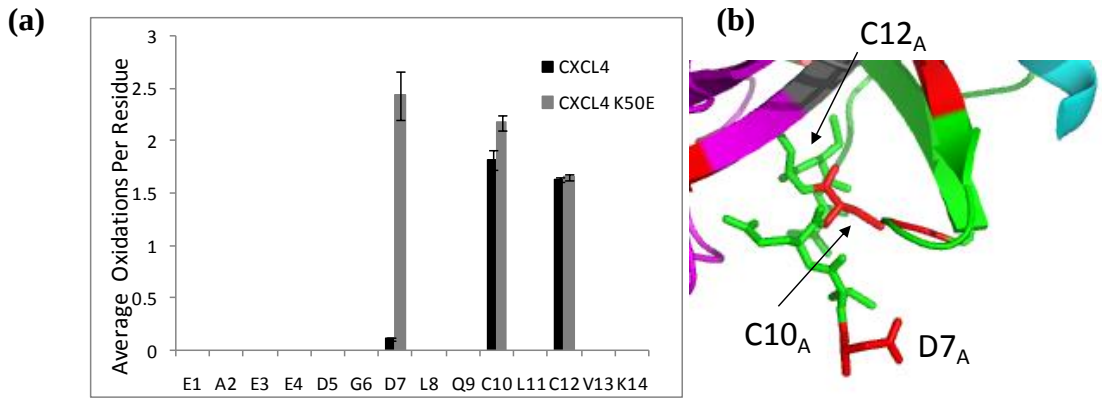

Figure 5.4.Residues involved in dimer-dimer interface by electrostatic interactions.(a) Extent of oxidation of CXCL4 compared to CXCL4 K50E at the residue level. Error bars represent±1 SD from a triplicate set of experiments. **(b)** Protected residues D7 is in the N-terminus of the sequence and the first residue in the crystal structure (PDB ID: 1RHP).

Peptide 15-22 is in the random-coil region of the crystal structure, and consists almost entirely of small aliphatic and polar residues which have poor reactivity to hydroxyl radicals. No residues in this region were found oxidized, probably due to the poor inherent reactivity of these residues compared to the very high reactivity of the adjacent cysteine residues, leading to the cysteine residues rapidly scavenging the radicals in the region.

Multiple sites of oxidation were identified in peptide 23-31, although the oxidation extent of this peptide is relatively low compared to others. Residues 28EVIK31 exhibits significant protection upon dimer-dimer association (**Figure 5.5**). A representative product-ion ETD spectrum

for singly-oxidized peptide 23-31 is shown in **Figure S5.2**. As we do not have sufficient ETD fragment ions to differentiate oxidation of these four amino acids, we must quantitate oxidation of this four amino acid stretch as a whole. Negatively charged residue E28A is adjacent to positively charged residue K250C shown in **Figure S5.5b**. Previous study show that E28A/K250C and E28C/K50A are involved in salt bridge pairs and contribute to the dimer-dimer interface [20]. The result of protection in the tetramer form and exposure in the K50E mutant by HRPF is consistent with previous study [20]. Other residues H23, I24, S26 and L27 in this peptide show no significant change of solvent accessibility indicating there's not a substantial conformational change occurring in the peptide upon disruption of the tetramer.

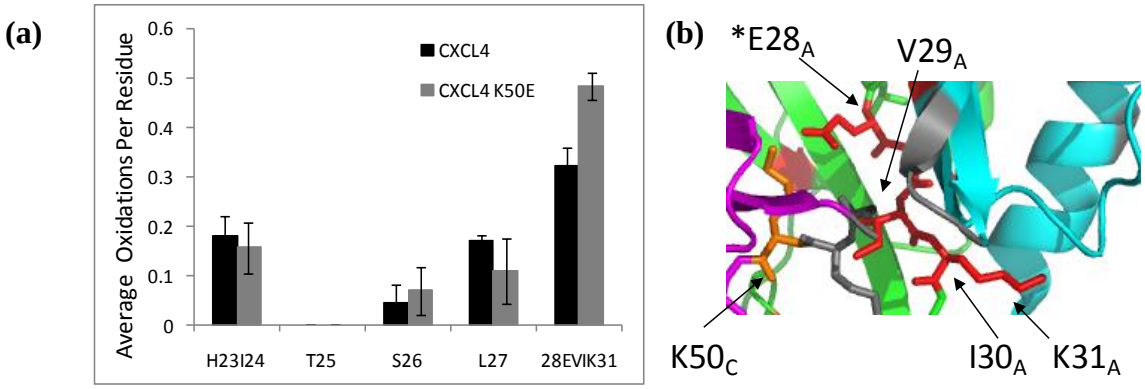

Figure 5.5. Residues involved in dimer-dimer interface by electrostatic interactions.**(a)** Extent of oxidation of CXCL4 compared to CXCL4 K50E at the residue level. Error bars represent±1 SD from a triplicate set of experiments. **(b)**Structure shows the protected residues^{28}EVIK^{31} are in the area of binding interface. Asterisk indicates the E28 residue identified as interacting with K250 in orange previously.

