OXIDATIVE SURFACE MAPPING AND MOLECULAR DYNAMICS SIMULATIONS: OPTIMIZATION OF METHODS FOR STUDYING MACROMOLECULAR INTERACTIONS

by

OLGA CHARVÁTOVÁ

(Under the Direction of Robert J. Woods)

ABSTRACT

For certain classes of biomolecular complexes, particularly those involving glycans binding to proteins, it is rarely possible to employ traditional methods such as x-ray crystallography or nuclear magnetic resonance to determine their structural features. Therefore, a contact surface mapping (footprinting) method that would be applicable in these cases could provide information for characterizing the 3D structures in these interactions. The protein surface mapping method described here employs hydroxyl radicals, generated by pulsed-laser induced photodissociation of hydrogen peroxide, to modify the amino acid side chains. Analysis of the extent and location of oxidation is subsequently determined by mass spectrometry. This data can then be combined with the predictions of an average solvent accessible surface area for each residue employing molecular dynamics simulations and molecular modeling to validate predicted 3D structures. Development of such a combined method is a contribution to the field of 3D structure modeling as well as experimental footprinting.

INDEX WORDS: β-lactoglobulin, ByOnic, footprinting, galectin-1, MD simulations, oxidative surface mapping, Shigella flexneri, solvent accessible surface area, SASA
OXIDATIVE SURFACE MAPPING AND MOLECULAR DYNAMICS SIMULATIONS:
OPTIMIZATION OF METHODS FOR STUDYING MACROMOLECULAR
INTERACTIONS

by

OLGA CHARVÁTOVÁ

M.S., The Institute of Chemical Technology, The Czech Republic, 2002

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

MASTER OF SCIENCE

ATHENS, GEORGIA

2010
OXIDATIVE SURFACE MAPPING AND MOLECULAR DYNAMICS SIMULATIONS:
OPTIMIZATION OF METHODS FOR STUDYING MACROMOLECULAR
INTERACTIONS

by

OLGA CHARVÁTOVÁ

Major Professor: Robert J. Woods
Committee: Michael G. Hahn
            Ron Orlando

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2010
ACKNOWLEDGEMENTS

I would like to thank my major professor Rob Woods for guidance through the graduate study at the University of Georgia as well as giving me a great opportunity to be a part of his research group and gain valuable experience in an experimental and computational biochemical laboratory.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Thesis Structure</td>
<td>1</td>
</tr>
<tr>
<td>Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>2  QUANTIFYING PROTEIN INTERFACE FOOTPRINTING BY HYDROXYL RADICAL</td>
<td>5</td>
</tr>
<tr>
<td>OXIDATION AND MOLECULAR DYNAMICS SIMULATION: APPLICATION TO GALECTIN-1</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>6</td>
</tr>
<tr>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>Experimental</td>
<td>11</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>18</td>
</tr>
<tr>
<td>Conclusions</td>
<td>26</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>28</td>
</tr>
<tr>
<td>3  PULSED ELECTRON BEAM WATER RADIOLYSIS FOR SUBMICROSECOND HYDROXYL</td>
<td>43</td>
</tr>
<tr>
<td>RADICAL PROTEIN FOOTPRINTING</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>44</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Observed mass changes from side chain oxidation</td>
<td>29</td>
</tr>
<tr>
<td>2.2</td>
<td>Synthetic peptides used in quantification studies</td>
<td>30</td>
</tr>
<tr>
<td>2.3</td>
<td>Oxidation level equivalents for mixtures of synthetic peptides</td>
<td>31</td>
</tr>
<tr>
<td>2.4</td>
<td>Per-residue oxidation levels and side chain $&lt;\text{SASA}&gt;$ values for galectin-1</td>
<td>32</td>
</tr>
<tr>
<td>2.5</td>
<td>Experimentally-measured and $&lt;\text{SASA}&gt;$-derived oxidation levels for reporter residues in the galectin-1 dimer interface</td>
<td>34</td>
</tr>
<tr>
<td>3.1</td>
<td>Average solvent accessibility of oxidized residues in $\beta$-Lactoglobulin.</td>
<td>70</td>
</tr>
<tr>
<td>S3.1</td>
<td>Reactions and parameters used in the transient absorption global fitting.</td>
<td>71</td>
</tr>
<tr>
<td>4.1</td>
<td>Oxidation products searched for by ByOnic.</td>
<td>97</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1: Peptide-1 before (A) and after oxidation (B), indicating oxidation of phenylalanine (m/z 799) and histidine (m/z 761 and 773). .................................................................35

Figure 2.2: MS spectrum of tryptic digestion of galectin-1 before oxidation (A), and after oxidation (B). ..............................................................................................................36

Figure 2.3: Tandem mass spectra of a tryptic peptide from galectin-1 in non-oxidized (A) and oxidized (B) forms indicating peaks associated with a 22 Da mass loss, typical of histidine oxidation.......................................................................................................37

Figure 2.4: Per-residue oxidation levels plotted on the solvent accessible surface of dimeric galectin-1 (A) ...............................................................................................................38

Figure 2.5: Solvent accessible surface area (<SASA>) for each of the 134 side chains in galectin-1......................................................................................................................39

Figure 2.6: Percentage of oxidation versus side chain <SASA> for selected amino acids. ..............40

Figure S2.1: A representative example of the label-free quantitation procedure, using LC-MS signal intensities..........................................................................................................41

Figure S2.2: Silver-stained SDS-PAGE gel of reduced galectin-1 (monomer molecular weight 14.5 kDa) indicating effect of multiple laser exposures and variation in hydrogen peroxide concentration........................................................................................................42

Figure 3.1: A 3 minute HSQC spectrum of galectin-3 in the absence (red) and presence (black) of 1% hydrogen peroxide (v/v). .............................................................................................72
Figure 3.2: LC-MS of the 10+ charge state of intact ubiquitin .....................................................73
Figure 3.3: Concentration profiles of species contributing most to the transient absorption signals at 250 nm based on the global fitting results. ..............................................................74
Figure 3.4: (a) Two views of X-ray crystal structure of β-lactoglobulin A with waters removed for clarity created using PyMOL (PDB identifier 1BSY). ........................................75
Figure S3.1: Schematic representation of the electron pulse protein oxidation setup. ..............76
Figure S3.2: LC-MS of the 10+ charge state of intact ubiquitin ..................................................77
Figure S3.3: LC-MS of the 13+ charge state of intact β-Lactoglobulin ......................................78
Figure S3.4: Transient absorption signals monitored after the pulse of electrons at 250nm ......79
Figure S3.5: Surface average solvent accessibility (<SASA>) value for the amino acid groups of β-lactoglobulin that contained at least one oxidized residue .............................................80
Figure 4.1: A peptide from galectin-1 protein digest .................................................................98
Figure 4.2: MALDI-TOF spectra of a peptide LHFNPR .............................................................99
Figure 4.3: Tandem mass spectra of synthetic peptide GLEWVAEIR and an identical peptide from a digest of the Shigella flexneri mAb sample .........................................................100
Figure 4.4: SDS gel of oxidized and non-oxidized Shigella flexneri mAb and galectin-1 .........101
Figure 4.5: MALDI spectra of GLEWVAEIR peptide with decreasing signal intensities after laser exposure ..............................................................................................................102
CHAPTER 1

INTRODUCTION

Thesis Structure

This thesis contains two chapters that have been adapted from articles published in peer-reviewed journals (chapters two and three). The first chapter is an introduction to the research area discussed in this document with a literature overview of the existing techniques related to structural studies of macromolecular interactions. Chapter four consists of unpublished experimental results relevant to the research area. Finally, the fifth chapter contains conclusions and discusses challenges of the existing methods.

Literature Review

Traditional structural methods such as x-ray crystallography or NMR spectroscopy are often difficult to apply to studies of macromolecular complexes that involve carbohydrates. The co-crystallization of some molecules such as flexible oligo- or polysaccharides can be a challenging task when x-ray crystallographic methods are intended to be used for 3D structure determinations. In the case of NMR studies, the large size of a protein-glycan complex can be a limitation. Therefore there is a need for alternative methods for ligand binding characterization in certain complex structures. For a number of macromolecular complexes, alternative techniques that are all based on the same principle of a partial shielding of the protein by the interacting macromolecule have already been established. Chemical derivatization methods,
such as biotinylation [1, 2] or acetylation [3, 4] have been employed to probe protein-nucleic acid, protein-protein, or protein-ligand interactions. Similarly, hydroxyl radical footprinting [5, 6] whereby the amino acid side chains are covalently modified has been applied to such studies. All the above listed chemical modifications are detected and analyzed by mass spectrometric techniques. However, none of these covalent labeling methods has been applied to complexes wherein one of the interacting molecules is a glycan. Hydrogen-deuterium exchange mass spectrometry (H/D-Ex MS) is a possible alternative, and its application to studies of a carbohydrate-protein complexes has been reported [7, 8]. As is obvious from the name of the approach, H/D-Ex is not a covalent labeling technique and therefore, certain conditions such as low temperature and low pH need to be maintained during the experiment in order to prevent the back exchange of the amide deuterons [9]. To detect the level of deuterium incorporation, mass spectrometric analysis needs to be performed typically on peptides of the digested protein. Due to these specific requirements of temperature and pH, proteolytic digestion is challenging to achieve [8] and the resulting peptides are often non-specific fragments which leads to complicated MS data analysis. Another complication when this method is applied arises from possible shielding of the backbone amide protons by amino acid side chains in both the free and bound form of the complex, which prevents the detection of ligand binding. Lastly, the time regime for proton exchange is long with respect to the kinetics of ligand binding, which degrades the sensitivity of the method for ligand footprint detection [9, 10].

If a contact region of the molecular complex is relatively small, such as in protein-carbohydrate complexes, high resolution analysis will require the modification of several types of amino acids if one of the covalent labeling techniques is employed. This requirement is met in methods that employ hydroxyl radicals to modify the side chains of the protein. Most amino
acids are reactive in an oxidative environment, and the extent of reactivity depends on the radical concentration [11], solvent accessibility of the residues [12], and the type of the amino acid [5, 13]. Several approaches to generation of hydroxyl radicals are known in the field. One of the established methods, but often difficult to carry out at the research site due to the enormous costs of the instrumentation, is synchrotron radiolysis of water [5, 14], which has been used in studies of protein-protein and protein-nucleic acid interactions. Recently, a pulsed electron beam water radiolysis technique [15] has been introduced (chapter three) and, similarly to synchrotron radiolysis, does not require the presence of hydrogen peroxide. Another approach is known as Fenton chemistry, whereby hydroxyl radicals are generated by contact between hydrogen peroxide and metal-bound ethylenediaminetetraacetic acid (EDTA) [16-18]. This method is not well suited to the study of carbohydrate-protein complexes due to the quite long incubation time. Protein-carbohydrate complexes form a dynamic system and therefore reaction needs to be fast before the ligand changes the conformation or dissociates from the protein. Finally, methods that employ photodissociation of dilute hydrogen peroxide induced by UV [19, 20], γ-irradiation [21] or high energy laser [11, 22, 23] are alternatives with relatively low demands on instrumentation and high efficiency of radical generation. The use of a pulsed laser allows fast protein irradiation by powerful short pulses on nanosecond timescale. Nevertheless, studies involving carbohydrate-protein interactions using this approach have not been reported by date. As already mentioned above, the interacting surfaces in complexes with glycans have relatively small size and therefore, the selected technique needs to be accurate and have high resolution. In contrast to larger complexes, wherein the shielded region might include many amino acids and therefore at least some residues involved in the binding might be completely shielded, carbohydrate
ligands would most likely yield only partial shielding. Therefore some form of quantitation of how much modification occurred on residues within the binding site needs to be established.

Computational simulations are often helpful tools for predicting 3D structure of the macromolecular complexes [24-26] and are widely used by scientists in the field of covalent labeling or H/D-Ex techniques for predictions of the amino acid solvent accessibility or for molecular docking [8, 27, 28]. When no experimentally-derived structure of a protein of interest is available, homology modeling [29-31] may be particularly useful in generating its 3D fold. GLYCAM-Web tools [32] can be employed to built the initial carbohydrate structure. For a protein–carbohydrate-ligand complex formation, either docking tools such as the computational docking algorithm AUTODOCK [33] or a manual alignment of the carbohydrate with the protein based on a similar crystal structure can be employed, prior to refinement of the complex by MD simulation [34]. Predictions of solvent accessibility of the amino acid side chains over time in the MD simulation can give valuable insight into features of the complex. Calculation of time-averaged solvent accessible surface area (<SASA> [35]) of each side chain demonstrates more realistic conditions (imitating behavior in solution) than SASA calculated from a single structure. When computational tools and oxidative surface mapping are combined, the resulting method should be well applicable for structural studies of macromolecular complexes including protein-carbohydrate interactions.

In the following chapters, results from the above introduced methods are presented. Specifically, the oxidative surface mapping method, molecular modeling, MD simulations, and <SASA> calculations have been employed in the studies of the carbohydrate-binding proteins.
CHAPTER 2

QUANTIFYING PROTEIN INTERFACE FOOTPRINTING BY HYDROXYL RADICAL OXIDATION AND MOLECULAR DYNAMICS SIMULATION: APPLICATION TO GALECTIN-1¹

Abstract

Biomolecular surface mapping methods offer an important alternative method for characterizing protein-protein and protein-ligand interactions in cases in which it is not possible to determine high resolution 3D structures of complexes. Hydroxyl radical footprinting offers a significant advance in footprint resolution compared to traditional chemical derivatization. Here we present results of footprinting performed with hydroxyl radicals generated on the nanosecond time scale by a laser-induced photodissociation of hydrogen peroxide. We applied this emerging method to a carbohydrate-binding protein, galectin-1. Since galectin-1 occurs as a homodimer, footprinting was employed to characterize the interface of the monomeric subunits. Efficient analysis of the MS data for the oxidized protein was achieved with the recently developed ByOnic software that was altered to handle the large number of modifications arising from side chain oxidation. Quantification of the level of oxidation has been achieved by employing spectral intensities for all of the observed oxidation states on a per-residue basis. The level of accuracy achievable from spectral intensities was determined by examination of mixtures of synthetic peptides, related to those present after oxidation and tryptic digestion of galectin-1. A direct relationship between side chain solvent accessibility and level of oxidation emerged that enabled the prediction of the level of oxidation given the 3D structure of the protein. The precision of this relationship was enhanced through the use of average solvent accessibilities computed from 10 ns molecular dynamics simulations of the protein.

Keywords: AMBER, ByOnic, galectin-1, MD simulations, oxidative surface mapping, solvent accessible surface area, SASA
Introduction

There has been increasing interest in the development of high-throughput methods for characterizing protein-ligand contact surfaces (footprinting). One approach employs hydrogen-deuterium exchange mass spectrometry (H/D-Ex MS) method [7, 28, 36-38], in which amide protons in the backbone of a protein are allowed to undergo exchange in deuterated water. To detect and quantify the level of deuterium incorporation by mass spectrometry typically necessitates proteolytic digestion under conditions that attenuate hydrogen back-exchange, such as low pH and low temperature. However, under these conditions, complete digestion of the protein is difficult to achieve [8]. Additionally, while deuterium exchange is altered in the presence or absence of a ligand, it is a complex phenomenon that depends heavily on protein folding kinetics and local amino acid sequence. It is also worth noting that the backbone amide protons are not necessarily differentially shielded by the presence of a ligand, that is, they may be shielded by side chains in both the free and bound form of a protein-ligand complex.

Chemical-derivatization methods, such as biotinylation [1, 2] or acetylation [3, 4], which focus on the side chains, have been employed to probe protein-protein, protein-ligand or protein-nucleic acid interactions. The advantage of these methods over deuterium-exchange is that the modification is irreversible, facilitating analysis, whereas perhaps the most notable limitation is that the only a few residues or types of residues will be labeled in any protein. Thus, the resolution of the contact surface is extremely low. An alternative chemical derivatization method that employs hydroxyl radicals to oxidize the side chains appears to offer several advantages. Firstly, hydroxyl radicals are highly reactive and result in covalent modifications that facilitate a high throughput proteomic type analysis of the product. Further, in contrast to H/D exchange, the extent of hydroxyl radical oxidation, for any given amino acid, appears to
depend directly on solvent accessibility [5, 12, 19]. Thus, this type of labeling should be useful in identifying ligand contact regions. Lastly, in contrast to other chemical derivatization methods, depending on the manner in which they are generated and on exposure times, hydroxyl radicals can react with all amino acid side chains. Methods to generate hydroxyl radicals in biological environments, include: direct synchrotron radiolysis of water [39-42], photodissociation of dilute hydrogen peroxide induced by high energy laser [11, 22, 23], UV [19, 20] or $\gamma$-irradiation [21]. Alternative methods include, Fenton chemistry [16-18], whereby hydroxyl radicals are generated by contact between hydrogen peroxide and metal-bound ethylenediaminetetraacetic acid, or by high voltage discharge within an electrospray ion source [43, 44]. In several methods, a major drawback is the availability or complexity of the instrumentation. While Fenton chemistry is an affordable method that has been utilized in the structural analysis of nucleic acids [45-47], and has been also applied to study proteins [48, 49] it also requires long incubation times and chemical additives. For a weak interaction, as is typical for many carbohydrate-protein complexes, a long incubation, or exposure time would allow the ligand to undergo multiple exchanges with the protein surface, during which the contact region would be exposed to reactive radicals. Additionally, rapid reaction is necessary to ensure that the oxidation of the amino acids take place before any unfolding or other conformational changes of the protein occurs. To minimize these issues, we believe pulsed laser-induced dissociation of hydrogen peroxide is advantageous due to the relatively rapid speed of hydroxyl radical generation.

Previous studies by Hambly [22, 23] and Aye [11] have demonstrated the utility of generating hydroxyl radicals by a single laser pulse of dilute hydrogen peroxide. The radical concentration achieved from a single pulse of a dilute peroxide solution (1% or less) is adequate
to achieve significant levels of protein surface oxidation [11, 22]. This is an important observation since the resulting oxidation leads to protein denaturation [11] and, thus, although subsequent irradiation would ensure high oxidation levels, it would not be appropriate for 3D structural studies. In the limit, multiple laser pulses can also lead to backbone cleavage [50].

High resolution footprinting necessitates not only the identification of the sites of oxidation (on a per amino acid residue basis), but also the quantitation of the level or rate of oxidation at each site. Mass spectrometry provides a powerful tool for this characterization, although since the majority of side chains can exhibit multiple oxidation states data analysis is challenging. In a typical protein sequencing analysis, the protein is digested by a specific protease, such as trypsin, and the resulting fragments sequenced by tandem MS methods. This approach relies heavily on knowledge of the masses of each amino acid and any post-translation modifications. Commonly used proteomics software, such as SEQUEST [51] or Mascot [52], are not optimized to handle the extensive number of permutations generated by the range of possible side chain oxidation states. To address this issue, we employed a program (ByOnic [53]), which has been specifically modified to handle the large number of amino acid modifications associated with hydroxyl radical oxidation. ByOnic combines de novo sequencing and database searching to match the peptide fragments to a sequence provided by the user.