Residue C36 and C52 in peptide 32-46 and 50-61 show significant protection from 2.389 to 0.051, and 1.869 to 0.142, respectively (**Figure 5.6**). Previous report indicates that C10 and C36, C12 and C52 formed two disulfide bonds to stabilize the monomer random-coil region [20].

Significant protection of C10 and its bond partner C36 suggests that there is a stable structure that is disrupted in the K50E mutant. Interestingly, all of these cysteine residues forming disulfide bonds exhibit significant change in solvent accessibility except C12 (**Figure 5.4**). C12, the disulfide bond partner of C52, shows no significant protection upon association.

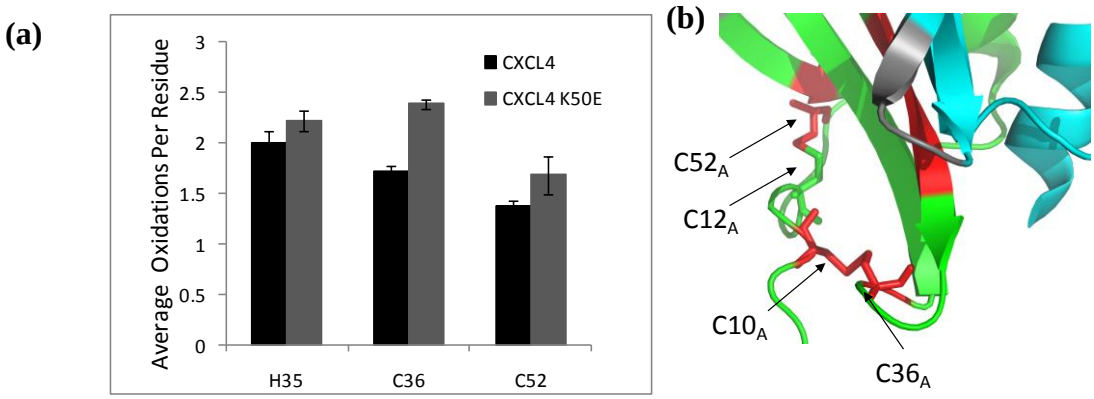

Figure 5. 6. Residues not involved in tetramer association show significant protections. (a) Residue-level extent of oxidation for residues H35, C36 and C52. Error bars represent±1 SD from a triplicate set of experiments. **(b)**Structural model showing C10 involved in the disulfide bond with C36, and C12 involved in disulfide bond with C52. C12 shows no significant change in solvent accessibility.

DISCUSSION

We have carried out a high spatial resolution hydroxyl radical protein footprinting analysis to study the dimer-dimer association interface of human CXCL4.These HRPF results revealed that residues exposed in the dimer while protected in tetramer form are mainly near the N-terminus (D7, C10, C36), as well as involved in electrostatic interactions in antiparallel β-sheet region (E28) (**Figure 5.7a**). The HRPF result indicates two dimers associate with each other through surface interactions of the N-terminus coil region and six-strand antiparallel β-sheets region. No residues

on the α-helices were found with significant protections upon dimer-dimer association, exposing an equatorial ring of positively charged amino acids derived from the carboxy-terminus α-helices to attract heparin molecules, which is consistent with previous reports [17].

Previous studies on the crystal structure of CXCL4 indicate that residues from N-terminus previously defined as a random coil or extended loop regions actually possess a very defined structural conformation due to interactions of N-terminal residues between the A to C and B to D dimers, respectively. These antiparallel β-sheet-like structures are positioned lateral to the β-bilayer motif and stabilize the tetrameric unit. Residues in the N-terminus form hydrogen bonds and contribute significant energy to the interaction between dimers AC, BD forming the ultimate tetramer structure. D7 was reported to be involved in β-sheet-type hydrogen bonds in crystal structure [20]. However, the dramatic change of oxidation extent in HRPF indicates that this residue is heavily exposed in dimer form. In the crystal structure, D7 is at the very end of N-terminus. It is possible that full-length N-terminal structure of CXCL4 in solution has D7 involved in the dimer-dimer interface. Negatively charged D7A is adjacent to positively charged K14C and able to form electrostatic interactions (**Figure 5.7b**). However, residue K14 failed to provide any information as it was not oxidized in HRPF samples. The strong protection of D7, along with significant protection of C10 and its disulfide partner C36 suggests that there is a stable structure in this region in the tetramer that is disrupted in the K50E mutant.

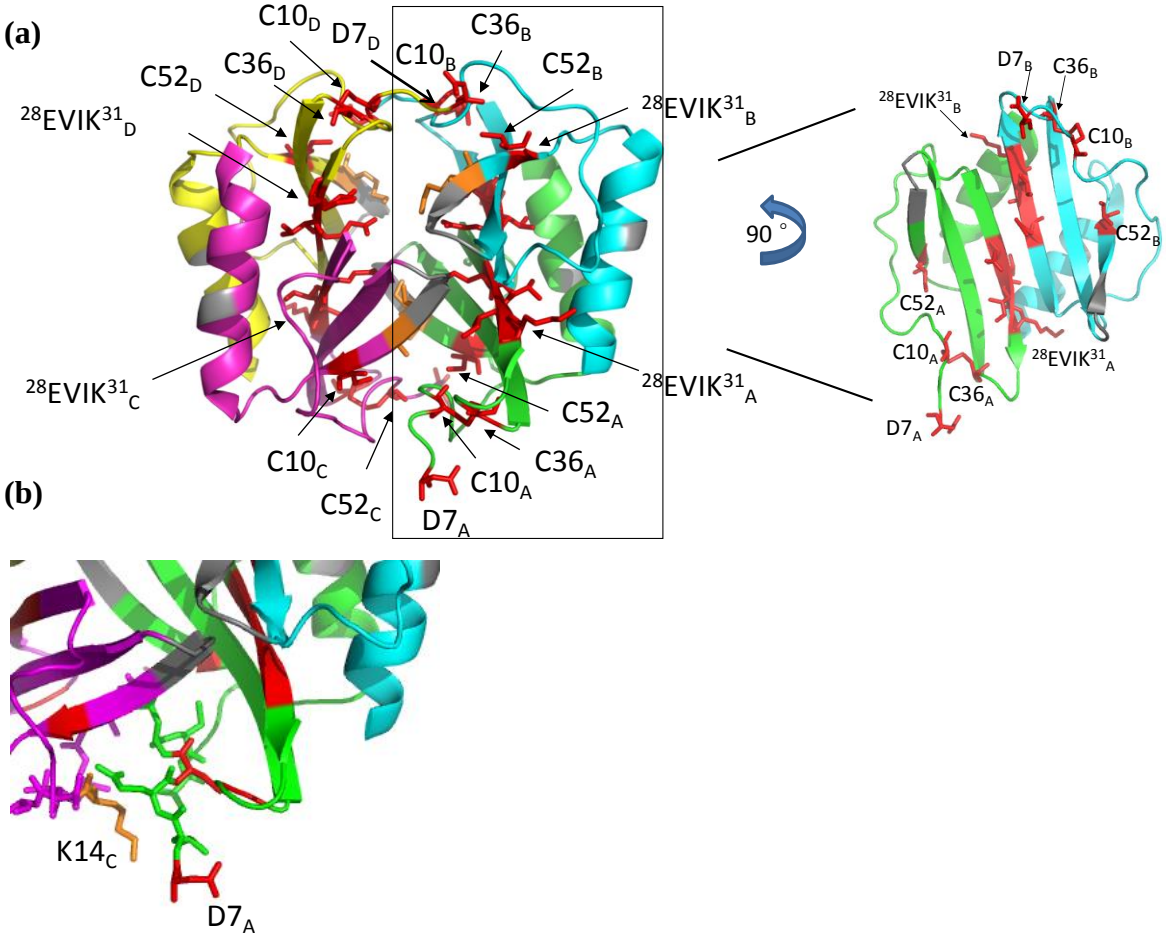

Figure 5.7. Model of dimer-dimer interface of CXCL4.(a) Dimers associate with each other through surface interactions of the N-terminal coil region and antiparallel β-sheets. Residues with significant protection upon tetramer formation are shown in red. The tetramer interface on the A and B molecules are shown on the right to provide a cutaway view for dimer-dimer association (PDB ID: 1RHP). **(b)**Negatively charged residue $D7_A$ (colored in red) in the N-terminus interacts with positively charged $K14_C$ (colored in orange), which was not oxidized by HRPF.