Although relative peptide abundances in a protein digest may be estimated by spectral counting, the quantitation of proteomics data benefits from more sophisticated treatments [54]. Here we employ a label-free approach to quantify the subtle differences in pre-residue oxidation levels arising from reaction with hydroxyl radical. Proteolysis of an oxidized protein leads to multiple instances of peptides with the same sequence, which differ only in the extent of oxidation. To quantify the level of oxidation at any given site requires the identification of all
peptides with oxidation at that site, using tandem MS. We compute the percentage of oxidation at each amino acid ($\% AA_{ox}$) from the LC-MS signal intensities of each peptide containing a specific oxidized amino acid ($I_{ox}$), relative to the total of all intensities associated with that peptide sequence ($I_{ox} + I_{non-ox}$) using the following straightforward relationship:

$$\% AA_{ox} = \frac{\sum I_{ox}}{\sum I_{ox} + \sum I_{non-ox}} \times 100$$  \hspace{1cm} (2.1)$$

To ascertain the suitability of our use of spectral intensities for quantitation of oxidation levels, as well as to indicate the extent to which a quantitative relationship exists between oxidation and side chain solvent accessibility, we have applied hydroxyl radical footprinting to a small globular protein for which extensive x-ray structural data exists. Galectin-1 is a homodimer member of the galectin family found in many animals including humans, which has the ability to bind oligosaccharide ligands containing galactosyl residues [55-59]. Galectins are the target of many studies because of their involvement in various biological processes such as cell adhesion, differentiation, regulation of cell growth, regulation of apoptosis, cell activation, mRNA splicing, inflammation, and induction of Ca$^{2+}$ dependent signal transduction pathways [60-63]. Due to its small size, and the availability of x-ray structures from various species, with various ligands, galectin-1 is a suitable protein for mass spectrometric analysis and surface mapping method development.

The interpretation of hydroxyl radical surface mapping data, in terms of 3D structural properties, is based on the assumption that those residues that are exposed to solvent will react more readily with hydroxyl radicals than buried or otherwise shielded residues. To assist in the analysis of the MS data, we have employed molecular dynamics (MD) simulations to predict the time-averaged solvent accessible surface area ($<SASA>$) of each side chain [28]. In contrast to x-ray structural data, MD simulations enable the generation of a realistic ensemble of side chain
orientations in solution at experimental temperature, which can be particularly important when employing homology models [25]. Additionally, MD refined side chain ensembles have been shown to lead to better correlation with experimentally-determined levels of oxidation [28], and so provide a robust basis for estimating <SASA> values. This computational approach also allows us to separate the galectin-1 homodimer into its two subunits and predict the solvent accessibility of the residues in the interface region. In addition, molecular modeling techniques [29-31] can be particularly useful in generating the 3D fold of proteins of interest when no experimentally-derived structure is available, and in this study we examine the extent to which a homology-based model of galectin-1 can be used to interpret oxidative footprinting data. In principle, oxidative surface mapping could provide a valuable alternative to NMR or x-ray analysis for validating predictions from such popular computational methods as homology modeling and automated ligand docking. To employ footprinting effectively in this regard, a quantitative understanding of the relationship between solvent accessibility and extent of oxidation must be developed. We are approaching this problem by quantifying the experimental data and by employing a robust method to estimate side chain <SASA>.

Experimental

Sample preparation

Galectin-1 expressed by Chinese hamster ovary (CHO) cells with the sequence characterized by Cho & Cummings [56], was purified as described previously [56], and stored under reducing conditions in presence of dithiothreitol (DTT) [64], Bio-Rad Laboratories (Hercules, CA). Prior to laser exposure, DTT was removed by dialysis using Spectra/Por RC Float-A-Lyzer, MW 3,500, Spectrum Laboratories, Inc. (Rancho Dominguez, CA). Galectin-1
concentration was adjusted to 0.25 mM by adding a ten times diluted phosphate buffer saline (0.1 × PBS), Mediatech, Inc. (Herndon, VA). Immediately prior to laser exposure, the protein solution was mixed with hydrogen peroxide (10%, by dilution of 30% solution w/w), Sigma-Aldrich (St. Louis, MO) to a final concentration of 1%, to a final volume of 8 µL (final protein concentration approximately 0.22 mM).

Synthetic peptides, in greater than 98% purity were purchased from Biomatic Corporation (Wilmington, DE) and stored as a lyophilized powder. Prior to laser exposure, each peptide was re-suspended in 25 mM ammonium bicarbonate buffer to a concentration of 0.75 mM. An aliquot (4.2 µL) of the peptide solution was added to a solution of hydrogen peroxide (10%) and 25 mM ammonium bicarbonate (both Sigma-Aldrich) and adjusted to a final concentration of 1% H₂O₂, and the total volume of 8 µL (final peptide concentration approximately 30 mM). Purified water (18.2 MΩ) for all re-suspensions and buffer preparations was obtained from an in-house NANOpure Diamond system.

**Laser photolysis**

Hydroxyl radicals were generated by exposure to radiation from a Compex 110 KrF excimer laser operating at 248 nm, Lambda Physik, Coherent Inc. (Santa Clara, CA). A microcentrifuge tube (1.5 mL) containing 8 µL of the protein, or peptide, hydrogen peroxide solution was aligned along the beam axis to ensure the maximal contact of the laser beam with the protein solution, at a distance of 12 cm from the laser source. Each sample was exposed to a single 30 ns laser pulse with the adjusted laser power of 110 mJ/pulse. In the case of multiple pulses, a pulse rate of 5 Hz was employed. Immediately after irradiation, the sample was flash-frozen in liquid nitrogen and lyophilized for 1.5 hr to remove any remaining H₂O₂ and water.
**Gel electrophoresis**

Prior to gel electrophoresis, the lyophilized samples were re-suspended in 0.1 × PBS buffer. A premixed running electrophoresis Laemmli sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% sodium dodecyl sulfate (SDS), 25% glycerol, 0.01% Bromophenol Blue) with addition of 5% reducing agent β-mercaptoethanol (both Bio-Rad Laboratories) was added to the PBS-sample solution in a 2:1 v/v ratio. A 4-10% Tris-Tricine SDS gel (Bio-Rad Laboratories) was run under two buffer systems, anode (0.2 M Tris-HCl, pH 8.9) and cathode (0.1 M Tris, 1 M Tricine, 1% SDS, pH 8.25) [65] to identify any possible protein fragments. Electrophoresis voltage and current conditions were 150 V and 100 A, respectively, and the length of the run was 1 hr 15 min. After completion of the run, the gel was washed in nanopure water and stained via a silver staining procedure, using SilverSNAP® Stain Kit II, Pierce (Rockford, IL) following the manufacturer’s protocol.

**Proteolytic digestion**

In order to prepare the sample for mass spectrometric analysis, the lyophilized protein (or peptide) was re-suspended in 15 µL (50 mM) ammonium bicarbonate (Sigma-Aldrich) digest buffer. Dithiothreitol (Bio-Rad Laboratories) was added to a final DTT concentration of 10 mM, and the solution heated in a water bath for 1 hr at 60° C. Iodoacetamide (IAA, Sigma-Aldrich) was added to the sample to give a final IAA concentration of 50 mM, and the sample incubated for 1 hr at room temperature. Finally, sequencing grade modified trypsin, Promega (Madison, WI) was added in a 1:30 (protease:protein) ratio. The pH of the mixture was determined by pHhydrion Papers, Micro Essential Laboratory (Brooklyn, NY) to be between 6 and 8. The sample was then incubated for 18 hrs at 37° C. Digestion was quenched by removing the trypsin
and salts from the solution by passing through a C-18 ZipTip, Millipore Corporation (Bedford, MA). The galectin-1 digest was eluted from the ZipTip with 10 μL of mixed solvent acetonitrile:water:formic acid (50:49.9:0.1 (v/v)), dried on a speedvac, and resuspended in 10 μL (0.1%, v/v) aqueous solution of formic acid (Sigma-Aldrich) and stored at 6° C before analyzed.

**ESI-FT and MALDI-TOF**

Single mass spectra were obtained by high resolution MALDI-TOF, 4700 Proteomics Analyzer, Applied Biosystems (Lincoln, CA). In a typical experiment, 0.5 μL of each sample was mixed with 0.6 μL of 10 mg/mL α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich) dissolved in ethanol, spotted on a MALDI plate, Applied Biosystems (Foster City, CA) and analyzed in a reflector positive mode with 4,200-fixed laser intensity.

Tandem mass spectrometry was performed on a Finnigan LTQ-FT hybrid linear ion trap/Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (7-Tesla) with an electrospray ionization source (ESI), Thermo Electron (San Jose, CA), connected to an Agilent 1100 capillary liquid chromatograph (Palo Alto, CA) through a C18 column, 15 cm × 150 μm, Micro-Tech Scientific (Vista, CA). To elute peptides from the column, mobile phases A, 99.9% H₂O/0.1% formic acid (v/v), and B, 99.9% acetonitrile/0.1% formic acid (v/v) were used. In a 130 min run, a gradient from 5-80% (v/v) B was employed at a flow rate of 1 μL/min. MS/MS spectra were acquired on the nine most abundant ion precursors from each MS scan. Raw data sets were converted into mzXML format, followed by conversion to pkl format, using ReAdW and mzXML2Other [66].
Peak assignment and quantitation of oxidation

Established software for identification of MS/MS peaks from proteolytic digests, such as Mascot [52] and SEQUEST [51] are not well suited for the analysis of data sets that include a large number of mass modifications, as arise from hydroxyl radical oxidation. Emerging programs, such as MyriMatch [67] and ByOnic [53], have more flexibility in this regard. For peak assignment we employed ByOnic to search against a database containing only the CHO galectin-1 sequence [56], augmented by common mass modifications, as associated with carbamidomethylation of cysteine and oxidation of methionine, as well as 14 potential mass shifts arising from side chain oxidation. This number of mass modifications leads to over 45 potential oxidation products (Table 2.1). Most of the expected oxidation products have been previously characterized by Chance’s laboratory [5, 6, 14] and some of the less common modifications have been described by Berlet et al. [68] and Kubota et al. [69]. ByOnic includes three features not found in most database-search programs that make it especially suitable for identifying multiply modified peptides. First, it recalibrates m/z measurements based on well-identified peaks from an initial pass through the data. We defined a well-identified peak to be a singly charged b- or y-ion from a peptide identification with p-value smaller than 0.001. Unlike SEQUEST or Mascot, ByOnic’s scoring algorithm counts a match with 0.1 Da error more heavily than one with 0.2 Da error; hence recalibration offers improved sensitivity and specificity. Second, to improve efficiency, each type of modification has a user-defined limit on the number of occurrences per peptide. For example, up to five modifications per peptide were allowed, with at most four oxidations on reactive residues (cysteine, methionine, histidine, tryptophan, leucine, isoleucine, valine, arginine, phenylalanine, and tyrosine), and at most one on less reactive residues. This is in contrast to other methods that allow a peptide to carry any
number of each type of enabled modification. Third, results are double-checked with a “wild-card” modification: this type of search allows any integer change, within user-settable mass limits, to any one residue in each peptide. A wild card can be enabled along with any other set of modifications, and often finds unanticipated chemical artifacts such as sodiation, over-alkylation [70], and carbamidomethylated N-termini. Therefore, ByOnic was chosen as the main database search tool for this study. In general searches were performed with the following search parameters: specificity for tryptic fragments, fragment ion tolerance of 0.35 Da (after recalibration), peptide tolerance varied with parent charges (1.5 Da per charge), and including all of the mass modifications listed in Table 2.1. All oxidation positions were confirmed by tandem MS.

Quantitation of the level of oxidation at each amino acid position was performed using an in-house program that first matched the peptide masses from the LC-MS spectra to the sequences previously identified by ByOnic. The level of oxidation for any given amino acid was then computed by summing the spectral intensities for all peptides containing that amino acid, using the LC-MS spectral intensities for the peptides. The percentage of oxidation on a per-residue basis was then readily computed from Equation 2.1 and a representative example of this procedure is presented in Figure S2.1. For reasons of sensitivity, levels of oxidation were computed from LC-MS data rather than from MALDI-TOF data. Although this required a more complex data analysis, it generated a more complete list of oxidation states and therefore led to a more complete understanding of the relationship between solvent accessibility and level of oxidation.
Molecular modeling

Since there are no crystal structures of CHO galectin-1, a homology model was generated that was subsequently employed in MD simulations. A template structure (pdb id: 1W6P [71]) for the homology model that has 90% sequence identity with CHO galectin-1 was identified by searching the National Center for Biotechnology Information (NCBI) BLAST database (www.ncbi.nlm.nih.gov) and retrieved from the Research Collaboratory for Structural Biology (RCSB) data base [72]. The model was built using the Molecular Operating Environment (MOE), Chemical Computing Group Inc. (www.chemcomp.com) program. In order to calculate the <SASA> of each side chain in galectin-1, a 10-ns MD simulation of the complete protein (homodimer), as well as of one subunit of the protein (monomer) was performed using the all-atom AMBER 9 [73] force field with the PARM99 protein parameters [74]. All histidine residues were considered as neutral and were protonated at the Nε position. Prior to the MD simulations, the protein was solvated by 10,600 TIP3P [75] water molecules for the homodimer, and 5,637 waters for monomer. Counter ions were added to neutralize charge (four Na+ per monomer) using the tLEAP module of AMBER. Each system was subjected to energy minimization (1000 steps of steepest descent and 1000 of conjugate gradient) and heated from 0 K to the simulation temperature of 300 K over 50 ps. The simulation was performed using the NPT ensemble, at 1 atm using a 2-fs integration time step. The atomic co-ordinates were stored every 10 ps for analysis, for a total of 1000 snapshots. SASA values were calculated for individual snapshots employing the NACCESS program [76], with <SASA> and standard deviations computed over the 10 ns simulation.
Results and Discussion

Effect of number of laser exposures, and hydrogen peroxide concentration on protein integrity

To ensure a high resolution footprint, it is necessary to minimize the duration of exposure to hydroxyl radical relative to ligand binding kinetics; yet, to observe oxidation an adequate level of hydroxyl radicals is essential. Laser induced radical formation has the potential to satisfy these requirements [11]. Additionally, the extent of side chain oxidation may be enhanced by increasing the number of laser exposures, or by increasing the concentration of peroxide [11]; however, extreme oxidation conditions can lead to protein disintegration via backbone cleavage [50]. To examine the effect of these variables on protein integrity, the experiment was performed with 1, 5, 15, and 25 laser exposures, and under a range of hydrogen peroxide concentrations and analyzed by SDS-PAGE.

In the case of galectin-1, backbone cleavages were observed (Figure S2.2) after multiple laser pulses, especially in cases of 15 and 25 shots, which on the SDS-PAGE gel can be seen as smaller fragments that migrated between 14 and 5 kDa. In addition, there is an obvious decrease in the intensity of galectin-1 at 14.5 kDa after 25 laser pulses which can be attributed to extensive protein degradation. In contrast, the SDS-PAGE analysis indicates that increasing the concentration of hydrogen peroxide does not affect protein integrity with a single laser pulse. However, to avoid the potential for protein denaturation [11], we elected to employ a single laser exposure with a 1% peroxide concentration.

Oxidation experiments applied to a synthetic peptide

To confirm that our experimental conditions were able to generate the expected oxidation states without leading to any backbone fragmentation (as suggested by the SDS-PAGE data) a
synthetic fragment of galectin-1 (LHFNPR, peptide-1) was subjected to hydroxyl radical oxidation. Peptide-1 corresponds to positions 44-49 in galectin-1 and with m/z 783.4 (molecular weight of 782.91 Da). An examination of the MALDI-TOF data for the oxidized peptide confirmed that there was no backbone cleavage. As expected, the most abundant product was the peptide with a single oxidation (mass change of +16 Da, m/z 799.4), while the second most abundant product contained two oxidations (+32 Da, m/z 815.4) (Figure 2.1). Subsequent MS/MS analysis confirmed that most of the expected oxidation states were present in the sample. The difference between the MALDI-TOF and LC-MS experiments was in the ability of the FT instrument to detect products of oxidation that were present in very low amounts. By tandem mass spectrometry, histidine oxidation states with mass changes of +16, -10 and -22 Da were identified, as well as leucine oxidation products with +16 and +14 mass changes, and proline with addition of 14 Da. Phenylalanine was found in three oxidation states with addition of 16, 32, and 48 Da. Some of these higher oxidation products were also identified in the MALDI-TOF spectrum, but in such low quantities that in other circumstances these would be most likely considered as noise.

Not all of the possible oxidation states were identified for the residues in peptide-1, for example neither arginine nor asparagine were oxidized. Nonetheless, the extent of oxidation was deemed sufficient to warrant application of this method to intact galectin-1. Prior to that, however, we continued with the examination of synthetic peptides in order to determine the extent to which the level of oxidation could be quantified from spectral intensities.
Quantitation of level of oxidation

In order to confirm that spectral intensities may be employed to quantify the extent of oxidation, as defined in Equation 2.1, samples of two additional synthetic peptides, which represent oxidized fragments of galectin-1, were prepared in known concentrations. In peptide-2 (LHYNPR), the phenylalanine has been replaced by a tyrosine in position three. This substitution results in a net change of 16 Da and is equivalent to an oxidation product of the phenylalanine. Similarly, peptide-3 (LFDFNPR), is also an oxidized form of peptide-1. In peptide-3, the histidine was replaced by an aspartic acid, which is a net mass change of -22 Da. At an advanced stage of histidine oxidation, the ring of the histidine side chain opens, and two amino groups and a carbonyl group are replaced by a carboxyl group, which results in a loss of 22 Da [5, 40]. A summary of the peptide molecular weights, amino acid replacements and the net molecular weight changes is presented in Table 2.2.

Mixtures of each of the peptides containing representative oxidations (peptide-2 and peptide-3) with the non-oxidized material (peptide-1) were prepared in the following ratios peptide-1:peptide-2 (60:40) and peptide-1:peptide3 (33:66). Additionally, a sample containing all three peptides was prepared in a ratio of peptide-1:peptide-2:peptide3 (50:25:25). From the peak intensities of the peptides in each sample, the levels of oxidation for phenylalanine and histidine were determined (Equation 2.1, Table 2.3). The intensities for all peptides in each sample were determined using both MALDI-TOF and LC-MS experiments. To be consistent with the proposed analysis of galectin-1, all MS data for the synthetic samples were processed using the ByOnic software, followed by intensity summation on a per-residue basis using an in-house program. In general, both mass spectrometric methods resulted in close agreement with each other and with the known concentrations. Quantitation using the LC-MS data was only
slightly less accurate than the high resolution MALDI data. In the case of the phenylalanine oxidation (peptide-1:peptide-2) a maximum error of 2% with respect to the known concentrations was observed, while for the histidine oxidation (peptide-1:peptide-3) a slightly higher error of 6% was noted. In the third test case, the ratio of all three peptides was determined. Again, there was generally good agreement between the known ratios and the intensity-determined values for each MS method. As seen in the simple mixtures, the measured level of phenylalanine oxidation (24% by MALDI-TOF, 26% by LC-MS) was in slightly better agreement with the known concentration (25%) than for the histidine oxidation product (22% by MALDI-TOF, 28% by LC-MS). The fact that the ESI and Maldi data were in good agreement with each other and with the expected values suggests that each of the peptides displayed similar ionization efficiencies, in spite of the fact that the “oxidation” of histidine to aspartate (-22) might be expected to impact the net charge. Generally, the dominant charge state for these peptides under the ionization conditions used was +1. Previous studies of a peptide containing both a histidine and an arginine side chain indicated that the first protonation occurred on exclusively on the arginine [77]. Eliminating the histidine would leave the arginine as the dominant basic site, and replacing it with an aspartate should not significantly alter the surface activity in the droplet for the ESI experiments [78, 79]. Therefore, under electrospray ionization conditions where the +1 charge state dominates, it is not surprising to find that the replacement of a histidine with an aspartate does not significantly alter the ionization efficiency of the arginine-containing peptide.
**Extent of oxidation of galectin-1 from single laser exposure**

An examination of the high resolution MALDI-TOF spectra of the tryptic digest of oxidized galectin-1 clearly indicated that a single laser pulse in the presence of 1% hydrogen peroxide induced significant oxidation of the protein (Figure 2.2). As a result of the oxidation, intensities of non-oxidized peptides decreased, for example in the peptides with m/z 1486, 1142, 1041 and 942, while numerous new peaks resulting from side chain oxidation appeared. To fully characterize the sites of oxidation, tandem mass spectrometry with further fragmentation of peptides was performed.