Residue E28 is in a salt bridge with K50, and mutation of K50 disrupts this salt bridge and results in exposure of E28, which is confirmed by our HRPF results. A lack of change in the solvent

accessibility of H23, I24, S26 and L27 clearly show that there is not a substantial conformational change occurring in the peptide upon disruption of the tetramer, indicating that the overall core dimer structure remains unperturbed in the K50E mutant. C10, C36 and C52 exhibit significant protection upon tetramer formation, with the most extensive protection occurring to C36. Cysteine residues are highly reactive to hydroxyl radicals, and can be significantly oxidized even at low levels of solvent exposure. Similarly, substantial changes in the hydroxyl radical footprint can be observed even with small changes in the solvent accessible surface area of these residues [27]. It should be noted that C12, the disulfide bond partner of C52, remains equally exposed in both the dimer and tetramer form. As the reactive atom in cysteine is the sulfur, this suggests that only the C52 side of the disulfide is protected in tetramer. The protection of only one member of a disulfide bond has been observed previously when the unaffected cysteine remains buried in the protein structure [31].

The CXCL4 K50E mutant was characterized structurally by HRPF in this study for the first time, providing detailed structural information to study the dimer-dimer interface of CXCL4. By comparing residues with significant change of solvent accessibility, we have proposed a model of the dimer-dimer interface of CXCL4 and its disruption by the K50E mutation. Our model suggests that forces to stabilize tetramer not only consist of two salt bridges pairs in antiparallel β-sheets region (E28A/K50C, E28C/K50A), but also electrostatic interactions from N-terminal residues (D7A/K14C, D7B/K14D, D7C/K14A, D7D/K14B). This model provides a more complete dimer-dimer association interface than crystal structure, as well as a deeper understanding of CXCL4 tetramer structure in solution.

ACKNOWLEDGEMENT

This research was supported by a grant from National Institute of General Medical Sciences as part of the Research Resource for Integrated Glycotechnology (P41GM103390).

SUPPLEMENTARY INFORMATION

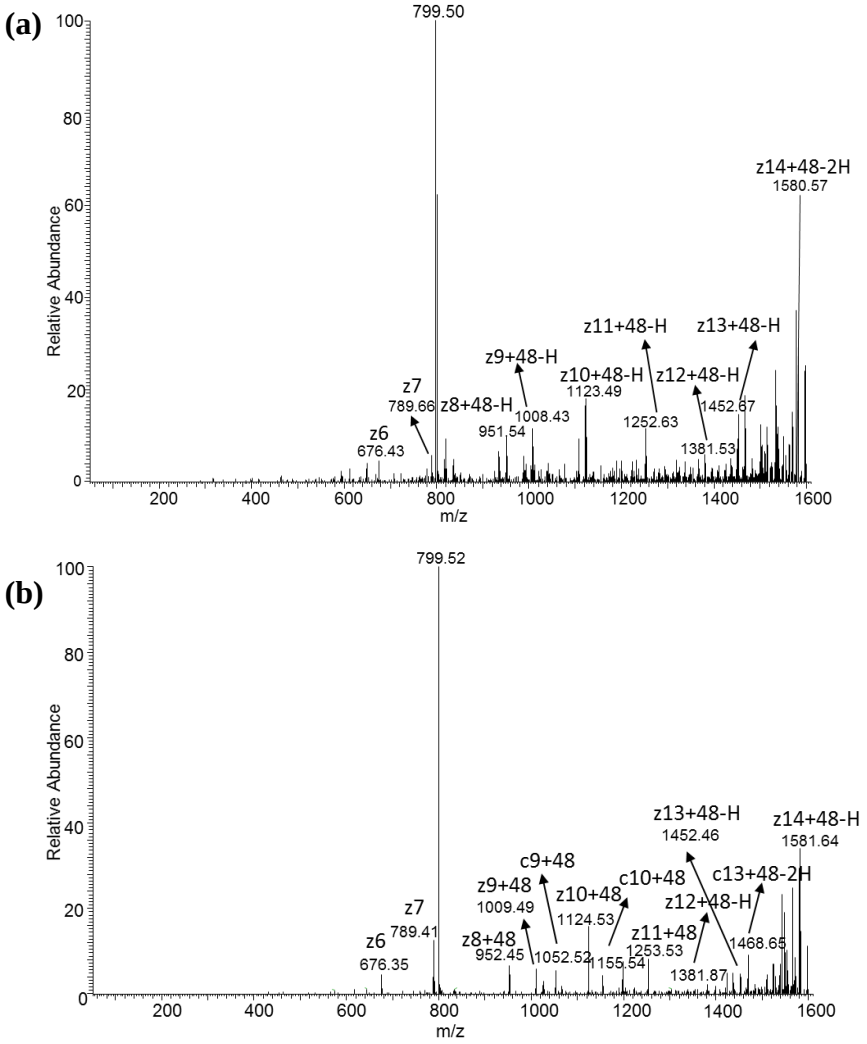

Figure S5.1. ETD spectrum of peptide EAEEDGD(+48)LQCLCVK for (a) CXCL4 and (b) CXCL4 K50E. The ETD spectrum shows unoxidizedz6, z7 and oxidized (+48Da) z ions from z8 to z14 while unoxidizedz1-z5 ions are absent in the spectrum, indicating of oxidation on D7. Oxidized c9, c10, c13 ions generated by ETD are also shown in the spectrum of PF4 K50E.

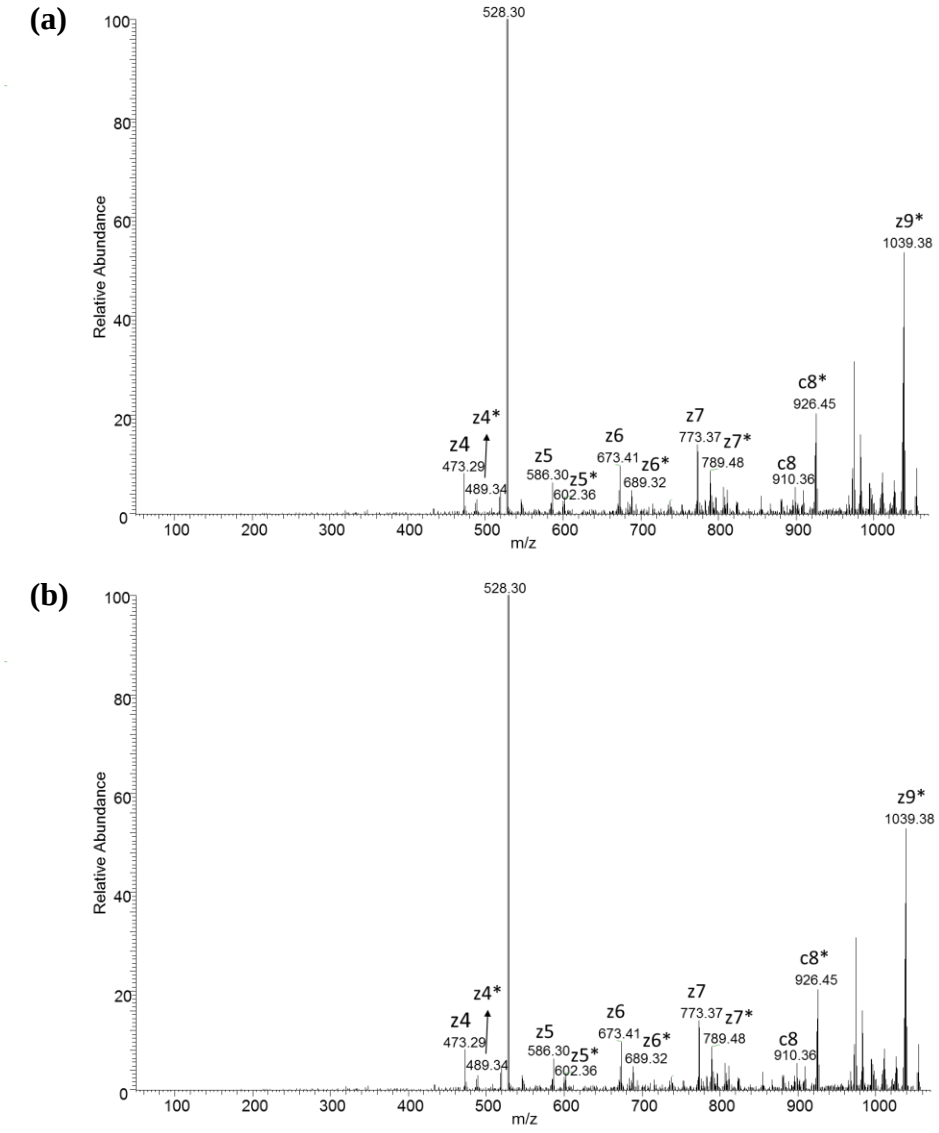

Figure S5.2. Representative ETD spectrum of singly-oxidized peptide HITSLEVIK for (a) CXCL4 and (b) CXCL44 K50E. The asterisks indicate the product ions that are oxidized (+16Da). The ETD spectrum shows unoxidized and oxidized c ions as pairs from z4 to z7 while unoxidized and oxidized z1, z2, z3 and z8 ions are absent in the spectrum. Unoxidized and oxidized c8 ions generated by ETD are also shown in the spectrum.

REFERENCES

1. Broekman, M.J., R.I. Handin, and P. Cohen, *Distribution of fibrinogen, and platelet factors 4 and XIII in subcellular fractions of human platelets.* Br J Haematol, 1975. **31**(1): p. 51-5.