Employing ByOnic to search against a database comprised of the sequence of galectin-1, allowing for 14 known mass changes from oxidation, led to the observation of 43 unique side chain oxidation products (Table 2.1). Oxidation of the intact protein led to the identification of additional oxidation products, relative to the data obtained for synthetic peptide-1, for example, for arginine and asparagine. An example of the appearance of an oxidized peptide in the tandem MS spectrum of a tryptic peptide from galectin-1 (FNAHGDANTIVCNSK) before and after laser exposure is presented in Figure 2.3. The MS/MS data clearly identified the oxidation of the histidine residue (-22 Da). Sequence coverage of galectin-1 as determined by MS/MS sequencing of tryptic fragments was satisfactory over triplicates of the experiments (approximately 85% on average). Of these identified residues, approximately 42% displayed oxidation after a single laser exposure (Figure 2.4A). A quantified list of the total oxidation level observed for each residue is presented in Table 2.4.

Because some residues will be on the protein surface, and therefore exposed to hydroxyl radicals, while others will be occluded in the core of the protein, it is important to determine the percentage of residues that might be expected to react, based on their exposure on the surface of
the protein, and on their chemical susceptibility to oxidation. In order to relate the oxidation data to the 3D structure of the protein, it is necessary to determine the accessibility of each amino acid side chain to hydroxyl radical. Since hydroxyl radical is essentially the same size as water, it is reasonable to assume that, for a given side chain, the extent of oxidation will be related to SASA of the side chain. This assumption has been qualitatively demonstrated by previous hydroxyl radical oxidation studies [20, 22]. While SASA values may be determined directly from any 3D structure, to account for side chain motion in solution at room temperature, we believe a more robust approach is to determine average <SASA> values by MD simulation of the protein, rather than from a single static structure, such as a homology model or x-ray structure. As indicated from the standard deviations in the <SASA> values computed from the MD data (Table 2.4), side chain motion can lead to significant variations, which are not taken into account when only a single structure is employed in the analysis.

**Relationship between side chain oxidation levels and solvent accessibility**

From 1000 snapshots extracted from a 10 ns MD simulation of the galectin-1 homodimer, the average <SASA> of each amino acid side chain for all 134 residues was calculated (Figure 2.5 and Table 2.4). To quantify the correlation between the observed oxidation levels and the calculated <SASA> values, the experimental and theoretical data were compiled for each amino acid side chain. By plotting oxidation level versus <SASA> value, for each side chain of the same type, a linear relationship between <SASA> and extent of oxidation was observed (Figure 2.6). Additionally, for most side chains, a minimum solvent exposure appeared to be required to enable any oxidation to occur. The level of exposure required varied according to the side chain, from a low of 2 Å² for leucine to a high of approximately 48 Å² for proline. The relationship
naturally also depends on the chemical susceptibility of the side chain to oxidation, and to a far lesser extent on the local primary sequence [80]. Additionally, under the conditions employed here, some side chains remained essentially unoxidized regardless of their exposure. From the slopes of the plots of \(<\text{SASA}>\) versus percent oxidation, the side chains could be divided into categories according to their reactivity (high, medium or low). For example, highly reactive residues such as phenylalanine could be characterized by 12.9 % oxidation/Å², whereas a medium reactivity residue, such as arginine or proline would result in only 1-2 % oxidation/Å². This simple analysis neglects the precise details of the atomic site of reactivity, but nevertheless demonstrated the direct dependence of oxidation level on side chain exposure. From each reactivity group, two plots of the oxidation level dependence on the \(<\text{SASA}>\) are shown in Figure 2.6. The extent to which these relationships are transferable to other proteins is currently under investigation. However, the analysis presented here agrees well with earlier characterization of side chain reactivity to hydroxyl radical [6, 14]. Generally highly reactive amino acids include cysteine, methionine, the aromatics and leucine. The medium reactivity category includes arginine and proline, while the remaining amino acids such as asparagine, aspartic acid, serine, lysine, glutamine, glutamic acid, threonine and alanine have low reactivities. Having derived a quantitative relationship between SASA and percent oxidation, it is possible in principle to predict the extent of oxidation that should be observed, given the 3D structure of a protein. To test the accuracy of such predictions, as well as the sensitivity of the experimental method to 3D structure, these techniques were applied to the characterization of the interface region between monomeric domains in galectin-1.
**Characterization of the galectin-1 dimer interface**

Most amino acid side chains were found to be reactive to some extent under the current experimental conditions (Table 2.1); however, the following residues were consistently observed to be sensitive to oxidation: cysteine, methionine, tryptophane, tyrosine, phenylalanine, histidine, leucine, valine, proline, arginine. Although an experiment, in which differences in oxidation levels between the monomer and dimer were observed, would be well suited to characterize the dimer interface, galectin-1 cannot be obtained in monomeric form. In contrast, computationally, both species are amenable to examination. Thus, by examining the difference in the computed <SASA> values from independent 10 ns MD simulations of the dimer and monomer, the interface region was readily identified (residues 1-9 and 128-143), in agreement with those residues identified in the x-ray structure of the dimer [71]. A subset of reporter residues in the interface (L4, V5, L9, I128, V131 and F133) were identified on the basis that they were predicted to be at least partially shielded from solvent in the interface, and that were likely to be reactive. The observed oxidation levels for these residues are presented in Table 2.5 and illustrated in Figure 2.4B. With the exception of F133, which was not identified in the MS/MS data, the oxidation levels were markedly lower for these interfacial residues than would be expected for exposed residues. Notably, the predicted oxidation levels for the dimer, based on the relationships derived above, were in good agreement with the observed values.

The observation that L4 and V5 are significantly oxidized (56% and 19%, respectively) despite being associated with the interface [71], suggests that easily oxidized side chains must be highly excluded from solvent in order to prevent their oxidation, under these reaction conditions. Both L4 and V5 have reduced solvent accessibilities relative to their <SASA> in the galectin-1 monomeric domain (approximately 75% and 50%, respectively); however, this is not sufficient
to prevent their partial oxidation. By employing the linear relationship between percentage oxidation and SASA it is possible to predict the level of oxidation for each of the residues in the interface (Table 2.5). Notably the predicted oxidation levels for L4 (44%) and V5 (18%) are in good agreement with the observed values. These relationships need to be confirmed by analysis of further proteins; however, they appear to offer a promising approach to relating observed oxidation to 3D structure.

As expected, residues in the interface with very low $<\text{SASA}>$ values were not found to be oxidized, for example L9, I128 and V131. Other residues that are associated with the dimer interface, but which have very low reactivity levels did not get oxidized regardless their solvent exposure (A1, A6, S7, N8, K129, E134), which illustrates the benefit of identifying a sub-set of reporter residues prior to attempting to interpret the oxidation data purely in terms of 3D structure effects.

**Conclusions**

A combined experimental and theoretical approach directly applicable to protein surface mapping has been presented. A linear relationship between the level of side chain oxidation and average $<\text{SASA}>$, derived from MD simulation data, has been derived. Notably, each side chain has been found to require a minimum level of exposure before it will be oxidized under the mild conditions employed here. By considering this minimal exposure requirement, with the known side chain reactivities, potential reporter groups may be identified. Knowledge of the expected reporter groups can be used to provide an estimate of the expected surface oxidation levels and is therefore of significance to the study of protein-protein and protein-ligand interactions. Although not all amino acid side chains react under the conditions presented here, the level of
surface coverage, in terms of those residues that were exposed to solvent and that were not inert to oxidation was approximately 70%. This level of coverage is far greater than would be achieved using traditional chemical derivatization methods, such as biotinylation or acetylation, wherein only a few residues can act as reporter groups. Equally significant to good coverage is the ability to quantify the level of oxidation and relate that directly to per-residue <SASA> values. This ability significantly elevates the level of footprint resolution, which is key to the practical application of this method in characterizing protein complexes. Such quantification facilitates the identification of occluded surfaces and should provide a powerful tool for determining the 3D structures of complexes that are not amenable to analysis by traditional experimental structural methods.

Recent studies have employed oxidative footprinting data to guide computational molecular docking, however it was concluded that footprinting data alone was insufficient to provide a definitive choice of model [27]. It is anticipated that the quantitative relationships derived here to relate oxidation levels to 3D structure will be valuable in such approaches. Additionally, we have demonstrated that given an accurate structure for the complex, it is possible to use <SASA> values to predict oxidation levels that are in good agreement with experimental data. Thus, data from oxidative footprinting are likely to be particularly useful in validating independently generated molecular complexes, by for example automated docking procedures.
Acknowledgements

We thank the National Institutes of Health (RR05357 and GM55230) for financial support. We would also like to thank to M. A. Duncan (UGA) for providing access to the excimer laser and T. Cheng (UGA) for helping with its set up. R.J.W wishes to thank M. R. Chance (Case Western) for useful discussions.
Table 2.1. Observed\(^a\) mass changes from side chain oxidation.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Potential mass changes</th>
<th>Associated modifications [6, 68]</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-16; +32, +48</td>
<td>hydroxy, sulfonic acid</td>
</tr>
<tr>
<td>M</td>
<td>+16; -32; +32</td>
<td>sulfoxide, aldehyde, sulfone</td>
</tr>
<tr>
<td>W</td>
<td>+16, +32, +48; +20</td>
<td>hydroxy, ring opening, hydroxykynurenin</td>
</tr>
<tr>
<td>Y</td>
<td>+16, +32, +48</td>
<td>hydroxy</td>
</tr>
<tr>
<td>F</td>
<td>+16, +32, +48</td>
<td>hydroxy</td>
</tr>
<tr>
<td>H</td>
<td>+16; +5, -10, -23, -22</td>
<td>oxo, ring opening</td>
</tr>
<tr>
<td>L</td>
<td>+16; +14</td>
<td>hydroxy, carbonyl</td>
</tr>
<tr>
<td>I(^b)</td>
<td>+16; +14</td>
<td>hydroxy, carbonyl</td>
</tr>
<tr>
<td>V</td>
<td>+16; +14</td>
<td>hydroxy, carbonyl</td>
</tr>
<tr>
<td>P</td>
<td>+16; +14; -30</td>
<td>hydroxy, carbonyl, pyrrolidone</td>
</tr>
<tr>
<td>R</td>
<td>-43; +16; +14</td>
<td>deguanidination, hydroxy, carbonyl</td>
</tr>
<tr>
<td>K(^b)</td>
<td>+16; +14</td>
<td>hydroxy, carbonyl</td>
</tr>
<tr>
<td>E</td>
<td>-30; +16; +14</td>
<td>decarboxylation, hydroxy, carbonyl</td>
</tr>
<tr>
<td>Q</td>
<td>+16; +14</td>
<td>hydroxy, carbonyl</td>
</tr>
<tr>
<td>D(^b)</td>
<td>-30; +16</td>
<td>decarboxylation, hydroxy</td>
</tr>
<tr>
<td>N(^b)</td>
<td>+16</td>
<td>hydroxy</td>
</tr>
<tr>
<td>S</td>
<td>+16; -2</td>
<td>hydroxy; - H(_2)O</td>
</tr>
<tr>
<td>T(^b)</td>
<td>-2</td>
<td>- H(_2)O</td>
</tr>
<tr>
<td>A(^b)</td>
<td>+16</td>
<td>hydroxy</td>
</tr>
</tbody>
</table>

\(^a\)Observed mass changes shown in boldface
\(^b\)Low level of oxidation observed (< 2%)
Table 2.2. Synthetic peptides used in quantification studies.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
<th>m/z</th>
<th>Mass Change</th>
<th>Oxidation Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LHF NPR</td>
<td>782.91</td>
<td></td>
<td>Non-oxidized</td>
</tr>
<tr>
<td>2</td>
<td>LHYNPR</td>
<td>798.91</td>
<td>+16 Da</td>
<td>F3Y</td>
</tr>
<tr>
<td>3</td>
<td>LDFNPR</td>
<td>760.85</td>
<td>-22 Da</td>
<td>H2D</td>
</tr>
</tbody>
</table>
Table 2.3. Oxidation level equivalents for mixtures of synthetic peptides.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pep-1 and Pep-2 F oxid – 40%</th>
<th>Pep-1 and Pep-3 H oxid – 66%</th>
<th>Pep-1, pep-2 and pep-3 F and H oxid – 25 % each</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m/z 799 oxidation (%)</td>
<td>m/z 761 oxidation (%)</td>
<td>m/z 799 oxidation (%)</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>41.9 (1.2)</td>
<td>64.1 (2.8)</td>
<td>24.5 (0.8)</td>
</tr>
<tr>
<td>LC-MS</td>
<td>42.8 (2.1)</td>
<td>66.4 (6.6)</td>
<td>23 (1.2)</td>
</tr>
</tbody>
</table>