2. Handin, R.I. and H.J. Cohen, *Purification and binding properties of human platelet factor four.* J Biol Chem, 1976. **251**(14): p. 4273-82.

3. Wolpe, S.D. and A. Cerami, *Macrophage inflammatory proteins 1 and 2: members of a novel superfamily of cytokines.* FASEB J, 1989. **3**(14): p. 2565-73.

4. Han, Z.C., et al., *Platelet factor 4 inhibits human megakaryocytopoiesis in vitro.* Blood, 1990. **75**(6): p. 1234-9.

5. Maione, T.E., et al., *Inhibition of angiogenesis by recombinant human platelet factor-4 and related peptides.* Science, 1990. **247**(4938): p. 77-9.

6. Conley, C.L., R.C. Hartmann, and J.S. Lalley, *The relationship of heparin activity to platelet concentration.* Proc Soc Exp Biol Med, 1948. **69**(2): p. 284-7.

7. Broxmeyer, H.E., et al., *Comparative-Analysis of the Human Macrophage Inflammatory Protein Family of Cytokines (Chemokines) on Proliferation of Human Myeloid Progenitor Cells - Interacting Effects Involving Suppression, Synergistic Suppression, and Blocking of Suppression.* Journal of Immunology, 1993. **150**(8): p. 3448-3458.

8. Bebawy, S.T., et al., *In vitro effects of platelet factor 4 on normal human neutrophil functions.* J Leukoc Biol, 1986. **39**(4): p. 423-34.

9. Sharpe, R.J., et al., *Growth inhibition of murine melanoma and human colon carcinoma by recombinant human platelet factor 4.* J Natl Cancer Inst, 1990. **82**(10): p. 848-53.

10. Sato, Y., M. Abe, and R. Takaki, *Platelet factor 4 blocks the binding of basic fibroblast growth factor to the receptor and inhibits the spontaneous migration of vascular endothelial cells.* Biochem Biophys Res Commun, 1990. **172**(2): p. 595-600.

11. Davie, E.W., K. Fujikawa, and W. Kisiel, *The Coagulation Cascade - Initiation, Maintenance, and Regulation.* Biochemistry, 1991. **30**(43): p. 10363-10370.

12. Loscalzo, J., B. Melnick, and R.I. Handin, *The interaction of platelet factor four and glycosaminoglycans.* Arch Biochem Biophys, 1985. **240**(1): p. 446-55.
13. Rybak, M.E., et al., *Interaction of platelet factor four with cultured vascular endothelial cells.* Blood, 1989. **73**(6): p. 1534-9.

14. Lee, G.M. and G.M. Arepally, *Heparin-induced thrombocytopenia.* Hematology Am Soc Hematol Educ Program, 2013. **2013**: p. 668-74.

15. Cai, Z., et al., *Atomic description of the immune complex involved in heparin-induced thrombocytopenia.* Nat Commun, 2015. **6**: p. 8277.

16. Mayo, K.H. and M.J. Chen, *Human platelet factor 4 monomer-dimer-tetramer equilibria investigated by 1H NMR spectroscopy.* Biochemistry, 1989. **28**(24): p. 9469-78.

17. Mayo, K.H., et al., *Heparin binding to platelet factor-4. An NMR and site-directed mutagenesis study: arginine residues are crucial for binding.* Biochem J, 1995. **312 (Pt 2)**: p. 357-65.

18. Ludlam, C.A., et al., *The release of a human platelet specific protein measured by a radioimmunoassay.* Thromb Res, 1975. **6**(6): p. 543-8.

19. St Charles, R., D.A. Walz, and B.F. Edwards, *The three-dimensional structure of bovine platelet factor 4 at 3.0-A resolution.* J Biol Chem, 1989. **264**(4): p. 2092-9.

20. Zhang, X., et al., *Crystal structure of recombinant human platelet factor 4.* Biochemistry, 1994. **33**(27): p. 8361-6.

21. Gerber, N., et al., *Receptor-binding conformation of the "ELR" motif of IL-8: X-ray structure of the L5C/H33C variant at 2.35 angstrom resolution.* Proteins-Structure Function and Genetics, 2000. **38**(4): p. 361-367.

22. Rauova, L., et al., *Ultralarge complexes of PF4 and heparin are central to the pathogenesis of heparin-induced thrombocytopenia.* Blood, 2005. **105**(1): p. 131-8.

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

24. Xu, G.H. and M.R. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics.* Chemical Reviews, 2007. **107**(8): p. 3514-3543.

25. Takamoto, K. and M.R. Chance, *Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 251-76.

26. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochem Biophys Res Commun, 2001. **287**(3): p. 614-21.