*aStandard deviation in parentheses
Table 2.4. Per-residue oxidation levels and side chain <SASA> values for galectin-1.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Oxidation [%]</th>
<th>&lt;SASA&gt; [Å²]</th>
<th>Residue</th>
<th>Oxidation [%]</th>
<th>&lt;SASA&gt; [Å²]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dimer</td>
<td>dimer</td>
<td>monomer</td>
<td>dimer</td>
<td>dimer</td>
</tr>
<tr>
<td>A1</td>
<td>0.0 (0.0)</td>
<td>57 (14)</td>
<td>82 (2)</td>
<td>K28</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A6</td>
<td>0.1 (0.1)</td>
<td>8 (4)</td>
<td>34 (4)</td>
<td>K36</td>
<td>0.1 (0.0)</td>
</tr>
<tr>
<td>A24</td>
<td>0.0 (0.0)</td>
<td>38 (8)</td>
<td>17 (6)</td>
<td>K63</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A27</td>
<td>0.0 (0.0)</td>
<td>10 (4)</td>
<td>0 (0)</td>
<td>K99</td>
<td>n.d.</td>
</tr>
<tr>
<td>A51</td>
<td>0.0 (0.0)</td>
<td>18 (6)</td>
<td>21 (4)</td>
<td>K107</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A55</td>
<td>0.0 (0.0)</td>
<td>40 (6)</td>
<td>46 (4)</td>
<td>K127</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A67</td>
<td>0.0 (0.0)</td>
<td>62 (4)</td>
<td>64 (2)</td>
<td>K129</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A76</td>
<td>0.0 (0.1)</td>
<td>44 (12)</td>
<td>41 (5)</td>
<td>L4</td>
<td>56.1 (12.4)</td>
</tr>
<tr>
<td>A94</td>
<td>n.d.</td>
<td>57 (5)</td>
<td>57 (3)</td>
<td>L9</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A116</td>
<td>0.0 (0.0)</td>
<td>35 (11)</td>
<td>27 (6)</td>
<td>L11</td>
<td>13.3 (8.6)</td>
</tr>
<tr>
<td>A121</td>
<td>1.9 (3.1)</td>
<td>11 (3)</td>
<td>11 (2)</td>
<td>L17</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A122</td>
<td>0.0 (0.0)</td>
<td>6 (5)</td>
<td>4 (1)</td>
<td>L32</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>A132</td>
<td>1.5 (2.6)</td>
<td>15 (6)</td>
<td>50 (4)</td>
<td>L34</td>
<td>0.0 (0.1)</td>
</tr>
<tr>
<td>C2</td>
<td>41.9 (9.7)</td>
<td>85 (9)</td>
<td>71 (4)</td>
<td>L41</td>
<td>12.5 (7.2)</td>
</tr>
<tr>
<td>C16</td>
<td>18.7 (4.0)</td>
<td>29 (10)</td>
<td>34 (9)</td>
<td>L43</td>
<td>20.7 (6.6)</td>
</tr>
<tr>
<td>C42</td>
<td>54.9 (7.8)</td>
<td>1 (2)</td>
<td>0 (1)</td>
<td>L96</td>
<td>n.d.</td>
</tr>
<tr>
<td>C60</td>
<td>5.8 (1.4)</td>
<td>2 (2)</td>
<td>0 (0)</td>
<td>L100</td>
<td>8.9 (7.8)</td>
</tr>
<tr>
<td>C88</td>
<td>n.d.</td>
<td>23 (8)</td>
<td>15 (5)</td>
<td>L112</td>
<td>n.d.</td>
</tr>
<tr>
<td>C130</td>
<td>0.0 (0.0)</td>
<td>25 (8)</td>
<td>68 (7)</td>
<td>M114</td>
<td>94.0 (2.5)</td>
</tr>
<tr>
<td>D26</td>
<td>0.0 (0.0)</td>
<td>100 (6)</td>
<td>88 (12)</td>
<td>M120</td>
<td>83.8 (10.2)</td>
</tr>
<tr>
<td>D37</td>
<td>2.6 (0.5)</td>
<td>55 (8)</td>
<td>50 (5)</td>
<td>N8</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>D54</td>
<td>0.0 (0.0)</td>
<td>23 (4)</td>
<td>16 (2)</td>
<td>N10</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>D64</td>
<td>0.0 (0.0)</td>
<td>65 (9)</td>
<td>76 (5)</td>
<td>N33</td>
<td>0.0 (0.1)</td>
</tr>
<tr>
<td>D92</td>
<td>n.d.</td>
<td>67 (6)</td>
<td>70 (5)</td>
<td>N39</td>
<td>0.9 (1.3)</td>
</tr>
<tr>
<td>D95</td>
<td>n.d.</td>
<td>45 (8)</td>
<td>38 (5)</td>
<td>N40</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>D102</td>
<td>1.5 (1.3)</td>
<td>98 (16)</td>
<td>72 (4)</td>
<td>N46</td>
<td>0.6 (1.1)</td>
</tr>
<tr>
<td>D123</td>
<td>1.0 (0.9)</td>
<td>38 (10)</td>
<td>65 (5)</td>
<td>N50</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>D125</td>
<td>0.0 (0.0)</td>
<td>69 (9)</td>
<td>29 (6)</td>
<td>N56</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>E15</td>
<td>0.0 (0.0)</td>
<td>68 (30)</td>
<td>37 (11)</td>
<td>N61</td>
<td>0.0 (0.1)</td>
</tr>
<tr>
<td>E22</td>
<td>0.0 (0.0)</td>
<td>86 (13)</td>
<td>67 (10)</td>
<td>N65</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>E71</td>
<td>0.1 (0.2)</td>
<td>66 (8)</td>
<td>49 (4)</td>
<td>N110</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>E74</td>
<td>10.4 (11.1)</td>
<td>37 (10)</td>
<td>30 (5)</td>
<td>N113</td>
<td>0.0 (0.1)</td>
</tr>
<tr>
<td>E86</td>
<td>n.d.</td>
<td>50 (14)</td>
<td>35 (8)</td>
<td>N118</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>E105</td>
<td>0.0 (0.0)</td>
<td>75 (10)</td>
<td>77 (9)</td>
<td>P13</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>E115</td>
<td>0.0 (0.0)</td>
<td>104 (15)</td>
<td>119 (9)</td>
<td>P25</td>
<td>55.1 (9.0)</td>
</tr>
<tr>
<td>E134</td>
<td>n.d.</td>
<td>95 (14)</td>
<td>77 (18)</td>
<td>P47</td>
<td>0.1 (0.1)</td>
</tr>
<tr>
<td>F30</td>
<td>0.1 (0.1)</td>
<td>7 (3)</td>
<td>2 (2)</td>
<td>P75</td>
<td>91.6 (7.5)</td>
</tr>
<tr>
<td>F45</td>
<td>18.5 (12)</td>
<td>7 (5)</td>
<td>2 (2)</td>
<td>P78</td>
<td>n.d.</td>
</tr>
<tr>
<td>F49</td>
<td>3.5 (2.8)</td>
<td>8 (3)</td>
<td>10 (2)</td>
<td>P81</td>
<td>n.d.</td>
</tr>
<tr>
<td>F77</td>
<td>94.8 (7.6)</td>
<td>14 (7)</td>
<td>15 (3)</td>
<td>P101</td>
<td>7.4 (4.1)</td>
</tr>
<tr>
<td>F79</td>
<td>n.d.</td>
<td>10 (6)</td>
<td>3 (1)</td>
<td>P109</td>
<td>35.9 (7.4)</td>
</tr>
<tr>
<td>Residue</td>
<td>Oxidation [%]</td>
<td>&lt;SASA&gt; [Å²]</td>
<td>Residue</td>
<td>Oxidation [%]</td>
<td>&lt;SASA&gt; [Å²]</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>dimer</td>
<td>dimer</td>
<td>monomer</td>
<td>dimer</td>
<td>dimer</td>
</tr>
<tr>
<td>F91</td>
<td>n.d.</td>
<td>4 (2)</td>
<td>1 (1)</td>
<td>Q80</td>
<td>n.d.</td>
</tr>
<tr>
<td>F106</td>
<td>n.d.</td>
<td>18 (6)</td>
<td>14 (2)</td>
<td>Q93</td>
<td>n.d.</td>
</tr>
<tr>
<td>F108</td>
<td>n.d.</td>
<td>5 (3)</td>
<td>0 (0)</td>
<td>R20</td>
<td>98.9 (0.4)</td>
</tr>
<tr>
<td>F126</td>
<td>0.0 (0.0)</td>
<td>6 (5)</td>
<td>6 (3)</td>
<td>R48</td>
<td>6.1 (1.7)</td>
</tr>
<tr>
<td>F133</td>
<td>n.d.</td>
<td>21 (9)</td>
<td>70 (13)</td>
<td>R73</td>
<td>n.d.</td>
</tr>
<tr>
<td>G3</td>
<td>0.0 (0.0)</td>
<td>8 (3)</td>
<td>23 (2)</td>
<td>R111</td>
<td>54.4 (9.1)</td>
</tr>
<tr>
<td>G14</td>
<td>0.0 (0.0)</td>
<td>36 (4)</td>
<td>31 (2)</td>
<td>S7</td>
<td>5.1 (1.4)</td>
</tr>
<tr>
<td>G21</td>
<td>0.0 (0.0)</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>S29</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>G35</td>
<td>0.0 (0.0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>S38</td>
<td>35.9 (2.0)</td>
</tr>
<tr>
<td>G53</td>
<td>0.0 (0.0)</td>
<td>38 (4)</td>
<td>41 (1)</td>
<td>S62</td>
<td>4.2 (0.4)</td>
</tr>
<tr>
<td>G66</td>
<td>0.0 (0.0)</td>
<td>19 (5)</td>
<td>10 (2)</td>
<td>S83</td>
<td>n.d.</td>
</tr>
<tr>
<td>G69</td>
<td>0.0 (0.0)</td>
<td>20 (3)</td>
<td>18 (1)</td>
<td>T57</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>G82</td>
<td>n.d.</td>
<td>14 (4)</td>
<td>18 (3)</td>
<td>T70</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>G103</td>
<td>0.0 (0.0)</td>
<td>31 (4)</td>
<td>23 (4)</td>
<td>T84</td>
<td>n.d.</td>
</tr>
<tr>
<td>G124</td>
<td>0.0 (0.0)</td>
<td>11 (6)</td>
<td>24 (2)</td>
<td>T90</td>
<td>n.d.</td>
</tr>
<tr>
<td>H44</td>
<td>4.5 (1.1)</td>
<td>33 (8)</td>
<td>27 (4)</td>
<td>T97</td>
<td>n.d.</td>
</tr>
<tr>
<td>H52</td>
<td>n.d.</td>
<td>116 (17)</td>
<td>113 (8)</td>
<td>V5</td>
<td>18.6 (8.6)</td>
</tr>
<tr>
<td>H72</td>
<td>92.1 (14.8)</td>
<td>74 (11)</td>
<td>50 (5)</td>
<td>V19</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>H104</td>
<td>5.0 (5.7)</td>
<td>62 (19)</td>
<td>87 (11)</td>
<td>V23</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>I58</td>
<td>0.0 (0.0)</td>
<td>2 (2)</td>
<td>0 (0)</td>
<td>V31</td>
<td>15.8 (1.3)</td>
</tr>
<tr>
<td>I89</td>
<td>n.d.</td>
<td>8 (5)</td>
<td>3 (2)</td>
<td>V59</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>I98</td>
<td>n.d.</td>
<td>3 (3)</td>
<td>0 (1)</td>
<td>V85</td>
<td>n.d.</td>
</tr>
<tr>
<td>I117</td>
<td>0.0 (0.0)</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>V87</td>
<td>n.d.</td>
</tr>
<tr>
<td>I128</td>
<td>0.0 (0.0)</td>
<td>14 (10)</td>
<td>13 (3)</td>
<td>V131</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>K12</td>
<td>0.0 (0.0)</td>
<td>107 (15)</td>
<td>97 (10)</td>
<td>W68</td>
<td>96.9 (3.3)</td>
</tr>
<tr>
<td>K18</td>
<td>0.0 (0.0)</td>
<td>64 (10)</td>
<td>109 (9)</td>
<td>Y119</td>
<td>17.3 (22.1)</td>
</tr>
</tbody>
</table>

\(^a\) Standard deviation in parentheses

\(^b\) Oxidation level not determined due to excessive uncertainty in measurement or peak assignment
Table 2.5. Experimentally-measured and \(<\text{SASA}>\)-derived oxidation levels for reporter residues in the galectin-1 dimer interface.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Oxidation [%]</th>
<th>&lt;SASA&gt; [Å²]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>measured</td>
<td>predicted</td>
</tr>
<tr>
<td>L4</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>L9</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>V5</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>V131</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>I128</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 2.1. Peptide-1 before (A) and after oxidation (B), indicating oxidation of phenylalanine (m/z 799) and histidine (m/z 761 and 773).
Figure 2.2. MS spectrum of tryptic digestion of galectin-1 before oxidation (A), and after oxidation (B). Oxidation leads to more facile proteolysis, formation of peaks from oxidation products, and a decrease in the abundance of non-oxidized peptides (see peak indicated by arrows).
Figure 2.3. Tandem mass spectra of a tryptic peptide from galectin-1 in non-oxidized (A) and oxidized (B) forms indicating peaks associated with a 22 Da mass loss, typical of histidine oxidation.
Figure 2.4. Per-residue oxidation levels plotted on the solvent accessible surface of dimeric galectin-1 (A). Galectin-1 dimer interface in detail. Reporter residues are indicated in shades of red according to the measured level of side chain oxidation. From left: ribbon structure of the homodimer, solvent accessible surface structure of the homodimer, and monomeric subunit illustrating interfacial residues (B).
Figure 2.5. Solvent accessible surface area (<SASA>) for each of the 134 side chains in galectin-1. Values for monomer (pink) and dimer (black) domains are reported. Shaded regions indicate the dimer interface (blue). Other regions of low exposure include the sheets associated with the β-sandwich structure (yellow).
Figure 2.6. Percentage of oxidation versus side chain <SASA> for selected amino acids. Highly reactive side chains (A, B) are characterized by steep slopes (%Ox/Å$^2$) and low minimal exposure values (Å$^2$). Moderately reactive side chains (C, D) require significant exposure before oxidization may occur. Side chains that are inert under these experimental conditions (E, F) show no sensitivity to level of exposure.
Example of quantitation:

\[ \text{EPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 1} \]
\[ \text{EPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 2} \]
\[ \text{EPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 3} \]
\[ \text{EPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 4} \]
\[ \text{EPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 5} \]
\[ \text{EPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 6} \]
\[ \text{DNGAWGTEHREPAPFPQPGSTVEVCITFDQADLTIK} \quad \text{Intensity 7} \]

\[ \%_{\text{ox}}(\text{Phe1}) = (I_2+I_4+I_6) / (I_1+I_2+I_3+I_4+I_5+I_6+I_7) \]
\[ \%_{\text{ox}}(\text{Phe2}) = (I_3+I_5+I_6+I_7) / (I_1+I_2+I_3+I_4+I_5+I_6+I_7) \]
\[ \%_{\text{ox}}(\text{Phe3}) = (I_3+I_4+I_5+I_6+I_7) / (I_1+I_2+I_3+I_4+I_5+I_6+I_7) \]
\[ \%_{\text{ox}}(\text{Ala1}) = 0 / (I_1+I_2+I_3+I_4+I_5+I_6+I_7) \]

**Figure S2.1.** A representative example of the label-free quantitation procedure, using LC-MS signal intensities.
**Figure S2.2.** Silver-stained SDS-PAGE gel of reduced galectin-1 (monomer molecular weight 14.5 kDa) indicating effect of multiple laser exposures and variation in hydrogen peroxide concentration. Lane 1: No laser exposure. Lane 2: 1 laser pulse, 1% H$_2$O$_2$. Lane 3: 1 laser pulse, 2% H$_2$O$_2$. Lane 4: 1 laser pulse, 3% H$_2$O$_2$. Lane 5: 5 laser pulses, 1% H$_2$O$_2$. Lane 6: 15 laser pulses, 1% H$_2$O$_2$. Lane 7: 25 laser pulses, 1% H$_2$O$_2$. 
CHAPTER 3

PULSED ELECTRON BEAM WATER RADIOLYSIS FOR SUBMICROSECOND HYDROXYL RADICAL PROTEIN FOOTPRINTING\(^2\)

Contributions: Computational analysis of solvent accessibilities of the protein residues: molecular dynamics simulations and average solvent accessibility surface area calculations.

Abstract

Hydroxyl radical footprinting is a valuable technique for studying protein structure, but care must be taken to ensure that the protein does not unfold during the labeling process due to oxidative damage. Footprinting methods based on submicrosecond laser photolysis of peroxide that complete the labeling process faster than the protein can unfold have been recently described; however, the mere presence of large amounts of hydrogen peroxide can also cause uncontrolled oxidation and minor conformational changes. We have developed a novel method for submicrosecond hydroxyl radical protein footprinting using a pulsed electron beam from a 2 MeV Van de Graaff electron accelerator to generate a high concentration of hydroxyl radicals by radiolysis of water. The amount of oxidation can be controlled by buffer composition, pulsewidth, dose, and dissolved nitrous oxide gas in the sample. Our results with ubiquitin and β-lactoglobulin A demonstrate that one submicrosecond electron beam pulse produces extensive protein surface modifications. Highly reactive residues that are buried within the protein structure are not oxidized, indicating that the protein retains its folded structure during the labeling process. Time-resolved spectroscopy indicates that the major part of protein oxidation is complete in a time scale shorter than that of large scale protein motions.
Introduction

Mass spectrometry is becoming a widely used technique for structural analysis of proteins. Radical-based surface mapping techniques, such as hydroxyl radical protein footprinting, followed by mass spectrometry analysis has become increasingly popular for studying protein structure, protein-protein and protein-ligand interaction interfaces [5]. Hydroxyl radicals have gained popularity as a labeling technique because they provide a fast, relatively nonspecific, covalent label that probes a variety of solvent accessible amino acid residues with one experiment [50]. Proteins experience structural distortion due to conformational changes, multimerization, ligand binding, and aggregation, and these structural changes expose different solvent accessible surfaces of the protein. By labeling a protein with hydroxyl radicals before and after inducing a structural change, a mass spectrometric comparison can be used to determine protection or exposure of specific residues of each protein conformation [5, 81]. However, common methods of production of hydroxyl radicals produce undesired side reactions that complicate analysis. Here a superior method with few uncontrolled side reactions is presented.

Hydroxyl radicals for labeling of proteins have been successfully produced by Fenton chemistry [49, 82], hydrogen peroxide (H$_2$O$_2$) photolysis [19, 20, 22], and water radiolysis by either gamma (γ) rays [83-87] or exposure to an X-ray synchrotron beam [5, 41, 81, 88-90]. Protein oxidation is a labeling process that is dependent on the solvent-accessible surface and the reactivity of the exposed residues [5, 35, 49, 91], although a small dependence on the local sequence has been noted [20]. 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 [21, 92, 93].
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. Most hydroxyl radical-generation techniques, including water radiolysis by
gamma (γ) rays and X-rays, are performed on a millisecond to minute timescale, making it more
likely that the protein will experience unfolding due to modifications [21, 92]. Most previous
strategies for ensuring that the native conformation is exclusively probed include the use of
circular dichroism to determine when conformational changes occur in order to label the protein
before structural changes can be detected [11], limiting the amount of oxidation to where the vast
majority of protein molecules have one or no oxidations [20, 49], or monitoring of the reaction
kinetics to ensure the oxidation of the protein [93] or its constituent peptides [41, 87, 91] follows
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, and
this limited amount of oxidation results in fewer oxidation sites and lower structural resolution.

A relatively new strategy for ensuring that the native conformation of the protein is
probed that still allows for high levels of oxidation is to ensure that all oxidation is completed on
a faster timescale than the protein can unfold by using a very short UV laser pulse to photolyse
H₂O₂. This can complete the labeling process on a sub-microsecond timescale [22, 23]. The
timescale for large scale protein motions, such as helix coiling/uncoiling, is in the long
microseconds to milliseconds range [94], therefore, oxidatively labeling a protein at or below
this time frame without the use of a precursor oxidant will prevent labeling of the oxidatively
unfolded conformation. However, hydrogen peroxide-based methods of oxidation, including
Fenton chemistry [49, 82, 95] and UV photolysis [11, 20, 22, 23], all have several problems
stemming from the presence of H₂O₂. The presence of H₂O₂ in a protein solution can induce
uncontrolled Fenton-like chemistry in redox-active metal-binding proteins [96, 97], and cysteine and methionine two-electron oxidation can proceed spontaneously [98]. Furthermore, H$_2$O$_2$ has different physical properties than water and may induce protein conformational changes independent of oxidation.

Here we demonstrate the importance of limiting the exposure to hydrogen peroxide using an NMR based assessment of its effects on a model cysteine containing protein (Galectin-3). We then introduce a novel pulsed electron beam water radiolysis technique for hydroxyl radical protein footprinting that does not require hydrogen peroxide. In this method, hydroxyl radicals are generated in less than 1 µs using a 2 MeV Van de Graaff electron accelerator. Using a single electron beam pulse to oxidize the protein sample and subsequently scavenging secondary oxidants using methionine amide, protein oxidation is complete before the protein conformation can change. Ubiquitin, a small protein that ionizes efficiently, was used as a model to determine if controllable protein oxidation could be achieved by this water radiolysis method. Furthermore, β-lactoglobulin A was used as a model protein for method development because of its sensitivity to oxidation-induced conformational changes [93] and the availability of a high-resolution X-ray crystal structure [99]. In order to account for the motion of the side chains, molecular dynamics (MD) simulations of the protein were performed. These simulations enable the average solvent accessibilities to be computed for each side chain, and thus in principle provide a more complete model for the protein surface, under the conditions of the experiment, than possible from a single static crystal structure [28, 35]. We also determined the half-life of the hydroxyl radical label in anoxic solution with the protein by time-resolved UV spectroscopy to ensure that the label is consumed on a timescale consistent with large scale protein motions.
The detected sites of oxidation are compared with residues known to be present on the surface of the natively-folded protein to ensure that we are not labeling an oxidatively-unfolded protein.

**Materials and Methods**

**Materials**

All solvents were purchased from Sigma-Aldrich (St. Louis, MO), unless otherwise noted, at the highest purity available and used as supplied without further purification. Methionine amide and ammonium phosphate were purchased from Bachem (Torrance, CA) and J.T. Baker (Phillipsburg, NJ), respectively. Deionized water (18 MΩ) was prepared in-house with a Millipore Mill-Q water purification system (Millipore Bedford, MA). Ubiquitin (Sigma-Aldrich) was prepared at 15 µM in 20 mM sodium phosphate buffer. β-lactoglobulin A (Sigma-Aldrich) was prepared at 20 µM in 20 mM sodium phosphate, ammonium bicarbonate, and ammonium phosphate (J.T.Baker, Phillipsburg, NJ) buffers.

**NMR spectroscopy**

Heteronuclear single quantum coherence (HSQC) spectra of the $^{15}$N-labeled galectin-3 samples were acquired at 600 MHz using a gradient enhanced version of a standard HSQC experiment (gNHSQC from the Varian BioPack). A small number of t1 points (32) and a minimal number of scans were used to reduce acquisition times to three minutes and still allow observation at a reasonable S/N ratio. Data were processed using nmrPipe and plotted using nmrDraw [100]. Unambiguous resonance assignments were taken from BMRB deposition no. 4909 [101].
**Electron pulse irradiation**

20 µM buffered (pH=7) protein solutions (ubiquitin or β-lactoglobulin A) were subjected to electron pulses from Radiation Laboratory 2 MeV Van de Graaff accelerator. Solutions were saturated prior irradiation either with air or a gas mixture of nitrous oxide/oxygen (N₂O/O₂; 4/1 v/v Mittler Supply Inc., ultrahigh purity, South Bend, IN). For initial studies with ubiquitin the dose was changed either by varying the beam current, via changes in heater settings of the dispenser cathode, at fixed electron pulsewidth (660 ns) or by varying the electron pulsewidth (480–660 ns) at fixed beam current. Typically to attain high levels of labeled protein, the dose had to be around 300 Gy achieved by a heater setting of 15 and we routinely used a dose in this range. A 250 µl syringe (Hamilton, Reno, NV, custom grinded to minimize the loss of electrons due to scattering in the glass wall) was placed in the front of the electron beam exit window. The beam was focused to a 2 mm diameter and was spatially adjusted to hit the very front volume of the syringe (Supplemental Data). From the glass coloration after irradiation we could determine the spread of the beam in the sample cell and estimate the total irradiated volume to be about 5 µl. After each pulse of electrons the piston was moved to release irradiated solution plus a 0.5 µL volume of unirradiated solution and refill the irradiation volume with the fresh protein solution. The released solution was directed to an empty 1 mL centrifuge vial or a 1 mL centrifuge vial prior filed with 50 µl of buffered (pH=7) 20 mM methionine amide (Bachem) to stop any chain oxidation processes. After 10 subsequent electron pulses the vial containing 100 µl of 10 µM oxidized protein solution was stopped, frozen and stored awaiting labeling analysis. In initial studies the frequency of pulses and corresponding flow in the irradiated volume was varied to establish contribution of secondary irradiation before contact with the quencher. Electron pulse frequency of 1 Hz was used for results presented in this report. The doses applied
to the protein solutions were estimated with the Fricke dosimeter [102]. Oxygen saturated solutions of Fricke dosimeter were irradiated at the same conditions like the protein samples (beam current, electron pulsewidth, cell position) after protein oxidation process was completed. Over twelve months of studies, the dose varied in the range of 30% at initially established accelerator settings i.e. the beam current and the electron pulsewidth.

**Time-resolved UV spectroscopy**

Pulse radiolysis experiments were performed using 100–1500 ns pulses from the Radiation Laboratory 8 MeV electron linac to obtain the range of doses comparable with doses applied in other protein oxidation experiments. Analyzing light from a pulsed 75 W xenon lamp (Photon Technology International, Birmingham, NJ) was selected using monochromator SPEX-270M. UV kinetics were measured at 250 nm where hydroxyl radicals extinction coefficient [103] is $\varepsilon=535 \, \text{M}^{-1} \, \text{cm}^{-1}$. Samples of protein were prepared by dissolving $\beta$-lactoglobulin A to a concentration of 20 $\mu$M in ammonium phosphate buffer (pH=7). Buffer solutions were prepared in deionized water (18.2 M$\Omega$-cm, Barnstead Nanopure System, Dubuque, IA). Prior to experiment, samples were bubbled with $\text{N}_2\text{O}$ (Mittler Supply Inc., Ultrahigh Purity) at ambient conditions resulting in a final concentration of $\text{N}_2\text{O}$ of 25 mM. To improve the signal-to-noise ratio, UV signals were averaged over 15 consecutive traces. Due to transient absorption of quartz in the flow sample cell after the pulse of electrons (pronounced mostly at the highest doses) the blank traces for a given dose were collected in the empty cell purged with argon gas (Mittler Supply Inc., Ultrahigh Purity). These blank traces were subtracted from the corresponding sample traces before kinetic analysis. The UV transient absorption traces were fit using a system of first order kinetic differential equations model (fitting code written in Igor Pro
5.00, WaveMetrics Inc., Lake Oswego, OR) that incorporates most important reactions expected after pulse radiolysis of β-lactoglobulin A in N₂O saturated solutions. The initial concentration of hydroxyl radicals after the radiation pulse was taken as a sum of initial yields of hydroxyl radicals and hydrated electrons given that hydrated electrons convert to hydroxyl radicals within the duration of the pulse in N₂O saturated solutions (see reaction (3.7)). Extinction coefficients of hydrated electrons and hydroxyl radicals at 250 nm are comparable within the experimental error [104]. Thus, their rapid inter-conversion occurring in the first couple hundreds of nanoseconds does not obstruct the overall kinetic analysis extending up to 40 µs after the irradiation pulse. A fitting program was setup to fit all collected UV traces at once fitting the reaction rate of hydroxyl radicals with protein and the resulting product extinction coefficient, while keeping most other (known) parameters of the model fixed. For simplicity the rate constant of the protein radical decay was assumed to be first order since this reaction is many orders of magnitude slower than the initial decay of hydroxyl radicals in the monitored transient absorption time window.

**LC-MS of oxidized protein**

Intact ubiquitin samples were analyzed using a hand pulled fused silica (Technologies, Phoenix, AZ) spray column (75 µm × 10 cm; tip 15 ± 1 µm) that was prepared by packing silica C₁₈ resins (Rainin Microsorb MV, 5 µm, 300 Å pore size) from a 50% isopropanol and 50% methanol slurry into the pulled fused silica capillary using a pressurized stainless steel bomb. Prior to reverse-phase HPLC, the column was equilibrated with 0.1% formic acid in water and the intact samples were loaded onto the column using a stainless steel bomb pressurized with nitrogen gas at 1,000 psi for 45 minutes. Liquid chromatography was initiated at a primary flow
rate of 4 µL through the Agilent 1100 Series reversed-phase HPLC system (Agilent Technologies, Waldbronn, Germany) that ran through a splitter and resulted in a flow rate of 400 nL/min over the column with a 10 min rinse in 95% Buffer A (0.1% formic acid in water) followed by a 20-min linear gradient of 5 to 95% Buffer B (0.1% formic acid in acetonitrile). The spectra were acquired by nano-electrospray ionization on a Thermo Finnigan LTQ-FT mass spectrometer (San Jose, CA). The capillary temperature was 250°C and the spray voltage was 2.2 kV for ubiquitin.

For all intact β-lactoglobulin A samples, 8 µL were injected using an Agilent autosampler module (Agilent Technologies) over an Agilent ZORBAX 300SB C_{18} (150 × 0.3 mm, 5 µM particles) reverse-phase column (Agilent Technologies). The HPLC, directly coupled to the LTQ-FT for mass spectrometry, was run using a linear gradient of 95% Buffer A to 95% Buffer B over 20 minutes at a flow rate of 2 µL/min, followed by a 10 minute wash with 95% Buffer B and an 80 minute wash with 95% Buffer A. The capillary temperature was 250°C and the spray voltage was 2.1 kV for β-lactoglobulin A.

**Tryptic digestion of oxidized β-Lactoglobulin A**

One β-lactoglobulin A irradiated with 480 ns electron pulses was selected for tryptic digestion to determine the sites and amounts of oxidation. Ammonium bicarbonate (50 uL, 50 mM) and DTT (25 µL, 25 mM) were added to the β-lactoglobulin A sample (50 µL) and the samples were incubated at 50°C for 3 hours to denature and reduce the protein. Iodoacetamide (25 µL, 90 mM) was added to the solution to carbamidomethylate the reduced disulfide bonds, and incubated in the dark at room temperature for 45 minutes. Sequencing grade modified trypsin (Promega, Madison, WI) was added (25 µL, 0.1 µg/µL) and incubated at 37°C overnight
while rotating to digest the protein samples. The samples were analyzed as is in triplicate and the remaining sample was stored at 0°C.

**LC-MS/MS of digested oxidized β-Lactoglobulin A**

8 µL of the oxidized β-lactoglobulin A tryptic digest were injected over an Agilent ZORBAX 300SB C<sub>18</sub> (150 × 0.3 mm, 5 µM particles) reverse-phase column (Agilent Technologies). The HPLC, interfaced to an LTQ-FT mass spectrometer, was run using a linear gradient of 95% Buffer A to 60% Buffer B over 47 minutes then to 90% Buffer B over 15 minutes at a flow rate of 2 µL/min, followed by a 5 minute wash with 90% Buffer B and an 55 minute wash with 95% Buffer A. The capillary temperature was 250°C, and the spray voltage was 2.3 kV for β-lactoglobulin A tryptic digest. After the LC-MS/MS experiment, the measured peptides were screened computationally for different modifications using MASCOT in conjunction with ProteIQ [105] and ByOnic [53]. Semi-tryptic peptides were included in the search and analysis to obtain better sequence coverage as well as to determine other sites of oxidation. All tandem mass spectra assignments and sites of oxidation verified manually due to sample and search space complexity.

**MD simulations and solvent accessibility calculations**

A crystal structure of β-lactoglobulin (PDB ID code 1BSY) [99] was retrieved from the Research Collaboratory for Structural Biology (RCSB) database [72]. In order to calculate the time-averaged solvent accessible surface area (<SASA>) of each side chain in β-lactoglobulin, a 10-ns MD simulation of the protein was performed using the all-atom AMBER 8 [106] force field with the PARM99 protein parameters [74]. Prior to the MD simulation, all histidine
residues were considered as neutral and were protonated at the Nε position. Nine Na⁺ ions were added to neutralize charge and the protein was solvated by 5,417 TIP3P [75] water molecules, using the Protein Builder component of the GLYCAM-Web tool (http://www.glycam.com [32]), which employs the tLEaP [106] module of AMBER. The energy was minimized using the SANDER module [106] of AMBER by performing 5000 steps of steepest descent followed by 5000 steps of conjugate gradient minimization, again using GLYCAM-Web. The water molecules were then subjected to a simulated annealing protocol in which they were heated from 5 K to 300 K over a period of 50 ps, held at 300 K for 100 ps, before being cooled to 5 K over a final 50 ps. Following solvent annealing, the entire system was heated from 5 K and to 300 K over 100 ps with no restraints applied to the coordinates. A production MD data set was then collected for 10 ns with the temperature held constant at 300 K. All simulations were performed using the NPT ensemble, at 1 atm using a 2 fs integration time step, with the SHAKE algorithm [107] treatment of all hydrogen-containing bonds, and a unit dielectric constant. The atomic coordinates were stored every 10 ps for analysis, for a total of 1000 snapshots. SASA values were calculated for individual snapshots employing the NACCESS program [76] and averaged over the 10 ns simulation by in-house program to obtain <SASA> values and standard deviations for each of the 162 residues.

Results and Discussion

Rapid NMR spectroscopy of proteins in solution with H₂O₂

H₂O₂ photolysis is one of the fastest methods for producing hydroxyl radicals, but there is concern about sensitivity of proteins to the physical and chemical effects that H₂O₂ can impart. Before questioning the use of methods dependent on H₂O₂ for transition metal-free protein
preparations, we felt it is important to assess the extent to which its presence affected conformation. Using time-resolved NMR spectroscopy we were able to observe changes in the conformation of a model protein, Galectin-3, in the presence of H$_2$O$_2$ versus H$_2$O. Galectin-3 is a small (15 kDa) stable protein that binds galactose terminated oligosaccharides. It is not suspected to be usually sensitive to hydrogen peroxide. However, it does have a cysteine residue that is not particularly solvent exposed. A three minute HSQC spectrum of galectin-3 was recorded in the presence and absence of H$_2$O$_2$ (Figure 3.1). The small shifts labeled (Figure 3.1) in the HSQC spectrum that can be interpreted as a conformational change. Many of the shifted peaks are assigned to resonances near the lactose binding site. Due to the very short time scale (three minutes) and known conformational flexibility of the binding site, direct oxidation by hydrogen peroxide is not thought to be the driving force behind this noted conformation change. The subtle shifts are more likely due to physical association of H$_2$O$_2$ versus water. Some of the perturbed resonances do, however, cluster near C173, the partially buried cysteine that is expected to be most sensitive to oxidative chemical events. At longer times, perturbations of additional resonances more indicative of rapid two-electron oxidation events are seen (data not shown). An LC-MS analysis of galectin-3 after a three minute H$_2$O$_2$ exposure determined the protein is not unusually sensitive to hydrogen peroxide, as no noticeable oxidation products were detected, while a sixty minute H$_2$O$_2$ exposure resulted in detectable protein oxidation (data not shown). Regardless of the origin of the shifts, the data illustrate that proteins are not necessarily probed in their native conformation when using methods that involve H$_2$O$_2$ to produce hydroxyl radicals even in the absence of peroxide-mediated oxidation events, and that these perturbations can occur on short time scales.
Protein oxidation

High energy electrons passing through the aqueous solution ionize/excite water molecules forming a number of transient species and stable products according to equation (3.1).

\[ \text{H}_2\text{O} \xrightarrow{\text{irr}} \text{e}_{\text{aq}}^-, \cdot \text{OH}, \text{H}^+, \text{OH}, \text{H}_2\text{O}_2, \text{H}_2 \]  

(3.1)

Among all species generated during water radiolysis, hydroxyl radicals (\(\cdot\text{OH}\)) and hydrated electrons (\(\text{e}_{\text{aq}}^-\)) are the most reactive with peptides and proteins [50]. When the radiolysis is performed in the presence of oxygen, superoxide radicals and its acidic form of hydroperoxyl radicals are generated in reaction (3.2) suppressing the effect of hydrated electrons on the proteins.

\[ \text{e}_{\text{aq}}^- (\text{H}^+) + \text{O}_2 \rightarrow \text{O}_2^-(\text{HO}_2^-) \]  

(3.2)

\[ \text{HO}_2^- \leftrightarrow \text{O}_2^- + \text{H}^+ \]  

(3.3)

Since the pKa of \(\text{HO}_2^-\) radical is around 4.8 (reaction (3.3)) and oxidation experiments are performed at buffered pH 7, most of the hydroperoxyl radicals are in the form of a superoxide radical (\(\text{HO}_2^-/\text{O}_2^-\)). \(\text{HO}_2^-/\text{O}_2^-\) radicals are known to react with amino acids at very low rate constants ranging from 10^{-10^2} \text{dm}^3\text{mol}^{-1} [108, 109], therefore one can expect very little effect of \(\text{HO}_2^-/\text{O}_2^-\) radicals with the metal free proteins that are subject of the current studies. \(\text{HO}_2^-/\text{O}_2^-\) radicals undergo comparatively rapid disproportionation to \text{H}_2\text{O}_2 and \text{O}_2 in reactions (3.4) and (3.5) via a pH dependent mechanism that also involves reaction (3.3).

\[ \text{HO}_2^- + \text{HO}_2^- \rightarrow \text{H}_2\text{O}_2 + \text{O}_2 \]  

(3.4)

\[ \text{HO}_2^- + \text{O}_2^- \rightarrow \text{H}_2\text{O}_2 + \text{O}_2 + \text{OH}^- \]  

(3.5)

On the basis of the kinetic equation for this decay mechanism [110, 111], one can estimate the rate constant for disproportionation at pH=7 to be about \(2k_{\text{obs}}=1.2 \times 10^6 \text{ dm}^3\text{mol}^{-1}\). This indicates that if left alone \(\text{HO}_2^-/\text{O}_2^-\) radicals can last milliseconds after the irradiation pulse;
however, it is not long enough to cause any significant oxidation of the available protein amino acids. Converting 'OH radicals leading to water and recovering initially consumed oxygen in reaction (3.6) is an additional reaction that consumes \( \text{HO}_2'/\text{O}_2'^- \) radicals before they can react with the protein.

\[
\text{HO}_2'/\text{O}_2'^- + '\text{OH} \rightarrow \text{H}_2\text{O} + \text{O}_2
\]  

(3.6)

Since the rate constant of reaction (3.6) is as fast as \( 1 \times 10^{10} \text{dm}^3\text{mol}^{-1} \), it inconveniently consumes some of important 'OH radicals for oxidation.

In the initial microseconds following pulse irradiation, the balance of the above reactions leads mostly to the very reactive 'OH radicals and relatively nonreactive \( \text{HO}_2'/\text{O}_2'^- \) radicals. 'OH radical reaction with amino acids is well established and proceeds with reaction rate constants [112] varying from \( 10^{10} \) to \( 10^7 \text{dm}^3\text{mol}^{-1} \). From a kinetic point of view the majority of free 'OH radicals that happen to be close to the protein surface will interact with the fastest reacting amino acids; those include amino acids containing sulfur as well as unsaturated/aromatic side groups. Amino acids with the aliphatic side groups react with 'OH radicals relatively slower. In oxygenated solutions protein surface radicals add oxygen forming the corresponding peroxyl radicals. The mechanism of initial sulfur centered radicals is not completely explained yet but leads to well characterized and recognizable products [113, 114]. The fate of the relatively fast surface located peroxyl radicals depends entirely on their nature and surroundings. Some of the aromatic peroxyl radicals will eliminate \( \text{HO}_2' \) radicals leaving behind a stable hydroxyl group [113] or a carbonyl group [115]. C-centered aliphatic peroxyl radicals can undergo a bimolecular recombination leading to the creation of hydroxyl and/or carbonyl group on the parent carbon atom, or transfer the radical to form a hydroperoxyl group that propagates oxidation to neighboring amino acids [115]. Those reactions are well understood and
summarized in the number of articles and textbooks [5, 113, 115]. In either case in diluted protein solution, sites of oxidation should be anchored to the place where the initial 'OH radical attack occurred. Chances are that some of the initially created peroxyl radicals can survive on the surface and do not undergo any of the mentioned transformations. In this case they can react with lengthy present HO$_2^–$/O$_2^–$ radicals producing hydroxyl group, oxygen and hydrogen peroxide [116]. The presence of oxygen is quite essential in the 'OH radical induced protein oxidation even though the fastest surface labeling occurs by way of hydroxylation with 'OH radicals [5].

Our initial experiments were performed with aerated buffered ubiquitin solutions in order to establish optimal doses and concentrations for maximum oxidation. We varied the pulsewidth and current of the electron beam, as well as, the dissolved gas in the sample. In order to double the concentration of hydroxyl radicals in the pulse irradiation we saturated our solutions with nitrous oxide/oxygen combination (N$_2$O/O$_2$ 4/1 v/v). In this case the oxygen concentration remained on the same level as in the air saturated solutions, but additionally we benefited from the well known reaction (3.7) of fast conversion of hydrated electrons to hydroxyl radicals [112]. Since the solubility of N$_2$O in water is about 20 times higher than oxygen, most hydrated electrons convert to hydroxyl radicals within the electron pulse, doubling the hydroxyl radical concentration.

$$e_{aq}^- + N_2O \rightarrow \cdot OH + OH^- + N_2 \quad (3.7)$$

The influence of gas composition on ubiquitin oxidation is presented in Figure 2. Due to the large dose applied, we see very substantial oxidation in both air and N$_2$O/O$_2$ saturated solutions. The solutions containing N$_2$O show higher abundance of heavily oxidized ubiquitin. The initial experiments showed that we can easily control any excess of labeled protein by
varying the dose and the gas composition. Additionally, we noted that the increase of the applied radiation dose does not increase the level of oxidation proportionally. It is mostly related to the oxygen uptake in the solution after the pulsed irradiation. In aerated or N₂O/O₂ saturated solutions, oxygen concentrations are on the order of 2.7×10⁻⁴ mol/dm³ that corresponds to the concentration of water radiolysis transient species (’OH, H’, e⁻_aq) achievable at roughly 500 Gy of absorbed dose. Considerably increasing the dose above 500 Gy decreases the amount of available oxygen and leads to undesired reactions of hydrated electrons with the protein surface amino acids. Lack of oxygen can distort the formation process of carbon centered peroxyl radicals as precursors of hydroperoxides [116] and other oxidation products on the protein surface. The quantitative studies of the dose dependence on the pulse radiolysis ’OH protein footprinting were not the subject of the current investigation but will be addressed in the future reports.

**LC-FTMS analysis of Ubiquitin and β-lactoglobulin A**

Sodium phosphate-buffered ubiquitin was used as a model to determine if protein oxidation could be achieved by this method. LC-FTMS of ubiquitin (Figure 2) shows the range of intact protein oxidation obtained with and without N₂O, and with and without methionine amide, a scavenger of secondary oxidants. Ubiquitin was also irradiated with various pulsewidths (480 or 660 ns) and electron beam amperages (doses) (Supplemental Data). In the absence of methionine amide (Figure 2b) no unoxidized ubiquitin is detected even at relatively low radiation doses despite the fact that 10% of the solution was unirradiated. However, in the presence of methionine amide (Figure 2c and d), unoxidized ubiquitin from the unirradiated fraction of the solution was still detected despite heavy amounts of oxidation. These results
indicate that the methionine amide is sufficient to prevent oxidation of the unirradiated ubiquitin even at high radiation dosages, while secondary oxidation sufficient to consume all of the unirradiated ubiquitin occurs in the absence of the methionine amide even at relatively low radiation dosages (Supplemental Data). Quenching is necessary to ensure the vast majority of labeling occurs on a microsecond time scale and before the protein can unfold, rather than by secondary oxidants (e.g. peroxides in the presence of metal ion traces of UV light) on the second to minute timescale. With electron beam pulsewidths under 700 ns, we detected an extensive amount of oxidation (Figure 2d) using a sub-microsecond electron pulse. By adjusting various parameters of the irradiation (pulsewidth, dissolve N₂O and dose), we are able to control the amount of oxidation from very little oxidation to very extensive oxidation (Supplemental Data).