27. Charvatova, O., et al., *Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1.* Journal of the American Society for Mass Spectrometry, 2008. **19**(11): p. 1692-1705.

28. Hambly, D. and M. Gross, *Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin.* International Journal of Mass Spectrometry, 2007. **259**(1-3): p. 124-129.

29. Chance, M.R., *Unfolding of apomyoglobin examined by synchrotron footprinting.* Biochemical and Biophysical Research Communications, 2001. **287**(3): p. 614-621.

30. Goldsmith, S.C., et al., *Synchrotron protein footprinting: A technique to investigate protein-protein interactions.* Journal of Biomolecular Structure & Dynamics, 2001. **19**(3): p. 405-418.

31. Li, Z.X., et al., *High Structural Resolution Hydroxyl Radical Protein Footprinting Reveals an Extended Robo1-Heparin Binding Interface.* Journal of Biological Chemistry, 2015. **290**(17): p. 10729-10740.

32. Maleknia, S.D., M. Brenowitz, and M.R. Chance, *Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry.* Analytical Chemistry, 1999. **71**(18): p. 3965-3973.

33. Watson, C., et al., *Pulsed Electron Beam Water Radiolysis for Submicrosecond Hydroxyl Radical Protein Footprinting.* Analytical Chemistry, 2009. **81**(7): p. 2496-2505.

34. Gau, B.C., et al., *Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding.* Analytical Chemistry, 2009. **81**(16): p. 6563-6571.

35. Li, X.Y., et al., *Improved Identification and Relative Quantification of Sites of Peptide and Protein Oxidation for Hydroxyl Radical Footprinting.* Journal of the American Society for Mass Spectrometry, 2013. **24**(11): p. 1767-1776.

36. Li, X., et al., *Supercharging by m-NBA Improves ETD-Based Quantification of Hydroxyl Radical Protein Footprinting.* J Am Soc Mass Spectrom, 2015. **26**(8): p. 1424-7.

37. Chen, J.W., et al., *Fast Photochemical Oxidation of Proteins and Mass Spectrometry Follow Submillisecond Protein Folding at the Amino-Acid Level.* Journal of the American Chemical Society, 2012. **134**(45): p. 18724-18731.

38. Kaur, P., et al., *Quantitative Protein Topography Analysis and High-Resolution Structure Prediction Using Hydroxyl Radical Labeling and Tandem-Ion Mass Spectrometry (MS).* Molecular & Cellular Proteomics, 2015. **14**(4): p. 1159-1168.

39. Allen, S.J., D.J. Hamel, and T.M. Handel, *A rapid and efficient way to obtain modified chemokines for functional and biophysical studies.* Cytokine, 2011. **55**(2): p. 168-73.

40. Saladino, J., et al., *Aliphatic Peptidyl Hydroperoxides as a Source of Secondary Oxidation in Hydroxyl Radical Protein Footprinting.* Journal of the American Society for Mass Spectrometry, 2009. **20**(6): p. 1123-1126.

CHAPTER 6

CONCLUSIONS

Hydroxyl radical protein footprinting (HRPF) coupled to liquid chromatography and tandem mass spectrometry (LC-MS/MS) has been developed over the last decade and has matured to a powerful method for analyzing protein structure and dynamics. It has been successfully applied in the analysis of protein structure, protein folding, protein dynamics, and protein–protein and protein–DNA interactions. The present work demonstrates the development and improvement of novel ETD-based high spatial resolution HRPF method, and its successful applications on identification of binding interfaces of protein-carbohydrate complexes and oligomerization of chemokines at residue level.

Hydroxyl radical protein footprinting is based on the idea that by monitoring the change in the amount of oxidation of amino acid side chains under different conditions, changes in protein structure can be determined by how the structural changes affect the solvent accessibility, and thus the amount of oxidation, of the different regions of the protein. Traditional HRPF quantitation was performed on the peptide level using LC-MS, allowing for general structural information of large scale protein interaction. However, this method yielded structural information with a spatial resolution limited by the size of the peptide. Electron-transfer dissociation (ETD) has proven to be a more reliable fragmentation technique for simultaneous identification and relative quantification of a variety of functionally different oxidation isomers. The development of high spatial resolution ETD-based HRPF methods necessitates not only the identification of the sites of oxidation on a residue level, but also the quantitation of the level of oxidation at each site within the peptide, providing structural information with higher spatial resolution and more accuracy.