In initial experiments on β-lactoglobulin A, a sodium phosphate buffer was irradiated and we detected sodium adduction (M + n22) in each sample despite multiple attempts to desalt the solution. This interfered with our data analysis (data not shown). Ammonium bicarbonate and ammonium phosphate buffers were used for subsequent experiments. Upon irradiation, the amount of oxidation in the ammonium bicarbonate-buffered samples was significantly lower, probably due to radical scavenging by the bicarbonate buffer. Also, an adduct or modification of unknown origin (observed mass shift of 174) is present in all ammonium bicarbonate-buffered β-lactoglobulin A samples (data not shown). Ammonium phosphate buffer gave the best results for intact β-lactoglobulin A, as it did not scavenge the radical, nor did it result in sodium or other adducts (Supplemental Data).
**Time-resolved UV absorbance spectra reveal the hydroxyl radical lifetime**

Since proteins can unfold significantly on the order of a few microseconds [117], labeling chemistry should be completed faster than that to prevent labeling of oxidatively unfolded protein. One way to observe the progress of labeling chemistry is to monitor the decay of hydroxyl radicals in the microsecond time scale after their formation during the pulse of electrons. The lifetime of hydroxyl radicals in aqueous solutions after each pulse of electrons can be estimated based on the known reaction rate constants of hydroxyl radicals with solutes [112]. In pure buffered water hydroxyl radicals recombine with each other since reactions with buffer components are usually much slower. The second order hydroxyl radical recombination (3.8) competes with slower first order reactions especially at higher doses when higher concentrations of hydroxyl radicals are produced.

\[ \cdot \text{OH} + \cdot \text{OH} \rightarrow \text{H}_2\text{O}_2 \]  

(3.8)

We performed time resolved pulse radiolysis experiments with transient absorption detection at 250 nm to look at the upper limit of hydroxyl radical life time in buffered N\textsubscript{2}O saturated (no oxygen present) solutions without proteins. The traces from these experiments are provided in Supplemental Data. The initial increase of signal corresponds to formation of hydroxyl radicals and is related to their concentration via an extinction coefficient. Analyzing the transient absorption one can see that the apparent signal decays very quickly and within 8 \( \mu \)s reaches the plateau of H\textsubscript{2}O\textsubscript{2} absorption. Reaction (3.9) of hydroxyl radicals with H\textsubscript{2}O\textsubscript{2} is relatively slow \((2.7 \times 10^7 \text{ mol}^{-1}\text{dm}^3)\) [112] and consumes only about 1% of the formed H\textsubscript{2}O\textsubscript{2} 40 microseconds after the irradiation pulse.

\[ \cdot \text{OH} + \text{H}_2\text{O}_2 \rightarrow \text{HO}_2^-/\text{O}_2^- + \text{H}_2\text{O} \]  

(3.9)
However, this minimal concentration of superoxide radical contributes some 20% to the final absorption, since the extinction coefficient of $\text{HO}_2^-/\text{O}_2^-$ radicals greatly exceeds the extinction coefficient of hydrogen peroxide. Contribution of $\text{HO}_2^-/\text{O}_2^-$ radical absorption increases, obviously, with the increase of the applied dose since reaction (3.9) proceeds more efficiently but, the final concentrations of formed $\text{HO}_2^-/\text{O}_2^-$ are always at a lower fraction than the initial $\cdot\text{OH}$ radical concentration. Extension of the pulsewidth from 400 ns to 1500 ns increases the dose almost 4 times but not the peak transient absorption of hydroxyl radicals (Supplemental Data). Reaction (3.8) effectively decreases the concentration of hydroxyl radicals during the 1500 ns pulse and the rate of hydroxyl radical formation approaches the rate of hydroxyl radical decay giving steady state concentration of hydroxyl radicals just only 50% higher than the peak concentration of hydroxyl radicals after 400 ns pulse. It is important to note that the final absorption of $\text{H}_2\text{O}_2$ and $\text{HO}_2^-/\text{O}_2^-$ after the completion of reaction (3.8) is more than 3 times higher for 1500 ns pulse than 400ns (Figure 3), which confirms that reaction (3.8) with reaction (3.9) is the main channel of hydroxyl radicals decay in the protein free $\text{N}_2\text{O}$ saturated buffered water. Based on the extinction coefficients at 250 nm of species present, we determined that the time for hydroxyl radicals to decay to 0.2 µM (1% of protein in oxidation experiments) is about 20.6 µs + pulsewidth of electrons. It is very fast but still comparable with the duration of some protein unfolding events [118]. Obviously, in the presence of protein, hydroxyl radicals will not stay free for as long because they will react with the protein surface groups. For most amino acids side changes, the reactions are very fast [112] and should contribute very effectively to the decrease of hydroxyl radical concentration. To confirm the effect of protein, we performed an experiment in which $\beta$-lactoglobulin A was added to the buffered solution and the decay of hydroxyl radicals was monitored. The UV traces in
Supplemental Data show that after the addition of 4 μM of protein the apparent UV signal reaches a higher plateau than in buffer alone suggesting that some other absorbing product is being formed in addition to H₂O₂ and HO₂/O₂⁻. In fact, a further increase of the protein concentration results in an even higher increase in the amplitude of the formed product. The apparent UV signal results from a sum of contributions from various transient species produced upon hydroxyl radical reaction with the protein surface and can be represented in symbolic reaction (3.10). The initial step of hydroxyl radical reaction with protein is formation of a protein radical:

\[ \text{OH} + \text{protein} \rightarrow \text{protein radical} \quad (3.10) \]

The protein radical in oxygen-free, N₂O saturated solution undergoes further reactions leading to the formation of the final oxidized protein (in case of radicals formed on the sulfur containing or unsaturated/aromatic residues) or decays in the bimolecular processes of disproportionation/dimerization [50, 113]. For the time being we are interested in the fate of hydroxyl radicals, and based on the several UV transient signals collected for different protein concentrations, we constructed a simple model to extract the portion of the UV signal that is related to the decay of hydroxyl radicals. The applied global fitting model allowed us to estimate the rate constant of reaction (3.9) and the overall extinction coefficient of the protein radical for the studied system. All other parameters used to obtain fitting results are tabulated in Supplemental Data. The time profiles of most contributing species absorbing at 250 nm resulting from global fitting are shown on Figure 3.3. This basic experiment proved that the lifetime of hydroxyl radicals decreases considerably after addition of protein. In addition to the lifetime of hydroxyl radicals, the transient absorption experiment has also shown the importance of properly choosing the concentration of the protein in the pulse irradiation experiment. Lower
concentrations of protein may be necessary for some experiments and can increase the extent of protein labeling [117], but at the same time the lower protein concentration prolongs the hydroxyl radical’s lifetime, increasing the chance of oxidizing a previously unfolded residue and necessitating the use of an exogenous quencher to limit the hydroxyl radical lifetime [22].

**LC-MS/MS analysis of oxidized β-Lactoglobulin A tryptic and semi-tryptic fragments**

The β-lactoglobulin A sample irradiated with the dose of 260 Gy using a pulse length of 480 ns, was denatured, reduced, carbamidomethylated, and digested using trypsin. The peptide mixture was loaded on a C18 capillary column for LC-MS/MS analysis to locate sites of oxidation. The overall sequence coverage, including all oxidized and unoxidized peptides identified in the LC-MS/MS run, was 100%, with manual verification of all tandem mass spectra assignments. Fourteen oxidation sites were identified for β-lactoglobulin A in these LC-MS/MS experiments; however, one semi-tryptic peptide was detected as completely oxidized and although oxidation could not be mapped to a specific site(s), peptide 108-ENSAEPEQSLACQCLVR-124 contains many reactive residues with large <SASA> values. The identified oxidation sites are shown on the X-ray crystal structure of β-lactoglobulin A [99] in Figure 3.4a.

The surface average solvent accessibility (<SASA>) value of each amino acid residue for all 162 residues was obtained from 1000 snapshots over a 10 ns MD simulation (Supplemental Data). While 10 ns is too short a period to simulate large-scale protein backbone motions, it is adequate to capture much of the motion of the side chains, and provides a useful approach for achieving statistically significant average SASA values (<SASA>) on a given time scale. In addition, the use of MD simulations on this time scale provides the opportunity to probe the
extent to which a model based essentially on the fully folded protein is able to describe the data from an experiment performed on the millisecond timescale. The $<\text{SASA}>$ values for each mapped oxidation site on β-lactoglobulin A are listed in Table 3.1. Using this method we are able to determine that, with the exception of M24, all amino acids determined to be oxidized under our conditions were predicted to be suitably solvent accessible by MD simulations.

The sulfur-containing residues (Met and Cys) are the most susceptible to oxidative modifications [98]. No cysteine residues were conclusively identified as sites of oxidation; even though they are chemically reactive, they are buried within the protein structure, and therefore should react only slowly with a hydroxyl radical if the protein retains its folded form during the labeling process. Three of the four methionine residues found in β-lactoglobulin A, Met7, Met24 and Met145, were identified as sites of oxidation with a mass shift of +16. Met24 has a very small $<\text{SASA}>$ value of 0.21 Å$^2$, but Met7 and Met145 have small to moderate solvent accessibility ($<\text{SASA}>$ value = 35.88 and 6.7 Å$^2$ respectively). Met107 also has a large $<\text{SASA}>$ value and was not detected as oxidized by our measurements. However, it is present on a large peptide that was determined to be the site of multiple semi-tryptic cleavages and unusual oxidation events; it is quite possible that Met107 was oxidized quite readily, but that we are unable to resolve and identify the oxidized form of the peptide from the MS/MS spectra. All other sites of oxidation identified had large $<\text{SASA}>$ values (Table 3.1), and can be clearly seen on the surface of the protein (Figure 3.4); importantly, all of the oxidized sites were predicted by MD simulation to be among the most highly accessible residues of each type within the protein (e.g. out of twenty-two total leucine residues, the four leucine residues oxidized were all among the five most solvent accessible according to MD simulations), with the exception of Met24 (Supplemental Data).
This leaves an unresolved question: why was Met24 oxidized under these conditions? One potential cause is incomplete scavenging of secondary oxidants, which have been noted to lead to uncontrolled oxidation of methionine [98]. However, these same secondary oxidants would also be capable of oxidizing cysteine, which is not detected. A more likely possibility is that the protein undergoes larger backbone motions, in the region of M24, on the time scale of the experiment, than are captured in the 10 ns MD simulation. In that case, the oxidation of M24 would be larger than predicted by the <SASA> values. It is important to note here that by using average SASA values, rather than those from a single protein structure, the possibility that this discrepancy is related to an anecdotal orientation of the side chain of M24 is greatly reduced.

Because most of the oxidized peptides contain more than one oxidation site, accurate quantitation of the amount of oxidation at individual sites is currently not possible. We were able to quantitate the amount of oxidized versus unoxidized peptide using the triplicate LC-MS/MS runs to determine the peak area and the standard deviation of the MS total ion chromatogram of a given m/z for each peptide and its oxidized form(s). The peak areas were used to calculate peptide fractional oxidation by dividing the peak area of the oxidized peptide by the sum of the peak areas of the oxidized and unoxidized forms of that peptide. The peptide fractional oxidation is shown for each β-lactoglobulin A peptide in Figure 3.4b and values are listed in Table 3.1. The sites of oxidation that could be determined by MS/MS analysis are also listed in Table 3.1; however, while these identified sites are major sites of oxidation for the peptide, they may not be the only substantial sites of oxidation. It is important to note that, as is all LC/MS experiments of complex mixtures, it is quite possible that major oxidation products exist that were not detected or that did not fragment sufficiently well to determine the site(s) of oxidation.
Conclusion

The overall purpose of this study was to demonstrate a method for hydroxyl radical footprinting of proteins in solution that is completed faster than large scale oxidation-induced conformational changes can occur, without requiring a precursor oxidant. Time-resolved NMR spectroscopy illustrates that the presence of hydrogen peroxide can result in uncontrolled oxidation or peroxide-induced conformational changes, causing the labeling of a non-native protein structure. Additionally, the presence of redox-active transition metals in solution with the hydrogen peroxide can lead to uncontrolled metal catalyzed oxidation [82], and in many cases it is not possible to completely purify redox-active transition metals from the protein-ligand preparation. The development of a hydrogen peroxide-free method of hydroxyl radical footprinting that is capable of completing the labeling reaction faster than the protein can unfold increases the applicability of hydroxyl radical footprinting technology to peroxide-sensitive proteins, while still allowing for extensive labeling of the native conformation. As shown from the results presented, the electron accelerator pulsed water radiolysis method is suitable for heavy, controlled oxidation of proteins and oxidation of solvent accessible residues. Using an electron beam pulse, controllable protein oxidation can be obtained by adjusting variables such as pulsewidth, dose, and dissolved gas. Time–resolved UV spectroscopy indicates that the most reactive radical species are consumed in less than 2 μs without a quencher, and this rapid timescale allows for extensive oxidation before the protein can unfold due to the modifications. Our ability to heavily oxidize the protein without concern for oxidation-induced unfolding allows us to detect a much greater amount of oxidation than previously-reported results on the same protein [20], allowing for higher resolution hydroxyl radical protein footprinting data. MS/MS analysis and MD simulations indicate that all oxidation sites except one (Met24) identified by
this method are moderately to highly solvent accessible. Additionally, as a rapid oxidation method, electron pulse water radiolysis has broad applicability in time-resolved structural studies, most notably UV-based spectroscopy methods where the high concentrations of hydrogen peroxide necessary for flash photolysis of peroxide can confound spectroscopy. Such a pulse labeling method will prove to be extremely useful in designing time-resolved spectroscopic studies of oxidation-induced protein unfolding, which is important for understanding the biophysical basis of oxidative stress-induced protein inactivation.
Acknowledgement

This research was supported by the National Center for Research Resources of the National Institute of Health (P41RR005351). The Notre Dame Radiation Laboratory is supported by the Office of Basic Energy Sciences at the United States Department of Energy. This is document number NDRL-4795 from the Notre Dame Radiation Laboratory.
Table 3.1. Average solvent accessibility of oxidized residues in β-Lactoglobulin.

<table>
<thead>
<tr>
<th>Peptide (residues)</th>
<th>Peptide Fractional Oxidation</th>
<th>Site(s) of Oxidation</th>
<th>&lt;SASA&gt; [Å²]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>0.3861 ± 0.1686</td>
<td>L1</td>
<td>49.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M7</td>
<td>35.88</td>
</tr>
<tr>
<td>23-40</td>
<td>0.1184 ± 0.0052</td>
<td>M24</td>
<td>0.21</td>
</tr>
<tr>
<td>41-60</td>
<td>0.0127 ± 0.0003&lt;sup&gt;a&lt;/sup&gt;</td>
<td>E51</td>
<td>123.54</td>
</tr>
<tr>
<td></td>
<td>0.0227 ± 0.0009</td>
<td>L57</td>
<td>40.63</td>
</tr>
<tr>
<td>84-91</td>
<td>0.0174 ± 0.0103</td>
<td>L87</td>
<td>156.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E89</td>
<td>88.65</td>
</tr>
<tr>
<td>92-100</td>
<td>0.0091 ± 0.0043</td>
<td>Y99</td>
<td>40.51</td>
</tr>
<tr>
<td>92-101</td>
<td>0.0437 ± 0.0218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>125-135</td>
<td>0.0124 ± 0.0072</td>
<td>V128</td>
<td>28.86</td>
</tr>
<tr>
<td>125-138</td>
<td>0.0310 ± 0.0040</td>
<td>L133</td>
<td>58.05</td>
</tr>
<tr>
<td>142-148</td>
<td>0.1607&lt;sup&gt;b&lt;/sup&gt;</td>
<td>M145</td>
<td>6.7</td>
</tr>
<tr>
<td>149-162</td>
<td>0.0590 ± 0.0057</td>
<td>F151</td>
<td>20.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E158</td>
<td>81.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I162</td>
<td>83.35</td>
</tr>
</tbody>
</table>

<sup>a</sup> Mass shift of -30 which is characteristic of oxidized glutamic acid residue. <sup>b</sup> Only seen in one replicate.
Table S3.1. Reactions and parameters used in the transient absorption global fitting.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Product extinction coefficient [M⁻¹ cm⁻¹]</th>
<th>Rate constant [M⁻¹ dm³]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH + OH → H₂O₂</td>
<td>26</td>
<td>5.0×10⁹</td>
</tr>
<tr>
<td>OH + H₂O₂ → O₂⁻ + H₂O + H⁺</td>
<td>1890</td>
<td>3.7×10⁷</td>
</tr>
<tr>
<td>OH + H → H₂O</td>
<td>non absorbing</td>
<td>7.0×10⁹</td>
</tr>
<tr>
<td>H + H → H₂</td>
<td>non absorbing</td>
<td>5.0×10⁹</td>
</tr>
<tr>
<td>OH + PO₄⁢⁻ → products</td>
<td>non absorbing</td>
<td>1.0×10⁵</td>
</tr>
<tr>
<td>OH + protein → protein radical</td>
<td>691</td>
<td>1.24×10¹⁰</td>
</tr>
<tr>
<td>H + protein → protein radical</td>
<td>691</td>
<td>8.7×10⁸</td>
</tr>
<tr>
<td>protein radical → products</td>
<td>non absorbing</td>
<td>2.9×10³</td>
</tr>
</tbody>
</table>
Figure 3.1. A 3 minute HSQC spectrum of galectin-3 in the absence (red) and presence (black) of 1% hydrogen peroxide (v/v).
Figure 3.2. LC-MS of the 10+ charge state of intact ubiquitin. Phosphate adducts were seen from sodium phosphate buffer. (a) Unirradiated sodium phosphate buffered ubiquitin. Ubiquitin irradiated with 800 ns pulsewidth in (b) air, (c) air + methionine amide, and (d) N₂O/O₂ + methionine amide.
Figure 3.3. Concentration profiles of species contributing most to the transient absorption signals at 250 nm based on the global fitting results. Dotted lines, solutions with no protein; dashed lines, solutions with 4µM of β-lactoglobulin; solid lines, 20µM of β-lactoglobulin.
Figure 3.4. (a) Two views of X-ray crystal structure of β-lactoglobulin A with waters removed for clarity created using PyMOL (PDB identifier 1BSY). Oxidized residues confirmed by LC-MS/MS are colored with red. (b) Quantitation of peptide fractional oxidation. All oxidation mass shifts are +16 and seen in at least twice in the triplicate runs unless otherwise noted. Error bars represent the standard deviation of identical triplicate runs.

\[ \text{Mass shift of -30 which is characteristic of oxidized glutamic acid residue} \]

\[ \text{Mass shift of +28 (two +14 shift from aliphatic residues)} \]

\[ \text{Only seen in one replicate} \]
**Figure S3.1.** Schematic representation of the electron pulse protein oxidation setup.
Figure S3.2. LC-MS of the 10+ charge state of intact ubiquitin. Phosphate adducts were seen from sodium phosphate buffer. The conditions of irradiation are listed to the right of each spectra.
Figure S3.3. LC-MS of the 13+ charge state of intact β-Lactoglobulin. (a) Unirradiated ammonium phosphate buffered β-Lactoglobulin. β-Lactoglobulin was irradiated for 200 ns (b) and 400 ns (c) in air + methionine amide.
Figure S3.4. Transient absorption signals monitored after the pulse of electrons at 250nm. Solid lines-signals recorded after 400ns pulse; dashed line-signal recorded after 1500ns pulse; white dotted line-represents global fit of the transient absorption signals. All solutions were fixed at pH=7 with ammonium phosphate buffer and contained 25mM N₂O, the concentration of β-lactoglobulin was varied like indicated on the figure.
Figure S3.5. Surface average solvent accessibility (<<SASA>>) value for the amino acid groups of β-lactoglobulin that contained at least one oxidized residue. The oxidized residues are colored red.
CHAPTER 4

FEATURES OF OXIDIZED PEPTIDE MIXTURES AND THE IMPACT OF PROTEIN IRRADIATION CONDITIONS ON PROTEIN INTEGRITY RELEVANT TO OXIDATIVE SURFACE MAPPING

Introduction

In this chapter, several phenomena that occur during oxidative labeling of proteins, but are rarely described in the literature, will be discussed. The studies were performed on carbohydrate binding proteins and synthetic peptides related to their sequences. The target proteins included galectin-1 (discussed in chapter two) and a monoclonal antibody (mAb) against Shigella flexneri (IgG). Both proteins have the ability to bind carbohydrate ligands therefore there is a potential benefit from the analysis of certain mass spectra and protein oxidative conditions for future characterization of protein-carbohydrate binding by oxidative surface mapping. The crystal structure of each protein or related species is known: for the mAb, there is a structure (pdb id 1M7I) [119] of the antigen fragment binding (Fab), the galectin-1 description was presented in chapter two. A synthetic peptide of the Shigella flexneri IgG Fab belongs to the complementarity determining region (CDR) of the antibody heavy chain. Its detailed study can give an insight into the possible amino acid oxidation in one critical region of the carbohydrate binding antibody. Studying isolated synthetic peptides, as opposed to a mixture of proteolytic peptides from protein digest, simplifies the data analysis, provides information as to which
residue is likely to be oxidized irrespective of the residue’s solvent accessibility, and indicates the type and abundance of likely modifications.

Oxidative changes that occur in the proteins and peptides are characterized by mass spectrometry. Even though a great effort was made to automate the process of mass spectra identifications and a search program ByOnic [53] has been adjusted for oxidative modification searches, in certain cases it is still difficult to fully resolve complicated mass spectra. This is because mixtures of isomeric peptides with oxidative products on different amino acids eluted from a reversed-phase column (prior entering mass spectrometer) within the same retention time. Mass spectra of such isomeric mixtures must be assigned manually as further shown on spectrum examples and discussed in the text. This phenomenon was only very recently mentioned in the literature [120, 121] but its relevance to footprinting studies has not yet been considered. Here some well-fragmented MS/MS spectra with a mixture of two to four peptides are presented with the majority of peaks with significant intensities assigned.

Another part of this work was an examination of oxidative conditions, particularly with respect to the impact of laser light and hydrogen peroxide on protein integrity during photodissociation. In order to achieve high resolution so that a significant number of the amino acids in the ligand binding site would be modified, it is necessary to adjust oxidative conditions to efficiently label as many residues as possible without affecting the structure of the protein. There are two technically different approaches to photooxidation: a single batch laser pulse with an Eppendorf tube aligned with the laser beam [11], and a syringe pump with a capillary tube flow system where a series of laser light pulses hit a UV-transparent window of the capillary through which the diluted sample is pumped and then collected in an Eppendorf tube [22, 23]. The amount of radicals introduced into the sample and efficient oxidation should be balanced in
order to prevent both under-oxidation, which would lower the resolution of the method, and over-oxidation, which might cause backbone cleavages of the protein [35, 50].

To characterize protein-ligand interactions, it is critical to ensure the 3D state of the complex is maintained during the oxidative labeling. The structure of the protein should stay folded in order for the binding site to be fully functional. Circular dichroism (CD) studies have been performed on ubiquitin and apomyoglobin previously to determine whether any structural changes take place after pulsed-laser irradiation of the protein and exposure to hydrogen peroxide. A single batch laser pulse was applied to an Eppendorf tube with protein solution and hydrogen peroxide aligned with the beam [11], indicating that protein conformational changes take place after the protein’s single laser pulse exposure. Pulsed laser irradiation using the flow system - whereby the oxidation should be more efficient due to the lower amount of the protein solution being exposed to the laser light at a time - has not been tested. In order to examine the process of the flow system oxidation, we have employed sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to look at the protein stability after laser exposure. In previous studies in which only a single batch laser pulse (not the flow system oxidation) was employed, there was no impact on protein integrity. However after multiple laser shots, loss of intact protein due to the occurrence of backbone cleavages was observed [35]. Currently, examination of the flow system laser exposure as well as the single batch laser pulse by SDS-PAGE is providing a greater insight into the process of laser-induced protein oxidation that should be a valuable contribution to the development of the covalent labeling method for mapping of the protein-carbohydrate ligand binding.
Experimental

Materials - proteins and peptides preparation

A starting material containing mAb against *Shigella flexneri* epitope was kindly provided by the Bundle laboratory, University of Alberta (Edmonton, AB, Canada). The mAb from 10 mL of ascites fluid was purified in series of steps resulting in 21 mg of pure antibody. Initially, the crude mouse ascites material was centrifuged to remove the precipitated solid material. The remaining liquid ascites was fractionated by employing protein A chromatography. A Biologic FPLC purification system, Bio-Rad Laboratories (Hercules, CA, USA) with a HiTrap Protein A HP column, GE Healthcare (Uppsala, Sweden) with two buffer system was employed to initially bind the immunoglobulin (IgG) to the protein A, wash away the unwanted protein contaminants, and elute the bound antibody in total cycle length of 75 minutes. The binding step was performed with a starting buffer of 50 mM Tris, 150 mM NaCl, pH 8 whereas the elution buffer was 100 mM citrate, pH 4.0. In order to avoid protein denaturation that could occur due to the low pH of elution buffer, 0.5 mL of 1M Tris buffer, pH 8 was added to the tubes prior to the 1 mL fractions of the eluting material were collected. Fractions with the eluted antibody were dialyzed overnight against the starting buffer and protein was concentrated to 3 mL by ultracentrifugation using an Amicon Ultra centrifugal filter device with 5,000 molecular weight cut-off, Millipore (Billerica, MA, USA). Protein concentrations were determined using a bicinchoninic acid (BCA) assay, Pierce (Rockford, IL, USA), and the concentration of the final IgG product was 7 mg/mL. The antibody was divided into aliquots of 1000 µL, frozen in liquid nitrogen, and stored at −80 °C prior to the experiments. The second protein under study, galectin-1 purified in our lab, was prepared and stored before the experiments as described previously [35, 56] and kept in stock concentration of 250 µM.
Synthetic peptides were purchased from Biomatik Corporation (Wilmington, DE, USA) in purity greater than 98% and stored as lyophilized powder at −20ºC until re-suspended prior to the experiments in purified water (18.2 MΩ) from NANOpure Diamond system, BLD science (Garner, NC, USA) to final stock concentration of 500 µM.

**Laser-induced photochemical oxidation of synthetic peptides and proteins**

Two synthetic peptides were examined for oxidation purposes. One related to the *Shigella flexneri* mAb sequence, GLEWVAEIR, and one related to galectin-1 sequence, LHFNPR. Prior to the laser exposure the peptides were diluted in nanopure water to 1-10 µM final concentrations from the 500 µM stock solutions with addition of hydrogen peroxide (5%, achieved by dilution of 30% solution w/w), Sigma-Aldrich (St. Louis, MO) right before the experiment with final concentration of 1% H₂O₂.

A high energy KrF excimer laser, EX100, GAM Laser, Inc. (Orlando, FL, USA) with a suitable wavelength of 248 nm and a pulse length of about 20 ns was used to photodissociate hydrogen peroxide (absorption maximum of H₂O₂ is at 250 nm [11]) to hydroxyl radicals for peptide oxidation. The energy of the laser was in a range between 55-65 mJ/pulse. There were two methodologies employed for the peptide oxidations. First, the Eppendorf tube was aligned with the laser beam and a single pulse was applied to the entire peptide solution (further referred to in the text as “single batch oxidation”). Second, a flow system was assembled with a capillary tube, Polymicro Technologies (Phoenix, AZ, USA) with an inner diameter (ID) 99 and outer diameter (OD) 170, and a UV transparent window was created by burning off the coating of the tubing. The capillary was coupled to a 100 µL Hamilton syringe, Hamilton Company (Reno, NV, USA) with a Harvard 22 syringe pump, Harvard Apparatus (Holliston, MA, USA). The
laser trigger was set to 20 Hz using an external pulse generator and the flow rate of peptide solution containing hydrogen peroxide was 12.2 µL/min. The flow rate was calculated to keep the efficiency of oxidation high but ensuring that every portion of the solution was hit by the laser beam only once. Oxidized peptide was collected into an Eppendorf tube with a quenching solution of 7 mM dithiothreitol (DTT), Bio-Rad Laboratories (Hercules, CA, USA) (this method is further referred to in the text as “flow system oxidation”). Immediately after oxidation, peptide solutions were flash-frozen in liquid nitrogen, lyophilized, and re-suspended in 99.9% H₂O / 0.1% formic acid prior to the mass spectrometry analysis.

Protein oxidation was performed similarly as the peptide oxidation in terms of concentrations, hydrogen peroxide amount and the oxidation methodologies. Protein solutions that were further studied by mass spectrometry analysis had DTT added to a final concentration of 7 mM prior to freezing in liquid nitrogen and lyophilization. The protein samples to be analyzed by SDS-PAGE were not treated with DTT in order to prevent dissociation of the protein subunits linked by disulfide bonds. Lyophilized proteins for mass spectrometry analysis were resuspended in 12 µL of 100 mM ammonium bicarbonate, Sigma-Aldrich (St. Louis, MO, USA) to prepare the proteins for proteolytic digestion. Protein samples for SDS-PAGE analysis were re-suspended in 7 µL of nanopure water, mixed well and centrifuged to ensure that all protein that was stuck on the tube sides was re-dissolved in the solution for accurate quantitative analysis.

**Proteolytic digestion of proteins**

Protein samples in 100 mM ammonium bicarbonate buffer with addition of DTT (added immediately after the oxidation) were heated for 1 hr 25 min at 67°C to break disulfide bonds
and promote the process of unfolding prior to proteolytic digestion. Once removed from the heated water bath, the samples were cooled down at a room temperature for 15 minutes. For the digestion, sequencing grade modified trypsin, Promega (Madison, WI, USA) was added at a 1:25 protease/protein ratio. The proper pH range of 6-8 was assured by pHyrion Paper, Micro Essential Laboratory (Brooklyn, NY, USA) and if below the range, pH was adjusted by addition of 0.5 µL 1M Tris, Bio-Rad Laboratories (Hercules, CA, USA). The digestion mixture was incubated in a microfilm sealed eppendorf tube for 20 hrs at 37ºC. When digestion was completed, the reaction was quenched by removing salts and trypsin from the solution through a 10 µL ZipTip, Millipore Corporation (Bedford, MA, USA) that contains C18 reversed-phase media. Peptides resulting from the protein digestion were eluted by 10 µL of the following solvent: 50% acetonitrile / 49.9% H₂O / 0.1% formic acid (v/v/v). Solution containing acetonitrile was removed by a speedvac system and the peptide mixture was re-suspended in 99.9% H₂O / 0.1% formic acid (v/v) for mass spectrometry analysis.

**Mass spectrometry analysis of digested proteins and synthetic peptides**

All peptides and digested protein mixtures were initially analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). A high resolution MALDI-TOF 4700 Proteomics Analyzer, Applied Biosystems (Foster City, CA, USA) was employed for acquisition of the mass spectra without further fragmentation of the ions. In order to get the molecules to the gas phase in the instrument, 0.5 µL of each peptide sample was spotted on a MALDI plate (Applied Biosystems) and covered with 0.6 µL of matrix solution (10 mg/mL α-cyano-4-hydroxycinnamic acid, Sigma-Aldrich) in 100% ethanol, and was
dried at room temperature for sample-matrix co-crystallization. Spectra were acquired in a reflector positive mode with laser intensity fixed at 5000.

Further sample separation with subsequent tandem mass spectrometry was achieved using a Vydac C18 column, Grace Davison Discovery Sciences (Deerfield, IL, USA) having a length of 150 mm, ID of 150 µm, and particle size of 5 µm, connected to an Agilent 1100 capillary liquid chromatograph (Palo Alto, CA, USA) and a hybrid Finnigan LTQ-FT (linear ion trap/Fourier transform ion cyclotron resonance) mass spectrometer with equipped with a 7 Tesla magnet and an electrospray ion source (ESI), Thermo Electron Corporation (San Jose, CA, USA). Samples were introduced onto the column via an autosampler and eluted by a gradient with mobile phases of A: 99.9% H₂O / 0.1% formic acid (v/v) and B: 99.9% acetonitrile / 0.1% formic acid (v/v) over 130 minutes with an increase of the B phase from 5% to 80% (v/v) and a flow rate of 0.8 µL/min. The instrument was set to acquire MS/MS spectra (ion trap) on the nine most abundant precursor ions (FTMS) from each MS scan with a repeat count of two and repeat duration of 5 s, and dynamics exclusion enabled for 300 s. Acquired data were extracted in Thermo RAW file format.

**Peptide assignment from tandem mass spectra**

Datasets acquired from the instrument were converted from Thermo RAW files to mzXML format, followed by a conversion to the pkI format by ReAdW and mzXML2Other respectively [66]. The database searches, post-processing, and data management were facilitated by our internally developed program, SurfMap Search Interface (SSI), which uses ByOnic to perform peptide assignments. The search set up allowing a queue of multiple data sets was selected through SSI starting with MicroMass pkI files with the user adjustable parameters that
included the fragment and precursor mass-to-charge (m/z) error tolerances, proteolytic digestion, a set of oxidation modifications specified in Table 4.1, recalibration, and a database selection. The m/z error tolerances, typically 0.4 for parent and 2.3 for fragment (in Daltons, Da), specificity for trypsin with allowance for certain semi-tryptic and non-tryptic peptides with high confidence (further discussed in the text) were passed into ByOnic for the initial pass through the database but reverted to the default settings for the second pass since recalibration was selected.

**Gel electrophoresis**

Lyophilized proteins were re-suspended in 7 µL of nanopure water and 14 µL of Bio-Rad premixed Laemmli sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% glycerol, 0.01% bromophenol blue) was added to yield a 2:1 (sample buffer : protein solution) ratio. All reagents for gel electrophoresis and the apparatus were from Bio-Rad Laboratories unless otherwise specified. Samples were boiled for 9 minutes and loaded onto a ready made gel. Two kinds of precast gels were used for the proteins: 4-20% precast polyacrylamide gel for *Shigella flexneri* IgG and 15% precast polyacrylamide gel for galectin-1, with a single running buffer (25 mM Tris, 192 mM Glycine, 0.1% (w/v) SDS, pH 8.3) for an optimal separation in the appropriate molecular weight regions for each protein. SeeBlue Plus2 Pre-Stained protein standards, Invitrogen (Carlsbad, CA, USA) were used as protein molecular weight markers. Gels were run for 1 hr 20 min in a Mini PROTEAN 3 Cell at a voltage of 150 V and a current of 100 mM on ice. After run completion, gels were removed from the electrophoresis set up, washed with water and stained using Coomassie Brilliant Blue R-250 for 45 minutes. Washed again with water and de-stained using a water / methanol / acetic acid (6:3:1, v/v/v) solution (reagent from Fisher
for 1 hour and left in nanopure water overnight before images were acquired.

**Results and Discussion**

*Manual assignment of non-standard spectra*

Even after the automated process of tandem mass spectra assignment, manual validation is often useful for assuring the correct peptide identifications. During this validation process, a mixed spectrum of a peptide from digested galectin-1 was observed and is further discussed below (Figure 4.1). During the oxidation process, many amino acids – and more than one on a given peptide - can get oxidized, which raises the likelihood of the presence of isomers that cannot be resolved by automatic peak assignment. Figure 4.1A shows a non-modified peptide, DSNNLCLHFNPR, correctly assigned by ByOnic software, and Figure 4.1B shows a mixed spectrum of two to three oxidized isomers of that peptide. The software was able to assign only one of the isomeric peptides. After manual validation up to two additional peptides were assigned from the same spectrum. ByOnic made a correct assignment, but it left a large number of fairly intense peaks unassigned. ByOnic has the capability of identifying more than one peptide per spectrum; it does this by knocking out all the peaks of the first identification, and then sending the remaining peaks through the bioinformatics pipeline as if they were a new spectrum. This approach works for a mixture of two unrelated peptides, but rarely works for a mixture of two isomers of the same peptide, due to the many shared peaks. Even after the manual validation, there is no way to prove whether the peptide, DSNNLCLHFNPR, with Leu(+16), Cys(+32), and His(+16) is present in the mixture of the two other isomers due to exactly sharing at the same fragment peaks. This phenomenon might be partially resolved by
future programs giving probability values for modification occurrence on a particular amino acid
in a peptide. As one can imagine, the occurrence of isomers in a single spectrum will likely
complicate the surface mapping process of a protein due to problems assigning a modification to
a specific residue.

**Synthetic peptides simulating biological systems conditions**

In order to study biological systems such as protein complexes by the oxidative surface
mapping method, it is crucial to develop conditions under which the protein gets sufficiently
oxidized while limiting the protein degradation. The LHFNPR model hexapeptide was tested for
the amount of oxidation with respect to the peptide concentration. The level of oxidation for the
peptide was related to the peptide concentration such that when lowering the concentration of the
peptide in the sample, more of the oxidized species with greater intensities were observed after a
static single laser pulse irradiation, shown on the MALDI-TOF spectra (Figure 4.2). Typical
oxidation products of Phe, Leu, and His were formed as known from previous studies [14, 35,
40, 42]. Under the standard conditions, any additional low molecular weight fragments were not
observed and therefore peptide bond cleavages after laser irradiation were unlikely in this
sample.

A peptide related to *Shigella flexneri* mAb sequence, GLEWVAEIR, was studied to
determine which oxidation products should be expected after the laser irradiation. The peptide
sequence was selected from a region of the antibody known to be crucial for the ligand binding, a
CDR, for validation of some of the key residues and their susceptibility to oxidation for the
identification of the suitable probes relevant in the ongoing footprinting studies. A ByOnic
database search was applied to the peptide against a database that contained only the
GLEWVAEIR peptide; manual assignment for a number of spectra was required as well. Mixtures of the oxidation products were found in the oxidized MS/MS spectra of the peptide. Similar to a peptide from digested galectin-1, the automated tools identified only one of the products of oxidation. The mixtures of isomers were common in synthetic peptide samples, but were also identified in tryptic digests of *Shigella flexneri* mAb, demonstrated on the GLEWVAEIR synthetic peptide (Figure 4.3A) with +48 Da modification (Figure 4.3B) and the peptide from oxidized mAb digestion (Figure 4.3C). The isomeric mixture from the mAb peptide was found to contain only two isomers, and in certain spectra it was possible to identify the main product in the mixture due to higher fragment ion intensities derived from one isomeric form. This was in contrast to the synthetic peptide where the intensities of different fragments that specify the various isomers varied significantly. In the digested mAb sample, only the two following products were found: GLEW(+48)VAEIR and GL(+16)EW(+32)VAEIR whereas in the synthetic peptide mixtures, in addition to the above-mentioned isomers, GL(+16)EW(+16)VAEI(+16)R or GLEW(+32)VAEI(+16)R products were identified. It is not possible to determine which of these two species with isoleucine oxidized was present or if both were present because the fragment peaks that would discriminate these two peptides were already present in the mixture (Figure 4.3B). Possibly low levels of other isomers were present as well because in certain spectra, relatively abundant fragment peaks were left unassigned.