Spatial resolution of HRPF is limited by the smallest portion of the protein for which oxidation amounts can be accurately quantitated. The ideal case is that oxidation on every single amino acid can be quantitated by ETD. Most of the tryptic peptides are doubly-charged. However, efficient ETD normally requires triply or higher peptide charge states, which limits its applicability

for residue level HRPF. The use of supercharging reagents m-NBA was demonstrated to improve sequence coverage of ETD spectra without affecting the ETD-based quantification of site-specific oxidation. This improvement is particularly helpful to quantitate those short peptides with different oxidation sites in real HRPF samples. In addition, different digestion enzymes such as Lys-C can also be used to increase the peptide charge states for ETD analysis.

There are approaches such as chemical derivatization coupled with LC-MS/MS that have been developed to structurally characterize glycosaminoglycans (GAGs). Protein - GAG interactions contribute to modulate protein activity, and are essential to many biological processes. However, the large size, heterogeneous structure, and inherent flexibility of GAG chains make the detailed structural interrogation of protein-GAG interactions difficult, particularly for full-length GAG chains. Unfractionated heparin is one kind of GAG with large, flexible chains and a variety of modification patterns. The power of high spatial resolution HRPF lies in its ability to handle large protein with diverse mixture of complex ligand, which may not be solved by other standard high resolution structural technologies such as X-ray Crystallography or Nuclear Magnetic Resonance Spectroscopy (NMR). The binding interface of the Robo1-heparin complex has been revealed by HRPF at residue level, and a model for Slit2-Robo1-heparin/HS ternary complex formation and signaling was proposed based on previous study and our data. Heparin affinity column and SPR analysis performed by our collators both support our finding. The model provides a testable hypothesis for the mechanism of Slit-Robo-heparin/HS interaction in signal transduction.

There are three important factors that contribute to this successful application of ETD-based high spatial resolution HRPF method: first, 100% sequence coverage of Robo1 is achieved by multiple enzyme digestions, which provides full residue information for the whole protein; second, the use of a combination of multiple enzyme digestions may avoid short peptides in Robo1

that are insufficient for ETD fragmentation. Otherwise, the supercharging reagent needs to be added in mobile phase for improvement; third, Robo1 protein is relatively well-oxidized during FPOP experiments. Nearly every peptide in Robo1 has multiple sites of oxidation, which can cover the most of the binding interface and provide much information on their changes in solvent accessibility. If the protein is poorly oxidized, the information for those binding residues will be lost. In this case, HRPF may not be an ideal technique to solve this problem. In fact, there are still a few basic binding residues in Robo1-heparin that are not oxidized or identified by HRPF. Other labeling methods that specifically modify basic residue need to be developed and applied to investigation of protein –GAG interactions.

Study of protein oligomerization is another important application of ETD-based high spatial resolution HRPF techniques. CXCL4 K50E mutant are firstly characterized by HRPF, and the model of dimer-dimer interface of wild type CXCL4 tetramer are proposed. A deeper understanding of characteristics of K50E mutant regarding its stability and dynamics as a dimer and the structural process involved in tetramer formation provides insights to understand the mechanism of this protein, as well as design treatments for heparin-induced thrombocytopenia (HIT).

The difficulties presented in this work exposed certain limitations of our current protocol for protein oxidation since relatively small amount of residues in CXCL4 show oxidation, which may not cover the full dimer-dimer binding sites. An improvement of protein labeling can be achieved by increasing the laser power, reducing the concentration of the sample exposed to hydroxyl radicals, and development of more efficient labeling protocols. The continued work has been focused on the identification of CXCL4 interactions with heparin and chondroitin sulfate (CS). However, precipitation is observed when CXCL4 binds with native CS or CS dp8, indicating

structural differences between CXCL4 with heparin and CS. Smaller sizes of fractionated CS should be further tested for this binding. Another investigation is performed by comparing unfractionated heparin interactions with wild type CXCL4 tetramer and CXCL4 K50E dimer. Wild type CXCL4 can form ultralarge complexes when binding with unfractionated heparin. While K50E mutant fails to form ultralarge complexes but smaller complexes when incubated with unfractionated heparin compared to wild type CXCL4. This phenomenon indicates that the capacity of CXCL4 to form tetramer is required to form ultralarge complexes. The detailed investigation of these interactions and binding interfaces is necessary to provide insights to clinical treatment of HIT.

The present work demonstrated the ability of HRPF to probe protein structure and dynamics at physiologically relevant protein concentrations for a wide range of solution conditions in the presence of multiple complex ligands, providing structural resolution at the single amino acid level. The improved resolution has enabled this technique to be a powerful tool for protein characterization where detailed structural information is required to determine biological functions. Future prospects of HRPF could be coupled with computational modeling and focused on detecting structures of some proteins, such as membrane proteins that are challenging to X-ray crystallography and NMR.