The mixed spectra were also studied with respect to the elution time. There were differences in terms of the fragment ion abundances eluted early in the chromatogram (40-70 min) versus the ones eluted late (90-125 min). Several oxidation products were present in a single peptide MS/MS spectrum, and the abundance of each product ion in such mixed spectrum varied with the retention time in the synthetic peptide samples. In the digested protein, only a
few of such mixed spectra per sample were identified, so there was not enough information for
the elution time based comparison. We generally find less oxidation products in peptides from
oxidized protein digests versus oxidized synthetic peptides, which could be a consequence of the
tertiary structure in the protein (protecting it, somewhat, from oxidation), as opposed to a linear
chain of a short synthetic peptide where all residues are solvent exposed. It might also be more
challenging to detect some of the oxidized peptides with a low abundance in protein digests due
to the larger diversity of peptides that result from the digestion of the protein. The oxidative
conditions for the protein and the peptide were kept the same but the relative concentration of the
particular peptide of interest in the digest mixture was lower than in the sample that contained
pure peptide only.

The synthetic peptide yielded a wide range of oxidation products on almost all residues,
excluding the terminal glycine. All other residues were observed at least once with a hydroxyl
group attached, except that for alanine it was not possible to prove whether it was modified or
not due to the fragment ion overlap with other residues. The most commonly oxidized amino
acid in this sequence was tryptophan with +16, +32, +48, and +14 Da modifications identified.
Tryptophan was also often modified in the protein sample, which is helpful for footprinting
studies in this case due to its presumed involvement in carbohydrate-ligand binding. Another
footprinting target, the second glutamic acid of the peptide sequence, is more challenging to find
oxidized in the protein digest samples. One of the reasons is that it was mostly found in a singly
charged form and the fragmentation pattern was not good enough, missing the key ions in the
spectrum, to confirm the site of oxidation. The synthetic peptide oxidized sample, a doubly
charged form of the peptide with a −30 Da loss, suggests the oxidation of glutamic acid, but after
a manual interpretation, it was not clear whether a decarboxylation of the C-terminal causes the
mass shift or whether this is another case of the mixed spectrum containing both –
decarboxylated peptide on the terminus and on the Glu. Nevertheless, for the surface mapping
studies, it would be beneficial to increase the amount of oxidation to have more sampling of the
oxidized species since the mixed spectra in the digested mixture are rather rare or less
complicated and some of the less reactive residues such as the target glutamic acid are rarely
found in the oxidized form.

Impact of a pulsed laser on the protein structure

Gel electrophoresis can be useful for evaluating the protein integrity after laser exposure.
Various conditions including single batch laser pulse and a flow system exposure with and
without hydrogen peroxide present were tested to find potential backbone cleavages or other
protein degradation processes during the irradiation procedure. When a flow system is used, it is
also possible that pumping the protein solution through a capillary could potentially cause
protein losses that appear to happen after employing the flow system method of oxidation as
indicated by mass spectrometry analysis of the digested protein. Both galectin-1 and the mAb
were subjected to the test. Seven samples of each protein were exposed to various conditions
including static and flow system irradiation with and without hydrogen peroxide, as indicated in
Figure 4.4, and loaded onto an SDS gel for running the electrophoreses. As a result, both
proteins reflected a similar trend: a slight loss of protein during sample handling for laser
irradiation compared to the protein sample from the stock solution. However, the results also
showed a significant decrease of the band intensities in cases when the laser is turned on and the
protein has going through the capillary regardless of whether hydrogen peroxide is present
(Figure 4.4, lanes 3 and 4). This suggests that not only the concentration of the protein or
hydrogen peroxide in the solution is important, but also a volume or area of the exposed material is crucial for setting up the irradiation conditions. Apparently, when too small amounts of the protein are exposed to the laser pulse at a time, laser over-exposure results in degradation, which is an unwanted condition for the structural studies of the molecular complexes. In figure 4.4C, the laser has been unfocused from 1mm to 3 mm spot and it improved the overall result such that no more protein losses were observed. A similar trend was observed when studying synthetic peptides by mass spectrometry. After the flow system irradiation with a relatively high energy per pulse (~60 mJ), the MALDI spectrum of the peptide (GLEWVAEIR) had about a ten-fold decrease in intensity compared to the non-modified or single batch pulse irradiated sample (Figure 4.5). This suggests a significant loss of material during the peptide sample oxidation and agrees with the visual evaluation of the protein samples on the SDS gels.

**Conclusions**

This study has relevance to protein surface mapping, and in our case sets a stage for mapping carbohydrate-protein interactions for anti-carbohydrate mAb. Various conditions of the protein oxidations have been tested. The results suggest that choosing the most suitable conditions for protein irradiation is crucial in order to prevent degradation of the protein structure while still yielding sufficient oxidation for binding site mapping.

Oxidation of synthetic model peptides revealed a complicated problem of isomeric mixtures being present in an oxidized single tandem mass spectrum. To date, the search programs are capable of assigning one peptide to one spectrum only, so that manual interpretation of mixed spectra of oxidized products is required. Even then, it is not always possible to identify all of the isomeric products due to fragment peak sharing. This certainly
makes the method challenging and an effort should be put towards software development that would be able to resolve such cases even though this phenomenon is less common in protein samples than in synthetic peptides. The model peptide study of the reactivity of the amino acids present in the CDR region of the mAb showed many oxidation products of the synthetic peptide but only mild oxidation of the protein sample in that region occurred. Therefore, an overall increase in oxidation when studying the intact antibody is desirable for surface mapping studies in order to maximize the chances of oxidizing footprinting target residues as well as to increase the resolution of the method.
Table 4.1. Oxidation products searched for by ByOnic. The table contains the chemical modifications of the amino acids [6, 13, 68, 122, 123] as well as how the side chains change after oxidation on the bases of elemental composition.

<table>
<thead>
<tr>
<th>Mass Shift (Da)</th>
<th>Affected Residues</th>
<th>Modification</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>+15.995</td>
<td>Met, Trp, Tyr, Phe, His, Arg, Ile, Leu, Val, Lys, Gln, Ser, Thr, Pro, Glu, Asp, Asn, Ala</td>
<td>Oxidation/ Hydroxylation</td>
<td>O</td>
</tr>
<tr>
<td>+31.990</td>
<td>Cys, Met, Trp, Tyr, Phe</td>
<td>Double Oxidation</td>
<td>O(2)</td>
</tr>
<tr>
<td>+47.985</td>
<td>Cys, Trp, Tyr, Phe</td>
<td>Triple Oxidation</td>
<td>O(3)</td>
</tr>
<tr>
<td>-43.053</td>
<td>Arg</td>
<td>Oxidation to glutamic semialdehyde</td>
<td>N(-3) C(-1) H(-5)</td>
</tr>
<tr>
<td>-22.030</td>
<td>His</td>
<td>Oxidation to aspartic acid</td>
<td>H(-1) N(-2) C(-2) O</td>
</tr>
<tr>
<td>-15.977</td>
<td>Cys</td>
<td>Oxidation to serine</td>
<td>O S(-1)</td>
</tr>
<tr>
<td>-31.970</td>
<td>Met</td>
<td>Aldehyde product</td>
<td>H(-4) C(-1) O S(-1)</td>
</tr>
<tr>
<td>-10.030</td>
<td>His</td>
<td>Pyrrole-ring opening</td>
<td>H(-2) N(-2) C(-1) O(2)</td>
</tr>
<tr>
<td>+5.000</td>
<td>His</td>
<td>Pyrrole-ring opening</td>
<td>H(-1) N(-1) C(-1) O(2)</td>
</tr>
<tr>
<td>+13.979</td>
<td>Trp, Ile, Leu, Pro, Lys, Glu, Gln, Val, Arg</td>
<td>Carbonyl</td>
<td>H(-2) O</td>
</tr>
<tr>
<td>-2.010</td>
<td>Ser, Thr</td>
<td>Oxidation and water loss</td>
<td>H(-2)</td>
</tr>
<tr>
<td>-30.0105</td>
<td>Asp, Glu, Pro</td>
<td>Decarboxylation</td>
<td>H(-2) C(-1) O(-1)</td>
</tr>
<tr>
<td>+19.990</td>
<td>Trp</td>
<td>Oxidation to hydroxykynurenin</td>
<td>C(-1) O(2)</td>
</tr>
<tr>
<td>-18.010</td>
<td>Glu</td>
<td>Oxidation to Pyroglutamate</td>
<td>H(-2) O(-1)</td>
</tr>
<tr>
<td>-17.026</td>
<td>Gln</td>
<td>Oxidation to Pyroglutamate</td>
<td>H(-3) N(-1)</td>
</tr>
</tbody>
</table>
Figure 4.1. A peptide from galectin-1 protein digest. A. A non-oxidized peptide DSNNLCLHFNP; MS/MS fragmentation. B. Oxidized peptide with a mixture of two to three isomers present in a single spectrum. Colors: blue – non-oxidized ions; green – oxidation of Cys (+48) and His (+16); pink – oxidation of Leu (+16) and Cys (+48); red – shared fragment ions for +16 and +48 Da mass shifts.
Figure 4.2. MALDI-TOF spectra of a peptide LHFNPR. Level of oxidation increases with lower peptide concentration. A. Peptide before oxidation. B-D. Peptide after oxidation; hydrogen peroxide concentration was 1% whereas peptide concentration varied (1-10 µM).
Figure 4.3. Tandem mass spectra of synthetic peptide GLEWVAEIR and an identical peptide from a digest of the *Shigella flexneri* mAb sample.  

**A.** Non-oxidized form of the synthetic peptide.  

**B.** Mixture of oxidation products with +48 Da mass shift; synthetic peptide.  

**C.** The mAb peptide with mixture of only two oxidation products.  

Colors: navy – fragment ions without modification; magenta – fragments with Trp(+48) modification; light blue – fragment ions specific to GLEW(+32)VAEI(+16)R or GL(+16)EW(+16)VAEIR(+16)R; red – fragment ions with +48 Da modification shared by several peptides; green – fragments with either +16 or +32 Da modifications specifying GL(+16)EW(+16)VAEIR peptide; some of the green +16 ions can be shared with GL(+16)EW(16)VAEI(+16)R (light blue) if present.
**Figure 4.4.** SDS gel of oxidized and non-oxidized *Shigella flexneri* mAb and galectin-1. 1. Control – protein with no treatment. 2. Protein going through capillary without H$_2$O$_2$ and laser off. 3. Protein going through capillary without H$_2$O$_2$ with laser on. 4. Protein going through capillary with 1% H$_2$O$_2$ and laser on. 5. Protein with 1% H$_2$O$_2$ without laser irradiation. 6. Protein without H$_2$O$_2$ with a static laser pulse. 7. Protein with 1% H$_2$O$_2$ and a static laser pulse. Concentration of the antibody sample during testing was 4 µM, galectin-1 15 µM. The arrows (lanes 3 and 4) indicate a significant loss of material when the protein is going through the flow system with the laser turned on. In figure C, the laser has been unfocused from 1 mm to 3 mm size spot.
Figure 4.5. MALDI spectra of GLEWVAEIR peptide with decreasing signal intensities after laser exposure.
CHAPTER 5

CONCLUSIONS AND RECOGNIZED CHALLENGES OF THE OXIDATIVE FOOTPRINTING METHOD

In this thesis, approaches for the characterization of specific structural properties of proteins or protein-ligand complexes have been discussed and further explored experimentally. Oxidative surface mapping, also often called oxidative footprinting, is an emerging approach that has the potential to become a high throughput screening method for the identification of critical binding residues in protein-ligand complexes. Even though oxidative footprinting is not a completely novel method, a practical application to resolving structurally unknown protein-ligand complexes might not be totally straightforward. Several existing approaches were discussed in the introductory chapter and yet none of these techniques are considered standard. Also, the existing literature often consists of reviews or test cases showing that the method has the potential to work, but actual cases in which a biological problem has been solved are not very common. As the laser photolysis and water radiolysis have been applied on three different proteins described in three previous chapters, several challenges, some of them applicable to related oxidative footprinting techniques have been recognized and are summarized here. It appears that in every step of the method, choices must be made to adjust the working protocol for the particular study. This might be the reason why no standard method has been established.

Laser induced photodissociation of hydrogen peroxide can be performed under various conditions and these might impact the overall protein structure. Therefore, certain optimizations
are needed and may vary for each protein. Examples of such optimizations have been discussed in chapters two and four. Certainly the instrumentation for performing the photooxidation reaction is quite specific. A high energy pulsed laser is required as well as a special capillary pump setup. If the laser power is set too high, it may burn the capillary through which the protein solution is pumped. Again, no standard kit is available for the task and the capillary pump and laser setup need to be optimized by the researcher. From our results of oxidized proteins, it appears that there are differences in the amount of oxidation among various proteins creating another optimization task dependent on the type and perhaps the size of the protein. If the protein is not sufficiently oxidized, it lowers the resolution of the method. On the other hand, over-oxidation can lead to backbone cleavages [50] and protein losses as described in the chapter four.

Once the protein is oxidized, sample preparation for mass spectrometry analysis varies among the groups using the technique. Some studies suggest flash freezing of the sample to stop the oxidation reaction [11], others add quenchers [22, 23] to remove residual hydrogen peroxide, and a recent study suggests that flash freezing causes even more protein oxidation [124]. Proteolytical digestion is the next step that needs to be optimized for every protein. If the traditional protease, which is trypsin for sequencing purposes, does not give satisfactory results such that the region of interest is not seen by mass spectrometry, additional proteases need to be considered.

Mass spectrometry can also be performed at different levels of sophistication, for example conventional mass spectrometry, usually by MALDI-TOF, can be used to find whether a peptide has been oxidized, but it might not always be possible to determine which of the amino acids carries the oxidation. The specific site of oxidation is usually resolved by tandem mass
spectrometry. For better resolution, electrospray instruments are usually employed, which is a technically more difficult and longer experiment than acquiring MALDI-TOF spectra. Variations are also possible within electrospray experiments: for example a reverse phase column can be used to pre-separate the peptides prior to entering the mass spectrometer and undergoing the MS/MS fragmentation. Also, different types of fragmentation, such as collisionally-induced dissociation (CID) or electron capture dissociation (ECD) can be used to quantify the oxidation abundance. That leads to another question in the procedure, which is the type of quantification method. It can in MALDI-TOF spectra use relative intensities from a single spectrum or in the electrospray mass spectrometry a summary of intensities over a period of time. There are also attempts to quantify using the fragment ions from MS/MS spectra, which would have the advantage of taking the peak intensity of the oxidized fragment ion instead of the intensity of the entire peptide. This procedure employs ECD and technically the experiment is more challenging due to manual peaks selection for the fragmentation rather than relatively well-established CID methods that in our setup were automatically performed by the mass spectrometer. A certain degree of background and experience in mass spectrometry is required for the last listed protocol, where as acquiring MALDI-TOF spectra and extracting peak intensities, for instance, is usually a less difficult task. Also, with the electrospray ion source, large data files are generated and with the high sensitivity of the instruments, a certain percentage of the data might result from sample contamination. It is usually necessary to pre-sort and assign the spectra with a database search program such as Mascot, ByOnic or X!Tandem [125]. A possible problem with the automated tools is related to the peptide oxidations. Assignment in oxidative surface mapping, while lacking some of the problems associated with proteomics, creates a significant computational challenge, as the modifications can potentially
result in many permutations of the same peptide due to isomers that result from modifications present on various amino acids in the peptide. The programs are often able to determine whether or not the particular peptide contains an oxidized residue, but the confidence of the correct assignment for the particular amino acid might not be sufficiently high. One of the reasons is that not all the mass spectra have complete fragmentation and therefore, the programs might miss the crucial peaks that would define the position of the modification. Development of software that would be suited just for oxidative modification searches could rapidly increase the efficiency of the method. ByOnic is one of the searching tools that has been modified to search for a set of oxidative modifications, but originally was not written for this task. Therefore, it might not be efficient and suitable for everyone in the footprinting field to use. For instance, it is not generally possible to search for specifically chosen modifications; instead, a set of amino acids and their mass shifts is pre-defined. It also lacks an option to view the assigned spectra for oxidative searches, but, as mentioned in chapter four, a visualization tool was developed (SurfMap Search Interface) that enables the spectra assignment visualization. An alternative search strategy is to use a standard search engine such as Mascot or X!Tandem in a multiple-pass fashion. Each pass searches the data with a limited number of user-defined oxidative modifications enabled. For comprehensive data analysis, the results of the passes can be combined and the output viewed by ProteoIQ software (BIOINQUIRE, www.bioinquire.com). As described above for ByOnic assignments, other search engines also struggle with similar uncertainty about the position of the modification in a peptide. Nevertheless, even with the drawbacks of these programs, it is a big progression in the field to be able to do a database search with over forty oxidative modifications at a time. One has to decide to which extent a manual validation of the results is necessary. The number of assignments for an average size protein per
run can go to several hundreds of identified peptides, which makes it difficult to verify the majority of them manually.

Another recently discovered challenge in oxidative footprinting is the occurrence of peptide mixtures in a single mass spectrum. These are the co-eluting isomers that were not separated even though they went through a reverse phase column prior to entering the mass spectrometer. What happens when a search program assigns an identification of such spectrum is that the software picks only one of the isomeric peptide and ignores the rest of them. From the experiments in chapter four, there were up to four peptides present in one spectrum. In such a case, only 25% of the automated assignment would be correct. Fortunately, it appears that such mixed spectra are not very common, as very recently investigated by Xi et al. [120]. For the future development of searching tools, it would be appropriate to consider that such cases can occur in order to be able to fully resolve the mixed spectra.

As a summary, a bullet-point list of the major recognized challenges of the oxidative surface mapping method follows:

- Oxidative footprinting lacks a standard method and conditions of each step in the workflow need to be adjusted for the specific case.
- Considerable system-specific tuning of experimental apparatus required, for example:
  - laser beam alignment and power adjustment
  - syringe pump and capillary setup
- Considerable system-specific optimization of experimental conditions required, such as:
  - optimization of protein proteolytic digestion
protein concentration for irradiation
quenching of oxidation

- The amount of oxidation varies among proteins; optimization of the amount of oxidation is needed for the protein in each study.
- The working protocol varies among the methods and needs to be carefully chosen; for example flash freezing of the protein can cause over-oxidation as a recent study suggests.
- The mass spectrometry technique needs to be chosen: MALDI or ESI with an adequate quantification method and proper MS/MS techniques such as CID or ECD.
- Database search programs suited for the task in case of employing MS/MS fragmentation and ESI have to be used together with manual validation which is time consuming.
- The occurrence of complicated MS/MS spectra with a mixture of isomers must be taken into consideration in the method and such spectra cannot be resolved by database searching tools alone.

Of all the challenges in the oxidative surface mapping workflow named above, the most important improvement that could be done is improvement of the automated tools for electrospray data analysis, because that is the most time consuming part of the method especially if extended manual validation is necessary. In conclusion, the author believes that presenting examples of surface mapping studies, and defining the challenges and describing some of the interesting phenomena in the method can lead to a more efficient workflow in future footprinting studies, as well as contribute to the field of covalent labeling techniques.
REFERENCES


